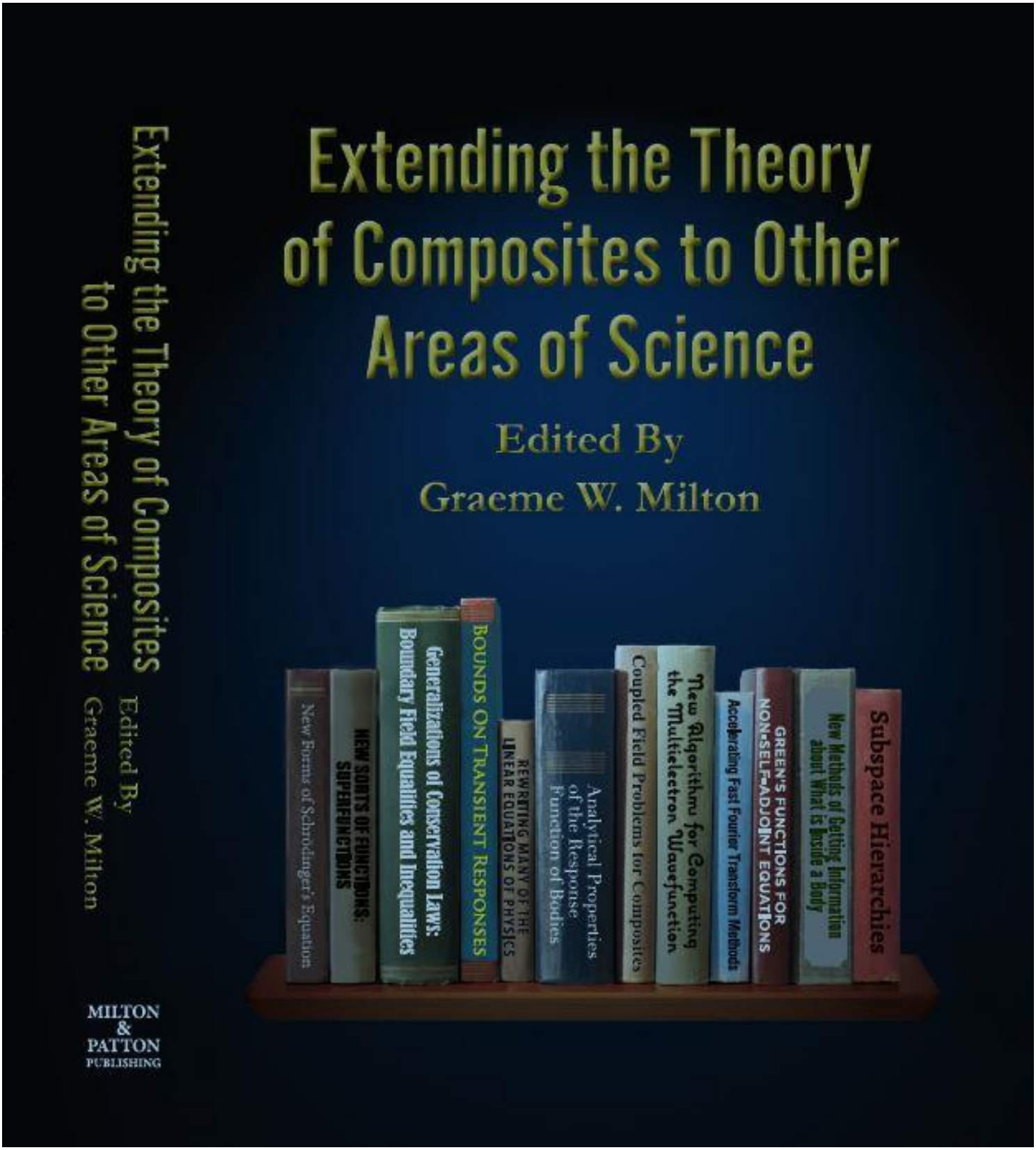


This is the electronic version. Please consider purchasing the printed version, or ask your library. Only $80, at

www.math.utah.edu/~milton/

This helps keep the costs of books low, compared with traditional publishers.

Extending the Theory of Composites to Other Areas of Science

Edited By

Graeme W. Milton

Library of Congress Cataloging-in-Publication Data

Milton, Graeme W. (Graeme Walter), 2016

Extending the Theory of Composites to Other Areas of Science /
Graeme W. Milton, editor
xx + 422 p. 24 cm.
Includes 38 illustrations.
Includes bibliographic references, author/editor index, and subject index.
ISBN-13 978-1-4835-6919-2

Printed on acid-free paper.

Print Production: `www.bookbaby.com`

The World-Wide Web site `http://www.math.utah.edu/books/milton` contains current information about this book, together with online resources for the book, including the BIBTEX bibliographic databases, figures, and videos.

© 2016 by Graeme W. Milton

Published by Milton-Patton Publishers
P.O. Box 581077
Salt Lake City, Utah, 85148, U.S.A.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing of the copyright holders.

Printed in the United States of America

10 09 08 07 06 05 04 03 02 01 00 10 9 8 7 6 5 4 3 2 1

`print ISBN-13: 978-1-4835-6919-2`

`print ISBN-10: 1-4835-6919-5`

`e-book ISBN-13: 978-1-4835-6920-8`

`e-book ISBN-10: 1-4835-6920-9`

Dedication

To my husband John, and my brothers David, Ian and Richard

Preface

The first book I wrote, *The Theory of Composites* (Milton (2002)), took me almost 14 years to complete. By contrast this book mainly took just 4 months (and another 4 months polishing) which is indicative of the concentrated effort it took, and motivation I had. I should emphasize that it is not just my work, and that this is a volume edited by me, rather than a monograph. Pivotally important contributions have been made by my coauthors on various chapters: Maxence Cassier, Ornella Mattei, Moti Milgrom, and Aaron Welters. They are all wonderful collaborators. I am also very grateful to Hervé Moulinec and Pierre Suquet: they helped clean up the formulation in **Chapter 8** of a method for accelerating Fast Fourier Transform schemes for computing the moduli and fields in composites, and did the numerical computations which showed that the scheme works in practice.

I said the book took 4 months to write, but that is not completely true as some chapters have been in the works for a while. In particular, the work with Moti Milgrom dates back over 23 years to our collaborations at the Courant Institute when I was an associate professor there. We had essentially finished that work, but for some reason never published it. As the subject matter fits closely with parts of this book, it seemed very appropriate to include it as **Chapter 9**. I had been thinking for many years off and on since 1987 about the subject matter of **Chapter 7**, on the algebraic properties of subspace collections. I published a draft of it in April of 2015 on arXiv (arXiv:1504.08061 [math.AG]). At that time I was missing the important ingredient of how to "multiply" subspace collections, which only became apparent to me at the end of July 2015 in conversations with Mihai Putinar and Aaron Welters during a visit to KAIST in Korea. The new approach to accelerating Fast Fourier Transform schemes by substituting a subspace collection with non-orthogonal subspaces in one with orthogonal subspaces was started at the end of 2013, with Hervé Moulinec and Pierre Suquet being involved in early 2014, and is presented in **Chapter 8**. Ornella Mattei came and visited me for six months in Spring of 2015, and we worked on a method for bounding the transient response of bodies, presented in **Chapter 6**. This work forms part of her Ph.D. thesis (Mattei 2016). Subsequently, I realized that the Dirichlet-to-Neumann map, governing the response of bodies, should share the same analytic properties as effective tensors of composites. This led to some new methods discussed in **Chapter 5** for the inverse problem of finding inclusions in a body from measurements of the response of the body to transient fields. An in depth and rigorous study of the analytic properties of the Dirichlet-to-Neumann map for electromagnetism was also initiated with Maxence Cassier and Aaron Welters earlier this past summer, and led to **Chapter 4**. In an intense collaboration they also placed the field equation recursion method on a rigorous basis for composites of two isotropic components, and this resulted in **Chapter 10**. The field equation recursion method uses the structure of subspace collections to obtain a continued fraction expansion for the effective tensor from which bounds on the effective tensor can be derived.

The material of the book varies among a very wide range of scientific topics that are connected to the theory of composites, or to which some aspects of the theory of composites can be applied.

In **Chapter 1** we review many of the linear equations of physics, and write them in a canonical form

appropriate to the theory of composites. Additionally we show how conservation laws, which have played a key role throughout the history of science, can be generalized to equalities which we call "boundary field equalities" and inequalities which we call "boundary field inequalities". These require some assumptions about the materials inside the body. The fields in the canonical form of the equations satisfy differential constraints, linked by a constitutive law involving a matrix $\mathbf{Z}(\mathbf{x})$ of material parameters that is Hermitian or the Hermitian part of $i\mathbf{Z}(\mathbf{x})$ is positive semi-definite. When written in the appropriate form, the scalar product of fields on the left and right of the constitutive law can be expressed as the divergence of some "supercurrent" $\mathbf{Q}(\mathbf{x})$. As a result of this we have the "key identity" that the integral of this scalar product can be expressed in terms of boundary fields. We remark that the canonical form of the equations we introduce here is similar in many respects to that of Strang (1986) [see also Strang (1988), and **Chapter 2** of Strang (2007)] where he proposes (see his figures 1.8, 2.2, 3.8, 3.9, 3.12, 3.16 and pages 126, 156) the form

$$e = b - Ax, \quad y = Ce, \quad A^T y = f, \tag{0.1}$$

where $y = Ce$ is the constitutive law, x are the potentials, b and f are sources, and in a discretization of the equations A with transpose A^T are matrices which represent the differential operators. In this discretization e, b, x and f are all represented by vectors. Then if $b = 0$ and f is non-zero except on the boundary nodes, we have

$$y \cdot e = -y \cdot Ax = -(A^T)y \cdot x = -f \cdot x. \tag{0.2}$$

This is equivalent to the key identity since the last term only involves terms at the boundary. Here we show that this format, appropriately generalized, applies to many of the linear equations of physics. The "boundary field equalities and inequalities" we derive in Sections 1.4 and 1.5 provide equalities and inequalities that the fields on the boundary of a body (possibly a body in "space-time") must satisfy given information about the partial differential equations governing the fields inside the body, and so provide generalizations of the concept of conservation laws. This information need not be complete: for example, in a body containing multiple materials, we can still get equalities and inequalities without knowing the orientation or the distribution of materials inside the body.

Chapter 2 reviews the abstract theory of composites, both for the effective tensor and for the associated "Y-tensor".

Chapter 3 shows that the problem of finding the Dirichlet-to-Neumann map, for acoustics, elastodynamics, or electromagnetism in an inhomogeneous body can be reformulated as the problem of finding the effective tensor (operator) associated with an abstract problem in the theory of composites. As a result of this, the Dirichlet-to-Neumann map is an operator-valued analytic function, and in fact Herglotz function, of the moduli of the tensors of the component materials, and many tools for bounding effective tensors extend to bounding the Dirichlet-to-Neumann map.

Chapter 4, with Maxence Cassier and Aaron Welters, studies in depth the analyticity properties of the Dirichlet-to-Neumann map for electromagnetism, first for a layered medium then for quite general bodies of N-isotropic phases, as functions of the $2N$ variables $\omega\varepsilon_j$ and $\omega\mu_j$, $j = 1, 2$ where ω is the frequency and ε_j and μ_j are the electrical permittivity and magnetic permeability of the phases. It is established that the map is an operator-valued Herglotz function of these $2N$ complex variables. The results are extended to the case when the phases are anisotropic, in which case the analyticity is as a function of the elements of each permittivity tensor and each permeability tensor, multiplied by the frequency. Also the Herglotz properties of the Dirichlet-to-Neumann map, as a function of frequency are established for bodies where the moduli $\omega\varepsilon(\mathbf{x}, \omega)$ and $\omega\mu(\mathbf{x}, \omega)$ are not piecewise constant but instead vary with position, and at each point $\mathbf{x}$ are Herglotz functions of the frequency ω.

Chapter 5 explores bounds on the Dirichlet-to-Neumann map and the associated inverse problem of what can be said about the distribution of materials inside a body from surface measurements, using the connection between Dirichlet-to-Neumann maps and effective tensors in composites. Various approaches are developed: the first, which may not be very useful, is to apply the simplest bounds on effective tensors to bound the associated Dirichlet-to-Neumann map. Improvements can be obtained by using the "translation method", which then couples the response at different frequencies or with different applied fields. For two-phase composites more can be said by looking for complex frequencies where the body is homogeneous (or has a homogeneous shear modulus in the case of elasticity). Then measurements of the transient response of the body, can be extrapolated back (using analyticity and representation formulas for the response) to obtain the Dirichlet-to-Neumann map (or bounds on this map) near or at these special frequencies. Since the body is nearly homogeneous, or has constant shear modulus, it is much easier to say something about the internal geometry from these results. Lastly for quasistatic electromagnetism in two-phase bodies we propose using matrix-valued Pick interpolation to interpolate the Dirichlet-to-Neumann map at a set of frequencies to obtain information about the geometry. It is not clear how successful and robust any of these methods will be, especially in the presence of measurement errors: this awaits numerical tests.

Chapter 6, with Ornella Mattei, uses representation formulas for the effective tensor of a composite as a function of the component moduli to derive bounds on the transient response of bodies. This is done in the context of antiplane elasticity, although the results apply immediately to the mathematically equivalent problem of two-dimensional conductivity, and can be easily extended to three-dimensional conductivity. Significantly, we found the volume fractions of the phases could almost be exactly determined from measurements of the transient response at certain times. This gives some hope that the method could be applied, as suggested in **Chapter 5**, to provide useful information about the geometry inside a two-phase body from its transient response. The work in this chapter formed part of the Ph.D. thesis of Ornella (Mattei 2016).

Chapter 7 develops the algebra of finite-dimensional subspace collections, like those appearing in the abstract theory of composites and in the theory of Y-tensors when all subspaces are finite-dimensional vector spaces. By relaxing the requirement that the subspaces are orthogonal we find that the associated effective modulus can be any homogeneous degree 1 function of the component moduli satisfying the normalization property that it takes the value 1 when the component moduli are all 1. There is a rich algebraic structure associated with subspace collections: operations of addition, subtraction, multiplication, division and substitution can all be defined. In many cases these are similar to the operations one can do on electrical networks. It is not clear where these ideas will lead, but certainly they represent a new mathematical direction.

In **Chapter 8** we show this algebra has important uses: in particular by substituting a collection with non-orthogonal subspaces in one with orthogonal subspaces we can accelerate Fast Fourier transform methods for computing the effective tensor and fields in periodic conducting composites, as demonstrated by the numerical results of Moulinec and Suquet.

Chapter 9, with Moti Milgrom, looks at the response of multiphase bodies and composites to a set of different fields, which may be electric fields, magnetic fields, temperature gradients, or concentration gradients, which interact in the components due to coupling terms in the constitutive laws. Particular attention is focused on the form the perturbation expansion takes in nearly homogeneous media. In composites the perturbation expansion coefficients allow one to recover the weight and normalization matrices which enter representation formulae for the relevant projection operators, and which enter bounds on the effective response tensor.

Chapter 10, with Maxence Cassier and Aaron Welters, develops a rigorous basis, using Fredholm theory, for the field equation recursion method for composites of two isotropic phases. The method is associated with orthogonal subspace collections, and uses a stratification of the Hilbert space, and an inductive procedure, to link together a sequence of associated effective tensors. Eliminating the intermediate effective tensors

from these relations results in a continued fraction expansion for the effective tensor of the original subspace collection. Appropriate truncations of this expansion lead to a sequence of nested bounds on the effective tensor that get tighter and tighter as the level of truncation is increased and correspondingly more and more series expansion coefficients are incorporated in the bounds. This is similar to the way truncations of continued fraction expansions of Herglotz or Stieltjes functions lead to bounds, but the advantage of the field equation recursion method is that it generalizes to multicomponent and polycrystalline composites.

Chapters 11, **12**, and **13** follow a different tack, and could be useful in quantum chemistry in computations of the wavefunction in multielectron systems. How useful is open to question, and needs to be explored numerically, but they do present novel approaches which could be refined and adapted in ways not anticipated here.

Chapter 11 presents a simple idea. When the integral of the square of Schrödinger's equation (with sources) is minimized over all (real) trial wavefunctions, with norm 1, the minimum is zero and only achieved when Schrödinger's equation is satisfied. By expanding out the square of Schrödinger's equation, and doing partial minimizations one arrives at a "density functional theory" for excited states, which I call Projection functional theory. Instead of just minimizing over the density it requires minimizing over three functions, each of which projects out information contained in the full wavefunction. It will still be very challenging to say something about what combinations of the three functions are realizable (meaning being associated with an electron wavefunction satisfying the required symmetries) and to determine what is an appropriate functional.

In **Chapter 12**, we begin with the form of Schrödinger's equation given in the first chapter, of fluxes and their derivatives connected by a matrix, through a constitutive law, to the wavefunction and its gradient. Then using the symmetry properties of the wavefunction we desymmetrize this equation: the potential entering the constitutive law needs not have the usual symmetries, but can be just a function of the coordinates $\mathbf{x}_1$ and $\mathbf{x}_2$ of two-electrons, and the fields on the left of the constitutive law need only be fluxes when they are symmetrized. This desymmetrized form should accelerate Fast Fourier transform methods for solving the equation: when going to real space from Fourier space and then back to Fourier space one need only do Fourier transforms on the variables $\mathbf{x}_1$ and $\mathbf{x}_2$. One does however need to keep track of the full wavefunction in Fourier space, which may make it prohibitive for systems with a large number of electrons.

Chapter 13 contains some miscellania: a minimizing variational principle for Schrödinger's equation when the energy is complex, and Q^*-convex quadratic forms for Schrödinger's equation. The latter might be useful for accelerating Fast Fourier transform methods, as discussed in **Chapter 8**.

Finally, **Chapter 14** is mainly concerned with Green's functions for non-self-adjoint operators. Following the ideas of Cherkaev and Gibiansky (1994) and Milton (1990) we start with the equation $\boldsymbol{\mathcal{L}}\mathbf{u} = \mathbf{f}$, and look at it together with the equation $\boldsymbol{\mathcal{L}}^\dagger\mathbf{u}' = \mathbf{f}'$, where $\boldsymbol{\mathcal{L}}^\dagger$ is the adjoint of $\boldsymbol{\mathcal{L}}$. By adding and subtracting equations we see (when $\boldsymbol{\mathcal{L}}$ has the same domain as $\boldsymbol{\mathcal{L}}^\dagger$), that one obtains an equation $\mathbf{MU} = \mathbf{F}$, where $\mathbf{F}$ involves $\mathbf{f}$ and $\mathbf{f}'$, $\mathbf{U}$ involves $\mathbf{u}$ and $\mathbf{u}'$, and $\mathbf{M}$ is self adjoint and involves $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{L}}^\dagger$. Thus one obtains an equation to solve with a self-adjoint operator $\mathbf{M}$ and one can use resolvents to obtain the Green function. We also give a brief review of results for Green's functions in infinite homogeneous media.

The chapters in the book need not be read sequentially. Some chapters may be skipped, or not, according to the reader's interests. The chapters with coauthors (**Chapter 4**, with Maxence Cassier and Aaron Welters, on analyticity of the Dirichlet-to-Neumann map for electromagnetism; **Chapter 6**, with Ornella Mattei, on bounds for the transient response of viscoelastic composites; **Chapter 8**, with Moti Milgrom, on the response of systems with coupled fields; and **Chapter 10**, with Maxence Cassier and Aaron Welters, on a rigorous approach to the field equation method) are essentially self-contained and can be read independently of the rest of the text. Most readers will want to read **Chapter 1**, as it sets the framework in the context of a wide variety of problems, and **Chapter 2** as it reviews the abstract theory of composites. Those readers primarily

interested in the Schrödinger equation may then just wish to jump to **Chapters 11**, **12**, and **13**. Alternatively, those readers primarily interested in inverse problems may wish to focus on **Chapters 3**, **4**, **5**, and **6**. Readers who are interested in numerical methods may be most interested in **Chapters 8**, **11**, and **12**. **Chapters 6**, **8**, **9**, and **10** are recommended for those readers whose interests lie mainly in the standard theory of composites. In general, we have tried to keep the book accessible to physicists, chemists, and engineers, but they should not be afraid to skip material they find too technical. Pure mathematicians may be most interested in **Chapters 7** and **8**.

We remark on some non-standard notations used in the book. Given a body Ω with boundary $\partial\Omega$ we will frequently need the volume and surface integrals

$$\int_\Omega f(\mathbf{x})\, d\mathbf{x}, \qquad \int_{\partial\Omega} f(\mathbf{x})\, dS. \tag{0.3}$$

To shorten notation and avoid repetition we will often leave off the $d\mathbf{x}$ and dS these being assumed when one knows the integral is over the body Ω or the surface dS. Also, in keeping with the notation in Milton (2002) which is the opposite of what is standard, when we write $\mathbf{B} = \nabla\mathbf{u}$ and $\mathbf{q} = \nabla\cdot\mathbf{A}$ where $\mathbf{A}$ and $\mathbf{B}$ are matrices and $\mathbf{u}$ and $\mathbf{q}$ are vectors, we mean in terms of components that

$$B_{ij} = \frac{\partial u_j}{\partial x_i}, \quad q_j = \frac{\partial A_{ij}}{\partial x_i} = \sum_i \frac{\partial A_{ij}}{\partial x_i}, \tag{0.4}$$

i.e., ∇ and $\nabla\cdot$ are associated with the first index: a matrix field which is the gradient of a potential has columns which are gradients of the potential components, and a matrix field which is divergence-free has columns which are divergence free. In (0.4) and throughout the book (unless otherwise stated) we use the Einstein summation convention that one sums over repeated indices.

As is inevitable in a book that is edited by me, rather than entirely authored by me, there are some inconsistencies. For example, in **Chapter 4** the notations $\mathbf{E}$, $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{B}$ are used to denote the electric field, electric displacement field, magnetic field, and magnetic induction entering Maxwell's equations, while in the rest of the book they are denoted by $\mathbf{e}$, $\mathbf{d}$, $\mathbf{h}$ and $\mathbf{b}$. Also in that chapter the inner product between fields in the Hilbert space is defined to be linear in the first field entering the inner product, and antilinear in the second field (common in mathematics) while in the rest of the book it is the opposite: the inner product is defined to be antilinear in the first field entering the inner product, and linear in the second field (common in physics and engineering). Also sometimes the same symbol is used to denote different quantities, thus $\boldsymbol{\sigma}$ could be either the conductivity or the stress [in Milton (2002) this ambiguity was avoided by using $\boldsymbol{\tau}$ for the stress, but $\boldsymbol{\sigma}$ is the more standard notation]. This should not cause confusion, taking into account the context in which these symbols are used. Mostly we use (a_1, a_2, a_3) for the components of a three-dimensional vector $\mathbf{a}$, but sometimes we use (a_x, a_y, a_z) which is more common in physics and engineering. Similarly, mostly we use (a_1, a_2) for the components of a two-dimensional vector $\mathbf{a}$, but sometimes we use (a_x, a_y), and the components of a second order two-dimensional symmetric tensor $\mathbf{M}$ may be denoted by M_{11}, M_{22}, and M_{12} or by M_{xx}, M_{yy}, and M_{xy}, or even by M_x, M_y, and M_{xy} where the last notation is used for brevity. We use a double dot ":" to denote a double contraction of indices: thus $\mathbf{A} : \mathbf{B}$ means $A_{ij}B_{ij}$. By contrast a single dot "$\cdot$" could mean the scalar product between two vectors, or some scalar-valued product between the fields on the left and right of the constitutive law, and its explicit definition depends on the problem under consideration as we will see in **Chapter 1**.

This book owes a lot to many people. Foremost to my husband, John Patton, who supported me in countless ways while I was writing this, in particular making great dinners, orchestrating the book publishing, and

for his understanding when I had to concentrate on writing the book in my office. Patrick Bardsley is thanked for his help with typing the initial draft of the manuscript from my notes — that saved me an enormous amount of time. I am also grateful for him for spotting a number of errors, and his help with some numerical computations. The book would not have been possible without the tireless and absolutely amazing efforts of Nelson Beebe who devoted a considerable amount of time and energy to assembling the various chapters together into the desired format, and for carefully checking the bibliography, putting it into a uniform format, and adding "doi" and "ISSN" information to almost every reference, and making other countless corrections. Thanks go to him too for providing reference to two interesting historical reviews of Density Functional Theory (Zangwill 2014; Jones 2015) as well as other papers too. He has always been a fountain of knowledge keeping me abreast of scientific news articles related to my research, and other interesting topics. I also appreciated the help of Fernando Guevara Vasquez and Hyeonbae Kang for providing useful references in Inverse Problems in response to my queries. I'm very grateful to Mihai Putinar, for bringing my attention to the vast literature connected with Nevanlinna–Pick interpolation, and for providing important references in the spectral theory of non-self-adjoint operators. Similarly I am grateful to Kirill Cherednichenko for clarifying the work that had been done on self-adjoint extensions of non-self-adjoint operators. I sent a very rough draft of manuscript to friends, and am grateful to Ping Sheng, Martin Wegener, and John Willis for their feedback. I thank Ross McPhedran for his suggestions for the book title. Particularly helpful were Elena Cherkaev, Hyeonbae Kang, Paul Martin, Ornella Mattei, and Aaron Welters in spotting (with eagle eyes) various typos and corrections to be made, and just as helpful were the valuable comments of Richard Craster, Michael Fisher, Fernando Guevara Vasquez, Davit Harutyunyan, Alexander Movchan, and Pierre Seppecher: amazingly, Davit Harutyunyan and Alexander Movchan gave many useful comments on every chapter. It was most welcome that Richard Craster and Michael Fisher suggested that I discuss conservation laws, which led me to recognize their generalization: the boundary field equalities and inequalities introduced in Sections 1.4 and 1.5. Besides well-known boundary field inequalities, such as the fact that the net flow of electrical energy into a passive body is non-negative, there are many others such as those given in Section 2 of Harutyunyan and Milton (2015b). I am additionally grateful to Paul Martin for feedback which lead to a restructuring of the first two chapters of the book, and to Richard James and Vikram Gavani for helpful remarks on the Projectional Functional Theory for the multielectron Schrödinger equation (**Chapter 11**).

I also want to thank some people in the administration of the University of Utah who were supportive of this endeavour: notably President David Pershing, Associate Vice President for faculty Amy Wildermuth, and the Associate Chairman of our mathematics department, Nicolas Korevaar. Also I am grateful for the support of my friends and colleagues, Andrej Cherkaev, Elena Cherkaev, Yekaterina Epshteyn, Fernando Guevara–Vasquez, and especially Alessandra Angelucci, Paul Bressloff, Annette MacIntyre, and Robert MacLeod. Lastly, I am deeply indebted to the National Science Foundation for continual support throughout my career, and to the Packard Foundation and the Sloan Foundation as well as my many mentors Ross McPhedran, Michael Fisher, George Papanicolaou, John Willis, Bob Kohn, Don Melrose, Graham Derrick, Al Sievers, David Mermin, and Michael Cross for helping launch my career.

Graeme W. Milton
April 18th, 2016
Salt Lake City, Utah

Contents

List of Figures

List of Tables

1 Canonical forms for many of the linear equations of physics and key identities

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

We manipulate many of the linear equations of physics into a canonical form, where the fields on the left and right side of the constitutive law satisfy differential constraints which ensure the key identity that their scalar product can be expressed as a divergence, and therefore integrated over a body and expressed in terms of boundary fields. Additionally the matrix $\mathbf{Z}(\mathbf{x})$ in the constitutive law is either Hermitian or the Hermitian part of $i\mathbf{Z}(\mathbf{x})$ is positive semidefinite. Using the fact that the differential constraints are preserved under complex conjugation of the fields, we obtain associated key identities which lead to conservation laws, and boundary field inequalities that generalize conservation laws. We show that there can be boundary field equalities that are not implied by conservation laws. Following the argument given in Section 2 of Harutyunyan and Milton (2015b), other boundary field inequalities can be generated using Q_C^*-convex functions, which are a generalization of quasiconvex functions. For equations in the time domain the constitutive law may be replaced by a convolution with respect to time, in which case the Fourier transform of the integral kernel should be such that at fixed real frequency the Hermitian part of i times this Fourier transform is positive, or negative, semidefinite, according to whether the frequency is positive or negative. With a regularizing imaginary part added to the matrix in the constitutive law, and with source terms added, we show how solutions in periodic media can be formulated as an abstract problem in the theory of composites.

1.1 Introduction

Here we show that many of the linear equations of physics can be written in unifying framework, namely the basic framework appropriate to the theory of composite materials, and in Section 1.4 we show how conservation laws can be generalized to "boundary field inequalities" that provide rigorous inequalities on the fields at the boundary of a body. In the unifying framework, an appropriately defined scalar product of the fields on the left and right hand sides of the constitutive law can be written as the divergence of some appropriately defined supercurrent leading to the key identity that the integral of this scalar product can be expressed in terms of boundary values of the fields. As mentioned in the preface, the canonical form of the equations we

introduce here is similar in many respects to that of Strang (1986) [see also Strang (1988), and **Chapter 2** of Strang (2007)] where he proposes (see his figures 1.8, 2.2, 3.8, 3.9, 3.12, 3.16 and pages 126, 156) the form

$$e = b - Ax, \quad y = Ce, \quad A^T y = f, \tag{1.1}$$

where $y = Ce$ is the constitutive law, x are the potentials, b and f are sources, and in a discretization of the equations A with transpose A^T are matrices which represent the differential operators. [The transpose of a matrix $\mathbf{A}$ has elements $\{\mathbf{A}^T\}_{ij} = \{\mathbf{A}\}_{ji}$.] In this discretization e, b, x and f are all represented by vectors. Then if $b = 0$ and f is non-zero except on the boundary nodes, we have

$$y \cdot e = -y \cdot Ax = -(A^T)y \cdot x = -f \cdot x. \tag{1.2}$$

This is equivalent to the key identity since the last term only involves terms at the boundary.

It is the purpose of this chapter to obtain the associated canonical forms of various equations, the key identities, and to initiate the subject of "boundary field equalities and inequalities". The advantage of this is that then we can apply some of the machinery that has been developed in the theory of composites to bear on a much wider class of problems. Presumably too, as kindly pointed out to me by Paul Martin, results in these other areas could have an impact on the theory of composites. My expertise is in the theory of composites so it is easier for me to see how tools in the theory of composites can have wider applications, rather than the reverse.

We mention that there has already been a successful cross-pollination of ideas between the theory of composites and some other areas of science and engineering. One is in topology optimization, where one adjusts the geometry of a body to optimize some combination of factors, such as weight, stiffness against some set of applied loads at the boundary, compliance against some other set of applied loads, or flux of heat through specified portions of the boundary in response to prescribed temperatures at the boundary, for example. Often a geometry which performs well has microstructure in some places, indicating that one should reformulate the problem where one allows not just pure materials in the design, but also composites with microstructure much smaller than the body, and possibly with microstructure on many length scales (Tartar 1975, 1987; Armand, Lurie, and Cherkaev 1984; Kohn and Strang 1986; Kohn 1992; Allaire, Bonnetier, Francfort, and Jouve 1997; Cherkaev 2000; Allaire 2002; Bendsøe and Sigmund 2004; Tartar (2009)). This then leads to the very difficult question of trying to estimate the combination of properties (effective tensors) that a composite built from given materials can exhibit. See the books of Nemat-Nasser and Hori (1999), Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009) for a survey of the many results in this area.

Another area where there has been successful cross-pollination has been in non-linear elasticity and related areas, and in particular in the theory of shape memory materials (see, for example, Coleman and Noll 1959; Khachaturyan 1966, 1983; Roytburd 1967, 1968, 1978, 1993; Ball 1977, 1996; Ball and Murat 1984; Ball and James 1987; 1992; James and Kinderlehrer 1989; Müller 1987, 1998; Bhattacharya 1993; Bhattacharya, Firoozye, James, and Kohn 1994; Bhattacharya and Kohn 1997; Kohn 1991; Luskin 1996; and also **Chapter 31** of Milton 2002). It is interesting that convexity of the energy, or free energy, which underpins much of thermodynamics (see, for example, Wightman 1979) is not appropriate to non-linear elasticity. In a gas-fluid system it is easy to see how the convexity arises: as illustrated in **Figure 1.1**(a), if the energy W is not a convex function of the density ρ, the system phase separates into say regions of gas having a low density and regions of fluid having high density, thus having a lower energy than a homogeneous substance of the same overall density. There are no "jump" conditions that the density has to satisfy across an interface. The geometry of this macroscopic gas-fluid mixture can be fairly arbitrary. Surface tension can play a role, but in the large volume limit its effect is usually fairly inconsequential except for the nucleation of phases. In

nonlinear elasticity the energy depends on the deformation. One supposes that a point in the body which is originally at $\mathbf{x}$ in the undeformed state gets moved to the point $\mathbf{X}(\mathbf{x})$. In hyperelasticity theory the local strain energy density W is just a function $W(\mathbf{F})$ of the deformation gradient $\mathbf{F} = \nabla\mathbf{X}(\mathbf{x})$. The following argument shows that $W(\mathbf{F})$ is not a convex function of $\mathbf{F}$ (Hill 1957; Coleman and Noll 1959, page 110). Convexity of $W(\mathbf{F})$ would imply

$$cW(\mathbf{F}_1) + (1-c)W(\mathbf{F}_2) \geq W(c\mathbf{F}_1 + (1-c)\mathbf{F}_2), \tag{1.3}$$

for all square matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ (of the required dimension) and for all weights $c \in [0,1]$. [The value of W along the line joining $\mathbf{F}_1$ and $\mathbf{F}_2$ must be below the tie-line joining $W(\mathbf{F}_1)$ and $W(\mathbf{F}_2)$.] For simplicity, suppose one is in two dimensions. Then the trivial deformation $\mathbf{X}(\mathbf{x}) = \mathbf{x}$ must have the same energy as the deformation $\mathbf{X}(\mathbf{x}) = -\mathbf{x}$ that corresponds to a 180° rotation of the material. Taking the weighted average of the corresponding deformation gradients, $\mathbf{F}_1 = \nabla\mathbf{X}(\mathbf{x}) = \mathbf{I}$ and $\mathbf{F}_2 = \nabla\mathbf{X}(\mathbf{x}) = -\mathbf{I}$, with weights c and $1-c$, where $c \in [0,1]$, gives $\nabla\mathbf{X}(\mathbf{x}) = (2c-1)\mathbf{I} + x_0$ which corresponds to a deformation where $\mathbf{X}(\mathbf{x}) = (2c-1)\mathbf{x} - \mathbf{x}_0$ in which $\mathbf{x}_0$ is constant. So when c is close to $1/2$ the material is greatly compressed which costs a tremendous amount of energy. (When $c = 1/2$ everything gets compressed to the point $\mathbf{x}_0$, but surely elasticity theory will not apply then.....neutron stars and black holes will be formed before that point.) So clearly (1.3) will be violated as the right hand side can be much larger than the left hand side: $W(\mathbf{F})$ cannot be a convex function of $\mathbf{F}$.

What replaces convexity is quasiconvexity, which is the requirement that for some region Ω, with volume $|\Omega|$,

$$\int_\Omega W(\mathbf{F}(\mathbf{x}))\,d\mathbf{x} \geq W(\langle\mathbf{F}\rangle) \text{ where } \langle\mathbf{F}\rangle = \frac{1}{|\Omega|}\int_\Omega F(\mathbf{x})\,d\mathbf{x}, \tag{1.4}$$

for all functions $\mathbf{F}(\mathbf{x})$ that are gradients $\mathbf{F}(\mathbf{x}) = \nabla\mathbf{X}(\mathbf{x})$ of functions $\mathbf{X}(\mathbf{x})$ that satisfy affine boundary conditions, i.e., for some matrix $\mathbf{F}_0$, $\mathbf{X}(\mathbf{x}) = \mathbf{x}\cdot\mathbf{F}_0$ for all $\mathbf{x}$ on the boundary $\partial\Omega$ of Ω, where $\mathbf{F}_0$ can be identified with $\langle\mathbf{F}\rangle$. If we remove the requirement that $\mathbf{F}(\mathbf{x})$ is a gradient then (1.4) is Jensen's inequality which holds if and only if $W(\mathbf{F})$ is convex. It turns out that if (1.4) is satisfied for one region Ω, then it is also satisfied for any other choice of Ω: quasiconvexity is independent of the shape of Ω. Alternatively, and equivalently, one could take Ω as a cube, and instead of affine boundary conditions require that $\mathbf{X}(\mathbf{x}) - \mathbf{x}\cdot\mathbf{F}_0$ satisfy periodic boundary conditions where again $\mathbf{F}_0$ can be identified with $\langle\mathbf{F}\rangle$. If the local elastic energy is not quasiconvex, "phase separation" again occurs as shown in **Figure 1.1**(b), but now the microstructure of the phases is restricted to elastic energy minimizing configurations. This time the continuity of $\mathbf{X}(\mathbf{x})$ across interfaces implies there are jump conditions: the jump in $\mathbf{F}(\mathbf{x})$ across any smooth interface must be a rank-one matrix, more precisely a matrix of the form $\mathbf{n}\otimes\mathbf{a} \equiv \mathbf{n}\mathbf{a}^T$ where $\mathbf{n}$ is the normal to the interface. In particular, by considering the possibility that the material "phase separates" into a stratified material with layers perpendicular to a unit vector $\mathbf{n}$ of two phases where $\mathbf{F}(\mathbf{x})$ takes the value $\mathbf{F}_1$ in phase 1 and $\mathbf{F}_2$ in phase 2 we see that a necessary condition for quasiconvexity is rank-one convexity, meaning that (1.3) holds for all $\mathbf{F}_1$ and $\mathbf{F}_2$ such that the jump $\mathbf{F}_1 - \mathbf{F}_2$ is a rank-one matrix of the form $\mathbf{n}\otimes\mathbf{a}$. In three-dimensions Šverák (1992) proved rank-one convexity is not a sufficient condition for quasiconvexity (he showed there are microstructures with lower energy than stratified ones, i.e., layered ones, or than ones with lamination on different length scales possibly with different directions of lamination), while in two-dimensions the question is presently still open. As an example of cross-pollination, Sverak's example led to an example (see **Section 31.9** in Milton 2002) of a composite with effective properties that could not be mimicked by a hierarchical laminate of seven given materials. A material with a local elastic energy that is not quasiconvex could phase separate into other elastic energy minimizing microstructures (composites). Strictly speaking one should refer to a sequence of energy minimizing microstructures since (ignoring surface energies) the infimum of the energy might only be approached and

not be achieved by any specific microstructure. Surface energies do play an important role in the selection of energy minimizing microstructures (Kohn and Müller 1992). If there are many possible energy minimizing microstructures available, then this is a way of minimizing hysteresis (Song, Chen, Dabade, Shield, and James 2013) and associated fatigue (Chluba, Ge, Lima de Miranda, Strobel, Kienle, Quandt, and Wuttig 2015) when one cycles through phase transformations. One could argue that the elastic energy minimizing microstructures are not thermodynamically stable: that over long time scales it would be energetically preferable for cracks to develop along interfaces, thus removing the restriction that $\mathbf{X}(\mathbf{x})$ is continuous across interfaces. While this is true, one may have to wait an extremely long time, and elasticity theory typically works well for the intermediate times during which one wants answers. Another thing one should always remember is that from a physical perspective it is never energy that is minimized in any closed system (since energy is conserved), rather it is entropy (a measure of disorder) which is maximized. Thermodynamics allows one to mathematically reformulate entropy maximization principles as energy or free energy minimization principles (see, for example, Callen 1960a, in particular figures 5.1 and 5.2).

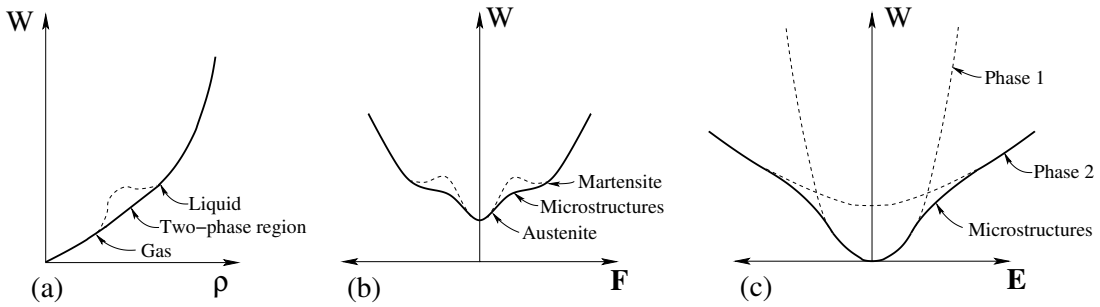


Figure 1.1: Three closely related problems, where convexification or quasiconvexification is important. For fluids, as in (a), thermodynamics tells us that the energy density W should be a convex function of the density ρ. Roughly speaking, a nonconvex portion, such as the dashed line in (a), should be replaced by a tie-line representing a mixture of gas and liquid. For elastic materials, as in (b), the elastic energy W should be a quasiconvex function of the deformation gradient $\mathbf{F} = \nabla\mathbf{X}$. If it contains a portion (the dashed line in figure (b)) which causes it to be nonquasiconvex, then it needs to be quasiconvexified. Portions where the original (microscopic) energy function differs from its quasiconvexification, correspond to microstructures. In the example shown, which is meant to be representative of the energy function of shape memory materials, these microstructures could be a mixture of Austenite and Martensite (whose crystal structures have different symmetries). Similarly in (c), suppose one is looking for two-phase microstructures which mininimize a sum W of energies and complementary energies, with some constant energy added to phase 2 to penalize it, thus acting as a Lagrange multiplier for the volume fraction. This then corresponds to quasiconvexifying an energy function which is the minimum of two quadratic "wells". The portions where the minimum of the two quadratic wells differs from its quasiconvexification again correspond to microstructures. These figures are schematic in that $\mathbf{F}$ and $\mathbf{E}$ really live in multidimensional spaces.

There is also a strong connection between the quasiconvexification of energies $W(\mathbf{F})$, or more generally functions $W(\mathbf{E})$ over fields $\mathbf{E}(\mathbf{x})$ subject to differential constraints, and the problem of bounding the set of all possible effective tensors of composites containing a given number of phases (perhaps in prescribed volume fractions). As follows from the work of Kohn (1991), Francfort and Milton (1994), Milton (1994), and Milton

and Cherkaev (1995) [see also Cherkaev and Gibiansky (1992,1993) who introduced the idea of bounding the set of possible effective tensors through sums of energies and complementary energies, and see **Chapters 30** and **31** of Milton (2002)] bounding the set of all possible effective tensors can be reformulated as a problem of quasiconvexification of an appropriate function $W(\mathbf{E})$. The key idea (Kohn 1991) is to use variational principles, expressing a sum of energies and complementary energies as a minimization over trial fields $\mathbf{E}(\mathbf{x})$ satisfying appropriate differential constraints with a prescribed average value $\langle \mathbf{E} \rangle = \mathbf{E}_0$, and rather than taking the minimum over trial fields first and then minimizing over all microstructures, one does the optimization over microstructures first (moving the phases to where they minimize the sum of energies and complementary energies for a fixed choice of the trial field $\mathbf{E}(\mathbf{x})$, which is a trivial problem) and then one minimizes over $\mathbf{E}(\mathbf{x})$ satisfying the differential constraints, with $\langle \mathbf{E} \rangle = \mathbf{E}_0$ (which is the quasiconvexification of a function which itself is a minimum of quadratic wells—one for each phase). This is illustrated in **Figure 1.1**(c).

1.2 Some linear static and quasistatic equations of physics

Consider the various static, or stationary, linear equations of physics which are relevant to calculating the effective moduli of composites. (By stationary we mean that the fields do not change with time, even though they describe a dynamic process such as the flow of electrons, heat, particles, or water). One set of basic equations are those for D.C. electrical conductivity. Let $V(\mathbf{x})$ be the electrical potential, $\mathbf{e}(\mathbf{x})$ the electric field, $\mathbf{j}(\mathbf{x})$ the electrical current, and $\boldsymbol{\sigma}(\mathbf{x})$ the matrix-valued local conductivity (which is a second order tensor). Then the conductivity equations take the form:

$$\mathbf{j}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})\mathbf{e}(\mathbf{x}), \qquad \nabla \cdot \mathbf{j} = 0, \qquad \mathbf{e} = -\nabla V. \tag{1.5}$$

The first equation is the constitutive equation. It links together the fields (in this case the current and electric fields) and contains all the information about the material properties (in this case through the conductivity tensor). The remaining two equations are the differential constraints. Strictly speaking they are differential equations, but they do constrain the functions $\mathbf{j}(\mathbf{x})$ and $\mathbf{e}(\mathbf{x})$. These differential constraints do not depend on the material under consideration (though $\boldsymbol{\sigma}(\mathbf{x})$ could be zero in a region of the material that is dielectric, and consequently $\mathbf{e}(\mathbf{x})$ and $V(\mathbf{x})$ may be undefined there unless one takes limits).

The same equations arise in many different physical problems: dielectrics, magnetism, thermal conduction, diffusion, flow in porous media, and antiplane elasticity. In each of these contexts the vector fields $\mathbf{j}(\mathbf{x})$ and $\mathbf{e}(\mathbf{x})$ and the tensor $\boldsymbol{\sigma}(\mathbf{x})$ entering the constitutive relation have the interpretations given in **Table 1.1** which is reproduced from page 19 of Milton (2002), which in turn is adapted from one of Batchelor (1974). For example, in dielectric materials the electric displacement field $\mathbf{d}(\mathbf{x})$, electric field $\mathbf{e}(\mathbf{x})$ and electrical potential $V(\mathbf{x})$ satisfy

$$\mathbf{d}(\mathbf{x}) = \boldsymbol{\varepsilon}(\mathbf{x})\mathbf{e}(\mathbf{x}), \qquad \nabla \cdot \mathbf{d} = 0, \qquad \mathbf{e} = -\nabla V, \tag{1.6}$$

where the dielectric tensor $\boldsymbol{\varepsilon}(\mathbf{x})$ is in general matrix-valued (unless the medium is isotropic or has cubic symmetry).

Notice that

$$\mathbf{j} \cdot \mathbf{e} = -\mathbf{j} \cdot \nabla V = \nabla \cdot \mathbf{Q}, \quad \text{where } \mathbf{Q}(\mathbf{x}) = -V(\mathbf{x})\mathbf{j}(\mathbf{x}), \tag{1.7}$$

which allows one to integrate the electrical power absorbed over a region and express it in terms of the flux of power into that region, giving the key identity

$$\int_\Omega \mathbf{j} \cdot \mathbf{e} = \int_{\partial\Omega} -\mathbf{n} \cdot [V(\mathbf{x})\mathbf{j}(\mathbf{x})], \tag{1.8}$$

Table 1.1: Equivalences with the conductivity problem.

Problem	$\mathbf{j}$	$\mathbf{e}$	$\boldsymbol{\sigma}$
Electrical conduction	Electrical current $\mathbf{j}$	Electric field $\mathbf{e}$	Electrical conductivity $\boldsymbol{\sigma}$
Dielectrics	Displacement field $\mathbf{d}$	Electric field $\mathbf{e}$	Electric permittivity $\boldsymbol{\varepsilon}$
Magnetism	Magnetic induction $\mathbf{b}$	Magnetic field $\mathbf{h}$	Magnetic permeability $\boldsymbol{\mu}$
Thermal conduction	Heat current $\mathbf{q}$	Temperature gradient $-\nabla T$	Thermal conductivity $\boldsymbol{\kappa}$
Diffusion	Particle current	Concentration gradient $-\nabla c$	Diffusivity $\mathbf{D}$
Flow in porous media	Weighted fluid velocity $\eta_\mu \mathbf{v}$	Pressure gradient ∇P	Fluid permeability $\mathbf{k}$
Antiplane elasticity	Stress Vector (τ_{13}, τ_{23})	Vertical Displacement gradient ∇u_3	Shear matrix $\boldsymbol{\mu}$

where $\mathbf{n}$ is the outwards normal to the surface $\partial\Omega$.

Another important equation is that for linear elasticity. Before discussing it, let us briefly review the meaning of tensors, for simplicity using Cartesian coordinates. [Those readers familiar with tensors can skip this paragraph, while those readers wanting a more complete introduction to tensors should see, for example, Lawden (1982), **Chapter 1** of Jog (2002) or **Chapter 21** of Riley, Hobson, and Bence (2002)]. The simplest tensor is a scalar which remains unchanged under rotation (such as the temperature or pressure). The next simplest is a vector $\mathbf{v}$ which has Cartesian elements v_j that under a rotation $\mathbf{R}$ transform to a new vector $\mathbf{v}'$ with elements

$$v_i' = R_{ij} v_j, \tag{1.9}$$

where the matrix with elements R_{ij} represents the rotation (satisfying $R_{ij}R_{ik} = \delta_{jk}$ where the δ_{jk} are elements of the identity tensor taking the value 1 if $i = j$ and zero otherwise). Here, and subsequently, we assume the rotation is a proper rotation, with determinant 1, thus excluding reflections: otherwise the vector may have an additional sign flip under reflections, depending on whether the vector is a pseudovector (axial vector) or not. Note that a 360° degree rotation leaves a vector invariant.

From vectors it is natural to move to second-order tensors and the easiest example of one is a linear map from vectors to vectors, like the conductivity or dielectric tensor. A second-order tensor $\mathbf{A}$ has Cartesian elements which are represented by a matrix A_{ij}. Under the rotation $\mathbf{R}$, $\mathbf{A}$ transforms to $\mathbf{A}'$ with elements

$$A_{ij}' = R_{ik} R_{jm} A_{km}, \tag{1.10}$$

implying $\mathbf{A}' = \mathbf{R}\mathbf{A}\mathbf{R}^T$. A second-order tensor needs not be associated with a map from vectors to vectors, so long as its elements satisfy these transformation rules — an example is the tensor of thermal expansion $\boldsymbol{\alpha}$.

The general rule is that an n-th order tensor $\boldsymbol{\mathcal{A}}$ has Cartesian elements $A_{abcdef\ldots}$ each indexed by n numbers $a, b, c, d, e, f, \ldots$ taking values from 1 to d in a d-dimensional space, and under rotation transforms to $\boldsymbol{\mathcal{A}}'$ with elements

$$A'_{\alpha\beta\gamma\delta\epsilon\phi\ldots} = R_{\alpha a}R_{\beta b}R_{\gamma c}R_{\delta d}R_{\epsilon e}R_{\phi f}\ldots A_{abcdef\ldots}. \tag{1.11}$$

This definition is a little misleading in that it fails to convey the fact that tensors are geometric objects, not a collection of elements which are merely used to represent the geometric object. But the definition does correctly convey the fact that tensors of different orders have different properties under rotation. A third-order tensor could represent a mapping from first to second-order tensors, or from second to first-order tensors, or needs not be associated with any map at all. By an abuse of notation we will also call a tensor, a matrix of tensors that may enter the constitutive law and couple fields together, even when the tensors in the matrix have different orders. This classification leaves out things (which we don't consider in this book) that behave under rotation in strange ways like the spin of an electron, which requires a 720° rotation to return to itself, and is thus essentially a "half-order" tensor. I like the example of Robert Palais, who simplified a related example of Misner, Thorne, and Wheeler (1973): see their **Figure 41.6**. If you stand on a belt and rotate the buckle by 720°, you can easily manipulate it (without twisting the buckle) so that the belt straightens out to the original belt before the 720° rotation- but you cannot do this with 360° rotations. Try it yourself! (See also `http://www.math.utah.edu/~palais/links.html`).

In linear elasticity the relevant fields are the stress field which is a second-order tensor $\boldsymbol{\sigma}(\mathbf{x})$ (not to be confused with the conductivity tensor which has the same symbol) measuring the local tension or compression forces in the material (if one makes a small slit in the material perpendicular to a vector $\mathbf{n}$ then one has to apply the opposing forces $\boldsymbol{\sigma}\mathbf{n}$ and $-\boldsymbol{\sigma}\mathbf{n}$ per unit area to the two surfaces of the slit to restore the materials to its state before the slit was made: since we are free to choose the direction of $\mathbf{n}$, such experiments allow us to measure $\boldsymbol{\sigma}(\mathbf{x})$), the strain field $\boldsymbol{\epsilon}(\mathbf{x})$ measuring the local stretching, also a second-order tensor, which is the symmetrized gradient of the displacement vector field $\mathbf{u}(\mathbf{x})$ that measures the displacement of the body relative to its original stress free state (roughly speaking, neglecting thermal vibrations, an atom at $\mathbf{x}$ gets displaced to $\mathbf{x}' = \mathbf{x}+\mathbf{u}(\mathbf{x})$, where $\mathbf{u}(\mathbf{x})$ is in some sense small). A beautiful way to see strains in a transparent material is to place it between crossed polarizers (oriented at 90° with respect to each other so ordinarily no light would pass through): the strain causes the plane of polarization of the light to rotate which results in a twisting rainbow of colors when one looks through the crossed polarizers. The relation between stress and strain involves material parameters which are contained in $\boldsymbol{\mathcal{C}}(\mathbf{x})$, the fourth-order elasticity tensor. Then ignoring body forces (such as gravitation) the equations take the form

$$\boldsymbol{\sigma}(\mathbf{x}) = \boldsymbol{\mathcal{C}}(\mathbf{x})\boldsymbol{\epsilon}(\mathbf{x}), \quad \nabla\cdot\boldsymbol{\sigma} = 0, \quad \boldsymbol{\epsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]/2. \tag{1.12}$$

Again we have

$$\boldsymbol{\sigma} : \boldsymbol{\epsilon} = \boldsymbol{\sigma} : \nabla\mathbf{u} = \nabla\cdot\mathbf{Q}, \quad \text{where } \mathbf{Q}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}), \tag{1.13}$$

where the first identity follows from the symmetry of the stress field. Here the scalar product $\mathbf{A} : \mathbf{B}$ of two second-order tensors $\mathbf{A}$ and $\mathbf{B}$, with elements A_{ij} and B_{ij} is defined to be $\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) = A_{ij}B_{ij}$. (The two dots in : refer to a double contraction of indices). This leads to the key identity

$$\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\epsilon} = \int_{\partial\Omega} \mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}). \tag{1.14}$$

One can get materials where there is a coupling between electricity, magnetism, and elasticity. As before the electric potential, electric field, stress field, strain field and elastic displacement field are $V(\mathbf{x})$, $\mathbf{e}(\mathbf{x})$,

$\boldsymbol{\sigma}(\mathbf{x})$, $\boldsymbol{\epsilon}(\mathbf{x})$ and $\mathbf{u}(\mathbf{x})$. Additionally, we let $\mathbf{d}(\mathbf{x})$ be the electric displacement field, we let $\mathbf{b}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ be the magnetic induction, and magnetic field, and we let $\psi(\mathbf{x})$ be the magnetic scalar potential (assuming there are no free currents). Then the constitutive equation takes the form

$$\begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \mathbf{d}(\mathbf{x}) \\ \mathbf{b}(\mathbf{x}) \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{\mathcal{S}}(\mathbf{x}) & \boldsymbol{\mathcal{D}}(\mathbf{x}) & \boldsymbol{\mathcal{Q}}(\mathbf{x}) \\ \boldsymbol{\mathcal{D}}^T(\mathbf{x}) & \boldsymbol{\varepsilon}(\mathbf{x}) & \boldsymbol{\beta}(\mathbf{x}) \\ \boldsymbol{\mathcal{Q}}^T(\mathbf{x}) & \boldsymbol{\beta}^T(\mathbf{x}) & \boldsymbol{\mu}(\mathbf{x}) \end{pmatrix}}_{\mathbf{L}(\mathbf{x})} \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \mathbf{e}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}, \tag{1.15}$$

where the elements of the local tensor $\mathbf{L}(\mathbf{x})$ are material parameters, whose meaning arises from the constitutive equation: for example, the compliance tensor $\boldsymbol{\mathcal{S}}(\mathbf{x})$ gives the strain $\boldsymbol{\epsilon}(\mathbf{x})$ in terms of the stress $\boldsymbol{\sigma}(\mathbf{x})$, when $\mathbf{h}(\mathbf{x})$ and $\mathbf{e}(\mathbf{x})$ are zero; $\boldsymbol{\mathcal{Q}}(\mathbf{x})$ gives the strain $\boldsymbol{\epsilon}(\mathbf{x})$ in terms of $\mathbf{h}(\mathbf{x})$ when the stress $\boldsymbol{\sigma}(\mathbf{x})$ and electric field $\mathbf{e}(\mathbf{x})$ are zero. The fields satisfy the differential constraints

$$\begin{aligned} \boldsymbol{\epsilon} &= [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]/2, \quad \nabla\cdot\boldsymbol{\sigma} = 0; \\ \nabla\cdot\mathbf{d} &= 0, \quad \mathbf{e} = -\nabla V; \\ \nabla\cdot\mathbf{b} &= 0, \quad \mathbf{h} = -\nabla\psi. \end{aligned} \tag{1.16}$$

When $\boldsymbol{\mathcal{Q}}(\mathbf{x}) = \boldsymbol{\beta}(\mathbf{x}) = 0$, one has the equations of static piezoelectricity, while when $\boldsymbol{\mathcal{D}}(\mathbf{x}) = \boldsymbol{\beta}(\mathbf{x}) = 0$, one has the equations of static magnetostriction.

Again we have

$$\begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \mathbf{d}(\mathbf{x}) \\ \mathbf{b}(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \mathbf{e}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix} = \nabla\mathbf{u}(\mathbf{x}) : \boldsymbol{\sigma}(\mathbf{x}) - \nabla V(\mathbf{x})\cdot\mathbf{d}(\mathbf{x}) - \nabla\psi(\mathbf{x})\cdot\mathbf{b}(\mathbf{x}) = \nabla\cdot\mathbf{Q}(\mathbf{x}), \tag{1.17}$$

where

$$\mathbf{Q}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}) - V(\mathbf{x})\mathbf{d}(\mathbf{x}) - \psi(\mathbf{x})\mathbf{b}(\mathbf{x}), \tag{1.18}$$

which gives the key identity

$$\int_\Omega \begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \mathbf{d}(\mathbf{x}) \\ \mathbf{b}(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \mathbf{e}(\mathbf{x}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix} = \int_{\partial\Omega} \mathbf{n}\cdot[\boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}) - V(\mathbf{x})\mathbf{d}(\mathbf{x}) - \psi(\mathbf{x})\mathbf{b}(\mathbf{x})]. \tag{1.19}$$

Some equations that do not at first seem to fit in this scheme, in fact do when the full equations are considered. Consider the Duhamel–Neumann equations of linear thermoelasticity (classic references include Boley and Weiner 1997; Nowacki 1986). We let $\theta = T - T_0$ be the change in temperature T measured from some constant base temperature T_0, and we let $\boldsymbol{\alpha}(\mathbf{x})$ be the symmetric second-order tensor of thermal expansion. Let $\boldsymbol{\mathcal{S}}(\mathbf{x})$ be the fourth-order compliance tensor, which is the inverse of the elasticity tensor $\boldsymbol{\mathcal{C}}(\mathbf{x})$ when $\theta = 0$. Then with $\boldsymbol{\epsilon}(\mathbf{x})$ and $\boldsymbol{\sigma}(\mathbf{x})$ representing the strain and stress as before, the equations take the form:

$$\boldsymbol{\epsilon}(\mathbf{x}) = \boldsymbol{\mathcal{S}}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{x}) + \boldsymbol{\alpha}(\mathbf{x})\theta. \tag{1.20}$$

This equation of thermal expansion is insufficient to describe the total thermoelastic state of the composite. One also needs to introduce $\varsigma(\mathbf{x})$ which is the increase in entropy per unit volume over the entropy of the

state where $\boldsymbol{\sigma} = \theta = 0$, and $c(\mathbf{x})$ the specific heat per unit volume at constant stress. This specific heat measures the amount of additional heat energy or, more precisely, entropy that is stored in the material when the temperature is increased at constant stress. Then, within a linear theory the complete description is provided (Chandrasekharaiah 1986) by the equations

$$\begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \varsigma(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mathcal{S}}(\mathbf{x}) & \boldsymbol{\alpha}(\mathbf{x}) \\ \boldsymbol{\alpha}(\mathbf{x}) & c(\mathbf{x})/T_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \theta \end{pmatrix}, \quad \text{with} \quad \nabla \cdot \boldsymbol{\sigma} = 0, \quad \boldsymbol{\epsilon} = [\nabla \mathbf{u} + (\nabla \mathbf{u})^T]/2, \tag{1.21}$$

where the second line of this matrix equation is interpreted to mean $\varsigma = \boldsymbol{\alpha} : \boldsymbol{\sigma} + c\theta/T_0$. Now $\varsigma(\mathbf{x})$ is not subject to any differential constraints, so we can write $\varsigma(\mathbf{x}) = \nabla \cdot \mathbf{r}(\mathbf{x})$ for some nonuniquely defined vector potential $\mathbf{r}(\mathbf{x})$. For example, we could let $\mathbf{r}(\mathbf{x}) = \nabla\phi(\mathbf{x})$ and solve Poisson's equation $\nabla^2\phi(\mathbf{x}) = \varsigma(\mathbf{x})$ subject to some boundary conditions to obtain $\phi(\mathbf{x})$ and hence $\mathbf{r}(\mathbf{x})$. So we have

$$\begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \varsigma(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \theta \end{pmatrix} = \nabla \mathbf{u}(\mathbf{x}) \cdot \boldsymbol{\sigma}(\mathbf{x}) + \theta \nabla \cdot \mathbf{r}(\mathbf{x}) = \nabla \cdot \mathbf{Q}(\mathbf{x}), \tag{1.22}$$

where

$$\mathbf{Q}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}) + \theta \mathbf{r}(\mathbf{x}), \tag{1.23}$$

which gives the key identity

$$\int_\Omega \begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \varsigma(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma}(\mathbf{x}) \\ \theta \end{pmatrix} = \int_{\partial\Omega} \mathbf{n} \cdot [\boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x}) + \theta \mathbf{r}(\mathbf{x})]. \tag{1.24}$$

One caveat is that $\mathbf{r}(\mathbf{x})$ is not directly measurable from experiments, and so the key-identity in this case has less utility.

We remark that these equations of static thermoelasticity are mathematically the same as the static equations of poroelasticity (Biot 1962, Norris 1992). Also the equations (1.20) are appropriate if there is swelling due to humidity (Schulgasser 1989) or if the material is prestressed during manufacture (in which case θ cannot be varied, and can arbitrarily be set to 1).

In quasistatics, for dielectrics or conducting materials, for viscoelastic materials, or for thermoviscoelastic materials when the applied fields (including the spatially constant temperature for thermoviscoelasticity) vary with time the constitutive laws in (1.6) and in (1.20) get replaced by convolutions,

$$\begin{aligned} \mathbf{d}(\mathbf{x}, t) &= \boldsymbol{\varepsilon}(\mathbf{x}, t) * \mathbf{e}(\mathbf{x}, t) \equiv \int_{-\infty}^{t} \boldsymbol{\varepsilon}(\mathbf{x}, t - t')\mathbf{e}(\mathbf{x}, t')\, dt', \\ \boldsymbol{\epsilon}(\mathbf{x}, t) &= \boldsymbol{\mathcal{S}}(\mathbf{x}, t) * \boldsymbol{\sigma}(\mathbf{x}, t) + \boldsymbol{\alpha}(\mathbf{x}, t) * \theta(t) \\ &\equiv \int_{-\infty}^{t} \boldsymbol{\mathcal{S}}(\mathbf{x}, t - t')\boldsymbol{\sigma}(\mathbf{x}, t') + \boldsymbol{\alpha}(\mathbf{x}, t - t')\theta(t')\, dt', \end{aligned} \tag{1.25}$$

where in defining the convolution say $\boldsymbol{\varepsilon}(\mathbf{x}, t) * \mathbf{e}(\mathbf{x}, t)$, the notation means a convolution of the function $\boldsymbol{\varepsilon}(\mathbf{x}, \cdot)$ of time up to time t with the function $\mathbf{e}(\mathbf{x}, \cdot)$ of time up to time t, as written in each final expression. These convolutions arise because the electrons (which cause the current or displacement field) take a while to respond to changes in the electric field, while the displacement $\mathbf{u}(\mathbf{x})$ takes a while to respond to changes in forces (stresses) and temperature variations.

If the variation with time is sufficiently slow (so that the body, or unit cell of periodicity under consideration is much smaller than the wavelength) then we can apply the quasistatic approximation, which means keeping

the differential constraints on the fields the same. The constitutive laws (1.20) become local in frequency space,

$$\widehat{\mathbf{d}}(\mathbf{x},\omega) = \widehat{\boldsymbol{\varepsilon}}(\mathbf{x},\omega)\widehat{\mathbf{e}}(\mathbf{x},\omega), \quad \widehat{\boldsymbol{\epsilon}}(\mathbf{x},\omega) = \widehat{\boldsymbol{\mathcal{S}}}(\mathbf{x},\omega)\widehat{\boldsymbol{\sigma}}(\mathbf{x},\omega) + \widehat{\boldsymbol{\alpha}}(\mathbf{x},\omega)\widehat{\theta}(\omega), \tag{1.26}$$

where

$$\widehat{f}(\mathbf{x},\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(\mathbf{x},t)\, dt \tag{1.27}$$

denotes the Fourier transform with respect to time of a function $f(\mathbf{x},t)$, which has the inverse transform

$$f(\mathbf{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\widehat{f}(\mathbf{x},\omega)\, d\omega. \tag{1.28}$$

The dependence of the moduli $\widehat{\boldsymbol{\varepsilon}}(\mathbf{x},\omega)$ or $\widehat{\boldsymbol{\mathcal{S}}}(\mathbf{x},\omega)$ on the frequency ω is often quite complicated in real materials, though may sometimes be well approximated by simple models in certain frequency regimes. As we will see later in Section 1.6 the imaginary parts of the tensors $\widehat{\boldsymbol{\varepsilon}}(\mathbf{x},\omega)$ and $\widehat{\boldsymbol{\mathcal{S}}}(\mathbf{x},\omega)$ are associated with energy loss (to heat) in the material: the material is called "lossy" or "viscoelastic" if these are nonzero. In general the thermal expansion tensor $\widehat{\boldsymbol{\alpha}}(\mathbf{x},\omega)$ can be frequency dependent and complex [even in composites when the constituent materials have thermal expansion coefficients which are not frequency dependent: see Berryman 2009]. At fixed frequency one can think of the real parts of the fields $e^{-i\omega t}\widehat{\mathbf{d}}(\mathbf{x},\omega)$, $e^{-i\omega t}\widehat{\mathbf{e}}(\mathbf{x},\omega)$, $e^{-i\omega t}\widehat{\boldsymbol{\epsilon}}(\mathbf{x},\omega)$, $e^{-i\omega t}\widehat{\boldsymbol{\sigma}}(\mathbf{x},\omega)$, and $e^{-i\omega t}\widehat{\theta}(\omega)$, as representing the physical displacement field, electric field, strain field, stress field, and temperature field. Because the quasistatic equations are exactly the same as the static equations, except that the fields and moduli are complex, we will drop the "hats" from the symbols.

Note that the thermoviscoelastic quasistatic equations ignore mechanical source terms of heat which are quadratic in the strain rate of deformation (Francfort and Suquet 1986). These may be important in some solids, but then the equations become nonlinear.

1.3 The canonical forms and their key identities

We have seen in the previous section that for statics, or for steady-state problems such as electrical or thermal conductivity where there is a flow of electrons or heat but the fields do not change with time, the constitutive equation takes the canonical form

$$\mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})\mathbf{E}(\mathbf{x}), \tag{1.29}$$

where the Hermitian part of the tensor $\mathbf{L}(\mathbf{x})$ is positive definite on an appropriate space (e.g., for elasticity it is positive definite on the space of symmetric matrices, not all matrices which include the antisymmetric ones which lie in the null-space of $\mathbf{L}(\mathbf{x})$). [A matrix $\mathbf{B}$ is Hermitian if $\mathbf{B} = \overline{\mathbf{B}}^T$, i.e., it is the complex conjugate of its transpose. It is positive definite if $\overline{\mathbf{v}}\cdot\mathbf{B}\mathbf{v} > 0$ for all nonzero vectors $\mathbf{v}$ in the space on which $\mathbf{B}$ acts, and positive semidefinite if $\overline{\mathbf{v}}\cdot\mathbf{B}\mathbf{v} \geq 0$, for all vectors $\mathbf{v}$ in the space on which $\mathbf{B}$ acts. A matrix $\mathbf{C}$ is anti-Hermitian if $\mathbf{C} = -\overline{\mathbf{C}}^T$. The Hermitian part of a matrix $\mathbf{L}$ is $(\mathbf{L} + \overline{\mathbf{L}}^T)/2$ and it is such that the difference between $\mathbf{L}$ and it, $(\mathbf{L} - \overline{\mathbf{L}}^T)/2$, is anti-Hermitian.] The differential constraints on the fields are such that with an appropriate definition of the dot product,

$$\mathbf{J}(\mathbf{x})\cdot\mathbf{E}(\mathbf{x}) = \nabla\cdot\mathbf{Q}(\mathbf{x}), \tag{1.30}$$

for some "supercurrent" $\mathbf{Q}(\mathbf{x})$, and straightforward integration by parts, leads to the key identity that

$$\int_\Omega \mathbf{J}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x}) = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{Q}(\mathbf{x}). \tag{1.31}$$

This structure arises quite naturally. It is a natural consequence of "energy or power" minimizing variational principles that these key identities hold for static equations. The identities are also key to the mathematical basis of homogenization theory (Bensoussan, Lions, and Papanicolaou 1978; Kozlov 1978; Papanicolaou and Varadhan 1982; Bakhvalov and Panasenko 1989; Golden and Papanicolaou 1983; Zhikov, Kozlov, and Oleinik 1994), and are essential components of the "div-curl" lemma and the method of compensated compactness, which underlie the homogenization theory of Tartar and Murat (Tartar, 1975, 1979a, 1979b, 2009; Murat 1978, 1981, 1987; Murat and Tartar 1985)

However in the wider context of wave equations, it is not generally clear that there should be such key identities. It is the aim of this chapter to discover them, leaving aside the physical question as to why they exist.

For quasistatics, at fixed frequency (sufficiently low that the wavelength associated with the full wave equations, which give rise to the quasistatic equations, is much larger than the size of the body), the constitutive law takes again the form (1.29) only now the fields $\mathbf{E}(\mathbf{x})$, $\mathbf{J}(\mathbf{x})$ and the tensor $\mathbf{L}(\mathbf{x})$ are complex (and generally frequency dependent). The real and imaginary parts of the fields satisfy the same differential constraints as for the static or steady-state equations. Consequently (1.30) and the key identity (1.31) still hold. Note that the dot product does not involve any complex conjugation: if $\mathbf{E}(\mathbf{x}) = \mathbf{E}'(\mathbf{x}) + i\mathbf{E}''(\mathbf{x})$ and $\mathbf{J}(\mathbf{x}) = \mathbf{J}'(\mathbf{x}) + i\mathbf{J}''(\mathbf{x})$, where $\mathbf{E}'(\mathbf{x})$, $\mathbf{E}''(\mathbf{x})$, $\mathbf{J}'(\mathbf{x})$, and $\mathbf{J}''(\mathbf{x})$ are real vectors, then

$$\mathbf{J} \cdot \mathbf{E} = [\mathbf{J}' + i\mathbf{J}''] \cdot [\mathbf{E}' + i\mathbf{E}''] = \mathbf{J}' \cdot \mathbf{E}' - \mathbf{J}'' \cdot \mathbf{E}'' + i(\mathbf{J}' \cdot \mathbf{E}'' + \mathbf{J}'' \cdot \mathbf{E}'). \tag{1.32}$$

This is the standard dot product for bivectors, which are complex valued vectors [see, for example, equation (2.1.6) in Boulanger and Hayes 1993]. With this definition, $\mathbf{E} \cdot \mathbf{E}$ is not necessarily real and $\mathbf{E} \cdot \mathbf{E} = 0$ has nonzero solutions for $\mathbf{E}$. Typically, which we assume to be the case, $\mathbf{L}(\mathbf{x})$ can be represented by a symmetric complex matrix. Thus these quasistatic problems involve complex symmetric operators, as reviewed in Garcia, Prodan, and Putinar (2014). Interestingly, as illustrated by their example 2.21, it is not always easy to detect if a matrix is unitarily equivalent to a complex symmetric one.

If at a given frequency the equations are in the correct canonical form, then by definition there exists a range of angles θ (possibly just one angle) such that $\mathrm{Im}[e^{i\theta}\mathbf{L}(\mathbf{x})]$ is positive semidefinite for all $\mathbf{x}$. Introducing

$$\widetilde{\mathbf{J}}(\mathbf{x}) = e^{i\theta}\mathbf{J}(\mathbf{x}), \quad \widetilde{\mathbf{E}}(\mathbf{x}) = \mathbf{E}(\mathbf{x}), \quad \widetilde{\mathbf{L}}(\mathbf{x}) = e^{i\theta}\mathbf{L}(\mathbf{x}), \quad \widetilde{\mathbf{Q}}(\mathbf{x}) = e^{i\theta}\mathbf{Q}(\mathbf{x}), \tag{1.33}$$

where θ is independent of $\mathbf{x}$, we see that the new fields and tensors still satisfy the same constitutive law and differential constraints. Dropping the tilde's we see that we can assume, without loss of generality, that $\mathrm{Im}[\mathbf{L}(\mathbf{x})]$ is positive semidefinite. When $\mathrm{Im}[\mathbf{L}(\mathbf{x})]$ is positive definite, rather than just positive semidefinite, one can then, for instance, apply the minimization variational principles of Cherkaev and Gibiansky (1994).

For full wave equations at fixed frequency we will see in Sections 1.7 and 1.8 that the constitutive relation takes the form

$$\mathcal{G} = \mathbf{Z}\mathcal{F}, \tag{1.34}$$

where $\mathcal{F} = \mathcal{F}(\mathbf{x})$, $\mathcal{G} = \mathcal{G}(\mathbf{x})$, and $\mathbf{Z} = \mathbf{Z}(\mathbf{x})$ are generally complex and the differential constraints are such that with an appropriate definition of the dot product,

$$\mathcal{G}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}) = \nabla \cdot \mathbf{Q}(\mathbf{x}), \tag{1.35}$$

for some supercurrent $\mathbf{Q}(\mathbf{x})$, implying the key identity

$$\int_\Omega \mathcal{G}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}) = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{Q}(\mathbf{x}). \tag{1.36}$$

If the equations are in the correct canonical form the tensor $\mathbf{Z}(\mathbf{x})/i$ has a positive semidefinite Hermitian part when the materials are lossy (absorb energy), and when the frequency ω is complex with positive imaginary part (so the fields increase exponentially with time) and $\mathbf{Z}(\mathbf{x})$ is Hermitian (though not positive semidefinite) when there is no loss and the frequency is real. When the materials are lossy or the frequency is complex, one can look for real angles θ such that the Hermitian part of $e^{i\theta}\mathbf{Z}(\mathbf{x})$ is positive definite. One can then, for example, apply the minimization variational principles of Milton, Seppecher, and Bouchitté (2009) and Milton and Willis (2010) which generalize those of Cherkaev and Gibiansky (1994). A distinction between these equations and those of quasistatics is that while in quasistatics one can find periodic solutions for $\mathbf{E}(\mathbf{x})$ and $\mathbf{J}(\mathbf{x})$ when $\mathbf{L}(\mathbf{x})$ is periodic, for wave equations one cannot generally (in the absence of sources) find periodic solutions for $\mathcal{G}(\mathbf{x})$ and $\mathcal{F}(\mathbf{x})$ when $\mathbf{Z}(\mathbf{x})$ is periodic and the material is lossy (so the imaginary part of $\mathbf{Z}(\mathbf{x})$ is nonzero and positive semidefinite). Hence if one wants to obtain periodic solutions it makes sense to include source terms: we will see an example of this in **Chapters 12** and **13** for the Schrödinger equation.

For dynamic problems that are not time-harmonic, the constitutive law will still be (1.34) but now the fields $\mathcal{G}$ and $\mathcal{F}$, and possibly the tensor $\mathbf{Z}$, depend not just on space, but on the time t as well, i.e., they are functions of $\underset{\sim}{\mathbf{x}} = (x_1, x_2, x_3, -t)$, where generally it proves convenient to put a minus sign in front of t. With an appropriate definition of the dot product, we still require

$$\mathcal{G}(\underset{\sim}{\mathbf{x}}) \cdot \mathcal{F}(\underset{\sim}{\mathbf{x}}) = \underset{\sim}{\nabla} \cdot \mathbf{Q}(\underset{\sim}{\mathbf{x}}), \quad \text{with } \underset{\sim}{\nabla} = \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix}, \tag{1.37}$$

for some supercurrent $\mathbf{Q}(\underset{\sim}{\mathbf{x}})$, which leads to the key identity

$$\int_{\underset{\sim}{\Omega}} \mathcal{G}(\underset{\sim}{\mathbf{x}}) \cdot \mathcal{F}(\underset{\sim}{\mathbf{x}}) = \int_{\partial\underset{\sim}{\Omega}} \underset{\sim}{\mathbf{n}} \cdot \mathbf{Q}(\underset{\sim}{\mathbf{x}}), \tag{1.38}$$

where now $\underset{\sim}{\Omega}$ is a body in space–time with outward normal $\underset{\sim}{\mathbf{n}}$. For wave equations the tensor $\mathbf{Z}$ is Hermitian although for parabolic equations such as thermal conduction, diffusion, and thermoelasticity $\mathbf{Z}/i$ can have a nonzero positive semidefinite Hermitian part: see Sections 1.13 and 1.14.

Generally it takes time for the field $\mathcal{G}$ to respond to the field $\mathcal{F}$, so in many cases the constitutive law gets replaced by a convolution in time:

$$\mathcal{G} = \mathbf{Z} * \mathcal{F}, \quad \text{i.e., } \mathcal{G}(\mathbf{x}, -t) = \int_{-\infty}^{t} \mathbf{Z}(\mathbf{x}, t' - t)\mathcal{G}(\mathbf{x}, -t')dt', \tag{1.39}$$

which in Fourier space becomes a local relation

$$\mathcal{G}(\mathbf{x}, \omega) = \mathbf{Z}(\mathbf{x}, \omega)\mathcal{F}(\mathbf{x}, \omega). \tag{1.40}$$

If the equations are in the right canonical form then the $\mathbf{Z}(\mathbf{x}, \omega)/i$ will have a positive semidefinite Hermitian part in lossy materials if ω is real and positive, and negative semidefinite in lossy materials if ω is real and negative. The key-identity (1.38) still holds but it is cautioned that $\mathcal{G}(\underset{\sim}{\mathbf{x}})$ inside a space–time body depends on the field $\mathcal{F}(\underset{\sim}{\mathbf{x}})$ outside that body, unless the body boundary does not depend on the time coordinate.

We emphasize that this new canonical form of the equations is not just to make them look pretty, although it does that. More importantly it casts them in a form similar to that used in the theory of composites, enabling us to apply appropriate machinery from the theory of composites to obtain new results. The key identity is of course just integration by parts, but the challenge is to write the equations so an appropriate dot product of the fields on the left and right hand sides of the constitutive law can be integrated by parts.

Of course many materials at a sufficiently small length scale have a response which is both nonlocal in space and in time. In this case the constitutive relation will get replaced by an integral transform, in both space and time. Also if the fields are sufficiently strong, the constitutive relations should be replaced by nonlinear ones. We do not treat these generalizations in this book.

1.4 Associated key identities, conservation laws, & boundary field inequalities

In all the examples presented in this chapter, with the exception of the Dirac equation, the fields $\mathbf{E}(\mathbf{x})$, $\mathcal{F}(\mathbf{x})$ and $\mathcal{F}(\underset{\sim}{\mathbf{x}})$ appearing on the right-hand side of the constitutive equation have the property that the complex conjugate fields, $\overline{\mathbf{E}(\mathbf{x})}$, $\overline{\mathcal{F}(\mathbf{x})}$ and $\overline{\mathcal{F}(\underset{\sim}{\mathbf{x}})}$ satisfy the same differential constraints as $\mathbf{E}(\mathbf{x})$, $\mathcal{F}(\mathbf{x})$ and $\mathcal{F}(\underset{\sim}{\mathbf{x}})$: here the overline denotes complex conjugation, and $\underset{\sim}{\mathbf{x}} \equiv (x_1, x_2, x_3, -t)$. [Similarly the complex conjugate fields $\overline{\mathbf{J}(\mathbf{x})}$, $\overline{\mathcal{G}(\mathbf{x})}$ and $\overline{\mathcal{G}(\underset{\sim}{\mathbf{x}})}$ satisfy the same differential constraints as the fields $\mathbf{J}(\mathbf{x})$, $\mathcal{G}(\mathbf{x})$ and $\mathcal{G}(\underset{\sim}{\mathbf{x}})$ appearing on the left-hand side of the constitutive equation.] Consequently we have the associated key identities

$$\begin{aligned}
\int_\Omega \mathbf{J}(\mathbf{x}) \cdot \overline{\mathbf{E}(\mathbf{x})} &= \int_{\partial\Omega} \mathbf{n} \cdot \widetilde{\mathbf{Q}}(\mathbf{x}), \\
\int_\Omega \mathcal{G}(\mathbf{x}) \cdot \overline{\mathcal{F}(\mathbf{x})} &= \int_{\partial\Omega} \mathbf{n} \cdot \widetilde{\mathbf{Q}}(\mathbf{x}), \\
\int_{\underset{\sim}{\Omega}} \mathcal{G}(\underset{\sim}{\mathbf{x}}) \cdot \overline{\mathcal{F}(\underset{\sim}{\mathbf{x}})} &= \int_{\partial\underset{\sim}{\Omega}} \underset{\sim}{\mathbf{n}} \cdot \widetilde{\mathbf{Q}}(\underset{\sim}{\mathbf{x}})
\end{aligned} \tag{1.41}$$

which hold for some appropriate supercurrent $\widetilde{\mathbf{Q}}(\mathbf{x})$ or $\widetilde{\mathbf{Q}}(\underset{\sim}{\mathbf{x}})$, defined according to the problem under consideration. (In the first two identities Ω is a body in space, while in the third $\underset{\sim}{\Omega}$ is a body in space–time with outward normal $\underset{\sim}{\mathbf{n}}$). For example, for conductivity, we have $\widetilde{\mathbf{Q}}(\mathbf{x}) = -V(\mathbf{x})\overline{\mathbf{j}(\mathbf{x})}$ and the expressions for $\widetilde{\mathbf{Q}}(\mathbf{x})$ for other problems are just as easy to obtain from the corresponding expression for $\mathbf{Q}(\mathbf{x})$ given in this chapter.

If for some nonzero complex constant λ, $\lambda\mathbf{L}(\mathbf{x})$ is Hermitian, or $\lambda\mathbf{Z}(\mathbf{x})$ is Hermitian, or $\lambda\mathbf{Z}(\underset{\sim}{\mathbf{x}})$ is Hermitian, then by substituting the constitutive equation in the equation in (1.41), multiplying it by λ, and taking the imaginary part of both sides we obtain the conservation laws

$$0 = \operatorname{Im} \int_{\partial\Omega} \mathbf{n} \cdot \widetilde{\mathbf{Q}}(\mathbf{x})/\lambda, \quad 0 = \operatorname{Im} \int_{\partial\underset{\sim}{\Omega}} \underset{\sim}{\mathbf{n}} \cdot \widetilde{\mathbf{Q}}(\underset{\sim}{\mathbf{x}})/\lambda \tag{1.42}$$

as appropriate. For example, for the quasistatic dielectric problem if the dielectric constant is real and symmetric we obtain the result that

$$0 = \operatorname{Im} \int_{\partial\Omega} -\mathbf{n} \cdot [V(\mathbf{x})\overline{\mathbf{d}(\mathbf{x})}] \tag{1.43}$$

which is equivalent to conservation of energy: no power is absorbed by the lossless dielectric medium. It might happen that the tensor entering the constitutive law is Hermitian only in one phase. In that case for (1.42) to hold the integral needs to be restricted to a sub-domain of that phase.

Conservation laws have of course played an enormously important role throughout history across the sciences. One important theorem is due to Noether (1918) which makes the connection in Hamiltonian systems between conservation laws and continuous symmetries, such as invariance under time or space translation, or spatial rotation (implying respectively the laws of conservation of energy, momentum, and angular momentum). In mechanics, conservation laws (which imply the invariance of certain integrals) have played an important role in determining information about the field near a crack–tip, or other singularity, from the far field (see, for example, Eshelby 1951; Cherepanov 1967; Rice 1968; Eshelby 1970; Atkinson and Craster 1992). Mathematically, conservation laws are also connected to null-Lagrangians, which are functionals whose Euler–Lagrange equations vanish. If the fields are such that the null-Lagrangian itself vanishes everywhere in Ω, and the null-Lagrangian involves gradients of a potential up to order k, then its integral provides a conservation law: this is because any such (C^1) null-Lagrangian is a divergence (Ball, Currie, and Olver 1981). A simple example in two-spatial dimensions with a two component vector field $\mathbf{u}(\mathbf{x})$ is the null-Lagrangian $\det(\nabla\mathbf{u})$ and we have

$$\det(\nabla\mathbf{u}) = \frac{\partial u_1}{\partial x_1}\frac{\partial u_2}{\partial x_2} - \frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} = \frac{\partial}{\partial x_1}\left[u_1\frac{\partial u_2}{\partial x_2}\right] + \frac{\partial}{\partial x_2}\left[-u_1\frac{\partial u_2}{\partial x_1}\right], \tag{1.44}$$

and clearly the quantity on the right is a divergence, so its integral just depends on boundary values and there is an associated conservation law if $\det(\nabla\mathbf{u}(\mathbf{x})) = 0$ in Ω. Another simple example of a null-Lagrangian in one dimension t is an exact integrand, a function $f(t, \mathbf{u}(t), d\mathbf{u}(t)/dt)$ of the form $\partial S/\partial t + (d\mathbf{u}(t)/dt)\cdot\nabla_{\mathbf{u}}S$ derived from a function $S(t,\mathbf{u})$ whose integral just depends on the values of S at the endpoints, irrespective of the choice of the function $\mathbf{u}(t)$:

$$\int_{t_1}^{t_2} [\partial S/\partial t + (d\mathbf{u}(t)/dt)\cdot\nabla_{\mathbf{u}}S]\, dt = S(t_2, \mathbf{u}(t_2)) - S(t_1, \mathbf{u}(t_1)). \tag{1.45}$$

Exact integrands have played an important role in the calculus of variations from its infancy (see, for example, **Chapter 1** of Young 2000) and null-Lagrangians continue to play an important role, such as through the polyconvex functions introduced by Ball (1977), which are convex functions whose arguments are null-Lagrangians. Null-Lagrangians of fields which are not necessarily gradients, but are subject to other differential constraints, have been characterized by Murat (1978, 1981, 1987) [see also Pedregal (1989)].

Now take a general complex matrix $\mathbf{A} = \mathbf{A}' + i\mathbf{A}''$ and a vector $\mathbf{E} = \mathbf{E}' + i\mathbf{E}''$, where the primed quantities denote real parts and the double primed quantities denote imaginary parts. Also let $\mathbf{A}'_s$ and $\mathbf{A}''_s$ denote the symmetric parts of $\mathbf{A}'$ and $\mathbf{A}''$, and let $\mathbf{A}'_a$ and $\mathbf{A}''_a$ denote the antisymmetric parts of $\mathbf{A}'$ and $\mathbf{A}''$:

$$\mathbf{A}'_s = [\mathbf{A}' + (\mathbf{A}')^T]/2, \quad \mathbf{A}''_s = [\mathbf{A}'' + (\mathbf{A}'')^T]/2, \quad \mathbf{A}'_a = [\mathbf{A}' - (\mathbf{A}')^T]/2, \quad \mathbf{A}''_a = [\mathbf{A}'' - (\mathbf{A}'')^T]/2. \tag{1.46}$$

Then the quantity

$$\begin{aligned}
\mathrm{Im}[(\mathbf{AE})\cdot\overline{\mathbf{E}}] &= \mathrm{Im}\{[(\mathbf{A}' + i\mathbf{A}'')(\mathbf{E}' + i\mathbf{E}'')]\cdot(\mathbf{E}' - i\mathbf{E}'')\} \\
&= (\mathbf{A}''\mathbf{E}')\cdot\mathbf{E}' + (\mathbf{A}''\mathbf{E}'')\cdot\mathbf{E}'' + (\mathbf{A}'\mathbf{E}'')\cdot\mathbf{E}' - (\mathbf{A}'\mathbf{E}')\cdot\mathbf{E}'' \\
&= (\mathbf{A}''_s\mathbf{E}')\cdot\mathbf{E}' + (\mathbf{A}''_s\mathbf{E}'')\cdot\mathbf{E}'' + (\mathbf{A}'_a\mathbf{E}'')\cdot\mathbf{E}' - (\mathbf{A}'_a\mathbf{E}')\cdot\mathbf{E}'' \\
&= \begin{pmatrix}\mathbf{E}' \\ \mathbf{E}''\end{pmatrix}\cdot\begin{pmatrix}\mathbf{A}''_s & \mathbf{A}'_a \\ -\mathbf{A}'_a & \mathbf{A}''_s\end{pmatrix}\begin{pmatrix}\mathbf{E}' \\ \mathbf{E}''\end{pmatrix}
\end{aligned} \tag{1.47}$$

will be nonnegative for all vectors $\mathbf{E}$ if and only if the matrix

$$\begin{pmatrix}\mathbf{A}''_s & \mathbf{A}'_a \\ -\mathbf{A}'_a & \mathbf{A}''_s\end{pmatrix} \tag{1.48}$$

is positive semidefinite. This condition is equivalent to requiring that the Hermitian part of $\mathbf{A}/i$ is positive semi-definite. (I am grateful to Aaron Welters for making this observation.)

So suppose there is some nonzero complex constant λ such that for all $\mathbf{x} \in \Omega$, the matrix in (1.48) is positive semidefinite with $\mathbf{A} = \lambda\mathbf{L}(\mathbf{x})$, or with $\mathbf{A} = \lambda\mathbf{Z}(\mathbf{x})$, or with $\mathbf{A} = \lambda\mathbf{Z}(\underset{\sim}{\mathbf{x}})$ as appropriate to the physical problem under consideration. Then by substituting the constitutive equation in (1.41), multiplying it by λ, and taking the imaginary part of both sides we obtain the inequalities

$$0 \le \text{Im} \int_{\partial\Omega} \mathbf{n} \cdot \widetilde{\mathbf{Q}}(\mathbf{x})/\lambda, \quad 0 \le \text{Im} \int_{\partial\underset{\sim}{\Omega}} \underset{\sim}{\mathbf{n}} \cdot \widetilde{\mathbf{Q}}(\underset{\sim}{\mathbf{x}})/\lambda, \tag{1.49}$$

as appropriate. We call these *boundary field inequalities*. They generalize the idea of a conservation law, since the main thing that distinguishes them from the integral form of a conservation law is the presence of the inequality rather than an equality. For example, for the quasistatic dielectric problem if the dielectric constant is symmetric with a positive semi-definite imaginary part we obtain the result that

$$0 \le \text{Im} \int_{\partial\Omega} -\mathbf{n} \cdot [V(\mathbf{x})\overline{\mathbf{d}(\mathbf{x})}], \tag{1.50}$$

which reflects the fact that the dissipation of electrical energy into heat is nonnegative within the body, and thus there must be a net flow of electrical energy into the body.

1.5 Other boundary field equalities and inequalities

Interestingly, there are *boundary field equalities* that do not arise from conservation laws. Just as the boundary field inequality (1.49) requires one to make some assumption about the medium inside the body (namely that the matrix in (1.48) is positive semidefinite), so too do we need to make some assumptions about the medium inside the body to obtain these other boundary field equalities. An example is a coupled field problem in a locally isotropic medium where the constitutive law takes the form

$$\begin{pmatrix} \mathbf{j}_1(\mathbf{x}) \\ \mathbf{j}_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} a(\mathbf{x})\mathbf{I} & c(\mathbf{x})\mathbf{I} \\ c(\mathbf{x})\mathbf{I} & b(\mathbf{x})\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1(\mathbf{x}) \\ \mathbf{e}_2(\mathbf{x}) \end{pmatrix}, \tag{1.51}$$

in which $a(\mathbf{x})$, $b(\mathbf{x})$, and $c(\mathbf{x})$ are scalars, and the fields are subject to the differential constraints that

$$\nabla \cdot \mathbf{j}_1 = 0, \quad \nabla \cdot \mathbf{j}_2 = 0 \quad \mathbf{e}_1 = -\nabla V_1, \quad \mathbf{e}_2 = -\nabla V_2. \tag{1.52}$$

Here the fields $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$ could represent electric fields, magnetic fields, temperature gradients, or concentration gradients, and the associated fluxes $\mathbf{j}_1(\mathbf{x})$ and $\mathbf{j}_2(\mathbf{x})$ could represent electrical currents, electrical displacement currents, magnetic induction fields, energy fluxes and particle currents. A classical example is thermoelectricity, although one has to be careful how one defines the fields [for the proper formulation see Callen (1960b) or Section 2.4 of Milton (2002)]. The assumption we make is that the matrix of coefficients

$$\mathbf{M}(\mathbf{x}) = \begin{pmatrix} a(\mathbf{x}) & c(\mathbf{x}) \\ c(\mathbf{x}) & b(\mathbf{x}) \end{pmatrix}, \tag{1.53}$$

entering the constitutive law satisfies the bounds

$$\beta\mathbf{I} \ge \mathbf{M}(\mathbf{x}) \ge \alpha\mathbf{I}, \quad \text{for some } \beta > \alpha > 0, \tag{1.54}$$

and can be diagonalized by a congruence transformation, i.e., there exists a 2×2 matrix $\mathbf{W}$ independent of $\mathbf{x}$ such that

$$\mathbf{W}\mathbf{M}\mathbf{W}^T = \begin{pmatrix} a'(\mathbf{x}) & 0 \\ 0 & b'(\mathbf{x}) \end{pmatrix}. \tag{1.55}$$

In particular, the diagonalization assumption is satisfied if the body only contains two phases (in a possibly unknown configuration), so that $\mathbf{M}(\mathbf{x})$ takes just two known values $\mathbf{M}_1 > 0$ and $\mathbf{M}_2 > 0$: these can be simultaneously diagonalized by taking $\mathbf{W} = \mathbf{Q}\mathbf{M}_1^{-1/2}$ where $\mathbf{Q}$ is the rotation (satisfying $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) that diagonalizes $\mathbf{M}_1^{-1/2}\mathbf{M}_2\mathbf{M}_1^{-1/2}$. One boundary field equality says that if we apply boundary potentials of the form

$$\begin{pmatrix} V_1(\mathbf{x}) \\ V_2(\mathbf{x}) \end{pmatrix} = \mathbf{W}^T \begin{pmatrix} f(\mathbf{x}) \\ 0 \end{pmatrix} \quad \text{for } \mathbf{x} \in \partial\Omega, \tag{1.56}$$

for some choice of $f(\mathbf{x})$, then

$$W_{21}[\mathbf{n} \cdot \mathbf{j}_1(\mathbf{x})] + W_{22}[\mathbf{n} \cdot \mathbf{j}_2(\mathbf{x})] = 0, \quad \text{for all } \mathbf{x} \in \partial\Omega, \tag{1.57}$$

where W_{21} and W_{22} are the matrix elements of the second row of $\mathbf{W}$ and $\mathbf{n}$ is the outwards normal to the boundary $\partial\Omega$. This boundary field equality is easily proved and is an immediate corollary of the ideas of Straley (1981) and Milgrom and Shtrikman (1989b, 1989a) [see also the developments in Milgrom (1990, 1997) and Chen (1995, 1997) and **Chapter 6** of Milton (2002)]. Introducing new potentials and new fluxes, which are linear combinations of the old ones:

$$\begin{pmatrix} V_1'(\mathbf{x}) \\ V_2'(\mathbf{x}) \end{pmatrix} = (\mathbf{W}^T)^{-1} \begin{pmatrix} V_1(\mathbf{x}) \\ V_2(\mathbf{x}) \end{pmatrix}, \quad \begin{pmatrix} \mathbf{j}_1'(\mathbf{x}) \\ \mathbf{j}_2'(\mathbf{x}) \end{pmatrix} = \mathbf{W} \begin{pmatrix} \mathbf{j}_1(\mathbf{x}) \\ \mathbf{j}_2(\mathbf{x}) \end{pmatrix}, \tag{1.58}$$

we see that these satisfy the *uncoupled* equations:

$$\nabla \cdot \mathbf{j}_1'(\mathbf{x}) = 0, \quad \mathbf{j}_1' = -a'(\mathbf{x})\nabla V_1', \quad \nabla \cdot \mathbf{j}_2'(\mathbf{x}) = 0, \quad \mathbf{j}_2' = -b'(\mathbf{x})\nabla V_2'. \tag{1.59}$$

So clearly if $V_2'(\mathbf{x}) = 0$ on $\partial\Omega$ (which is implied by (1.56)) then $\mathbf{j}_2'(\mathbf{x}) \cdot \mathbf{n} = 0$ on $\partial\Omega$ (which implies (1.57)).

In two-dimensions other things can be said. Suppose that $c(\mathbf{x}) = 0$ and that $b(\mathbf{x}) = \alpha^2/a(\mathbf{x})$ where $\alpha > 0$ is a real constant. The key observation, following the ideas of Keller (1964), Dykhne (1970), and Mendelson (1975), is that the two-dimensional fields

$$\tilde{\mathbf{e}} = \alpha^{-1}\mathbf{R}_\perp \mathbf{j}_1, \quad \tilde{\mathbf{j}} = \alpha \mathbf{R}_\perp \mathbf{e}_1, \quad \text{where } \mathbf{R}_\perp = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.60}$$

are respectively divergence free and curl-free. (By curl-free in two-dimensions we mean that $\partial\tilde{e}_2/\partial x_1 - \partial\tilde{e}_1/\partial x_2 = 0$.) Here $\mathbf{R}_\perp$ is the matrix for a 90° rotation. Now instead of prescribing the potentials $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ at the boundary $\partial\Omega$ one can equivalently prescribe the tangential values $\mathbf{t} \cdot \mathbf{e}_1(\mathbf{x})$ and $\mathbf{t} \cdot \mathbf{e}_2(\mathbf{x})$ of the electric fields $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$, where $\mathbf{t} = \mathbf{R}_\perp \mathbf{n}$ is the vector tangential to $\partial\Omega$. By integrating $\mathbf{t} \cdot \mathbf{e}_1(\mathbf{x})$ and $\mathbf{t} \cdot \mathbf{e}_2(\mathbf{x})$ along $\partial\Omega$, from one point on $\partial\Omega$, one can recover (up to constants) $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ at the boundary $\partial\Omega$, and conversely from the tangential derivatives of $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ along $\partial\Omega$ one can obtain $\mathbf{t} \cdot \mathbf{e}_1(\mathbf{x})$ and $\mathbf{t} \cdot \mathbf{e}_2(\mathbf{x})$. Now if we prescribe $\mathbf{t} \cdot \mathbf{e}_2(\mathbf{x}) = \alpha^{-1}\mathbf{n} \cdot \mathbf{j}_1(\mathbf{x})$ the equations will be solved with $\mathbf{e}_2(\mathbf{x}) = \tilde{\mathbf{e}}$ and $\mathbf{j}_2(\mathbf{x}) = \tilde{\mathbf{j}}$. Thus we obtain the boundary field equality that on Ω

$$\mathbf{n} \cdot \mathbf{j}_2(\mathbf{x}) = -\alpha \mathbf{t} \cdot \mathbf{e}_1(\mathbf{x}) \quad \text{when} \quad \mathbf{t} \cdot \mathbf{e}_2(\mathbf{x}) = \alpha^{-1}\mathbf{n} \cdot \mathbf{j}_1(\mathbf{x}). \tag{1.61}$$

Boundary field equalities are in a sense analogous to exact (microstructure independent) relations satisfied by the effective tensors of composites. There is a very general theory of such exact relations (Grabovsky 1998; Grabovsky and Sage 1998; Grabovsky and Milton 1998; Grabovsky, Milton, and Sage 2000; Grabovsky 2004; see also **Chapter 17** in Milton (2002)).

Just as there is a connection between null-Lagrangians and conservation laws, so too is there a connection between Q_C^*-convex functions and boundary field inequalities. A Q_C^*-convex function $f(\mathbf{E}) = \overline{\mathbf{E}} \cdot \mathbf{T}\mathbf{E}$ is a quadratic function of the vector or tensor $\mathbf{E}$, involving the Hermitian matrix $\mathbf{T}$, such that its volume average is non-negative whenever $\mathbf{E} = \mathbf{E}(\mathbf{x})$ is C-periodic and satisfies the appropriate differential constraints, which we write as $\mathbf{E} \in \mathcal{E}$ where $\mathcal{E}$ is the space of C-periodic fields satisfying these differential constraints. [A more precise definition, which differs slightly from the one given in Milton (2013b), is provided later in equation (8.76)]. Using the ideas of Tartar and Murat (Murat 1978, 1981, 1987; Tartar 1979a); see also Milton (2013b); by taking Fourier transforms it is easy to obtain algebraic conditions for a quadratic function to be Q_C^*-convex: in essence it suffices to consider all C-periodic test fields $\mathbf{E}(\mathbf{x})$ that are pure sinusoidal waves in space. Convex functions are Q_C^*-convex, but the interesting applications generally come from Q_C^*-convex which are not convex. Q_C^*-convexity generalizes the notion of quasiconvexity that beginning with the work of Morrey (1952,1966) has been fundamental in proving the existence of minimizers of functionals, such as those that occur in elasticity theory (Ball 1977; Ball and James 1987): for an excellent introduction to the subject see the book of Dacorogna (2007). As shown by Milton (2013b) (see also the addendum to that paper) Q_C^*-convexity enables one to obtain inequalities of the form

$$\int_\Omega f(\mathbf{E}(\mathbf{x})) \geq f_0, \tag{1.62}$$

which holds for all fields $\mathbf{E}(\mathbf{x})$ satisfying appropriate differential constraints, which we write as $\mathbf{E} \in \mathcal{E}_\Omega$, where the constant f_0 just depends on fields at the boundary $\partial\Omega$ and Ω is a region that lies inside the unit cell of periodicity. To obtain such an inequality we look for one solution $\mathbf{E}_0(\mathbf{x})$ of the Euler–Lagrange equations:

$$\mathbf{J}_0(\mathbf{x}) = \mathbf{T}\mathbf{E}_0(\mathbf{x}), \quad \mathbf{J}_0 \in \mathcal{J}_\Omega, \quad \mathbf{E}_0 \in \mathcal{E}_\Omega. \tag{1.63}$$

where $\mathcal{J}_\Omega$ and $\mathcal{E}_\Omega$ are spaces of fields respectively satisfying differential constraints appropriate to the equations at hand, and such that the associated key identity holds: the integral over Ω of $\overline{\mathbf{E}_0} \cdot \mathbf{J}_0$, which we label f_0, can be obtained from boundary values of the superflux $\mathbf{n} \cdot \widetilde{\mathbf{Q}}(\mathbf{x})$ associated with $\mathbf{J}_0(\mathbf{x})$ and $\mathbf{E}_0(\mathbf{x})$. If $\mathbf{E}(\mathbf{x})$ is a field satisfying the same boundary conditions on Ω as $\mathbf{E}_0(\mathbf{x})$ then we define $\delta\mathbf{E}(\mathbf{x})$ to be $\mathbf{E}(\mathbf{x}) - \mathbf{E}_0(\mathbf{x})$ inside Ω and to be zero in that part of C which is outside Ω. Then we extend $\delta\mathbf{E}(\mathbf{x})$ to be C-periodic. Provided the boundary conditions on Ω, and the definitions of the spaces $\mathcal{E}$ and $\mathcal{E}_\Omega$ ensure that $\delta\mathbf{E}(\mathbf{x})$ satisfies the differential constraints appropriate to fields in $\mathcal{E}$, by the Q_C^*-convexity of f it follows that

$$0 \leq \int_C f(\delta\mathbf{E}(\mathbf{x})) = \int_\Omega f(\delta\mathbf{E}(\mathbf{x})). \tag{1.64}$$

Since f is quadratic, and $\mathbf{T}$ is Hermitian, we have

$$f(\delta\mathbf{E}(\mathbf{x})) = f(\mathbf{E}(\mathbf{x})) - f(\mathbf{E}_0(\mathbf{x})) - \overline{\delta\mathbf{E}(\mathbf{x})} \cdot \mathbf{J}_0(\mathbf{x}) - \overline{\mathbf{J}_0(\mathbf{x})} \cdot \delta\mathbf{E}(\mathbf{x}), \tag{1.65}$$

and the associated key identity implies

$$\int_\Omega \overline{\delta\mathbf{E}(\mathbf{x})} \cdot \mathbf{J}_0(\mathbf{x}) = 0, \quad \int_\Omega \overline{\mathbf{J}_0(\mathbf{x})} \cdot \delta\mathbf{E}(\mathbf{x}) = 0, \tag{1.66}$$

since the boundary fields on Ω associated with $\delta\mathbf{E}(\mathbf{x})$ vanish. So we conclude that

$$\int_\Omega f(\mathbf{E}(\mathbf{x})) \geq \int_\Omega f(\mathbf{E}_0(\mathbf{x})) = \int_\Omega \overline{\mathbf{E}_0} \cdot \mathbf{J}_0 = f_0, \tag{1.67}$$

which establishes (1.62). We emphasize that even with specified boundary conditions, there may be many fields $\mathbf{E}_0(\mathbf{x})$ satisfying (1.63).

The most interesting choices of $\mathbf{T}$ are the extremal translations (page 87 of Milton 1990; Allaire and Kohn 1994) which have the property that $f(\mathbf{E})$ loses its Q_C^*-convexity whenever a nonzero positive semidefinite matrix is subtracted from $\mathbf{T}$. For such choices of $\mathbf{T}$ one expects there to be many fields $\mathbf{E}_0(\mathbf{x})$ satisfying (1.63) that are compatible with the boundary conditions of interest. An extreme example is if we are in two-spatial dimensions, $\mathbf{u}(\mathbf{x})$ is a two-component vector field with components $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$, $\mathbf{E} = \nabla\mathbf{u}$, and $f(\mathbf{E})$ is the null-Lagrangian $\det(\mathbf{E})$ as in (1.44). Then the equations (1.63) can be expressed in the form

$$\underbrace{\begin{pmatrix}\mathbf{j}_1(\mathbf{x})\\ \mathbf{j}_2(\mathbf{x})\end{pmatrix}}_{\mathbf{J}_0(\mathbf{x})} = \underbrace{\begin{pmatrix}0 & \mathbf{R}_\perp/2\\ \mathbf{R}_\perp^T/2 & 0\end{pmatrix}}_{\mathbf{T}} \underbrace{\begin{pmatrix}\mathbf{e}_1(\mathbf{x})\\ \mathbf{e}_2(\mathbf{x})\end{pmatrix}}_{\mathbf{E}_0(\mathbf{x})}, \quad \nabla\cdot\mathbf{j}_i = 0, \quad \mathbf{e}_i = \nabla u_i,\ i = 1, 2, \tag{1.68}$$

where $\mathbf{R}_\perp$ with transpose $\mathbf{R}_\perp^T$ is the matrix for a 90° rotation given by (1.60). These equations are clearly satisfied for any choice of fields $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$, due to the fact that $\mathbf{R}_\perp$ maps curl-free fields to divergence-free fields. We remark that with extremal choices of $\mathbf{T}$ it is not clear that the left hand side of (1.62) is even bounded below as $\mathbf{E}(\mathbf{x})$ varies over those fields in $\mathcal{E}_\Omega$ satisfying desired boundary conditions. However it is bounded below (by f_0) if we can find solutions of the Euler–Lagrange equations (1.63) compatible with the desired boundary conditions. One could, for example, look for a superposition of plane-wave solutions to these Euler–Lagrange equations that satisfy the desired boundary conditions, or we could choose our boundary conditions to match those of a superposition of these plane waves.

Inequalities of the form (1.62) can be obtained by other means too. In particular, if $\mathbf{T}$ is positive semidefinite, so that $f(\mathbf{E})$ is a convex function of $\mathbf{E}$, then Jensen's inequality says that (1.62) holds for any boundary conditions with $f_0 = |\Omega| f(\langle\mathbf{E}\rangle)$ where $|\Omega|$ is the volume of Ω, and $\langle\cdot\rangle$ denotes a volume over Ω. It is frequently the case that the differential constraints on $\mathbf{E}(\mathbf{x})$ allow one to calculate $\langle\mathbf{E}\rangle$, and hence f_0 from boundary values: for instance, this is clearly true if $\mathbf{E}(\mathbf{x}) = \nabla\mathbf{u}(\mathbf{x})$ for some vector (or scalar) potential $\mathbf{u}(\mathbf{x})$.

Now, following the argument given in Section 2 of Harutyunyan and Milton (2015b), suppose, for simplicity, that $\mathbf{L}(\mathbf{x})$ is Hermitian and that we can find a constant $c > 0$ such that $\mathbf{L}(\mathbf{x}) - c\mathbf{T}$ is positive semi-definite for all $\mathbf{x} \in \Omega$. [See also the related papers of Kang, Kim, and Milton (2012), Milton and Nguyen (2012), Kang, Milton, and Wang (2014), Kang and Milton (2013), Kang, Kim, Lee, Li, and Milton (2014), and Thaler and Milton (2015) that implicitly derive boundary field inequalities for bodies containing two phases, with a known volume fraction of one phase: these are used in an inverse fashion to derive bounds on the volume fraction.] Such a constant c is easiest to find if the body contains N-phases and we know the value of $\mathbf{L}(\mathbf{x})$ inside each phase, although the phase geometry may be unknown. Then, using (1.62) and the associated key identity (1.41), we obtain the boundary field inequality,

$$0 \leq \int_\Omega \overline{\mathbf{E}(\mathbf{x})} \cdot \mathbf{L}(\mathbf{x})\mathbf{E}(\mathbf{x}) - cf(\mathbf{E}(\mathbf{x})) \leq -cf_0 + \int_{\partial\Omega} \mathbf{n}\cdot\widetilde{\mathbf{Q}}(\mathbf{x}). \tag{1.69}$$

Such boundary field inequalities, which generalize those in Section 2 of Harutyunyan and Milton (2015b), place constraints on the Dirichlet-to-Neumann map that maps the potential on the boundary $\partial\Omega$ to the associated flux through $\partial\Omega$. The inequality (1.69) is a natural extension of the translation method (also known

as the compensated compactness method) introduced by Murat and Tartar (Tartar 1979b; Murat and Tartar 1985; Tartar 1985) and independently by Lurie and Cherkaev (1982, 1984) and the associated comparison bounds (see, for example, **Chapter 24** in Milton 2002) in the theory of composites. These boundary field inequalities can be used in an inverse fashion to tell us something about the moduli inside Ω. For example, if the boundary field inequalities are violated for some value of c then this tells us that $\mathbf{L}(\mathbf{x}) - c\mathbf{T}$ cannot be positive semi-definite everywhere inside Ω, or alternatively that the equations have broken down. By taking a variety of possible $\mathbf{T}$ and boundary conditions, one can narrow down the range of possible values that $\mathbf{L}(\mathbf{x})$ takes inside the body Ω.

In some problems it may happen that we know that $\mathbf{E}(\mathbf{x})$ within Ω only takes values in a compact set $\mathcal{K}$. Then defining

$$f^+ = \max_{\mathbf{E}\in\mathcal{K}} f(\mathbf{E}), \tag{1.70}$$

we have from (1.62) the boundary field inequality

$$|\Omega| f^+ \geq f_0, \tag{1.71}$$

in which $|\Omega|$ is the volume of Ω. Here f_0 depends on the boundary fields, so if (1.71) is not satisfied for some boundary fields, then the occurrence of those boundary fields signals that $\mathbf{E}(\mathbf{x})$ has in fact taken values outside $\mathcal{K}$ somewhere within Ω. An example is polycrystalline plasticity, where the stress $\boldsymbol{\sigma}(\mathbf{x})$ must take values within the yield set appropriate to the crystal at $\mathbf{x}$. Suppose for simplicity there are only a finite number n of crystal grains in Ω, perhaps in an unknown geometry, with respective yield sets $\mathcal{K}_i$, $i = 1, 2, \ldots, n$. Then clearly within Ω,

$$\boldsymbol{\sigma}(\mathbf{x}) \in \mathcal{K}, \quad \mathcal{K} = \bigcup_{i=1}^{n} \mathcal{K}_i, \quad \nabla \cdot \boldsymbol{\sigma} = 0. \tag{1.72}$$

The associated boundary field inequalities can then be useful for determining which loads on the boundary $\partial\Omega$ will cause the material to yield. Similar ideas have been used, for example, by Kohn and Little (1998) in determining bounds on the homogenized yield set and stem from the theory of compensated compactness (Tartar 1979b) and the related translation method (see, for example, **Chapter 24** of Milton 2002 and references therein). Another example is if the body contains crystals of shape memory material. One may be interested in deformations that do not leave any residual stress. Then it is the strain field $\boldsymbol{\epsilon}(\mathbf{x})$ which lies in some set $\mathcal{K}$ and satisfies the differential constraint that $\boldsymbol{\epsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]$ for some displacement field $\mathbf{u}$. In this context similar ideas have been used, for example, by Bhattacharya and Kohn (1997) to bound the set of recoverable strains of polycrystalline shape-memory materials.

Boundary field inequalities, like conservation laws, should have many important applications in analysis. The main problem is to pick useful ones among the plethora of possibilities. In this connection, recent advances (Nesi and Rogora 2007; Milton 2013b; Harutyunyan and Milton 2015a; Harutyunyan and Milton 2015b; Harutyunyan and Milton 2016) on characterizing extremal Q_C^*-convex and quasiconvex functions may help. (For fields $\mathbf{E} = \nabla\mathbf{u}$, Q_C^*-convexity and quasiconvexity are equivalent, and the extremal ones are those that lose their quasiconvexity whenever a convex function is subtracted from it.)

1.6 Consequences of passivity

Let us suppose the frequency ω is complex, $\omega = \omega' + i\omega''$, where $\omega'' > 0$ so that the fields which are modulated by $e^{-i\omega t} = e^{\omega'' t} e^{-i\omega' t}$ grow with time. The physical electric and magnetic fields are

$$\begin{aligned} \mathbf{e}_R(\mathbf{x},t) = \mathrm{Re}[e^{-i\omega t}\mathbf{e}(\mathbf{x})] &= e^{\omega'' t}[e^{-i\omega' t}\mathbf{e}(\mathbf{x}) + e^{+i\omega' t}\overline{\mathbf{e}}(\mathbf{x})]/2, \\ \mathbf{h}_R(\mathbf{x},t) = \mathrm{Re}[e^{-i\omega t}\mathbf{h}(\mathbf{x})] &= e^{\omega'' t}[e^{-i\omega' t}\mathbf{h}(\mathbf{x}) + e^{+i\omega' t}\overline{\mathbf{h}}(\mathbf{x})]/2, \end{aligned} \tag{1.73}$$

where $\overline{a}$ is the complex conjugate of a, and the complex electric field $\mathbf{e}(\mathbf{x})$ and complex magnetic field $\mathbf{h}(\mathbf{x})$ do not depend on time. Now the total flow of energy into the body up to time T should be positive and by Poynting's theorem, this implies

$$\begin{aligned} 0 &\le \int_{-\infty}^{T}\int_{\partial\Omega} \mathbf{n}\cdot\mathbf{h}_R(\mathbf{x},t)\times\mathbf{e}_R(\mathbf{x},t) = \int_{-\infty}^{T}\int_{\partial\Omega} \mathbf{n}\cdot\{\mathrm{Re}[e^{-i\omega t}\mathbf{h}(\mathbf{x})]\times\mathrm{Re}[e^{-i\omega t}\mathbf{e}(\mathbf{x})]\} \\ &= \int_{-\infty}^{T}\int_{\partial\Omega} e^{2\omega'' t}\mathbf{n}\cdot[\mathbf{h}(\mathbf{x})\times\overline{\mathbf{e}}(\mathbf{x}) + \overline{\mathbf{h}}\times\mathbf{e}(\mathbf{x}) + e^{-2i\omega' t}\mathbf{h}(\mathbf{x})\times\mathbf{e}(\mathbf{x}) + e^{2i\omega' t}\overline{\mathbf{h}}(\mathbf{x})\times\overline{\mathbf{e}}(\mathbf{x})]/4 \\ &= \frac{e^{2\omega'' T}}{4\omega''}\int_{\partial\Omega}\mathbf{n}\cdot\{\mathrm{Re}[\mathbf{h}(\mathbf{x})\times\overline{\mathbf{e}}(\mathbf{x})]\} + e^{2\omega'' T}\,\mathrm{Re}\left\{\frac{e^{-2i\omega' T}}{4(\omega''-i\omega')}\int_{\partial\Omega}\mathbf{n}\cdot[\mathbf{h}(\mathbf{x})\times\mathbf{e}(\mathbf{x})]\right\}. \end{aligned} \tag{1.74}$$

If ω' is nonzero then the second term oscillates sinusoidally with time T, multiplied by $e^{2\omega'' T}$, and so changes sign with time. So a necessary condition for this to remain positive for all time is that

$$\begin{aligned} 0 &\le \mathrm{Re}\int_{\partial\Omega}\mathbf{n}\cdot[\mathbf{h}(\mathbf{x})\times\overline{\mathbf{e}}(\mathbf{x})] = \mathrm{Re}\int_{\Omega}\nabla\cdot[\mathbf{h}(\mathbf{x})\times\overline{\mathbf{e}}(\mathbf{x})] \\ &= \mathrm{Re}\int_{\Omega}\overline{\mathbf{e}}\cdot(\nabla\times\mathbf{h}) - \mathbf{h}\cdot(\nabla\times\overline{\mathbf{e}}) = \mathrm{Re}\int_{\Omega}[-\overline{\mathbf{e}}\cdot(i\omega\boldsymbol{\varepsilon}\mathbf{e}) - \mathbf{h}\cdot\overline{i\omega\boldsymbol{\mu}\mathbf{h}}] \\ &= \int_{\Omega}\mathbf{e}\cdot\mathrm{Im}(\omega\boldsymbol{\varepsilon})\overline{\mathbf{e}} + \mathbf{h}\cdot\mathrm{Im}(\omega\boldsymbol{\mu})\overline{\mathbf{h}}, \end{aligned} \tag{1.75}$$

where $\boldsymbol{\varepsilon}(\mathbf{x})$ and $\boldsymbol{\mu}(\mathbf{x})$ are the electric permittivity and magnetic permeability tensors and we have used the Maxwell equations,

$$\nabla\times\mathbf{h} = -i\omega\boldsymbol{\varepsilon}\mathbf{e}, \quad \nabla\times\overline{\mathbf{e}} = i\omega\boldsymbol{\mu}\mathbf{h}. \tag{1.76}$$

Clearly (1.75) will be nonnegative provided

$$\mathrm{Im}(\omega\boldsymbol{\varepsilon}(\mathbf{x},\omega)) \ge 0, \quad \mathrm{Im}(\omega\boldsymbol{\mu}(\mathbf{x},\omega)) \ge 0, \tag{1.77}$$

for all $\mathbf{x}\in\Omega$ and for all ω in the upper half plane $\mathrm{Im}\,\omega = \omega'' > 0$. Conversely, by considering a small region Ω with boundary conditions chosen so that the interior fields $\mathbf{e}$ and $\mathbf{h}$ have almost constant desired values, with either $|\mathbf{e}|\gg|\mathbf{h}|$ or $|\mathbf{e}|\gg|\mathbf{h}|$, (assuming $\boldsymbol{\varepsilon}(\mathbf{x},\omega)$ and $\boldsymbol{\mu}(\mathbf{x},\omega)$ are smooth) we see that the conditions (1.77) are not only sufficient, but also necessary.

Similar arguments can be applied to elasticity. The physical velocity, and traction at the surface of a body when the fields grow at a complex frequency $\omega = \omega' + i\omega''$ with $\omega'' > 0$ are

$$\begin{aligned} \mathbf{v}_R(\mathbf{x},t) &= \frac{\partial}{\partial t}\mathrm{Re}[e^{-i\omega t}\mathbf{u}(\mathbf{x})] = e^{\omega'' t}[-i\omega e^{-i\omega' t}\mathbf{u}(\mathbf{x}) + i\overline{\omega}e^{+i\omega' t}\overline{\mathbf{u}}(\mathbf{x})]/2, \\ \mathbf{t}_R(\mathbf{x},t) &= \mathrm{Re}[e^{-i\omega t}\mathbf{n}\cdot\boldsymbol{\sigma}] = e^{\omega'' t}[e^{-i\omega t}\mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x}) + e^{+i\omega t}\mathbf{n}\cdot\overline{\boldsymbol{\sigma}}(\mathbf{x})]/2, \end{aligned} \tag{1.78}$$

where $\mathbf{u}(\mathbf{x})$ and $\boldsymbol{\sigma}(\mathbf{x})$ are the complex displacement field and complex stress field, which are independent of time. Hence the total work done on the body up to time $t = T$, which must be positive is

$$\begin{aligned} 0 &\leq \int_{-\infty}^{T}\int_{\partial\Omega} \mathbf{v}_R(\mathbf{x},t)\cdot\mathbf{t}_R(\mathbf{x},t) \\ &= \int_{-\infty}^{T}\int_{\partial\Omega} e^{2\omega'' t}[e^{-i\omega t}\mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x}) + e^{+i\omega t}\mathbf{n}\cdot\overline{\boldsymbol{\sigma}}(\mathbf{x})][-i\omega e^{-i\omega' t}\mathbf{u}(\mathbf{x}) + i\overline{\omega}e^{+i\omega' t}\overline{\mathbf{u}}(\mathbf{x})]/4 \\ &= \frac{e^{2\omega'' T}}{4\omega''}\int_{\partial\Omega}\{\mathrm{Re}[i\mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x})\overline{\omega}\overline{\mathbf{u}}(\mathbf{x})]\} \\ &\quad -e^{2\omega'' T}\,\mathrm{Re}\left\{\frac{e^{-2i\omega' T}}{4(\omega''-i\omega')}\int_{\partial\Omega}[i\omega\mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x})\mathbf{u}(\mathbf{x})]\right\}. \end{aligned} \tag{1.79}$$

Again because the second term oscillates (unless $\omega' = 0$) a necessary condition for this to be nonnegative for all T is that

$$\begin{aligned} 0 &\leq \int_{\partial\Omega}\mathrm{Re}[i\mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x})\overline{\omega}\overline{\mathbf{u}}(\mathbf{x})] \\ &= \int_{\Omega}\mathrm{Re}[i\boldsymbol{\sigma}(\mathbf{x}) : \nabla\overline{\omega}\overline{\mathbf{u}}(\mathbf{x}) + i(\nabla\cdot\boldsymbol{\sigma}(\mathbf{x}))\cdot\overline{\omega}\overline{\mathbf{u}}(\mathbf{x})] \\ &= \omega\overline{\omega}\int_{\Omega}\boldsymbol{\epsilon}(\mathbf{x}) : \mathrm{Im}[-\boldsymbol{\mathcal{C}}/\omega]\overline{\boldsymbol{\epsilon}} + \mathbf{u}\cdot\mathrm{Im}[\omega\boldsymbol{\rho}]\overline{\mathbf{u}}, \end{aligned} \tag{1.80}$$

where $\boldsymbol{\mathcal{C}}(\mathbf{x},\omega)$ is the elasticity tensor, $\boldsymbol{\rho}$ is the density, and we have used the elastodynamic equations that

$$\boldsymbol{\sigma} = \boldsymbol{\mathcal{C}}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]/2, \quad i(\nabla\cdot\boldsymbol{\sigma}(\mathbf{x})) = \omega\mathbf{p}(\mathbf{x}) = -i\omega^2\boldsymbol{\rho}\mathbf{u}, \tag{1.81}$$

in which $\mathbf{p}(\mathbf{x})$ is the complex momentum density.

The nonnegativity of (1.80) will be ensured provided the moduli are such that

$$\mathrm{Im}(-\boldsymbol{\mathcal{C}}(\mathbf{x},\omega)/\omega) \geq 0, \quad \mathrm{Im}(\omega\boldsymbol{\rho}(\mathbf{x},\omega)) \geq 0, \tag{1.82}$$

hold for all $\mathbf{x} \in \Omega$ and for all ω in the upper half plane $\mathrm{Im}\,\omega = \omega'' > 0$, and conversely by considering small bodies with almost constant fields $\boldsymbol{\sigma}(\mathbf{x})$ or $\mathbf{u}(\mathbf{x})$, one sees that the conditions (1.82) are also necessary. Complex, frequency dependent, elasticity tensors arise naturally in the theory of viscoelasticity (Christensen 2003). Complex, frequency dependent, effective mass-density tensors also arise naturally in the theory of elastic metamaterials (Sheng, Zhang, Liu, and Chan 2003; Movchan and Guenneau 2004; Liu, Chan, and Sheng 2005; Milton, Briane, and Willis 2006; Milton and Willis 2007). The associated mathematical theory was first developed by Zhikov (2000), Section 8.1: see also Bouchitté and Felbacq (2004) and Smyshlyaev (2009).

1.7 The time-harmonic acoustic, Schrödinger, elastic, and electromagnetic equations

It was recognized by Milton, Seppecher, and Bouchitté (2009) that a similar structure holds true for the equations of acoustics, elastodynamics and electromagnetism in the case when the fields vary harmonically with

time, i.e., have a time-dependence of $e^{-i\omega t}$ where ω is the frequency. The constitutive equation takes the form

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\mathcal{F}(\mathbf{x}), \tag{1.83}$$

where, in the absence of sources, $\mathcal{G}(\mathbf{x})$ and $\mathcal{F}(\mathbf{x})$ satisfy the differential constraints that

$$\mathcal{F} = \begin{pmatrix} \nabla\mathbf{u} \\ \mathbf{u} \end{pmatrix}, \quad \mathcal{G} = \begin{pmatrix} \mathbf{G} \\ \nabla\cdot\mathbf{G} \end{pmatrix}, \tag{1.84}$$

in which the potential $\mathbf{u}(\mathbf{x})$ is scalar- or vector-valued field, while $\mathbf{G}(\mathbf{x})$ is, correspondingly, a vector or second-order tensor field. When $\mathbf{u}(\mathbf{x})$ is a vector field, these differential constraints imply

$$\mathcal{G}(\mathbf{x})\cdot\mathcal{F}(\mathbf{x}) \equiv \mathbf{G}(\mathbf{x}) : \nabla\mathbf{u}(\mathbf{x}) + \mathbf{u}(\mathbf{x})\cdot[\nabla\cdot\mathbf{G}(\mathbf{x})] = \nabla\cdot\mathbf{Q}(\mathbf{x}), \tag{1.85}$$

where the supercurrent $\mathbf{Q}(\mathbf{x})$ is the vector field

$$\mathbf{Q}(\mathbf{x}) = \mathbf{G}(\mathbf{x})\mathbf{u}(\mathbf{x}), \tag{1.86}$$

thus giving the key identity

$$\int_\Omega \mathcal{G}(\mathbf{x})\cdot\mathcal{F}(\mathbf{x}) = \int_{\partial\Omega} \mathbf{n}\cdot\mathbf{G}(\mathbf{x})\mathbf{u}(\mathbf{x}). \tag{1.87}$$

When $\mathbf{u}(\mathbf{x})$ is a scalar field $u(\mathbf{x})$ and $\mathbf{G}(\mathbf{x})$ is a vector field, the same identity holds once we define

$$\mathcal{G}(\mathbf{x})\cdot\mathcal{F}(\mathbf{x}) \equiv \mathbf{G}(\mathbf{x})\cdot\nabla u(\mathbf{x}) + u(\mathbf{x})[\nabla\cdot\mathbf{G}(\mathbf{x})]. \tag{1.88}$$

For acoustics we can make the identifications

$$\mathbf{u} = P, \quad \mathbf{G} = -i\mathbf{v}, \quad \mathbf{Z} = \begin{pmatrix} -(\omega\boldsymbol{\rho})^{-1} & 0 \\ 0 & \omega/\kappa \end{pmatrix}, \tag{1.89}$$

where P is the complex pressure, $\mathbf{v}$ the complex velocity, $\boldsymbol{\rho}$ the complex density tensor, and κ the complex bulk modulus [see Dukhin and Goetz (2009) and references therein]. With these substitutions, the constitutive law (1.83) implies

$$-i\mathbf{v} = -(\omega\boldsymbol{\rho})^{-1}\nabla P, \quad -i\nabla\cdot\mathbf{v} = (\omega/\kappa)P, \tag{1.90}$$

which on eliminating $\mathbf{v}$ leads to the familiar acoustics equation

$$\nabla\cdot\boldsymbol{\rho}^{-1}\nabla P + \omega^2\kappa^{-1}P = 0. \tag{1.91}$$

The multielectron Schrödinger equation, for time-harmonic fields with a time-dependence $e^{-iEt/\hbar}$ where E is the energy and $\hbar$ is Planck's constant divided by 2π, is equivalent to the acoustic equation and can be written in the same form, with the (generally complex-valued) wavefunction $\psi(\mathbf{x})$ playing the role of the pressure,

$$\begin{pmatrix} \mathbf{q}(\mathbf{x}) \\ \nabla\cdot\mathbf{q}(\mathbf{x}) \end{pmatrix} = \underbrace{\begin{pmatrix} -\mathbf{A} & 0 \\ 0 & E - V(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} \nabla\psi(\mathbf{x}) \\ \psi(\mathbf{x}) \end{pmatrix}, \tag{1.92}$$

where $V(\mathbf{x})$ is the potential and $\mathbf{A}$ in the simplest approximation is $\hbar^2\mathbf{I}/(2m)$ in which m is the mass of the electron, but it may take other forms to take into account the reduced mass of the electron, or mass polarization

terms due to the motion of the atomic nuclei. Here $\mathbf{x}$ lies in a multidimensional space $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ where following, for example, Parr and Weitao (1994), each $\mathbf{x}_i$ represents a pair $(\mathbf{r}_i, s_i)$ where $\mathbf{r}_i$ is a three dimensional vector associated with the position of electron i and s_i denotes its spin (taking discrete values $+1/2$ for spin up or $-1/2$ for spin down). Accordingly, ∇ represents the operator

$$\nabla = (\nabla_1, \nabla_2, \ldots, \nabla_N), \quad \text{where } \nabla_j = \left(\frac{\partial}{\partial r_1^{(j)}}, \frac{\partial}{\partial r_2^{(j)}}, \frac{\partial}{\partial r_3^{(j)}} \right). \tag{1.93}$$

Typically one thinks of the matrix $\mathbf{Z}(\mathbf{x})$ in (1.92) as being Hermitian, but it can have a positive semidefinite imaginary part if $E = E' + iE''$ is complex with a positive imaginary part E'', corresponding to having fields with a time-dependence $e^{E''t/\hbar} e^{-iE't/\hbar}$ which is increasing with time. The key restriction on $\psi(\mathbf{x})$ for the multielectron Schrödinger equation, is that it must be antisymmetric when we interchange any pair $\mathbf{x}_i$ and $\mathbf{x}_j$, with $i \neq j$, (i.e., when we interchange both position and spin). The wavefunction $\psi(\mathbf{x})$ has the physical interpretation that $\overline{\psi(\mathbf{x})}\psi(\mathbf{x})$ gives the joint probability density of finding electrons at the points $\mathbf{r}_1$, $\mathbf{r}_2$, …, $\mathbf{r}_N$ with corresponding spins s_1, s_2, …, $\mathbf{s}_N$ when a measurement of electron positions and electron spins is taken (and this measurement will destroy the wavefunction).

For elastodynamics in the absence of sources we identify $\mathbf{u}$ with the displacement, and

$$\mathcal{G}(\mathbf{x}) = \begin{pmatrix} -\boldsymbol{\sigma}/\omega \\ i\mathbf{p} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} -\boldsymbol{\mathcal{C}}/\omega & 0 \\ 0 & \omega\boldsymbol{\rho} \end{pmatrix}, \quad \mathcal{F} = \begin{pmatrix} \nabla\mathbf{u} \\ \mathbf{u} \end{pmatrix}, \tag{1.94}$$

where $\boldsymbol{\sigma}(\mathbf{x})$ is the complex stress tensor, $\mathbf{p}(\mathbf{x})$ is the complex momentum density, $\boldsymbol{\mathcal{C}}(\mathbf{x})$ the fourth-order elasticity tensor, and $\boldsymbol{\rho}(\mathbf{x})$ the complex density tensor. With these substitutions, the constitutive law (1.83) implies

$$\boldsymbol{\sigma}(\mathbf{x}) = \boldsymbol{\mathcal{C}}\nabla\mathbf{u}, \quad i\mathbf{p}(\mathbf{x}) = \nabla \cdot (-\boldsymbol{\sigma}(\mathbf{x})/\omega) = \omega\boldsymbol{\rho}\mathbf{u}, \tag{1.95}$$

thus reducing to the familiar elastodynamic equation,

$$\nabla \cdot \boldsymbol{\mathcal{C}}\nabla\mathbf{u} + \omega^2\boldsymbol{\rho}\mathbf{u} = 0. \tag{1.96}$$

In elastodynamics (or whenever the potential $\mathbf{u}$ is vector-valued) one could take any constant second-order tensor $\mathbf{M}$ and redefine

$$\mathcal{G}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}) \equiv \mathbf{G}(\mathbf{x}) : [(\nabla\mathbf{u}(\mathbf{x}))\mathbf{M}] + [\mathbf{u}(\mathbf{x})\mathbf{M}] \cdot [\nabla \cdot \mathbf{G}(\mathbf{x})] = \nabla \cdot [\mathbf{G}(\mathbf{x})(\mathbf{u}(\mathbf{x})\mathbf{M})]. \tag{1.97}$$

In this way one sees that there is not just one key identity but a whole multitude of them, parameterized by $\mathbf{M}$. As the identities are linear in $\mathbf{M}$ it suffices to give them for a basis of $\mathbf{M}$, yielding d^2 identities where $d = 2$ or 3 is the dimensionality of the space.

For electromagnetism at fixed frequency ω (in the absence of "free current" sources, but allowing for conduction currents $\boldsymbol{\sigma}\mathbf{e}$, which are instead incorporated in the term $\boldsymbol{\varepsilon}\mathbf{e}$) the basic equations for the electric field $\mathbf{e}(\mathbf{x})$, electric displacement field $\mathbf{d}(\mathbf{x})$, magnetizing field $\mathbf{h}(\mathbf{x})$ and magnetic field $\mathbf{b}(\mathbf{x})$ are

$$\nabla \times \mathbf{e} = i\omega\mathbf{b}, \quad \nabla \times \mathbf{h} = -i\omega\mathbf{d}, \quad \mathbf{d} = \boldsymbol{\varepsilon}\mathbf{e}, \quad \mathbf{b} = \boldsymbol{\mu}\mathbf{h}, \tag{1.98}$$

where $\boldsymbol{\varepsilon}(\mathbf{x})$ is the complex-valued electric permittivity tensor and $\boldsymbol{\mu}(\mathbf{x})$ is the complex-valued magnetic permeability tensor. We have chosen units of dimensions so the speed of light c is 1. Maxwell's equations can also be cast as an "elasticity type equation":

$$\nabla \cdot (\boldsymbol{\mathcal{C}}\nabla\mathbf{e}) - \omega^2\boldsymbol{\varepsilon}\mathbf{e} = 0, \tag{1.99}$$

where now $\mathcal{C}$ has elements

$$C_{ijk\ell} = e_{ijm}e_{k\ell n}\{\boldsymbol{\mu}^{-1}\}_{mn}, \tag{1.100}$$

and e_{ijm} is the Levi-Civita (alternating) tensor [$e_{ijm} = 1$ (-1) if (i,j,m) is an even (odd) permutation of $(1,2,3)$ and is zero otherwise]. Thus electromagnetism can also be cast in the form (1.83), with fields satisfying (1.84) with $\mathbf{u}$ being replaced by the electric field. Note however there is an error in equation (4.2) of Milton, Seppecher, and Bouchitté (2009): the left hand side should not have a minus sign. The correct form is (1.99) as originally correctly stated by Milton, Briane, and Willis (2006). Fortunately this does not effect much of the analysis in that paper which is mostly based on the equations (4.16) to (4.20) there, which are correct.

For electromagnetism, at fixed frequency, the fields and tensor entering the constitutive law are most naturally taken as

$$\mathcal{G}(\mathbf{x}) = \begin{pmatrix} -i\mathbf{h} \\ \omega\mathbf{d} \end{pmatrix}, \quad \mathbf{Z}(\mathbf{x}) = \begin{pmatrix} -[\omega\boldsymbol{\mu}(\mathbf{x})]^{-1} & 0 \\ 0 & \omega\boldsymbol{\varepsilon}(\mathbf{x}) \end{pmatrix}, \quad \mathcal{F}(\mathbf{x}) = \begin{pmatrix} i\omega\mathbf{b} \\ \mathbf{e} \end{pmatrix}, \tag{1.101}$$

where the fields $\mathcal{F}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ are subject to the differential constraints that

$$\mathcal{F}(\mathbf{x}) = \begin{pmatrix} \nabla\times\mathbf{e} \\ \mathbf{e} \end{pmatrix}, \quad \mathcal{G}(\mathbf{x}) = \begin{pmatrix} -i\mathbf{h} \\ i\nabla\times\mathbf{h} \end{pmatrix}, \tag{1.102}$$

for some fields $\mathbf{e}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$. Now we have

$$\mathcal{G}(\mathbf{x})\cdot\mathcal{F}(\mathbf{x}) = -i\mathbf{h}\cdot\nabla\times\mathbf{e} + i\mathbf{e}\cdot\nabla\times\mathbf{h} = \nabla\cdot\mathbf{Q}(\mathbf{x}), \tag{1.103}$$

where

$$\mathbf{Q}(\mathbf{x}) = -i\mathbf{e}\times\mathbf{h}. \tag{1.104}$$

This implies the key identity

$$\int_\Omega \mathcal{G}(\mathbf{x})\cdot\mathcal{F}(\mathbf{x}) = \int_{\partial\Omega} -i\mathbf{n}\cdot(\mathbf{e}\times\mathbf{h}), \tag{1.105}$$

which is essentially Poynting's theorem, apart from the appearance of $\mathbf{h}$ rather than $\overline{\mathbf{h}}$ on the right-hand side of the equation. (The associated key identity is Poynting's theorem.)

1.8 The time-harmonic thermoacoustic equations

A more challenging set of time-harmonic equations to express in the desired form are those of linearized thermoacoustics, which incorporate thermal and viscous losses into the equations of acoustics. The source of these equations that I am using is a COMSOL Acoustics Module User's Guide (COMSOL 2013), equations (7-5), page 286. Some related theory can be found in Pierce (1981).

The time-harmonic linearized thermoacoustic equations involve the density fluctuations, temperature fluctuations, pressure fluctuations, stress, and velocity fields which are the real parts of $e^{-i\omega t}\rho(\mathbf{x})$, $e^{-i\omega t}\theta(\mathbf{x})$, $e^{-i\omega t}P(\mathbf{x})$, $e^{-i\omega t}\boldsymbol{\sigma}(\mathbf{x})$, and $e^{-i\omega t}\mathbf{v}(\mathbf{x})$, where the complex fields, in the absence of source terms, satisfy

$$i\omega\rho = -\rho_0(\nabla\cdot\mathbf{v}), \quad i\omega\rho_0\mathbf{v} = \nabla\cdot\boldsymbol{\sigma}, \quad i\omega(\rho_0 C_p\theta - T_0\alpha_0 P) = \nabla\cdot[k(\mathbf{x})\nabla\theta], \tag{1.106}$$

representing the equations of conservation of mass, momentum, and energy: where ρ_0 and T_0 are the background density and temperature; C_p is the heat capacity at constant pressure; α_0 is the coefficient of thermal expansion at constant pressure; and $k(\mathbf{x})$ is the thermal conductivity. Additionally one has the relations

$$\begin{aligned} \rho &= \rho_0(\beta_T P - \alpha_0\theta), \quad \boldsymbol{\sigma} = -P\mathbf{I} + \boldsymbol{\mathcal{D}}\nabla\mathbf{v}, \\ \boldsymbol{\mathcal{D}}\nabla\mathbf{v} &= \mu[\nabla\mathbf{v} + (\nabla\mathbf{v})^T] + \left(\mu_B - \frac{2}{3}\mu\right)(\nabla\cdot\mathbf{v})\mathbf{I}, \end{aligned} \tag{1.107}$$

where the first equation is the linearization of the equation of state linking pressure, density, and temperature, while the second and third equations give the constitutive law for the stress in a fluid, in terms the velocity gradient and pressure fields. Here $\boldsymbol{\mathcal{D}}$ is the isotropic fourth-order tensor of viscosity moduli, μ and μ_B are the dynamic shear and bulk viscosities (for a discussion of bulk viscosity see Dukhin and Goetz 2009) and β_T is the isothermal compressibility. We first eliminate the density from these equations to get

$$s \equiv -(\nabla\cdot\mathbf{v}) - i\omega\beta_T P + i\omega\alpha_0\theta = 0, \tag{1.108}$$

and we can use this to express P in terms of the other variables (including s which is zero),

$$P = \frac{is}{\omega\beta_T} + \frac{i}{\omega\beta_T}\nabla\cdot\mathbf{v} + \frac{\alpha_0\theta}{\beta_T}, \tag{1.109}$$

which we can use to eliminate P from the other equations (in favor of s):

$$\begin{aligned} \boldsymbol{\sigma} &= \boldsymbol{\mathcal{D}}\nabla\mathbf{v} - \frac{is\mathbf{I}}{\omega\beta_T} - \frac{i\mathbf{I}}{\omega\beta_T}\nabla\cdot\mathbf{v} - \frac{\alpha_0\theta\mathbf{I}}{\beta_T}, \\ \nabla\cdot[k(\mathbf{x})\nabla\theta] &= i\omega\rho_0 C_p\theta + \frac{s\alpha_0 T_0}{\beta_T} + \frac{\alpha_0 T_0}{\beta_T}\nabla\cdot\mathbf{v} - i\omega\frac{\alpha_0^2 T_0\theta}{\beta_T}. \end{aligned} \tag{1.110}$$

Hence we can rewrite the thermoacoustic equations (without sources) as

$$\begin{pmatrix} i\boldsymbol{\sigma} \\ i\nabla\cdot\boldsymbol{\sigma} \\ \mathbf{q} \\ \nabla\cdot\mathbf{q} \\ -iP \end{pmatrix} = \underbrace{\begin{pmatrix} i\boldsymbol{\mathcal{D}}(\mathbf{x}) + \frac{\mathbf{I}\otimes\mathbf{I}}{\omega\beta_T} & 0 & 0 & \frac{-i\alpha_0 T_0\mathbf{I}}{\beta_T} & \frac{\mathbf{I}}{\omega\beta_T} \\ 0 & -\omega\rho_0 & 0 & 0 & 0 \\ 0 & 0 & ik(\mathbf{x})T_0 & 0 & 0 \\ \frac{i\alpha_0 T_0\mathbf{I}\cdot}{\beta_T} & 0 & 0 & \omega\frac{\alpha_0^2 T_0^2}{\beta_T} - \omega\rho_0 C_p T_0 & \frac{i\alpha_0 T_0}{\beta_T} \\ \frac{\mathbf{I}\cdot}{\omega\beta_T} & 0 & 0 & \frac{-i\alpha_0 T_0}{\beta_T} & \frac{1}{\omega\beta_T} \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} \nabla\mathbf{v} \\ \mathbf{v} \\ \nabla\theta/T_0 \\ \theta/T_0 \\ s \end{pmatrix}. \tag{1.111}$$

The matrix $\mathbf{Z}(\mathbf{x})$ entering this constitutive law is such that the Hermitian part of $\mathbf{Z}/i$ is positive semidefinite as desired. Furthermore, recalling that s is zero, we have the key identity,

$$\begin{pmatrix} i\boldsymbol{\sigma} \\ i\nabla\cdot\boldsymbol{\sigma} \\ \mathbf{q} \\ \nabla\cdot\mathbf{q} \\ -iP \end{pmatrix} \cdot \begin{pmatrix} \nabla\mathbf{v} \\ \mathbf{v} \\ \nabla\theta/T_0 \\ \theta/T_0 \\ s \end{pmatrix} = \nabla\cdot[i\boldsymbol{\sigma}\mathbf{v} + \mathbf{q}\theta/T_0]. \tag{1.112}$$

1.9 The time-dependent acoustic equation

Now let us consider the acoustic equations, removing the assumption of time-harmonicity. We begin by assuming that the bulk modulus $\kappa(\mathbf{x})$ and density $\rho(\mathbf{x})$ only depend on the spatial variable $\mathbf{x}$ and not on time t. Then the time-dependent equations of acoustics for the pressure field $P(\mathbf{x},t)$ and velocity field $\mathbf{v}(\mathbf{x},t)$ can be written as

$$\begin{pmatrix} \frac{\partial \mathbf{v}}{\partial t} \\ \nabla\cdot\mathbf{v}\end{pmatrix} = \underbrace{\begin{pmatrix} -\rho(\mathbf{x})^{-1}\mathbf{I} & 0 \\ 0 & \kappa(\mathbf{x})^{-1}\end{pmatrix}}_{\mathbf{Z}(\mathbf{x})}\begin{pmatrix}\nabla P \\ -\frac{\partial P}{\partial t}\end{pmatrix}. \tag{1.113}$$

Here we assume ρ and κ do not depend on frequency, but it's okay if they do: then the constitutive law gets replaced by a convolution in time. Writing out these equations enables us to check that indeed they are the equations of acoustics:

$$\kappa^{-1}\frac{\partial^2 P}{\partial t^2} = -\frac{\partial \nabla\cdot\mathbf{v}}{\partial t} = \nabla\rho^{-1}\nabla P. \tag{1.114}$$

Also, taking the dot product

$$\begin{aligned}\begin{pmatrix} \frac{\partial \mathbf{v}}{\partial t} \\ \nabla\cdot\mathbf{v}\end{pmatrix}\cdot\begin{pmatrix}\nabla P \\ -\frac{\partial P}{\partial t}\end{pmatrix} &= (\nabla P)\cdot\left(\frac{\partial \mathbf{v}}{\partial t}\right) - \left(\frac{\partial P}{\partial t}\right)(\nabla\cdot\mathbf{v}) + P\frac{\partial \nabla\cdot\mathbf{v}}{\partial t} - P\frac{\partial \nabla\cdot\mathbf{v}}{\partial t} \\ &= \begin{pmatrix}\nabla \\ -\frac{\partial}{\partial t}\end{pmatrix}\cdot\begin{pmatrix} P\frac{\partial \mathbf{v}}{\partial t} \\ P\nabla\cdot\mathbf{v}\end{pmatrix} \\ &= \underset{\sim}{\nabla}\cdot\mathbf{Q},\end{aligned} \tag{1.115}$$

where $\underset{\sim}{\nabla}$ is the 4 dimensional gradient with $x_4 = -t$, and $\mathbf{Q} = [(P\partial_t\mathbf{v})^T, P\nabla\cdot\mathbf{v}]^T$. This is a divergence so the key identity

$$\int_{\underset{\sim}{\Omega}}\begin{pmatrix} \frac{\partial \mathbf{v}}{\partial t} \\ \nabla\cdot\mathbf{v}\end{pmatrix}\cdot\begin{pmatrix}\nabla P \\ -\frac{\partial P}{\partial t}\end{pmatrix} = \int_{\partial\underset{\sim}{\Omega}}\mathbf{n}_x\cdot P\frac{\partial \mathbf{v}}{\partial t} - n_t P\nabla\cdot\mathbf{v} \tag{1.116}$$

holds. Now $\underset{\sim}{\Omega}$ is a body in space–time with outward normal $\underset{\sim}{\mathbf{n}} = (\mathbf{n}_x, n_t)$.

A further simplification can be made. Suppose we are in three (spatial) dimensions. Let $x_4 = -t$, and set

$$\underset{\sim}{\nabla} = \begin{pmatrix}\nabla \\ \frac{\partial}{\partial x_4}\end{pmatrix} = \begin{pmatrix}\nabla \\ -\frac{\partial}{\partial t}\end{pmatrix}, \quad j_i = \frac{\partial v_i}{\partial t}, \quad \text{for } i = 1,\dots 3, \quad j_4 = \nabla\cdot\mathbf{v}. \tag{1.117}$$

Then, with $\underset{\sim}{\mathbf{x}} = (x_1, x_2, x_3, -t)$, the equations (1.113) can be rewritten as

$$\underset{\sim}{\nabla}\cdot\mathbf{j}(\underset{\sim}{\mathbf{x}}) = 0, \quad \mathbf{j}(\underset{\sim}{\mathbf{x}}) = \mathbf{Z}(\underset{\sim}{\mathbf{x}})\underset{\sim}{\nabla}P(\underset{\sim}{\mathbf{x}}), \tag{1.118}$$

so they look like a "space–time" version of the conductivity equations, with $\mathbf{j}(\underset{\sim}{\mathbf{x}})$ satisfying the conservation law $\underset{\sim}{\nabla}\cdot\mathbf{j}(\underset{\sim}{\mathbf{x}}) = 0$. Of course, there are important differences: most notably the matrix entering the constitutive law is real but not positive definite, as it should be for a wave equation.

Now, by direct analogy with the transformation of the conductivity equations under affine transformations (see, for example, Section 8.3 of Milton (2002)), under the Galilean transformation

$$\underset{\sim}{\mathbf{x}}' = \mathbf{A}\underset{\sim}{\mathbf{x}}, \quad \text{with } \mathbf{A} = \begin{pmatrix}\mathbf{I} & \mathbf{w} \\ 0 & 1\end{pmatrix}, \tag{1.119}$$

where $\underset{\sim}{\mathbf{x}}' = (x_1', x_2', x_3', -t')$, $\underset{\sim}{\mathbf{x}} = (x_1, x_2, x_3, -t)$, and $\mathbf{w}$ is the velocity of the moving frame of reference, the equations transform to

$$\underset{\sim}{\nabla}' \cdot \mathbf{j}'(\underset{\sim}{\mathbf{x}}') = 0, \quad \mathbf{j}'(\underset{\sim}{\mathbf{x}}') = \mathbf{Z}'(\underset{\sim}{\mathbf{x}}')\underset{\sim}{\nabla}' P'(\underset{\sim}{\mathbf{x}}'), \tag{1.120}$$

in which

$$\begin{aligned} \mathbf{Z}' &= \mathbf{A}\mathbf{Z}\mathbf{A}^T = \begin{pmatrix} -\rho(\mathbf{x})^{-1}\mathbf{I} + \kappa(\mathbf{x})^{-1}\mathbf{w}\mathbf{w}^T & \kappa(\mathbf{x})^{-1}\mathbf{w} \\ \kappa(\mathbf{x})^{-1}\mathbf{w}^T & \kappa(\mathbf{x})^{-1} \end{pmatrix}, \\ \mathbf{j}'(\underset{\sim}{\mathbf{x}}') &= \mathbf{A}\mathbf{j}(\underset{\sim}{\mathbf{x}}), \quad P'(\underset{\sim}{\mathbf{x}}') = P(\underset{\sim}{\mathbf{x}}), \quad \underset{\sim}{\nabla}' P'(\underset{\sim}{\mathbf{x}}') = (\mathbf{A}^T)^{-1}\underset{\sim}{\nabla} P(\underset{\sim}{\mathbf{x}}). \end{aligned} \tag{1.121}$$

Thus in a moving fluid the tensor $\mathbf{Z}$ will not be diagonal, and the density term will not be isotropic. One expects in a fluid where there is a small fluid velocity $\mathbf{v}(\underset{\sim}{\mathbf{x}})$ superimposed on a large macroscopic velocity $\mathbf{w}(\underset{\sim}{\mathbf{x}})$, that the tensor $\mathbf{Z}$ will take the form

$$\mathbf{Z}(\mathbf{x}, t) = \begin{pmatrix} -[\boldsymbol{\rho}(\mathbf{x}, t)]^{-1} & \mathbf{k}(\mathbf{x}, t) \\ [\mathbf{k}(\mathbf{x}, t)]^T & [\kappa(\mathbf{x}, t)]^{-1} \end{pmatrix}, \tag{1.122}$$

with a generally anisotropic density $\boldsymbol{\rho}$. Note that this form is preserved under the transformations of special relativity and not just Galilean transformations (in fact we are free to choose any transformation matrix $\mathbf{A}$). Anisotropic effective mass densities may seem unfamiliar but Willis (1985) found the effective density in a composite should be a nonlocal anisotropic density operator, Schoenberg and Sen (1983) found anisotropic layered fluids had an anisotropic effective density, and Milton, Briane, and Willis (2006) found simple models exhibiting anisotropic density: see their **Figure 3**. The coupling one sees in (1.122), is similar to that one sees in the Willis equations of elastodynamics (developed in Willis 1981a, 1981b and explicitly stated in Willis 1997), or in the bianisotropic equations of electromagnetism (Serdikukov, Semchenko, Tretkyakov, and Sihvola 2001).

1.10 The equations of elastodynamics and piezoelectricity in the time domain

We rewrite the equations of elastodynamics in the form

$$\begin{pmatrix} \frac{\partial \boldsymbol{\sigma}}{\partial t} \\ \nabla \cdot \boldsymbol{\sigma} \end{pmatrix} = \underbrace{\begin{pmatrix} -\boldsymbol{\mathcal{C}}(\mathbf{x}) & 0 \\ 0 & \boldsymbol{\rho}(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} -\frac{1}{2}\left[\nabla \mathbf{v} + \nabla \mathbf{v}^T\right] \\ \frac{\partial \mathbf{v}}{\partial t} \end{pmatrix}, \tag{1.123}$$

where $\boldsymbol{\sigma}$ is the stress and $\mathbf{v}(t) = \frac{\partial \mathbf{u}}{\partial t}$ is the velocity. Once again taking dot products we find a divergence form:

$$\begin{aligned} \begin{pmatrix} \frac{\partial \boldsymbol{\sigma}}{\partial t} \\ \nabla \cdot \boldsymbol{\sigma} \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{2}\left[\nabla \mathbf{v} + \nabla \mathbf{v}^T\right] \\ \frac{\partial \mathbf{v}}{\partial t} \end{pmatrix} &= -\left(\frac{\partial \boldsymbol{\sigma}}{\partial t}\right) : \nabla \mathbf{v} + (\nabla \cdot \boldsymbol{\sigma}) \cdot \left(\frac{\partial \mathbf{v}}{\partial t}\right) + \boldsymbol{\sigma} : \frac{\partial (\nabla \mathbf{v})}{\partial t} \\ &\qquad - \boldsymbol{\sigma} : \frac{\partial (\nabla \mathbf{v})}{\partial t} \\ &= \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma} \frac{\partial \mathbf{v}}{\partial t} \\ \boldsymbol{\sigma} : \nabla \mathbf{v} \end{pmatrix} = \underset{\sim}{\nabla} \cdot \mathbf{Q}, \end{aligned} \tag{1.124}$$

where we have used the fact that $\boldsymbol{\sigma}$ is symmetric, $\underset{\sim}{\nabla}$ is the 4 dimensional gradient with $x_4 = -t$, and $\mathbf{Q} = [(\boldsymbol{\sigma}\partial_t\mathbf{v})^T, (\boldsymbol{\sigma}\nabla\cdot\mathbf{v})^T]^T$. Again the key identity applies: Ω is now a space–time body.

Guided by the form (1.118) of the acoustic equations, set

$$\underset{\sim}{\nabla} = \begin{pmatrix} \nabla \\ \frac{\partial}{\partial x_4} \end{pmatrix} = \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix}, \quad J_{ik} = -\frac{\partial \sigma_{ik}}{\partial t}, \quad \text{for } i,k = 1,2,3, \quad J_{4k} = -\{\nabla\cdot\boldsymbol{\sigma}\}_k, \tag{1.125}$$

which defines the 4×3 matrix-valued field $\mathbf{J}(\mathbf{x})$. Then, noting that, due to the symmetry of $\boldsymbol{\mathcal{C}}$, we can replace $[\nabla\mathbf{v} + (\nabla\mathbf{v})^T]/2$ by $\nabla\mathbf{v}$ in (1.123), the equations take the form

$$\underset{\sim}{\nabla}\cdot\mathbf{J} = 0, \quad \mathbf{J} = \mathbf{Z}\underset{\sim}{\nabla}\mathbf{v}. \tag{1.126}$$

So we see that under the Galilean transformation (1.119) the new fields, based on the transformation laws

$$\mathbf{J}'(\underset{\sim}{\mathbf{x}}') = \mathbf{A}\mathbf{J}(\underset{\sim}{\mathbf{x}}), \quad \underset{\sim}{\nabla}'\mathbf{v}(\underset{\sim}{\mathbf{x}}') = (\mathbf{A}^T)^{-1}\underset{\sim}{\nabla}\mathbf{v}(\underset{\sim}{\mathbf{x}}), \quad \text{where } \mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{w} \\ 0 & 1 \end{pmatrix}, \tag{1.127}$$

for $\mathbf{J}$ and $\underset{\sim}{\nabla}\mathbf{v}$, become

$$\begin{aligned} \begin{pmatrix} \frac{\partial\boldsymbol{\sigma}'}{\partial t'} \\ \nabla'\cdot\boldsymbol{\sigma}' \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\mathcal{I}} & \mathbf{w}\mathbf{I} \\ 0 & \mathbf{I} \end{pmatrix}\begin{pmatrix} \frac{\partial\boldsymbol{\sigma}}{\partial t} \\ \nabla\cdot\boldsymbol{\sigma} \end{pmatrix} = \begin{pmatrix} \frac{\partial\boldsymbol{\sigma}}{\partial t} + \mathbf{w}(\nabla\cdot\boldsymbol{\sigma})^T \\ \nabla\cdot\boldsymbol{\sigma} \end{pmatrix}, \\ \begin{pmatrix} -\nabla'\mathbf{v}' \\ \frac{\partial\mathbf{v}'}{\partial t'} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\mathcal{I}} & 0 \\ \mathbf{I}\mathbf{w}^T & \mathbf{I} \end{pmatrix}^{-1}\begin{pmatrix} -\nabla\mathbf{v} \\ \frac{\partial\mathbf{v}}{\partial t} \end{pmatrix} = \begin{pmatrix} -\nabla\mathbf{v} \\ \frac{\partial\mathbf{v}}{\partial t} + \mathbf{w}^T\nabla\mathbf{v} \end{pmatrix}, \end{aligned} \tag{1.128}$$

in which $\boldsymbol{\mathcal{I}}$ is the fourth-order identity tensor, $\mathbf{I}$ is the second-order identity tensor, and $\mathbf{w}\mathbf{I}$ is a third-order tensor with elements $w_i\delta_{jk}$ (in which δ_{jk} is one if $j = k$ and zero otherwise). These fields are now linked by the new constitutive tensor

$$\begin{aligned} \mathbf{Z}'(\underset{\sim}{\mathbf{x}}') &= \begin{pmatrix} \boldsymbol{\mathcal{I}} & \mathbf{w}\mathbf{I} \\ 0 & \mathbf{I} \end{pmatrix}\mathbf{Z}(\mathbf{x})\begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{I}\mathbf{w}^T & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} -\boldsymbol{\mathcal{C}}(\mathbf{x}) + \mathbf{w}\boldsymbol{\rho}(\mathbf{x})\mathbf{w}^T & \mathbf{w}\boldsymbol{\rho}(\mathbf{x}) \\ \boldsymbol{\rho}(\mathbf{x})\mathbf{w}^T & \boldsymbol{\rho}(\mathbf{x}) \end{pmatrix}. \end{aligned} \tag{1.129}$$

Under more general transformations, such as a Galilean transformation followed by a rotation, the fields in (1.128) will be also multiplied on the right by rotation matrices (and the transformation matrix $\mathbf{A}$ would also need to be adjusted). However if it is a pure Galilean transformation then there is no multiplication on the right by a rotation matrix: note that since $\mathbf{u}(\mathbf{x})$ represents an assumed small difference between the position of a particle in the undeformed state (now moving) and the position of the particle in the deformed state, $\mathbf{v}$ being the difference between two velocities is itself invariant under a Galilean transformation.

Note that the new stress field is not symmetric, and does not just depend on the symmetric part of $\nabla\mathbf{v}$, but also on its antisymmetric part.

Significantly, this analysis shows that the stress tensor $\boldsymbol{\sigma}(\mathbf{x},t)$ does not have the symmetries, nor the transformation law assumed in the theory of relativity: consequently this casts doubt on the general theory of relativity in its standard form, unless stress is taken to have some different meaning there. The need for

nonsymmetric stress-energy-momentum tensors with accompanying modifications to Einstein's gravitational theory has been noted before: see Hehl, von der Heyde, Kerlick, and Nester (1976) and references therein.

It has been argued (see appendix A in Martin, Parodi, and Pershan 1972 and also Lautrup 2005) that it is only the divergence of the stress (which equals the body force) that has a physical meaning and one can replace a nonsymmetric stress by a symmetric one, while keeping the divergence of the stress the same. However this seems unnatural as the constitutive law then becomes very nonlocal and stress fields then exist in regions that are void of material.

One expects in an elastic medium where there is a small material velocity $\mathbf{v}(\mathbf{x}, t)$ superimposed on a large macroscopic velocity $\mathbf{w}(\mathbf{x}, t)$ that the tensor $\mathbf{Z}$ will take the form

$$\mathbf{Z}(\mathbf{x}, t) = \begin{pmatrix} -\boldsymbol{\mathcal{C}}(\mathbf{x}, t) & \boldsymbol{\kappa}(\mathbf{x}, t) \\ [\boldsymbol{\kappa}(\mathbf{x}, t)]^T & \boldsymbol{\rho}(\mathbf{x}, t) \end{pmatrix}, \tag{1.130}$$

where the density $\boldsymbol{\rho}$ could be anisotropic, and $\boldsymbol{\mathcal{D}}$ need not satisfy the usual elastic symmetries. In general if $\mathbf{Z}(\mathbf{x}, t)$ was nonlocal, i.e., the constitutive relation involved an integral kernel, then the effective one would too, and again there would be a coupling. These nonlocal equations, with couplings, are known as the Willis equations (developed in Willis 1981a, 1981b and explicitly stated in Willis 1997), although his equations retain a symmetric stress field and a stress that only depends on the symmetrized displacement gradient. An explicit model exhibiting (over a very narrow frequency band) local Willis type couplings, but with a nonsymmetric stress, was constructed by Milton (2007). An example of a mechanism for producing a material with Willis type couplings is shown in **Figure 1.2**. The Willis equations keep their form under arbitrary spatial curvilinear coordinate transformations, even when the constitutive relation is nonlocal (see Appendix B and Appendix C of Milton, Briane, and Willis 2006). Due to this, materials with a Willis type constitutive law could be useful for elasticity cloaking (Milton, Briane, and Willis 2006) using the elastic analog of transformation optics cloaking techniques (Dolin 1961; Greenleaf, Lassas, and Uhlmann 2003a; Greenleaf, Lassas, and Uhlmann 2003b; Pendry, Schurig, and Smith 2006). By contrast the normal equations of elastodynamics do not keep their form unless one allows for stress fields that are nonsymmetric and elasticity tensors that lose their minor symmetries, i.e., $C_{ijk\ell} \neq C_{jik\ell}$ and $C_{ijk\ell} \neq C_{ij\ell k}$, but which keep the major one $C_{ijk\ell} = C_{k\ell ij}$ (Brun, Guenneau, and Movchan 2009; Norris and Shuvalov 2011; see also Guevara Vasquez, Milton, Onofrei, and Seppecher 2013).

For piezoelectricity in the absence of body forces and free charges the dynamic equations (Auld 1973; see also Norris 1994) are

$$\rho u_{i,tt} = \sigma_{ij,j}, \quad d_{j,j} = 0, \quad \sigma_{ij} = C_{ijk\ell}\epsilon_{k\ell} - a_{kij}e_k, \quad d_i = a_{ik\ell}\epsilon_{k\ell} + \varepsilon_{ik}e_k, \tag{1.131}$$

where the Einstein summation convection is assumed, $\mathbf{d}(\mathbf{x})$ and $\mathbf{e}(\mathbf{x})$ are the electric displacement and electric fields, and the third-order tensor $\mathbf{a}(\mathbf{x})$ couples the electromagnetic and elastic fields. These equations can be rewritten in the desired form

$$\begin{pmatrix} \frac{\partial \boldsymbol{\sigma}}{\partial t} \\ \nabla \cdot \boldsymbol{\sigma} \\ \frac{\partial \mathbf{d}}{\partial t} \end{pmatrix} = \underbrace{\begin{pmatrix} -\boldsymbol{\mathcal{C}}(\mathbf{x}) & 0 & -\mathbf{a}(\mathbf{x}) \\ 0 & \rho(\mathbf{x}) & 0 \\ -\mathbf{a}^T(\mathbf{x}) & 0 & \boldsymbol{\varepsilon}(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} -\frac{1}{2}\left[\nabla \mathbf{v} + \nabla \mathbf{v}^T\right] \\ \frac{\partial \mathbf{v}}{\partial t} \\ \frac{\partial \mathbf{e}}{\partial t} \end{pmatrix}, \tag{1.132}$$

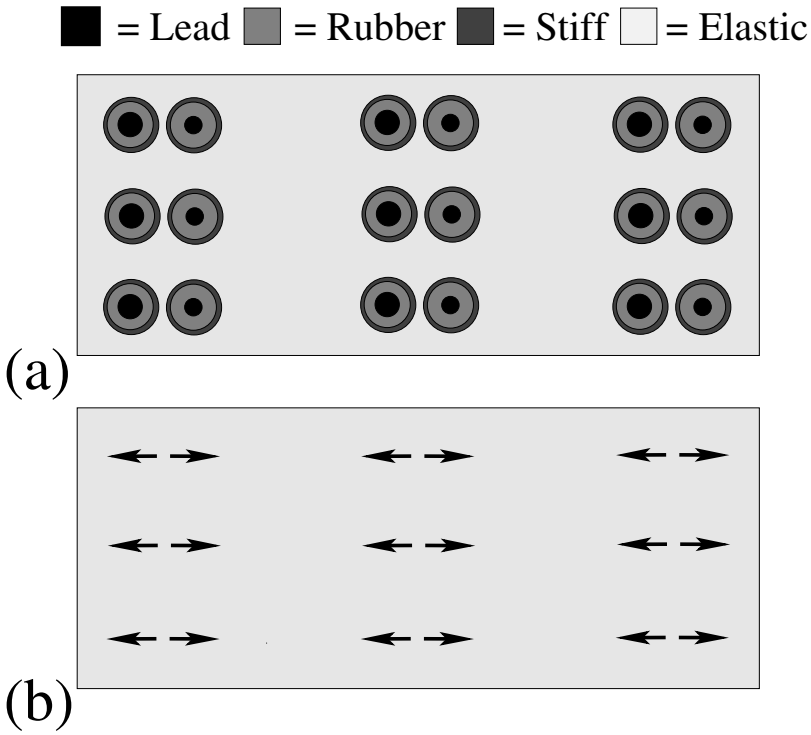


Figure 1.2: A mechanism for producing a material with Willis type couplings. The lead balls, surrounded by rubber, and coated by a shell of stiff material have a different amount of lead on the left and right sides of each pair, as shown in (a). This leads to the material in the shell on the left having a negative effective mass $-m$ and the material in the shell on the right having an almost equal and opposite positive effective mass $+m$, for a suitably tuned frequency. At this frequency these oscillating effective masses generate an array of oscillating force dipoles acting in the matrix as in (b) for one moment in time. Just like an array of electrical dipoles gives rise to an average polarization field, so too does an array of force dipoles give rise to an average stress field when the strain field is zero. Thus the time-harmonic acceleration of the material gives rise to a time-harmonic oscillating average stress when the average strain is zero, which is a characteristic feature of a Willis material.

with $\mathbf{Z}(\mathbf{x})$ being Hermitian. The key identity,

$$\begin{pmatrix} \frac{\partial \boldsymbol{\sigma}}{\partial t} \\ \nabla \cdot \boldsymbol{\sigma} \\ \frac{\partial \mathbf{d}}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -\frac{1}{2}\left[\nabla \mathbf{v} + \nabla \mathbf{v}^T\right] \\ \frac{\partial \mathbf{v}}{\partial t} \\ \frac{\partial \mathbf{e}}{\partial t} \end{pmatrix} = \underset{\sim}{\nabla} \cdot \mathbf{Q}, \tag{1.133}$$

holds with

$$\mathbf{Q} = \begin{pmatrix} \boldsymbol{\sigma} \frac{\partial \mathbf{v}}{\partial t} - \frac{\partial V}{\partial t} \frac{\partial \mathbf{d}}{\partial t} \\ \boldsymbol{\sigma} \nabla \cdot \mathbf{v} \end{pmatrix}, \tag{1.134}$$

where $V(\mathbf{x}, t)$ is the potential associated with $\mathbf{e}(\mathbf{x}, t)$: $\mathbf{e}(\mathbf{x}, t) = -\nabla V(\mathbf{x}, t)$. Of course, for these equations

to be valid the body must be small compared with the wavelength of the electromagnetic radiation associated with the fields $\mathbf{e}(\mathbf{x}, t)$ and $\mathbf{d}(\mathbf{x}, t)$.

1.11 The dynamic equations for vibrating thin plates and moderately thick plates

The dynamic plate equations can be written in the form

$$\begin{pmatrix} \frac{\partial \mathbf{M}}{\partial t} \\ -\nabla \cdot (\nabla \cdot \mathbf{M}) \end{pmatrix} = \underbrace{\begin{pmatrix} -\boldsymbol{\mathcal{D}}(\mathbf{x}) & 0 \\ 0 & h(\mathbf{x})\rho(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} -\nabla\nabla v \\ \frac{\partial v}{\partial t} \end{pmatrix}, \tag{1.135}$$

where $\mathbf{M}(\mathbf{x}, t)$ is the bending moment tensor, $\mathcal{D}(\mathbf{x})$ is the fourth-order tensor of plate rigidity coefficients, $h(\mathbf{x})$ is the plate thickness, $\rho(\mathbf{x})$ is the density, and $v = \partial w/\partial t$ is the velocity of the vertical deflection $w(\mathbf{x}, t)$ of the plate. Note that the matrix $\mathbf{Z}(\mathbf{x})$ is Hermitian, and we have the key identity,

$$\begin{pmatrix} \frac{\partial \mathbf{M}}{\partial t} \\ -\nabla \cdot (\nabla \cdot \mathbf{M}) \end{pmatrix} \cdot \begin{pmatrix} -\nabla\nabla v \\ \frac{\partial v}{\partial t} \end{pmatrix} = \underbrace{\begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix}}_{\underset{\sim}{\nabla}} \cdot \underbrace{\begin{pmatrix} -\frac{\partial v}{\partial t}\nabla \cdot \mathbf{M} - \frac{\partial \mathbf{M}}{\partial t}\nabla v \\ -(\nabla \cdot \mathbf{M}) \cdot (\nabla v) \end{pmatrix}}_{\mathbf{Q}}. \tag{1.136}$$

For moderately thick plates we need to replace these equations by those of Mindlin (1951) (see also Larsen, Laksafoss, Jensen, and Sigmund (2009)). We now assume the plate material is locally isotropic. The equilibrium equations for a small plate element read as

$$\begin{aligned} \frac{\rho h^3}{12}\frac{\partial^2 \psi_x}{\partial t^2} &= T_x - M_{x,x} - M_{xy,y}, \\ \frac{\rho h^3}{12}\frac{\partial^2 \psi_y}{\partial t^2} &= T_y - M_{y,y} - M_{xy,x}, \\ \rho h \frac{\partial^2 w}{\partial t^2} &= T_{x,x} + T_{y,y}. \end{aligned} \tag{1.137}$$

Here w is the out of plane deflection; ψ_x and ψ_y are the angles of rotation; M_x, M_y and M_{xy} are the bending moments (we use the abbreviated notation M_x and M_y to denote the components M_{xx} and M_{yy} of the tensor field $\mathbf{M}(\mathbf{x})$); T_x and T_y are the shear forces; $\rho = \rho(x, y)$ is the plate density; and $h = h(x, y)$ is the plate thickness. The relation which links the bending moments and shear forces to the deflection and rotation angles takes the form

$$\begin{pmatrix} M_x \\ M_y \\ M_{xy} \\ T_x \\ T_y \end{pmatrix} = -\mathbf{D} \begin{pmatrix} \psi_{x,x} \\ \psi_{y,y} \\ \psi_{x,y} + \psi_{y,x} \\ \psi_x - w_{,x} \\ \psi_y - w_{,y} \end{pmatrix}, \tag{1.138}$$

where the subscript comma denotes differentiation and the stiffness matrix $\mathbf{D}$ is

$$\mathbf{D} = \mathbf{D}(x,y) = \begin{pmatrix} D & \nu D & 0 & 0 & 0 \\ \nu D & D & 0 & 0 & 0 \\ 0 & 0 & \frac{1-\nu}{2} D & 0 & 0 \\ 0 & 0 & 0 & kGh & 0 \\ 0 & 0 & 0 & 0 & kGh \end{pmatrix}. \tag{1.139}$$

Here $D = Eh^3/[12(1-\nu^2)]$ is the flexural rigidity; E is the Young modulus; G is the shear modulus; ν is the Poisson ratio; and k is a shear correction factor taking the value $5/6$ for a plate. We now can rewrite the equations in the canonical form:

$$\underbrace{\begin{pmatrix} \dot{M}_x \\ \dot{M}_y \\ \dot{M}_{xy} \\ \dot{T}_x \\ \dot{T}_y \\ T_x - M_{x,x} - M_{xy,y} \\ T_y - M_{y,y} - M_{xy,x} \\ T_{x,x} + T_{y,y} \end{pmatrix}}_{\mathcal{G}} = \mathbf{Z}(x,y) \underbrace{\begin{pmatrix} \dot{\psi}_{x,x} \\ \dot{\psi}_{y,y} \\ \dot{\psi}_{x,y} + \dot{\psi}_{y,x} \\ \dot{\psi}_x - w_{,x} \\ \dot{\psi}_y - w_{,y} \\ \dot{\psi}_{x,t} \\ \dot{\psi}_{y,t} \\ \dot{w}_t \end{pmatrix}}_{\mathcal{F}}, \tag{1.140}$$

where the dot denotes differentiation with respect to time: $\dot{\psi}_x = \partial\psi_x/\partial t$, $\dot{\psi}_x = \partial\psi_x/\partial t$, and $\dot{w} = \partial w/\partial t$; and the matrix $\mathbf{Z}(x,y)$ entering the constitutive law takes the block-diagonal Hermitian form

$$\mathbf{Z}(x,y) = \begin{pmatrix} -\mathbf{D}(x,y) & 0 & 0 & 0 \\ 0 & \rho(x,y)[h(x,y)]^3/12 & 0 & 0 \\ 0 & 0 & \rho(x,y)[h(x,y)]^3/12 & 0 \\ 0 & 0 & 0 & \rho(x,y)h(x,y) \end{pmatrix}. \tag{1.141}$$

Now note that with $\mathbf{T} = (T_x, T_y)$, $\dot{\boldsymbol{\psi}} = (\dot{\psi}_x, \dot{\psi}_y)$ and $\nabla = (\partial/\partial x, \partial/\partial y)$ we have the identities

$$\begin{aligned}
-\frac{\partial \mathbf{T}}{\partial t} \cdot \nabla \dot{w} + (\nabla \cdot \mathbf{T}) \frac{\partial \dot{w}}{\partial t} &= \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -\dot{w} \frac{\partial \mathbf{T}}{\partial t} \\ -\dot{w} \nabla \cdot \mathbf{T} \end{pmatrix}, \\
\mathbf{T} \cdot \frac{\partial \dot{\boldsymbol{\psi}}}{\partial t} + \frac{\partial \mathbf{T}}{\partial t} \cdot \dot{\boldsymbol{\psi}} &= \frac{\partial}{\partial t} [\mathbf{T} \cdot \dot{\boldsymbol{\psi}}], \\
\frac{\partial M_x}{\partial t} \frac{\partial \dot{\psi}_x}{\partial x} - \frac{\partial M_x}{\partial x} \frac{\partial \dot{\psi}_x}{\partial t} &= \begin{pmatrix} \frac{\partial}{\partial x} \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -M_x \frac{\partial \dot{\psi}_x}{\partial t} \\ -M_x \frac{\partial \psi_x}{\partial x} \end{pmatrix}, \\
\frac{\partial M_y}{\partial t} \frac{\partial \dot{\psi}_y}{\partial y} - \frac{\partial M_y}{\partial y} \frac{\partial \dot{\psi}_y}{\partial t} &= \begin{pmatrix} \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -M_y \frac{\partial \dot{\psi}_y}{\partial t} \\ -M_y \frac{\partial \psi_y}{\partial y} \end{pmatrix}, \\
\frac{\partial M_{xy}}{\partial t} \frac{\partial \dot{\psi}_x}{\partial y} - \frac{\partial M_{xy}}{\partial y} \frac{\partial \dot{\psi}_x}{\partial t} &= \begin{pmatrix} \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -M_{xy} \frac{\partial \dot{\psi}_x}{\partial t} \\ -M_{xy} \frac{\partial \psi_x}{\partial y} \end{pmatrix}, \\
\frac{\partial M_{xy}}{\partial t} \frac{\partial \dot{\psi}_y}{\partial x} - \frac{\partial M_{xy}}{\partial x} \frac{\partial \dot{\psi}_y}{\partial t} &= \begin{pmatrix} \frac{\partial}{\partial x} \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} -M_{xy} \frac{\partial \dot{\psi}_y}{\partial t} \\ -M_{xy} \frac{\partial \psi_y}{\partial x} \end{pmatrix}.
\end{aligned} \tag{1.142}$$

These imply the key identity

$$\mathcal{G} \cdot \mathcal{F} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \mathbf{Q}, \tag{1.143}$$

with a supercurrent

$$\mathbf{Q} = \begin{pmatrix} -\dot{w} \frac{\partial T_x}{\partial t} - M_x \frac{\partial \dot{\psi}_x}{\partial t} - M_{xy} \frac{\partial \dot{\psi}_y}{\partial t} \\ -\dot{w} \frac{\partial T_y}{\partial t} - M_y \frac{\partial \psi_y}{\partial t} - M_{xy} \frac{\partial \psi_x}{\partial t} \\ -\dot{w} \nabla \cdot \mathbf{T} + \mathbf{T} \cdot \dot{\boldsymbol{\psi}} - M_x \frac{\partial \dot{\psi}_x}{\partial x} - M_y \frac{\partial \dot{\psi}_y}{\partial y} - M_{xy} \frac{\partial \dot{\psi}_x}{\partial y} - M_{xy} \frac{\partial \dot{\psi}_y}{\partial x} \end{pmatrix}. \tag{1.144}$$

1.12 The Biot wave equations of poroelasticity in the time domain

The Biot equations of poroelasticity (Biot 1962; see also Norris 1994) describe the dynamic motion of porous media containing fluids. One of the great successes of the theory was its prediction of an additional compressional sound mode, the Biot slow wave, which was experimentally observed by Plona (1980). In the absence of sources, the equations are comprised of equations of motion

$$\rho u_{i,tt} + \rho_f w_{i,tt} = \sigma_{ij,i}, \quad \rho_f u_{i,tt} + \widehat{m}_{ij} * w_{i,tt} = -P_{,i}\,, \tag{1.145}$$

where $\mathbf{u}(\mathbf{x}, t)$ is the displacement in the solid phase, $\mathbf{w}(\mathbf{x}, t)$ is the relative fluid displacement, $\boldsymbol{\sigma}$ is the stress in the solid, P is the fluid pressure, ρ and ρ_f are the solid and fluid densities, the $\widehat{m}_{ij}$ are viscodynamic convolution operators, in the time domain (the asterisk $*$ denoting a convolution)-and thus are represented by local operators $m_{ij}(\omega)$ in the frequency domain. The constitutive equations read

$$\sigma_{ij} = C_{ijk\ell} \epsilon_{k\ell} + M_{ij} \zeta, \quad P = M_{ij} \epsilon_{ij} + M \zeta, \tag{1.146}$$

where $\boldsymbol{\epsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]$ is the strain, $\zeta(\mathbf{x})$ is the increment of fluid content which measures the net fluid flow per unit volume in or out of a region, the $M_{ij} = M_{ji}$ are elastic moduli coupling the solid and fluid deformation, and the $C_{ijk\ell}$ are the elastic moduli of the drained porous solid. The scalar M should not be confused with the tensor $\mathbf{M}$ of elastic moduli. Finally, because the fluid is treated as incompressible, conservation of fluid mass implies

$$s \equiv \zeta_{,t} + w_{i,it} = 0, \tag{1.147}$$

where we have introduced the constant field s, which is zero if there is no net source of fluid.

We can use the relation between s and $\zeta_{,t}$ to eliminate ζ in favor of s:

$$\zeta_{,t} = -\nabla\cdot\mathbf{w}_{,t} + s, \quad \frac{\partial P}{\partial t} = \mathbf{M} : \nabla\mathbf{v} - M\nabla\cdot\mathbf{w}_{,t} + Ms, \quad \frac{\partial \boldsymbol{\sigma}}{\partial t} = \boldsymbol{\mathcal{C}}\nabla\mathbf{v} - \mathbf{M}(\nabla\cdot\mathbf{w}_{,t}), \tag{1.148}$$

where $\mathbf{v} = \partial\mathbf{u}/\partial t$ is the velocity, $\mathbf{w}_t = \partial\mathbf{w}/\partial t$, and in the last equation we have used the fact that $s = 0$ to eliminate s from it. Hence the Biot equations can be rewritten as

$$\begin{pmatrix} \frac{\partial\boldsymbol{\sigma}}{\partial t} \\ \nabla\cdot\boldsymbol{\sigma} \\ -\nabla P \\ -\frac{\partial P}{\partial t} \\ M\zeta_{,t} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\mathcal{C}} & 0 & 0 & \mathbf{M} & 0 \\ 0 & \rho & \rho_f & 0 & 0 \\ 0 & \rho_f & \widehat{m}_{ij}* & 0 & 0 \\ \mathbf{M} & 0 & 0 & M & M \\ 0 & 0 & 0 & M & M \end{pmatrix} \begin{pmatrix} -\nabla\mathbf{v} \\ \frac{\partial\mathbf{v}}{\partial t} \\ \frac{\partial\mathbf{w}_t}{\partial t} \\ -\nabla\cdot\mathbf{w}_t \\ s \end{pmatrix}. \tag{1.149}$$

Using the fact that s is zero, we have the key identity

$$\begin{pmatrix} \frac{\partial\boldsymbol{\sigma}}{\partial t} \\ \nabla\cdot\boldsymbol{\sigma} \\ -\nabla P \\ -\frac{\partial P}{\partial t} \\ M\zeta_{,t} \end{pmatrix} \cdot \begin{pmatrix} -\nabla\mathbf{v} \\ \frac{\partial\mathbf{v}}{\partial t} \\ -\frac{\partial\mathbf{w}_t}{\partial t} \\ -\nabla\cdot\mathbf{w}_t \\ s \end{pmatrix} = \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\sigma}\frac{\partial\mathbf{v}}{\partial t} - P\frac{\partial\mathbf{w}_t}{\partial t} \\ \boldsymbol{\sigma}\nabla\cdot\mathbf{v} - P\nabla\cdot\mathbf{w}_t \end{pmatrix}. \tag{1.150}$$

1.13 The equations of thermal conduction and diffusion in the time domain

Now let us consider thermal conduction with a time-dependent temperature $T(\mathbf{x}, t)$ and heat flux $\mathbf{q}(\mathbf{x}, t)$. Let $\mathbf{k}(\mathbf{x})$ be the second-order tensor of heat conduction, and let $\alpha(\mathbf{x})$ be the product of the heat capacity per unit mass $C_p(\mathbf{x})$ and the mass density $\rho(\mathbf{x})$. For diffusion we let $T(\mathbf{x}, t)$ denote the particle concentration, $\mathbf{q}(\mathbf{x}, t)$ the particle current, $\mathbf{k}(\mathbf{x})$ be the second-order diffusivity tensor, and we take $\alpha(\mathbf{x}) = 1$. Then the reformulated equations for thermal conductivity and diffusion read as

$$\begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t} \end{pmatrix} = \underbrace{\begin{pmatrix} i\mathbf{k}(\mathbf{x}) & 0 & 0 \\ 0 & 0 & -\frac{i\alpha(\mathbf{x})}{2} \\ 0 & \frac{i\alpha(\mathbf{x})}{2} & 0 \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} \nabla T \\ \frac{\partial T}{\partial t} \\ T \end{pmatrix}. \tag{1.151}$$

To see the connection with the standard equations of thermal conductivity, let us write the equations out explicitly:

$$\begin{aligned} \mathbf{q}_x &= i\mathbf{k}(\mathbf{x})\nabla T, \\ q_t &= -\frac{i\alpha(\mathbf{x})}{2}T, \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t} &= i\nabla\cdot\mathbf{k}(\mathbf{x})\nabla T - \frac{i\alpha(\mathbf{x})}{2}\frac{\partial T}{\partial t} = \frac{i\alpha(\mathbf{x})}{2}\frac{\partial T}{\partial t}. \end{aligned} \tag{1.152}$$

The latter equation implies

$$\alpha(\mathbf{x})\frac{\partial T}{\partial t} = \nabla\cdot\mathbf{k}(\mathbf{x})\nabla T, \tag{1.153}$$

which is the heat or diffusion equation. Clearly the matrix $\mathbf{Z}(\mathbf{x})$ entering the constitutive law has the desired feature that the real matrix $\mathbf{Z}(\mathbf{x})/i$ has a Hermitian part (symmetric part as it is real) which is positive semidefinite. However the Hermitian part of $e^{i\theta}\mathbf{Z}(\mathbf{x})$ is never positive definite for any choice of θ, and is positive semidefinite only when $e^{i\theta} = 1/i$.

The key identity,

$$\begin{aligned} \begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} \nabla T \\ \frac{\partial T}{\partial t} \\ T \end{pmatrix} &= \mathbf{q}_x\cdot\nabla T + q_t\frac{\partial T}{\partial t} + T\nabla\cdot\mathbf{q}_x + T\frac{\partial q_t}{\partial t} \\ &= \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{q}_x T \\ -Tq_t \end{pmatrix} \end{aligned} \tag{1.154}$$

holds.

1.14 The equations of thermoelasticity in the time domain

The equations of thermoelasticity (Chandrasekharaiah 1986; see also Norris 1994) take the form

$$\begin{aligned} \rho\frac{\partial^2\mathbf{u}}{\partial t^2} &= \nabla\cdot\boldsymbol{\sigma}, \quad \sigma_{ij} = C_{ijk\ell}\epsilon_{k\ell} - \beta_{ij}\theta, \\ \rho S &= (\rho c/\theta_0)\theta + \beta_{ij}\epsilon_{ij}, \quad \nabla\cdot\mathbf{q} + \theta_0\rho\frac{\partial S}{\partial t} = 0, \quad \mathbf{q} + \tau\frac{\partial\mathbf{q}}{\partial t} = -\mathbf{k}\nabla\theta, \end{aligned} \tag{1.155}$$

where $\boldsymbol{\sigma}$ is the stress; $\boldsymbol{\epsilon} = [\nabla\mathbf{u} + (\nabla\mathbf{u})^T]$ is the strain; $\mathbf{q}$ is the heat flux; S is the entropy change; θ is the change in temperature above the ambient temperature θ_0; c is the specific heat per unit mass at constant temperature; the $\beta_{ij} = \beta_{ji}$ are essentially coefficients of thermal expansion, $\mathbf{k}$ is the (matrix-valued) thermal conductivity tensor, and τ is a thermal relaxation time. We rewrite the last relation in (1.155) as

$$\mathbf{q} = -\boldsymbol{\kappa} * \nabla\theta, \tag{1.156}$$

where $\boldsymbol{\kappa}$ is a convolution operator in time (and the asterisk $*$ represents a convolution). This allows for a more general spectrum of relaxation times. The thermoelasticity equations can now be written as

$$\begin{pmatrix} \frac{i\partial\boldsymbol{\sigma}}{\partial t} \\ i\nabla\cdot\boldsymbol{\sigma} \\ i\mathbf{q} \\ i\rho S\theta_0 \\ i\left(\nabla\cdot\mathbf{q}+\frac{\partial\rho S\theta_0}{\partial t}\right) \end{pmatrix} = \underbrace{\begin{pmatrix} -i\boldsymbol{\mathcal{C}}\frac{\partial}{\partial t} & 0 & 0 & i\boldsymbol{\beta}\theta_0 & 0 \\ 0 & i\rho\frac{\partial}{\partial t} & 0 & 0 & 0 \\ 0 & 0 & i\theta_0\boldsymbol{\kappa}* & 0 & 0 \\ -i\boldsymbol{\beta}\theta_0 & 0 & 0 & 0 & -i\theta_0\rho c \\ 0 & 0 & 0 & i\theta_0\rho c & -i\theta_0\rho c\frac{\partial}{\partial t} \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} -\nabla\mathbf{u} \\ \frac{\partial\mathbf{u}}{\partial t} \\ -\nabla\theta/\theta_0 \\ -\frac{\partial\theta}{\partial t}/\theta_0 \\ -\theta/\theta_0 \end{pmatrix}. \tag{1.157}$$

Note that the operator $i\partial/\partial t$ which enters three of the diagonal elements of $\mathbf{Z}(\mathbf{x})$ is formally self-adjoint. We have the key identity

$$\begin{pmatrix} \frac{i\partial\boldsymbol{\sigma}}{\partial t} \\ i\nabla\cdot\boldsymbol{\sigma} \\ i\mathbf{q} \\ i\rho S\theta_0 \\ i\left(\nabla\cdot\mathbf{q}+\frac{\partial\rho S\theta_0}{\partial t}\right) \end{pmatrix} \cdot \begin{pmatrix} -\nabla\mathbf{u} \\ \frac{\partial\mathbf{u}}{\partial t} \\ -\nabla\theta/\theta_0 \\ -\frac{1}{\theta_0}\frac{\partial\theta}{\partial t} \\ -\theta/\theta_0 \end{pmatrix} = \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} i\boldsymbol{\sigma}\frac{\partial\mathbf{u}}{\partial t} - i\mathbf{q}\theta/\theta_0 \\ i\boldsymbol{\sigma}\nabla\cdot\mathbf{u} + i\rho S\theta \end{pmatrix}. \tag{1.158}$$

1.15 Maxwell's equations in the time domain

Guided by (1.101) and (1.102), it might seem that an appropriate form for Maxwell's equations in the time domain would be

$$\begin{pmatrix} \frac{\partial\mathbf{h}}{\partial t} \\ \nabla\times\mathbf{h} \end{pmatrix} = \underbrace{\begin{pmatrix} -\boldsymbol{\mu}^{-1} & 0 \\ 0 & \boldsymbol{\varepsilon} \end{pmatrix}}_{\mathbf{Z}} \begin{pmatrix} \nabla\times\mathbf{e} \\ \frac{\partial\mathbf{e}}{\partial t} \end{pmatrix}. \tag{1.159}$$

In this form the key identity does not hold. Alternatively, we can write the system as

$$\begin{pmatrix} -\frac{\partial\mathbf{h}}{\partial t} \\ \nabla\times\mathbf{h} \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{\mu}^{-1} & 0 \\ 0 & \boldsymbol{\varepsilon} \end{pmatrix}}_{\tilde{\mathbf{Z}}} \begin{pmatrix} \nabla\times\mathbf{e} \\ \frac{\partial\mathbf{e}}{\partial t} \end{pmatrix}, \tag{1.160}$$

and then the key identity will hold. To show this, introduce the antisymmetric matrices

$$\underline{\mathbf{h}} = \begin{pmatrix} 0 & -h_3 & h_2 \\ h_3 & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{pmatrix}, \quad \underline{\mathbf{e}} = \begin{pmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_1 & 0 \end{pmatrix}, \tag{1.161}$$

so that $\nabla\times\mathbf{h} = \nabla\cdot\underline{\mathbf{h}}$ and $\nabla\times\mathbf{e} = \nabla\cdot\underline{\mathbf{e}}$ (where the divergence acts on the first index). Then one can check that

$$\begin{aligned} -\frac{\partial\mathbf{h}}{\partial t}\cdot(\nabla\times\mathbf{e}) + \frac{\partial\mathbf{e}}{\partial t}\cdot(\nabla\times\mathbf{h}) &= -\frac{\partial\mathbf{h}}{\partial t}\cdot(\nabla\cdot\underline{\mathbf{e}}) + \frac{\partial\mathbf{e}}{\partial t}\cdot(\nabla\cdot\underline{\mathbf{h}}) + \mathbf{h}\cdot\frac{\partial\nabla\cdot\underline{\mathbf{e}}}{\partial t} - \mathbf{h}\cdot\frac{\partial\nabla\cdot\underline{\mathbf{e}}}{\partial t} \\ &= \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix}\cdot\mathbf{Q}(\mathbf{x}), \end{aligned} \tag{1.162}$$

where

$$\mathbf{Q} = \begin{pmatrix} \mathbf{h} \cdot \frac{\partial \mathbf{e}}{\partial t} \\ \mathbf{h} \cdot (\nabla \cdot \underline{\mathbf{e}}) \end{pmatrix}, \tag{1.163}$$

in which we have used the identity

$$\frac{\partial \mathbf{e}}{\partial t} \cdot (\nabla \cdot \underline{\mathbf{h}}) = \mathrm{Tr}[(\nabla \mathbf{h})^T \frac{\partial \underline{\mathbf{e}}}{\partial t}], \tag{1.164}$$

that in index notation takes the form

$$\frac{\partial e_\ell}{\partial t} \frac{\partial}{\partial x_k}[-e_{k\ell i} h_i] = h_{i,k}[-e_{ik\ell} \frac{\partial e_\ell}{\partial t}], \tag{1.165}$$

and is satisfied because $e_{k\ell i} = e_{ik\ell}$, where $e_{ik\ell}$ is the completely antisymmetric Levi-Civita tensor taking the value $+1$ (or -1) if $ik\ell$ is an even (or odd) permutation of 123, and is zero otherwise.

However, this does not seem like the right form for the constitutive equation as when the response is non-local in time, and one Fourier transforms the constitutive equation in time one sees that the Fourier transform of $\widetilde{\mathbf{Z}}$ entering (1.160) will generally *not* have positive semi-definite imaginary part.

Instead let us look for a solution using the formulation of special relativity theory including the vector potential $\mathbf{\Phi}$ of the electromagnetic field. We set $x_4 = t$ and the fields we take are essentially those in C of table 2 on page 55 of Post (1962) (with the rows and columns changed to account for the different choice of coordinates, and with my $\mathbf{G}$ being essentially Post's $-\mathfrak{G}$), namely

$$\mathbf{F} = \begin{pmatrix} 0 & b_3 & -b_2 & e_1 \\ -b_3 & 0 & b_1 & e_2 \\ b_2 & -b_1 & 0 & e_3 \\ -e_1 & -e_2 & -e_3 & 0 \end{pmatrix}, \tag{1.166}$$

and

$$\mathbf{G} = \begin{pmatrix} 0 & -h_3 & h_2 & d_1 \\ h_3 & 0 & -h_1 & d_2 \\ -h_2 & h_1 & 0 & d_3 \\ -d_1 & -d_2 & -d_3 & 0 \end{pmatrix}. \tag{1.167}$$

Introducing the electromagnetic potentials $\mathbf{\Phi} = (\Phi_1, \Phi_2, \Phi_2)$ and V such that

$$\mathbf{b} = \nabla \times \mathbf{\Phi}, \quad \mathbf{e} = -\nabla V - \frac{\partial \mathbf{\Phi}}{\partial t}, \tag{1.168}$$

and letting $\Phi_4 = -V$ and $\mathbf{\Phi}^0 = (\Phi_1, \Phi_2, \Phi_2, \Phi_4) = (\Phi_1, \Phi_2, \Phi_2, -V)$, Maxwell's equations now read

$$\begin{aligned} F_{ij} &= \Phi^0_{j,i} - \Phi^0_{i,j}, \\ G_{ij,j} &= 0, \end{aligned} \tag{1.169}$$

where we have assumed there are no free current sources. We also assume there are no conduction currents in the body such as resulting from $\mathbf{j} = \boldsymbol{\sigma}\mathbf{e}$, but at the end of this section we will return to this point and incorporate them.

We check for a divergence form:

$$\begin{aligned}\int_{\underset{\sim}{\Omega}} F_{ij}G_{ij} &= \int_{\underset{\sim}{\Omega}} (\Phi^0_{j,i} - \Phi^0_{i,j})G_{ij} \\ &= 2\int_{\underset{\sim}{\Omega}} (\Phi^0_j G_{ij})_{,i} \\ &= 2\int_{\partial\underset{\sim}{\Omega}} \underset{\sim}{n}_i (G_{ij}\Phi^0_j)dS, \end{aligned} \tag{1.170}$$

where we have denoted $\underset{\sim}{\mathbf{n}}$ as the 4 dimensional vector normal to the boundary of the space–time region $\underset{\sim}{\Omega}$. So the key identity holds.

Now look at the constitutive law. In engineering notation it can be written as

$$\begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = \underbrace{\begin{pmatrix} -[\boldsymbol{\mu}(\mathbf{x})]^{-1} & 0 \\ 0 & \boldsymbol{\varepsilon}(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}} \begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix}, \tag{1.171}$$

without destroying the key identity. Now the differential constraints (1.168) (1.169) take the form

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix} = \boldsymbol{\Theta} \begin{pmatrix} \boldsymbol{\Phi} \\ V \end{pmatrix}, \quad \boldsymbol{\Theta}^\dagger \begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = 0, \tag{1.172}$$

where the operator $\boldsymbol{\Theta}$ and its formal adjoint $\boldsymbol{\Theta}^\dagger$ are given by

$$\boldsymbol{\Theta} = \begin{pmatrix} \nabla\times & 0 \\ -\frac{\partial}{\partial t} & -\nabla \end{pmatrix}, \quad \boldsymbol{\Theta}^\dagger = \begin{pmatrix} \nabla\times & \frac{\partial}{\partial t} \\ 0 & \nabla\cdot \end{pmatrix}. \tag{1.173}$$

We can check directly that the key identity holds:

$$\begin{aligned}\int_{\underset{\sim}{\Omega}} \begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix} &= \int_{\underset{\sim}{\Omega}} -\mathbf{h}\cdot(\nabla\times\boldsymbol{\Phi}) - \mathbf{d}\cdot\nabla V - \mathbf{d}\cdot\frac{\partial\boldsymbol{\Phi}}{\partial t} - \frac{\partial\mathbf{d}}{\partial t}\cdot\boldsymbol{\Phi} + (\nabla\times\mathbf{h})\cdot\boldsymbol{\Phi} \\ &= \int_{\partial\underset{\sim}{\Omega}} \mathbf{n}_x\cdot(\mathbf{h}\times\boldsymbol{\Phi}) - \mathbf{n}_x\cdot(V\mathbf{d}) - n_t\mathbf{d}\cdot\boldsymbol{\Phi}, \end{aligned} \tag{1.174}$$

where $\underset{\sim}{\mathbf{n}} = (\mathbf{n}_x, n_t)$ is the normal to the boundary of the space–time body $\underset{\sim}{\Omega}$, and we have used the fact that $\partial\mathbf{d}/\partial t = \nabla\times\mathbf{h}$. The form (1.171) with differential constraints (1.172) are similar but not quite the same as those given in pages 207–210 of Strang (1986), who has a positive definite tensor in the constitutive law at the sacrifice of having operators in the differential constraints which are not formal adjoints.

In general there will be a convolution in time and possibly space, so $\mathbf{Z}$ gets replaced by a convolution operator $\mathbf{K}$:

$$\begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = \mathbf{K} * \begin{pmatrix} \mathbf{b} \\ \mathbf{e} \end{pmatrix}. \tag{1.175}$$

Suppose we only have a convolution in time. Then the constitutive relation in the frequency domain reads as

$$\begin{pmatrix} -\mathbf{h}(\mathbf{x},\omega) \\ \mathbf{d}(\mathbf{x},\omega) \end{pmatrix} = \underbrace{\begin{pmatrix} -[\boldsymbol{\mu}(\mathbf{x},\omega)]^{-1} & 0 \\ 0 & \boldsymbol{\varepsilon}(\mathbf{x},\omega) \end{pmatrix}}_{\mathbf{Z}(\mathbf{x},\omega)} \begin{pmatrix} \mathbf{b}(\mathbf{x},\omega) \\ \mathbf{e}(\mathbf{x},\omega) \end{pmatrix}. \tag{1.176}$$

Here we can allow for conduction currents by incorporating an additional term $i\sigma(\mathbf{x})/\omega$ into the complex permittivity $\boldsymbol{\varepsilon}(\mathbf{x},\omega)$. Now $\mathbf{Z}(\mathbf{x},\omega)$ has the properties that

$$\mathbf{Z}(\mathbf{x},\omega) = \overline{\mathbf{Z}(\mathbf{x},-\overline{\omega})}, \tag{1.177}$$

where the overline denotes complex conjugation, and

$$\operatorname{Im}(\mathbf{Z}(\mathbf{x},\omega)) \geq 0 \quad \text{when } \operatorname{Re}(\omega) \geq 0 \text{ and } \operatorname{Im}(\omega) \geq 0. \tag{1.178}$$

The first property follows from the fact that it is a Fourier transform of a real integral kernel. It implies that $\operatorname{Im}[\boldsymbol{\varepsilon}(\mathbf{x},\omega)] = 0$ and $\operatorname{Im}[\boldsymbol{\mu}(\mathbf{x},\omega)] = 0$ when $\omega = ip$ with p real and positive. Also from the fact that

$$\begin{aligned} \operatorname{Im}[\omega\boldsymbol{\varepsilon}(\mathbf{x},\omega)] &\geq 0 \quad \text{when } \operatorname{Im}\omega \geq 0, \\ \operatorname{Im}[\omega\boldsymbol{\mu}(\mathbf{x},\omega)] &\geq 0 \quad \text{when } \operatorname{Im}\omega \geq 0, \end{aligned} \tag{1.179}$$

we see that $\operatorname{Im}[\boldsymbol{\varepsilon}(\mathbf{x},\omega)] \geq 0$ and $-\operatorname{Im}\{[\boldsymbol{\mu}(\mathbf{x},\omega)]^{-1}\} \geq 0$ in the limit when ω approaches the positive real axis. (Strictly speaking these could be positive measures in this limit). Furthermore as $|\omega| \to \infty$ both $\boldsymbol{\varepsilon}(\mathbf{x},\omega)$ and $\boldsymbol{\mu}(\mathbf{x},\omega)$ approach the permittivity and permeability of free space, $\varepsilon_0\mathbf{I}$ and $\mu_0\mathbf{I}$, since the electrons (because of their inertia) can't respond quickly enough to the rapidly oscillating fields. Finally since $\boldsymbol{\varepsilon}(\mathbf{x},\omega)$ and $\boldsymbol{\mu}(\mathbf{x},\omega)$ are analytic functions of ω when $\operatorname{Im}\omega > 0$ it follows that for any fixed real vector $\mathbf{v}$ both $\operatorname{Im}[\mathbf{v}\cdot\boldsymbol{\varepsilon}(\mathbf{x},\omega)\mathbf{v}]$ and $-\operatorname{Im}\{\mathbf{v}\cdot[\boldsymbol{\mu}(\mathbf{x},\omega)]^{-1}\mathbf{v}\}$ are harmonic functions of ω which must take their infimum in the quadrant $\operatorname{Re}(\omega) \geq 0$ and $\operatorname{Im}(\omega) \geq 0$ as $|\omega| \to \infty$ or at the boundaries of the quadrant $\operatorname{Re}(\omega) \geq 0$ and $\operatorname{Im}(\omega) \geq 0$. This establishes (1.178).

Let $\tau = \omega^2$, and $\omega = \sqrt{\tau}$ where the square root is defined with a branch cut (in the τ plane) just below the negative real axis, so the cut complex τ plane (with the cut along the negative real axis) gets mapped to the upper half plane $\operatorname{Im}(\omega) \geq 0$, and the lower half τ plane gets mapped to the quadrant $\operatorname{Re}(\omega) \geq 0$ and $\operatorname{Im}(\omega) \geq 0$. Then we have

$$\operatorname{Im} \underbrace{\mathbf{Z}(\mathbf{x},\sqrt{\tau})}_{\mathbf{S}(\tau)} \geq 0 \quad \text{when } \operatorname{Im}\tau \leq 0. \tag{1.180}$$

Additionally (1.177) implies $\mathbf{S}(\tau)$ is real symmetric, i.e., $\mathbf{S}(\tau) = \overline{\mathbf{S}(\overline{\tau})}$, and as $\tau \to \infty$, $\mathbf{S}(\tau)$ approaches

$$\mathbf{S}_\infty = \begin{pmatrix} -\mathbf{I}/\mu_0 & 0 \\ 0 & \varepsilon_0\mathbf{I} \end{pmatrix}. \tag{1.181}$$

So $\mathbf{S}(\tau)$ has the integral representation

$$\mathbf{S}(\tau) = \mathbf{S}_\infty + \int_0^\infty \frac{d\mathbf{M}(\tau')}{\tau' + \tau}, \tag{1.182}$$

where $\mathbf{M}(\tau')$ is a 6×6 matrix-valued positive measure (which separates into two 3×3 blocks) when $\mathbf{Z}$ has the form (1.171). The measure must be such that the integral converges which is the case if

$$\int_0^\infty \frac{d\mathbf{M}(\tau')}{\tau' + 1} \tag{1.183}$$

is bounded. The function $\mathbf{S}(\tau) - \mathbf{S}_\infty$ is a matrix-valued Stieltjes function of τ. There are various definitions of Stieltjes functions in the literature. Berg (2008) defines a Stieltjes function $f(z)$, mapping the nonnegative

real axis $z \geq 0$ to the nonnegative real axis, to be a function of the form

$$f(z) = a + \int_0^\infty \frac{d\mu(z')}{z + z'}, \tag{1.184}$$

where $a \geq 0$ and μ is a positive measure on the nonnegative real axis (satisfying $\int 1/(1+z')d\mu(z') < \infty$ to ensure convergence of the above integral), and then points out (1.184) allows one to define $f(z)$ for all z in the cut complex plane, where the cut extends along the negative real z' axis. In a matrix-valued Stieltjes function, a (which is the value of $f(z)$ as $z \to \infty$) is replaced by a positive semidefinite matrix $\mathbf{A}$ and $d\mu(z')$ is replaced by a positive semidefinite matrix-valued measure $d\boldsymbol{\mu}(z')$.

If the medium is moving then it follows (Kaplan and Murnaghan 1930; Post 1962; see also Section 3 of Milton, Briane, and Willis (2006)) that there are *couplings*, i.e., the constitutive relation has the form

$$\begin{pmatrix} -\mathbf{h}(\mathbf{x},t) \\ \mathbf{d}(\mathbf{x},t) \end{pmatrix} = \begin{pmatrix} -[\boldsymbol{\mu}(\mathbf{x},t)]^{-1} & \boldsymbol{\kappa}(\mathbf{x},t) \\ [\boldsymbol{\kappa}(\mathbf{x},t)]^T & \boldsymbol{\varepsilon}(\mathbf{x},t) \end{pmatrix} \begin{pmatrix} \mathbf{b}(\mathbf{x},t) \\ \mathbf{e}(\mathbf{x},t) \end{pmatrix}. \tag{1.185}$$

Naturally one still expects this relation if the medium is moving in different directions with different velocities in different areas. In fact such equations are known as the "bianisotropic equations" of electromagnetism (Serdikukov, Semchenko, Tretkyakov, and Sihvola 2001). Note the fields and tensors could depend on time. Bianisotropic constitutive laws provide one way of understanding chiral effects in electromagnetism, as an alternative to including field gradients in the constitutive law with $\mathbf{d}(\mathbf{x},t)$ depending linearly on $\mathbf{e}(\mathbf{x},t)$ and $\nabla\mathbf{e}(\mathbf{x},t)$.

1.16 A canonical form for Schrödinger's equation in the time domain

To reformulate Schrödinger equation in the time domain for a system with N electrons, let

$$\psi(\mathbf{x},t) = \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, t) \tag{1.186}$$

be the time-dependent wavefunction at time t which is antisymmetric when we interchange $\mathbf{x}_j$ and $\mathbf{x}_k$. Here $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ where each $\mathbf{x}_i = (\mathbf{r}_i, s_i)$ represents a combination of a spatial coordinate $\mathbf{r}_i$ and a spin coordinate s_i, and the joint probability density $\rho(\mathbf{x},t)d\mathbf{r} = \overline{\psi(\mathbf{x},t)}\psi(\mathbf{x},t)d\mathbf{r}$, where $d\mathbf{r} = d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N$, has the quantum mechanical interpretation of being the joint probability density of finding the one electron in the neighborhood $d\mathbf{r}_1$ of $\mathbf{r}_1$ with spin s_1, and a second electron in the neighborhood $d\mathbf{r}_2$ of $\mathbf{r}_2$ with spin s_2, ..., and the remaining electron in the neighborhood $d\mathbf{r}_N$ of $\mathbf{r}_N$ with spin s_N at time t. Let the (time-independent) potential be $V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ and denote Planck's constant divided by 2π as $\hbar$. Then the canonical form of Schrödinger equation for a system with N electrons is

$$\begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t} \end{pmatrix} = \underbrace{\begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & 0 & -\frac{i\hbar}{2} \\ 0 & \frac{i\hbar}{2} & -V \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \begin{pmatrix} \nabla\psi \\ \frac{\partial\psi}{\partial t} \\ \psi \end{pmatrix}. \tag{1.187}$$

where $\mathbf{q}_x(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, t)$ is a $3N$ component vector field, $q_t(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, t)$ is a scalar field, and $\mathbf{A}$ in the simplest approximation is $\hbar^2\mathbf{I}/(2m)$ where m is the mass of the electron, but it may take other forms to

take into account the reduced mass of the electron, or mass polarization terms due to the motion of the atomic nuclei. Computing the dot product we find the usual divergence form and the key identity holds:

$$\begin{aligned}\int_\Omega \begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t}\end{pmatrix}\cdot\begin{pmatrix}\nabla\psi \\ \frac{\partial\psi}{\partial t}\\ \psi\end{pmatrix} &= \int_\Omega \mathbf{q}_x\cdot\nabla\psi + q_t\frac{\partial\psi}{\partial t} + \psi\nabla\cdot\mathbf{q}_x + \psi\frac{\partial q_t}{\partial t}\\ &= \int_\Omega \nabla\cdot(\mathbf{q}_x\psi) + \frac{\partial}{\partial t}(\psi q_t)\\ &= \int_\Omega \underset{\sim}{\nabla}\cdot\mathbf{Q}, \end{aligned} \tag{1.188}$$

where $\mathbf{Q} = (\mathbf{q}_x^T\psi, -q_t\psi)^T$. Since this is a divergence it can be expressed in boundary terms as

$$\int_{\partial\Omega}(\mathbf{n}_x\cdot\mathbf{q}_x + n_t\cdot q_t)\psi. \tag{1.189}$$

Note when we take "dot" products we do not conjugate the left hand side.

Also from the constitutive law we have

$$\begin{aligned}\mathbf{q}_x &= -\mathbf{A}\nabla\psi,\\ q_t &= -\frac{i\hbar}{2}\psi,\\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t} &= -\nabla\cdot\mathbf{A}\nabla\psi - \frac{i\hbar}{2}\frac{\partial\psi}{\partial t} = \frac{i\hbar}{2}\frac{\partial\psi}{\partial t} - V\psi,\end{aligned} \tag{1.190}$$

where the last equation is the Schrödinger equation. We remark in passing that another quite different formulation of the Schrödinger equation has recently given by Ajaib (2015): it uses matrices with special properties in a first order equation, and in this sense is reminiscent of the Dirac equation.

We also have the associated key identity:

$$\begin{aligned}\overline{\begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla\cdot\mathbf{q}_x + \frac{\partial q_t}{\partial t}\end{pmatrix}}\cdot\begin{pmatrix}\nabla\psi \\ \frac{\partial\psi}{\partial t}\\ \psi\end{pmatrix} &= \nabla(\overline{\mathbf{q}_x}\psi) + \frac{\partial}{\partial t}(\psi\overline{q_t})\\ &= \nabla\cdot\left(-(\nabla\overline{\psi})\psi\frac{\hbar^2}{2}\right) + \frac{\partial}{\partial t}\left(-\psi\left(\frac{i\hbar}{2}\right)\overline{\psi}\right),\end{aligned} \tag{1.191}$$

where we have assumed $\mathbf{A} = \mathbf{I}\frac{\hbar^2}{2}, m = 1$. Now take the imaginary part of both sides. We have

$$\mathrm{Im}\left[\overline{\begin{pmatrix}\nabla\psi\\ \frac{\partial\psi}{\partial t}\\ \psi\end{pmatrix}}\cdot\underbrace{\begin{pmatrix}-\mathbf{A} & 0 & 0\\ 0 & 0 & -\frac{i\hbar}{2}\\ 0 & \frac{i\hbar}{2} & -V\end{pmatrix}}_{\mathbf{Z}}\begin{pmatrix}\nabla\psi\\ \frac{\partial\psi}{\partial t}\\ \psi\end{pmatrix}\right] = 0, \tag{1.192}$$

since $\mathbf{Z}$ is Hermitian. This gives

$$\frac{\hbar^2}{4i}\nabla\cdot[\overline{\psi}(\nabla\psi) - (\nabla\overline{\psi})\psi] - \frac{\hbar}{2}\frac{\partial}{\partial t}\underbrace{(\psi\overline{\psi})}_{\rho} = 0, \tag{1.193}$$

which implies

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0, \tag{1.194}$$

with $\mathbf{J} = \frac{i\hbar}{2}[\psi\nabla\overline{\psi} - \overline{\psi}\nabla\psi]$ and $\rho = \psi\overline{\psi}$, which is the well-known conservation law for the probability density ρ.

1.17 Schrödinger equation for a single electron in a magnetic field

Now we take $\hbar = 1$ and we reformulate the Schrödinger equation for the single electron wavefunction $\psi(\mathbf{x}, t)$ in a magnetic field. Let $\mathbf{\Phi}(\mathbf{x}) = (\Phi_1(\mathbf{x}), \Phi_2(\mathbf{x}), \Phi_3(\mathbf{x}))$ be the time-independent magnetic potential, with $\mathbf{b} = \nabla \times \mathbf{\Phi}$ the magnetic induction, $V(\mathbf{x})$ the time-independent electric potential, e is the charge on the electron, and m its mass. Then the canonical form for the Schrödinger equation in a magnetic field is then

$$\begin{pmatrix} \mathbf{q}_x \\ q_t \\ \nabla \cdot \mathbf{q}_x + \frac{\partial q_t}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{-\mathbf{I}}{2m} & 0 & \frac{ie\mathbf{\Phi}}{2m} \\ 0 & 0 & -\frac{i}{2} \\ \frac{-ie\mathbf{\Phi}}{2m} & +\frac{i}{2} & -eV \end{pmatrix} \begin{pmatrix} \nabla\psi \\ \frac{\partial\psi}{\partial t} \\ \psi \end{pmatrix}. \tag{1.195}$$

The key identity still holds and

$$\begin{aligned} \mathbf{q}_x &= -\frac{1}{2m}\nabla\psi + \frac{ie\mathbf{\Phi}}{2m}\psi, \\ q_t &= -\frac{i}{2}\psi, \\ \nabla \cdot \mathbf{q}_x + \frac{\partial q_t}{\partial t} &= -\frac{\nabla^2\psi}{2m} + \frac{i\nabla\cdot(e\mathbf{\Phi}\psi)}{2m} - \frac{i}{2}\frac{\partial\psi}{\partial t} = -\frac{ie\mathbf{\Phi}}{2m}\nabla\psi + \frac{i}{2}\frac{\partial\psi}{\partial t} - eV\psi. \end{aligned} \tag{1.196}$$

Thus we have

$$i\frac{\partial\psi}{\partial t} = \frac{1}{2m}[i\nabla + e\mathbf{\Phi}]^2\psi + eV\psi, \tag{1.197}$$

which is the Schrödinger equation in a magnetic field.

1.18 Rewriting the Dirac equation

The Dirac equation for the electron can be represented using the 4×4 matrices

$$\begin{aligned} i\boldsymbol{\gamma}^0 &= \begin{pmatrix} i\mathbf{I}_2 & 0 \\ 0 & -i\mathbf{I}_2 \end{pmatrix}, \qquad \boldsymbol{\gamma}^1 = \begin{pmatrix} 0 & \boldsymbol{\sigma}_x \\ -\boldsymbol{\sigma}_x & 0 \end{pmatrix}, \\ \boldsymbol{\gamma}^2 &= \begin{pmatrix} 0 & \boldsymbol{\sigma}_y \\ -\boldsymbol{\sigma}_y & 0 \end{pmatrix}, \qquad \boldsymbol{\gamma}^3 = \begin{pmatrix} 0 & \boldsymbol{\sigma}_z \\ -\boldsymbol{\sigma}_z & 0 \end{pmatrix}, \end{aligned} \tag{1.198}$$

where $\mathbf{I}_2$ is the 2×2 identity matrix $\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_z$ are the Pauli matrices so that

$$\begin{aligned} i\mathbf{I}_2 &= \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \qquad \boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \boldsymbol{\sigma}_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{1.199}$$

Setting the speed of light $c = 1$, Dirac's equation for the electron wavefunction $\boldsymbol{\psi}$, which is a 4 component vector, takes the form

$$i\hbar\boldsymbol{\gamma}^\mu \frac{\partial \boldsymbol{\psi}}{\partial x_\mu} = m\boldsymbol{\psi}, \tag{1.200}$$

in which $(x_1, x_2, x_3, x_4) = (x, y, z, t)$ and m is the mass of the electron. It can be reformulated as

$$\underbrace{\begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_t \\ \underset{\sim}{\nabla} \cdot \mathbf{q} \end{pmatrix}}_{\mathcal{F}} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & \boldsymbol{\gamma}^1 \\ 0 & 0 & 0 & 0 & \boldsymbol{\gamma}^2 \\ 0 & 0 & 0 & 0 & \boldsymbol{\gamma}^3 \\ 0 & 0 & 0 & 0 & i\boldsymbol{\gamma}^0 \\ -\boldsymbol{\gamma}^1 & -\boldsymbol{\gamma}^2 & -\boldsymbol{\gamma}^3 & -i\boldsymbol{\gamma}^0 & \frac{2m}{\hbar}\mathbf{I} \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} i\frac{\partial \boldsymbol{\psi}}{\partial x} \\ i\frac{\partial \boldsymbol{\psi}}{\partial y} \\ i\frac{\partial \boldsymbol{\psi}}{\partial z} \\ \frac{\partial \boldsymbol{\psi}}{\partial t} \\ \boldsymbol{\psi} \end{pmatrix}}_{\mathcal{G}}. \tag{1.201}$$

Here $\underset{\sim}{\nabla} = (i\frac{\partial}{\partial x}, i\frac{\partial}{\partial y}, i\frac{\partial}{\partial z}, \frac{\partial}{\partial t})$ and $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_t)$ is a 4×4 matrix. Thus

$$\underset{\sim}{\nabla} \cdot \mathbf{q} = i\frac{\partial \mathbf{q}_1}{\partial x} + i\frac{\partial \mathbf{q}_2}{\partial y} + i\frac{\partial \mathbf{q}_3}{\partial z} + \frac{\partial \mathbf{q}_t}{\partial t} \tag{1.202}$$

is a 4-component vector. Note that the matrix $\mathbf{Z}$ entering the constitutive relation is Hermitian because $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \boldsymbol{\gamma}^3, i\boldsymbol{\gamma}^0$ are all anti-Hermitian.

Now let us verify this corresponds to the Dirac equation:

$$\begin{aligned} \mathbf{q}_1 &= \boldsymbol{\gamma}^1\boldsymbol{\psi}, \quad \mathbf{q}_2 = \boldsymbol{\gamma}^2\boldsymbol{\psi}, \quad \mathbf{q}_3 = \boldsymbol{\gamma}^3\boldsymbol{\psi}, \quad \mathbf{q}_t = i\boldsymbol{\gamma}^0\boldsymbol{\psi}, \\ \underset{\sim}{\nabla} \cdot \mathbf{q} &= i\nabla \cdot \mathbf{q}_x + \frac{\partial \mathbf{q}_t}{\partial t} \\ &= i\left[\boldsymbol{\gamma}^1\frac{\partial \boldsymbol{\psi}}{\partial x} + \boldsymbol{\gamma}^2\frac{\partial \boldsymbol{\psi}}{\partial y} + \boldsymbol{\gamma}^3\frac{\partial \boldsymbol{\psi}}{\partial z}\right] + i\boldsymbol{\gamma}^0\frac{\partial \boldsymbol{\psi}}{\partial t} \\ &= -i\boldsymbol{\gamma}^1\frac{\partial \boldsymbol{\psi}}{\partial x} - i\boldsymbol{\gamma}^2\frac{\partial \boldsymbol{\psi}}{\partial y} - i\boldsymbol{\gamma}^3\frac{\partial \boldsymbol{\psi}}{\partial z} - i\boldsymbol{\gamma}^0\frac{\partial \boldsymbol{\psi}}{\partial t} + \frac{2m}{\hbar}\boldsymbol{\psi}, \end{aligned} \tag{1.203}$$

and from the last two lines we see this is indeed equivalent to (1.200).

The equations (1.201) can be abbreviated as

$$\begin{pmatrix} \mathbf{q}_x \\ \mathbf{q}_t \\ i\nabla \cdot \mathbf{q}_x + \frac{\partial \mathbf{q}_t}{\partial t} \end{pmatrix} = \mathbf{Z} \begin{pmatrix} i\nabla\boldsymbol{\psi} \\ \frac{\partial \boldsymbol{\psi}}{\partial t} \\ \boldsymbol{\psi} \end{pmatrix}. \tag{1.204}$$

Checking the key identity we find

$$\begin{aligned} \begin{pmatrix} \mathbf{q}_x \\ \mathbf{q}_t \\ i\nabla \cdot \mathbf{q}_x + \frac{\partial \mathbf{q}_t}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} i\nabla\boldsymbol{\psi} \\ \frac{\partial \boldsymbol{\psi}}{\partial t} \\ \boldsymbol{\psi} \end{pmatrix} &= i\mathbf{q}_x \cdot \nabla\boldsymbol{\psi} + \mathbf{q}_t \cdot \frac{\partial \boldsymbol{\psi}}{\partial t} + i(\nabla \cdot \mathbf{q}_x)\boldsymbol{\psi} + \frac{\partial \mathbf{q}_t}{\partial t}\boldsymbol{\psi} \\ &= \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} i\mathbf{q}_x\boldsymbol{\psi} \\ -\mathbf{q}_t\boldsymbol{\psi} \end{pmatrix} = \underset{\sim}{\nabla} \cdot \mathbf{Q}(\mathbf{x}, t). \end{aligned} \tag{1.205}$$

1.19 Adding sources, regularizing the equations, and making them into periodic composite problems

Source terms are easily handled for any of the problems we have discussed, in a similar way that the source term $\boldsymbol{\alpha}(\mathbf{x})\theta$ is handled in the thermoelasticity equation (1.20).

Suppose $\mathbf{h}(\mathbf{x}, t)$ is the source term and θ is a constant scalar, and we have the relation

$$\mathcal{G}(\mathbf{x}, t) = \mathbf{Z}(\mathbf{x}, t)\mathcal{F}(\mathbf{x}, t) + \theta\mathbf{h}(\mathbf{x}, t). \tag{1.206}$$

Then we can reformulate this as

$$\begin{pmatrix} \mathcal{G}(\mathbf{x}, t) \\ \nabla \cdot \mathbf{r}(\mathbf{x}, t) \end{pmatrix} = \begin{pmatrix} \mathbf{Z}(\mathbf{x}, t) & \mathbf{h}(\mathbf{x}, t) \\ \mathbf{h}(\mathbf{x}, t)^T & c(\mathbf{x}, t) \end{pmatrix} \begin{pmatrix} \mathcal{F}(\mathbf{x}, t) \\ \theta \end{pmatrix}. \tag{1.207}$$

Here $\mathbf{r}(\mathbf{x}, t)$ is a field not subject to differential constraints, and $c(\mathbf{x}, t)$ can be chosen as we please. The main observation is this: since $\mathcal{G} \cdot \mathcal{F}$ can be written as a divergence, i.e., $\mathcal{G}(\mathbf{x}, t) \cdot \mathcal{F}(\mathbf{x}, t) = \underset{\sim}{\nabla} \cdot \mathbf{Q}(\mathbf{x}, t)$, we have the key identity

$$\int_\Omega \begin{pmatrix} \mathcal{G}(\mathbf{x}, t) \\ \nabla \cdot \mathbf{r}(\mathbf{x}, t) \end{pmatrix} \cdot \begin{pmatrix} \mathcal{F}(\mathbf{x}, t) \\ \theta \end{pmatrix} = \int_{\partial\Omega} \mathbf{n} \cdot [\mathbf{Q}(\mathbf{x}, t) + \theta\mathbf{r}(\mathbf{x}, t)]. \tag{1.208}$$

Again, however, we have the caveat that this may not be so useful as $\mathbf{r}(\mathbf{x}, t)$ cannot be directly obtained from boundary measurements.

In many of the equations we have discussed the matrix $\mathbf{Z}(\mathbf{x}, t)$ is Hermitian, or such that the Hermitian part of $\mathbf{Z}(\mathbf{x}, t)/i$ is positive semidefinite, rather than positive definite. To regularize the problem we can add a small imaginary part $i\delta\mathbf{I}$ to $\mathbf{Z}(\mathbf{x}, t)$. Also in the presence of source terms we can assume $c(\mathbf{x}, t)$ has an imaginary part. That is, instead of (1.207) we can consider the equations

$$\underbrace{\begin{pmatrix} \mathcal{G}(\mathbf{x}, t) \\ \nabla \cdot \mathbf{r}(\mathbf{x}, t) \end{pmatrix}}_{\underline{\mathcal{G}}} = \underbrace{\begin{pmatrix} \mathbf{Z}(\mathbf{x}, t) + i\delta\mathbf{I} & \mathbf{h}(\mathbf{x}, t) \\ \mathbf{h}(\mathbf{x}, t)^T & c(\mathbf{x}, t) \end{pmatrix}}_{\underline{\mathbf{Z}}} \underbrace{\begin{pmatrix} \mathcal{F}(\mathbf{x}, t) \\ \theta \end{pmatrix}}_{\underline{\mathcal{F}}}. \tag{1.209}$$

We may now look for solutions to these equations where $\underline{\mathcal{G}(\mathbf{x}, t)}$, $\underline{\mathcal{F}(\mathbf{x}, t)}$, $\underline{\mathbf{Z}(\mathbf{x}, t)}$ and $\mathbf{Q}(\mathbf{x}, t)$ are periodic functions with a common unit cell of periodicity Ω in space–time. Here we are focusing on wave equations: for static or quasistatic equations one expects $\mathbf{Q}(\mathbf{x}, t)$ to be a sum of a linear part plus a periodic part. Due to the periodicity of $\mathbf{Q}$,

$$\int_\Omega \mathcal{G}(\mathbf{x}, t) \cdot \mathcal{F}(\mathbf{x}, t) = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{Q}(\mathbf{x}, t) = 0, \tag{1.210}$$

since at opposing points across the cell $\mathbf{Q}(\mathbf{x}, t)$ takes the same value, while the outward normal $\mathbf{n}$ has opposite directions. Hence $\mathcal{G}(\mathbf{x}, t)$ and $\mathcal{F}(\mathbf{x}, t)$ lie in orthogonal subspaces.

Now let us define the space $\mathcal{E}$ to consist of all Ω-periodic fields $\mathbf{E}$ that are square integrable in the unit cell, of the form

$$\mathbf{E} = \begin{pmatrix} \mathcal{F}(\mathbf{x}, t) \\ 0 \end{pmatrix}, \tag{1.211}$$

where $\mathcal{F}(\mathbf{x}, t)$ satisfies the differential constraints appropriate to that field. We define $\mathcal{J}$ to consist of all Ω-periodic fields $\mathbf{J}$ that are square integrable in the unit cell, of the form

$$\mathbf{J} = \begin{pmatrix} \mathcal{G}(\mathbf{x}, t) \\ S(\mathbf{x}) \end{pmatrix}, \tag{1.212}$$

where $\mathcal{G}(\mathbf{x}, t)$ satisfies the complementary differential constraints appropriate to it, and the Ω-periodic field $S(\mathbf{x})$ has zero average value over the unit cell, but is not subject to any differential constraints. The space $\mathcal{U}$ consists of all constant fields of the form

$$\mathbf{U} = \begin{pmatrix} 0 \\ c \end{pmatrix}, \tag{1.213}$$

where c is a constant. These three spaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are mutually orthogonal and the problem of solving (1.209) is equivalent to the following: given $\mathbf{E}_0 \in \mathcal{U}$ (with $c = \theta$), find fields $\mathbf{E} \in \mathcal{E}$, $\mathbf{J} \in \mathcal{J}$, and $\mathbf{J}_0 \in \mathcal{U}$ such that

$$\mathbf{J}_0 + \mathbf{J}(\mathbf{x}, t) = \underline{\mathbf{Z}}(\mathbf{x}, t)[\mathbf{E}_0 + \mathbf{E}(\mathbf{x}, t)], \tag{1.214}$$

which as we will see in the next chapter is a problem in the abstract theory of composites. A similar formulation can be applied to time-harmonic problems, with ω replacing t, and with the fields $\underline{\mathcal{G}(\mathbf{x}, \omega)}$, $\underline{\mathcal{F}(\mathbf{x}, \omega)}$, $\underline{\mathbf{Z}(\mathbf{x}, \omega)}$ and $\mathbf{Q}(\mathbf{x}, \omega)$ being periodic functions with a common unit cell of periodicity Ω in space. In this setting quasiperiodicity (Bloch wave) conditions are natural on the cell of periodicity of $\underline{\mathbf{Z}(\mathbf{x}, \omega)}$, so to allow for this one should take Ω to include many unit cells of $\underline{\mathbf{Z}(\mathbf{x}, \omega)}$.

We may, for example, solve (1.214) using Fast Fourier Transform methods (Moulinec and Suquet, 1994, 1998; Eyre and Milton 1999; Moulinec and Silva 2014; see also Section 14.11 of Milton 2002), although the main interest will probably be in finding solutions when the regularization parameter δ entering (1.209) is small, and in this case the convergence of the Fast-Fourier transform algorithms may be slow.

2 Composites and the associated abstract theory

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

We provide an introduction to composites and review the accompanying abstract theory, where effective tensors are associated with a linear operator acting in a Hilbert, or finite-dimensional vector space with an inner product, that is decomposed into three orthogonal subspaces. The purpose of the chapter is to provide the necessary background for the reader in a handy place relative to the rest of the text. It is shown that convergent series expansions can be developed for the fields, and minimization variational principles developed for the effective tensor. We also review the theory of the $\mathbf{Y}_*$ tensor. This tensor is associated with a Hilbert, or finite-dimensional vector space with an inner product, that is decomposed into two orthogonal subspaces in two different ways, with a linear operator and accompanying $\mathbf{Y}_*$ tensor each acting on subspaces which are orthogonal complements. The transformation from the effective tensor to the $\mathbf{Y}_*$ tensor had its origins in the theory of composites for simplifying bounds, and for preserving analytic properties. The $\mathbf{Y}_*$ tensor is also naturally associated with resistor networks and the set of batteries that power the network.

2.1 An introduction to composite materials

The purpose of this chapter is to provide a brief introduction to composites and the associated abstract theory, for those readers not familiar with it, in a convenient place relative to the rest of the text.

Theoretical studies of composites have a long history: see the excellent review of Landauer (1978) who mentions that in 1837 Faraday had proposed a model of dielectrics consisting of metallic inclusions separated from each other by insulating material, and in 1846 Mossotti had submitted a paper "Analytical discussion of the influence which the action of a dielectric medium exerts on the distribution of the electricity on the surfaces of several electric bodies dispersed in it", but which was not published until 1850, as in the meantime he had fought in the first war for Italian independence at the head of a battalion of students at the University of Pisa, and was subsequently made prisoner.

Examples of composites include porous rock containing oil or salt water (of obvious interest to the oil industry); suspensions or colloids such as clouds, fog, mist, rain, dusty air, aerosols, or milk (which has butterfat globules in a watery solution), flour in water; silts or clays; carbon fiber composites; reinforced concrete; fiberglass in resin; wood; paper; sheep and steel wool; cotton; artificial and natural opals (consisting

of periodic arrays of silica spheres); polycrystalline materials such as metals, basalt and granite; sintered materials; foams; shape memory materials; sea ice containing pockets of sea water in ice; snow; ceramics; materials with microcracks; magnets (where magnetic domains cause inhomogeneities); liquid crystals with a spatially varying order-parameter; air (where thermal fluctuations cause inhomogeneities as manifested in the twinkling of stars); bone and biological tissue; and even chocolate chip ice-cream. Alloys, gels, glasses, and rubbers are generally not regarded as composites, as the inhomogeneities are on the atomic or molecular scale rather than on a larger scale, and one needs quantum equations rather than classical equations to predict their macroscopic properties.

In general a composite is periodic or statistically homogeneous on some length scale: if it is not we call it an inhomogeneous body. Roughly speaking, statistical homogeneity means that correlation functions, such as the probability that a polyhedron will land with all vertices in one phase when dropped randomly in say a sufficiently large cubic sample of the composite (called a Representative Volume Element (RVE) : see Hashin 1983 and Nemat-Nasser and Hori 1999), will be essentially independent of where that cubic sample was taken from (within a range determined by the length scale at which the material is statistically homogeneous). This definition is not so precise, but this is fitting as most materials are only composites in an approximate sense. Assumptions such as periodicity or translational invariance of ensembles of materials [See the books of Beran (1968); Bensoussan, Lions, and Papanicolaou (1978); Bakhvalov and Panasenko (1989); and Zhikov, Kozlov, and Oleinik (1994), and the article of Kozlov (1978)] lend precision to the meaning of a composite material and the associated definitions of effective tensors, but it is to be keep in mind that these are idealizations. We will see, in **Chapter 3**, that the abstract theory of composites also applies to inhomogeneous bodies, without the assumption of periodicity or statistical homogeneity.

In the theory of composites one is interested in the effective moduli of materials. The effective moduli govern how the material responds on the macroscopic scale: for a material which is periodic on the microscale, it behaves almost the same as a homogeneous material with moduli the same as the effective moduli. The mathematical framework for giving a precise meaning for this is homogenization theory. To describe a sequence of say conducting materials with finer and finer microstructure, we take a periodic function $\boldsymbol{\sigma}(\mathbf{x})$, where $\mathbf{x}$ takes values in 3-dimensions, and consider the sequence $\boldsymbol{\sigma}_\epsilon(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x}/\epsilon)$ of conductivity tensor fields parameterized by a real positive number ϵ and consider what equation the associated electrical potential $V_\epsilon(\mathbf{x})$ satisfies in the limit as $\epsilon \to 0$. To fix things so that the solutions are unique we could, for example, assume $\mathbf{x}$ is constrained to lie inside a body Ω which is independent of ϵ, and require that $V_\epsilon(\mathbf{x})$ takes an ϵ-independent value $V_0(\mathbf{x})$ at the boundary $\partial\Omega$. Additionally we require that the matrix representing the conductivity tensor $\boldsymbol{\sigma}(\mathbf{y})$ is symmetric and real, and that $\boldsymbol{\sigma}(\mathbf{y})$ is bounded, i.e., there exists a constant $\beta > 0$ such that

$$\beta\mathbf{I} \geq \boldsymbol{\sigma}(\mathbf{y}) \quad \text{for all } \mathbf{y}, \tag{2.1}$$

(where the inequality holds in the sense of quadratic forms, implying that $\beta\mathbf{I} - \boldsymbol{\sigma}(\mathbf{y})$ is a positive semidefinite matrix for all $\mathbf{y}$) and that $\boldsymbol{\sigma}(\mathbf{y})$ is coercive, i.e., there exists a constant $\alpha > 0$ such that

$$\boldsymbol{\sigma}(\mathbf{y}) \geq \alpha\mathbf{I} \quad \text{for all } \mathbf{y}. \tag{2.2}$$

Then it can be shown that if $V_\epsilon(\mathbf{x})$ satisfies

$$\nabla \cdot \boldsymbol{\sigma}(\mathbf{x}/\epsilon)\nabla V_\epsilon(\mathbf{x}) = -f(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega, \quad V_\epsilon(\mathbf{x}) = V_0(\mathbf{x}) \text{ on } \partial\Omega, \tag{2.3}$$

where the source of current $f(\mathbf{x})$ is assumed to be independent of ϵ, then the potential $V_\epsilon(\mathbf{x})$ converges (strongly) to the potential $V_*(\mathbf{x})$ satisfying the homogenized equation :

$$\nabla \cdot \boldsymbol{\sigma}_* \nabla V_*(\mathbf{x}) = -f(\mathbf{x}), \quad \text{for all } \mathbf{x} \in \Omega, \quad V_*(\mathbf{x}) = V_0(\mathbf{x}) \text{ on } \partial\Omega, \tag{2.4}$$

where $\boldsymbol{\sigma}_*$ is the effective conductivity tensor. One remark is that while $V_\epsilon(\mathbf{x})$ converges to $V_*(\mathbf{x})$, $\nabla V_\epsilon(\mathbf{x})$ only converges to $\nabla V_*(\mathbf{x})$ in a weak sense, as fluctuations in the electric field still remain. The weak convergence implies the volume integral of $g(\mathbf{x})\nabla V_\epsilon(\mathbf{x})$ converges to $g(\mathbf{x})\nabla V_*(\mathbf{x})$ for all (ϵ-independent) smooth scalar-valued test functions $g(\mathbf{x})$ with compact support contained in Ω. One can generalize the result to functions $\boldsymbol{\sigma}_\epsilon(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x}, \mathbf{x}/\epsilon)$ which depend on both the "slow variable" $\mathbf{x}$, and the "fast variable" $\mathbf{x}/\epsilon$, where $\boldsymbol{\sigma}(\mathbf{x}, \mathbf{y})$ is periodic in $\mathbf{y}$: in this case the associated effective tensor $\boldsymbol{\sigma}_*(\mathbf{x})$ depends on position, and at a point $\mathbf{x}_0$, $\boldsymbol{\sigma}_*(\mathbf{x}_0)$ is the same effective tensor as that associated with the periodic material $\boldsymbol{\sigma}(\mathbf{x}_0, \mathbf{x}/\epsilon)$. One can also consider materials where there are inhomogeneities on multiple, well-separated, length scales: a model example is where the local conductivity takes the form $\boldsymbol{\sigma}_\epsilon(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x}, \mathbf{x}/\epsilon, \mathbf{x}/\epsilon^2, \mathbf{x}/\epsilon^3, \ldots, \mathbf{x}/\epsilon^n)$ where $\boldsymbol{\sigma}(\mathbf{x}, \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \mathbf{y}^{(3)}, \ldots, \mathbf{y}^{(n)})$ is periodic in each variable $\mathbf{y}^{(j)}$, $j = 1, 2, \ldots, n$. Thus homogenization is a bit like quantum mechanics: one is interested in functions of one variable $\mathbf{x}$ (such as the electron density in quantum mechanics) but to obtain it one needs to solve equations involving functions of multiple variables (such as the electron wavefunction in quantum mechanics). In quantum mechanics the number of dimensions is related to the number of particles, while in homogenization it is related to the number of length scales. Periodic homogenization is described in the books of Bensoussan, Lions, and Papanicolaou (1978); Sanchez-Palencia (1980); Bakhvalov and Panasenko (1989); Persson, Persson, Svanstedt, and Wyller (1993); and Zhikov, Kozlov, and Oleinik (1994). The two-scale and multiscale treatments of Nguetseng (1989), Allaire (1992), and Allaire and Briane (1996), provide a rigorous basis for the method.

Homogenization theory also applies to random, but statistically homogeneous, materials: these can be treated by considering an ensemble of such materials and taking ensemble averages rather than volume averages. [See the books of Beran (1968); Bensoussan, Lions, and Papanicolaou (1978); Bakhvalov and Panasenko (1989); and Zhikov, Kozlov, and Oleinik (1994), and the article of Kozlov (1978).] It has been shown by Papanicolaou and Varadhan (1982) and Golden and Papanicolaou (1983) that the ensemble averaged definition of the effective conductivity tensor agrees with the more physical definition where a cubic sample of the composite is taken and then $\boldsymbol{\sigma}_*$ is obtained in an infinite volume limit as the size of the cube tends to infinity.

In fact, amazingly, homogenization theory in the framework of $G-$, $H-$, or $\Gamma-$convergence (which we not fully describe as they will not be needed in this book) applies to any real symmetric matrix-valued sequence $\boldsymbol{\sigma}_\epsilon(\mathbf{x})$ which is bounded and coercive (with constants β and α that are independent of ϵ) provided one takes an appropriate subsequence. [See for example, the articles of De Giorgi (1984), Allaire (1997), and Murat and Tartar (1997), the books of Buttazzo (1989), Dal Maso (1993), Zhikov, Kozlov, and Oleinik (1994), and Attouch (1984), and the lecture notes of Raitums (1997). Allaire (1997), in particular, provides an excellent short summary of the different approaches.] Roughly speaking, the restriction to a subsequence allows one to remove parts of the sequence where in some region, for example, $\boldsymbol{\sigma}_\epsilon(\mathbf{x})$ oscillates as ϵ varies but is independent of $\mathbf{x}$. This general perspective is beautiful, but has the difficulty that in physical experiments one is given a single material and it is difficult to imagine what an associated sequence $\boldsymbol{\sigma}_\epsilon(\mathbf{x})$ could be.

A suspension of bubbles in water is very good at screening sound (and hence has been used to cloak the sound of submarines): it behaves as an effective material with a large damping to compression oscillations (essentially the water near the bubbles is sheared, and the shear viscosity of the water is converted to a compressional viscosity of the bubbly fluid). It's easy to do an experiment yourself: take an empty wine glass (filled with air) — it rings when struck with a table knife. Fill it with water — it again rings. But add Alka–Seltzer and wait until it effervesces — the composite of air bubbles in water only produces a dull thud when the glass is struck by the knife. The brilliant reds of old stained glass windows, and the colors of the Roman Lycurgus cup, (`https://en.wikipedia.org/wiki/Lycurgus_Cup`) come from the effective electromagnetic properties of suspensions of tiny gold and silver particles, not from any chemical interactions. A sponge rubber behaves as an effective elastic material, the holes only becoming evident on closer inspection. An array

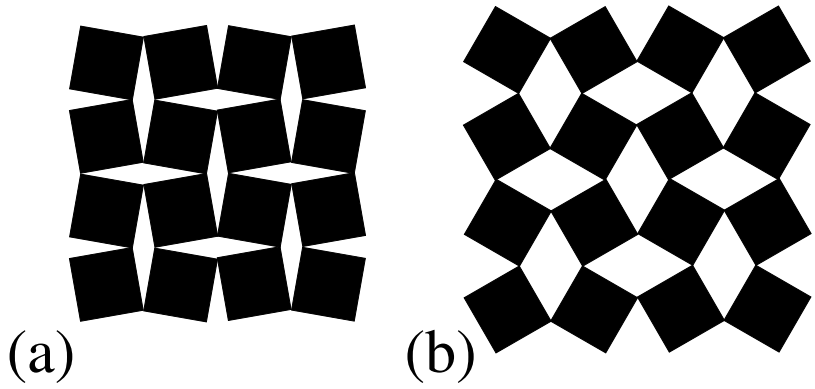


Figure 2.1: Grima and Evans (2000) realized that an array of rotating rigid squares, as shown in (a), attached at appropriate corners by flexible junctions would have a Poisson's ratio of -1 over a finite range of deformation, i.e., a dilation would be the only easy mode of deformation. [An earlier, but more complicated, model having this property is the hexagonal spoked model of Milton (1992).] Under stretching the material deforms from (a) to (b). Their model simplified a related one given in Figure 4 of Sigmund (1995). Three dimensional materials for which a dilation is the only easy mode of deformation over a finite deformation range have been constructed by Sigmund (1995), Milton (2013a), Bückmann, Schittny, Thiel, Kadic, Milton, and Wegener (2014) and Milton (2015b).

of parallel conducting wires in a dielectric medium has an effective conductivity which is highly anisotropic, being greatest in the direction of the wires. The effective properties of rock containing oil and salt water are important to the oil industry, a high effective conductivity indicating the presence of salt water rather than oil. The effective properties of polycrystalline rocks are important to geophysicists in seismic studies. Metals can be treated to control their polycrystalline structure to obtain desired properties The effective properties of carbon fiber composites are important in applications. The theory of composites is discussed in the books of Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009).

There are surprises in the theory of composites: it is possible to combine materials which narrow when stretched to obtain (auxetic) composites which widen when stretched (Lakes 1987; Milton 1992; Greaves, Greer, Lakes, and Rouxel 2011). A beautiful example of a composite which does this is illustrated in **Figure 2.1**. It is also possible to combine three materials (or two plus void) all of which by themselves expand when heated to obtain a composite which contracts when heated, or alternatively which expands more than the constituent materials (Lakes 1996; Sigmund and Torquato 1996, 1997). An example showing how one can get negative thermal expansion from positive thermal expansion is shown in **Figure 2.2**. One can combine materials with positive Hall-coefficient to obtain a composite with negative Hall-coefficient (Briane and Milton 2009, Kadic, Schittny, Bückmann, Kern, and Wegener 2015), thereby destroying the argument that in classical physics it is the sign of the Hall coefficient which tells one the sign of the charge carrier; one can combine nonmagnetic materials to produce artificial magnetism and composites with negative magnetic permeability (Schelkunoff and Friis 1952, Pendry, Holden, Robbins, and Stewart 1999); as shown in **Figure 2.3** one can combine materials with positive mass density to obtain composites with negative effective mass density (Sheng, Zhang, Liu, and Chan 2003; Liu, Chan, and Sheng 2005); and as shown in **Figure 2.4** one can obtain materials with anisotropic and even complex effective mass density (Schoenberg and Sen 1983; Mil-

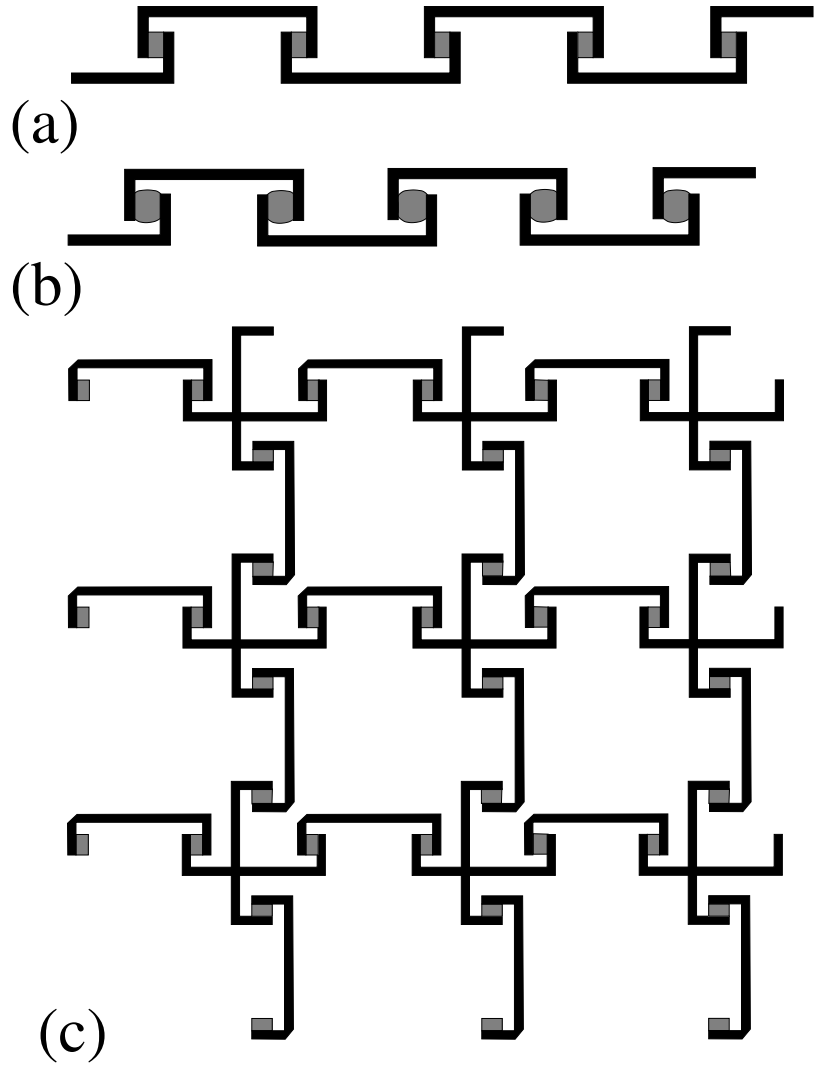


Figure 2.2: Figures (a) and (b) show a mechanism whereby one can get negative thermal expansion from positive thermal expansion. The black regions are rigid and do not expand much when heated while the shaded regions expand a lot, thus shortening the length of the structure from (a) to (b) as it is heated. Figure (c) shows how one can extend this idea to two dimensions, where the white region between the elements is void, or material that is easily compressed. The microstructure can also clearly be extended to three dimensions.

ton, Briane, and Willis 2006; Bückmann, Kadic, Schittny, and Wegener 2015). In fact, it follows directly from the work of Movchan and Guenneau (2004) that there is a close link between negative effective mass density and negative magnetic permeability: in cylindrical geometries the same Helmholtz equation underlies both

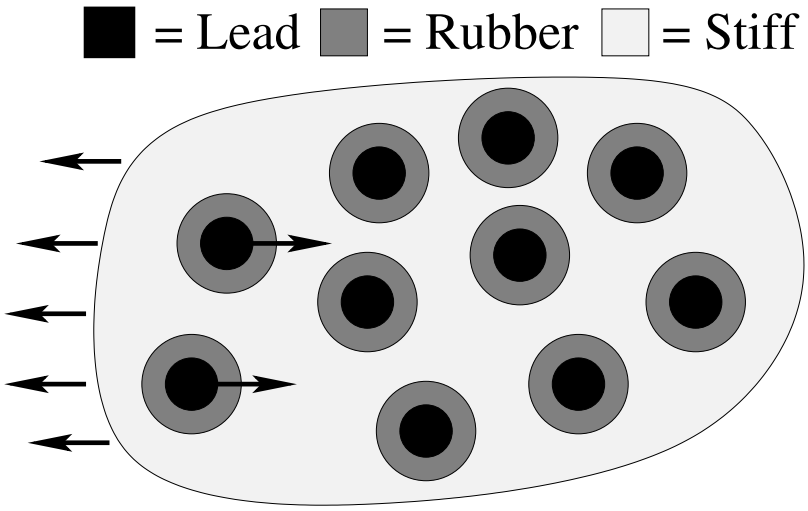


Figure 2.3: As shown by Sheng, Zhang, Liu, and Chan (2003) and Liu, Chan, and Sheng (2005), negative effective mass density can be achieved in composites consisting of inclusions of coated spheres, with a core of lead surrounded by rubber, in a stiff light matrix. As the material is oscillated above the resonant frequency the lead balls move out of phase with the velocity $\mathbf{v}$ of the stiff light matrix, which from the outside is the only velocity which matters. Thus when the lead balls are moving in the direction of the arrows pointing to the right, the average momentum $\mathbf{p}$ is also in this direction, while the matrix is moving in the direction of the arrows pointing to the left. Hence $\mathbf{p}$ and $\mathbf{v}$ are in opposite directions: the effective mass is negative.

antiplane elasticity and transverse electric (TE) or transverse magnetic (TM) electromagnetism.

Composites which have unexpected properties outside the range of naturally occurring materials are frequently called metamaterials: see Cai and Shalaev (2010) for an introduction to the topic of optical metamaterials, and see Banerjee (2011) and Craster and Guenneau (2013) for an introduction to acoustic and elastic metamaterials.

We mention too, that part of the reason for the surge of interest in these metamaterials arises because three-dimensional lithography and printing techniques (Pendry and Smith 2004; Kadic, Bückmann, Stenger, Thiel, and Wegener 2012; Bückmann, Stenger, Kadic, Kaschke, Frölich, Kennerknecht, Eberl, Thiel, and Wegener 2012; Bückmann, Schittny, Thiel, Kadic, Milton, and Wegener 2014; Meza, Das, and Greer 2014) now allow one to tailor beautiful structures with desired properties. Still there are limitations: usually one wants the cell size to be small and this restricts the size of the overall sample, since in three dimensions the number of cells scales as the cube of the sample side. For this reason metasurfaces may hold more promise for practical applications.

To determine the effective properties of composites it usually suffices to consider periodic materials (i.e., take a large cubic sample of the material and periodically extend it). So one looks for solutions of (1.5) where $\mathbf{j}(\mathbf{x})$, $\boldsymbol{\sigma}(\mathbf{x})$ and $\mathbf{e}(\mathbf{x})$ are periodic functions with the same unit cell of periodicity Ω as the composite. The potential $V(\mathbf{x})$ however is not periodic: it has a linear part $-\langle \mathbf{e} \rangle \cdot \mathbf{x}$ plus a periodic part: here the angular brackets denote a volume average over Ω. It is the relation between the volume averaged current field $\langle \mathbf{j} \rangle$ and

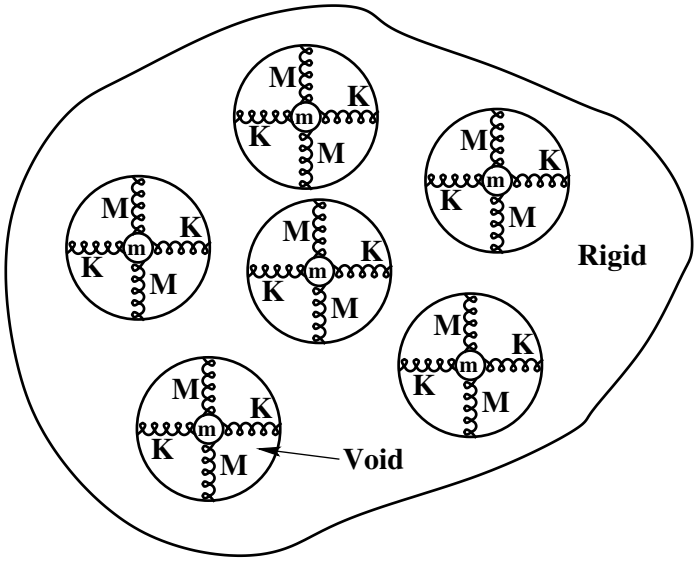


Figure 2.4: This figure, which is a variant of Figure 3 in Milton, Briane, and Willis (2006), shows how it is possible to get metamaterials with an anisotropic and possibly complex effective mass density. The body is treated as a black box, and one is interested in how its momentum is related to its velocity under time-harmonic vibrations at a fixed frequency ω. If the spring constants K and M are different, then the inertial responses to vibrations in the horizontal and vertical directions differ: the effective mass density is anisotropic. If the springs in the cavities have some viscosity, causing energy loss, then the effective mass density will be complex-valued.

the volume averaged electric field $\langle \mathbf{e} \rangle$ which determines the effective conductivity tensor $\boldsymbol{\sigma}_*$;

$$\langle \mathbf{j} \rangle = \boldsymbol{\sigma}_* \langle \mathbf{e} \rangle. \tag{2.5}$$

So the recipe for obtaining $\boldsymbol{\sigma}_*$ is to compute the periodic fields $\mathbf{e}(\mathbf{x})$ and $\mathbf{j}(\mathbf{x})$ for a basis set of independent "applied fields" $\langle \mathbf{e} \rangle$ (d applied fields in d dimensions), find $\langle \mathbf{j} \rangle$ for each of them, which then through the linear relation (2.5) determines the effective tensor $\boldsymbol{\sigma}_*$. In experiments applying the field $\langle \mathbf{e} \rangle$ is achieved by inserting a large sample of the material (with many unit cells) between parallel conducting plates. If the unit cell has some reflection symmetries, then it may suffice to put a single cell between the plates, with no flux of current through the sides of the unit cell.

For conduction in a magnetic field the Hall effect causes the conductivity tensor to be nonsymmetric (Landau, Lifshitz, and Pitaevskiĭ 1984), and convection enhanced diffusion also can be cast as a problem with a nonsymmetric conductivity tensor (Fannjiang and Papanicolaou 1994). We remark in passing that the nonlinear equations of conduction similarly apply to a wide variety of problems (Milgrom 2002).

2.2 The history of the abstract theory of composites

This section can easily be skipped by those readers not interested in historical details. The mathematical content will be revisited in subsequent sections.

Kohler and Papanicolaou (1982), Papanicolaou and Varadhan (1982) and Golden and Papanicolaou (1983) formulated the conductivity problem in composites with random microstructure. They introduced the Hilbert space of square integrable, stationary, fields with zero average value and observed that the nonlocal operator $\mathbf{\Gamma} = \nabla(\nabla^2)^{-1}\nabla\cdot$ was a projection onto the subspace of curl-free fields. Fokin (1982) formulated the conductivity problem in terms of two projections onto orthogonal subspaces. Dell'Antonio, Figari, and Orlandi (1986) considered the response of bodies under quite general boundary conditions and formulated the equations for conductivity and elasticity in terms of appropriate projections, which for periodic boundary conditions reduce to the projections onto the space of vector fields which are gradients of periodic scalar potentials, or for elasticity are the symmetrized gradient of a periodic vector potential.

The general formulation applicable to many problems in composites, not just conductivity, where the Hilbert space is split into three-orthogonal subspaces $\mathcal{U}, \mathcal{E}$ and $\mathcal{J}$ each appropriate to the problem at hand was presented by Milton (1987a, 1990).

By contrast the theory of the Y-tensor $\mathbf{Y}_*$ developed from many different directions. Walpole (1966) noticed, for example, that the bounds of Hashin and Shtrikman (1963) and Hill (1963) on the effective bulk modulus κ^* of an isotropic 3-dimensional composite of 2 isotropic phases, having bulk moduli κ_1 and κ_2, and shear moduli μ_1 and μ_2, with $\mu_1 \geq \mu_2$ could be expressed as

$$[\sum_{i=1}^{2} f_i(\kappa_2^* + \kappa_i)]^{-1} - \kappa_2^* \leq \kappa_* \leq [\sum_{i=1}^{2} f_i(\kappa_1^* + \kappa_i)]^{-1} - \kappa_1^*, \tag{2.6}$$

where $\kappa_2^* = 4\mu_2/3$, $\kappa_1^* = 4\mu_1/3$, and f_1 and f_2 are the volume fractions of the two phases (with $f_1 + f_2 = 1$). Milton (1981b, 1982) found the conductivity bounds of Beran (1965) could be simplified to

$$\begin{aligned} \sigma_* &\leq f_1\sigma_1 + f_2\sigma_2 - \frac{f_1 f_2(\sigma_1 - \sigma_2)^2}{f_2\sigma_1 + f_1\sigma_2 + 2(\zeta_1\sigma_1 + \zeta_2\sigma_2)}, \\ 1/\sigma_* &\leq f_1/\sigma_1 + f_2/\sigma_2 - \frac{f_1 f_2(1/\sigma_1 - 1/\sigma_2)^2}{f_2/\sigma_1 + f_1/\sigma_2 + (\zeta_1/\sigma_1 + \zeta_2/\sigma_2)/2}, \end{aligned} \tag{2.7}$$

and similarly the bulk modulus bounds of Beran and Molyneux (1966) could be simplified to

$$\begin{aligned} \kappa_* &\leq f_1\kappa_1 + f_2\kappa_2 - \frac{f_1 f_2(\kappa_1 - \kappa_2)^2}{f_2\kappa_1 + f_1\kappa_2 + 4(\zeta_1\mu_1 + \zeta_2\mu_2)/3}, \\ 1/\kappa_* &\leq f_1/\kappa_1 + f_2/\kappa_2 - \frac{f_1 f_2(1/\kappa_1 - 1/\kappa_2)^2}{f_2/\kappa_1 + f_1/\kappa_2 + 3(\zeta_1/\mu_1 + \zeta_2/\mu_2)/4}, \end{aligned} \tag{2.8}$$

where the nonnegative parameters ζ_1 and ζ_2 (with $\zeta_1 + \zeta_2 = 1$) are given by

$$\zeta_1 = 1 - \zeta_2 = \frac{9}{2f_1 f_2}\int_0^\infty dr \int_0^\infty ds \int_{-1}^{+1} du \frac{f_{111}(r,s,u)}{rs} P_2(u), \tag{2.9}$$

in which $P_2(u) = (3u^2 - 1)/2$ is a Legendre polynomial, and $f_{111}(r,s,u)$ is the probability that all three vertices of a triangle having side lengths r and s, and included angle $\cos^{-1} u$ land in phase 1 when thrown

randomly in the composite. (The conductivity bounds of Beran (1965) were also simplified by Torquato (1980)). Part of the reason I chose this form for the simplified bounds was that I realized when formulating it that, for example, the second bound in (2.7) could be expressed equivalently as

$$\sigma_* \geq f_1\sigma_1 + f_2\sigma_2 - \frac{f_1f_2(\sigma_1-\sigma_2)^2}{f_2\sigma_1 + f_1\sigma_2 + 2(\zeta_1/\sigma_1 + \zeta_2/\sigma_2)^{-1}} \tag{2.10}$$

thus connecting it to the form of the upper bound in (2.7). Also the form (2.7) of the bounds displays a nice symmetry in the sense that when σ_j is replaced by $1/\sigma_j$, for $j = 1, 2, *$ in the upper bound it almost maps to the lower bound apart from the factor of 2. Values for the parameter $\zeta_1 = 1 - \zeta_2$, and an associated one $\eta_1 = 1 - \eta_2$ which enters the simplified bounds on the effective shear modulus (McCoy 1970; Milton 1981b, 1982; Milton and Phan-Thien 1982; Gibiansky and Torquato 1995), are given for various microgeometries in the references found in the section 15.6 of Milton (2002) and in the book of Torquato (2002): for recent results with impressive numerical simulations see Gillman, Amadio, Matouš, and Jackson (2015) and Hlushkou, Liasneuski, Tallarek, and Torquato (2015).

Berryman (1982) (see also Berryman and Milton (1988)) introduced a transformation which for n-phase elastic composites took the form

$$K(x) = [\sum_{i=1}^{n} f_i(4x/3 + \kappa_i)]^{-1} - 4x/3, \tag{2.11}$$

and realized the bounds of Hashin and Shtrikman (1963) and Hill (1963), and those of Beran and Molyneux (1966) as given in (2.8), could be written when $n = 2$ as

$$K(\mu_2) \leq K((\zeta_1/\mu_1 + \zeta_2/\mu_2)^{-1}) \leq \kappa_* \leq K((\zeta_1\mu_1 + \zeta_2\mu_2)) \leq K(\mu_1). \tag{2.12}$$

For multiphase composites with $n \geq 3$ it is not clear if the transformation (2.11) has any relation with the Y-tensor.

For isotropic two-component conducting composites with isotropic component conductivities σ_1 and σ_2 with $\sigma_1 > \sigma_2$, Milton and Golden (1985) found a sequence of transformations which preserved analytic properties. The first transformation in this sequence is the transformation from σ_* to σ_*^1 given by

$$\begin{aligned} \sigma_*^1 &= \sigma_1\sigma_2(\sigma_*\langle\sigma^{-1}\rangle - 1)/[2(\langle\sigma\rangle - \sigma_*)] \\ &= \{-f_2\sigma_1 - f_1\sigma_2 + f_1f_2(\sigma_1-\sigma_2)^2[f_1\sigma_1 + f_2\sigma_2 - \sigma_*]^{-1}\}/2, \end{aligned} \tag{2.13}$$

in terms of which the bounds of Hashin and Shtrikman (1962) and those of Beran (1965) as given by (2.7) reduce to

$$\sigma_2 \leq (\zeta_1/\sigma_1 + \zeta_2/\sigma_2)^{-1} \leq \sigma_*^1 \leq \zeta_1\sigma_1 + \zeta_2\sigma_2 \leq \sigma_1. \tag{2.14}$$

It was found (Milton 1986a) that the related matrix transformation simplified the translation method conductivity bounds of Murat and Tartar (1985) and Lurie and Cherkaev (1986) on the effective conductivity tensor $\boldsymbol{\sigma}_*$ of possibly anisotropic three-dimensional composites of two isotropic phases mixed in fixed proportions.

For multiphase conducting composites (and related problems) one could stratify the Hilbert space, and obtain a sequence of effective tensors which were linked by this sort of transformation (Milton 1987a, 1987b, 1991). Independently Cherkaev and Gibiansky (1992) recognized that the translation bounds for two phase composites with commuting tensors $\mathbf{L}_1$, $\mathbf{L}_2$ and $\mathbf{L}_*$ generally simplified when expressed in terms of the tensor

$$\mathbf{Y}_* = -f_2\mathbf{L}_1 - f_1\mathbf{L}_2 + f_1f_2(\mathbf{L}_1 - \mathbf{L}_2)[f_1\mathbf{L}_1 + f_2\mathbf{L}_2 - \mathbf{L}_*]^{-1}(\mathbf{L}_1 - \mathbf{L}_2). \tag{2.15}$$

Later their simplification was found to be valid even when $\mathbf{L}_1$, $\mathbf{L}_2$ and $\mathbf{L}_*$ did not commute (Milton 1991). A simple physical interpretation of the tensor $\mathbf{Y}_*$ for two-phase composites was given by Gibiansky and Milton (1993) [see equation (2.99) in Section 2.9]. The complete abstract formulation of the $\mathbf{Y}_*$ tensor problem is given in **Chapters 19**, **20**, and **29** of Milton (2002).

2.3 The abstract setting for defining effective tensors

Here we review the general abstract theory of composites, see also Section 12.7 of Milton 2002. The setting is a Hilbert or finite-dimensional vector space $\mathcal{H}$ (over the complex numbers) which has some inner product $(\mathbf{P}_1, \mathbf{P}_2)$ defined for all $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{H}$, having the usual properties that

$$(\mathbf{P}_1, \mathbf{P}_2) = \overline{(\mathbf{P}_2, \mathbf{P}_1)}, \quad (\mathbf{P}_1, \mathbf{P}_1) > 0 \quad \text{for all } \mathbf{P}_1 \neq 0. \tag{2.16}$$

where the overline denotes complex conjugation. The Hilbert space has the decomposition

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}, \tag{2.17}$$

where the subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are mutually orthogonal with respect to this inner product. [By the definition of a subspace of $\mathcal{H}$ these are required to remain closed under multiplication by complex numbers, i.e., if $\mathbf{P}$ is in one of the subspaces then so is $\lambda\mathbf{P}$ for all complex numbers λ.] Then we suppose we are given an operator $\mathbf{L}$ which maps fields in $\mathcal{H}$ to fields in $\mathcal{H}$. Given a field or a vector $\mathbf{E}_0 \in \mathcal{U}$ the problem is to find fields or vectors

$$\mathbf{J}_0 \in \mathcal{U}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{E} \in \mathcal{E}, \tag{2.18}$$

such that

$$\mathbf{J}_0 + \mathbf{J} = \mathbf{L}(\mathbf{E}_0 + \mathbf{E}). \tag{2.19}$$

Assuming that there is a unique solution for the fields $\mathbf{J}_0$, $\mathbf{E}$ and $\mathbf{J}$ for all $\mathbf{E}_0 \in \mathcal{U}$, then since $\mathbf{J}_0$ depends linearly on $\mathbf{E}_0$ we can write

$$\mathbf{J}_0 = \mathbf{L}_*\mathbf{E}_0, \tag{2.20}$$

which defines the effective operator $\mathbf{L}_*$ which maps $\mathcal{U}$ to $\mathcal{U}$, or to some subspace of it if $\mathbf{L}_*$ is singular. There is also the dual problem: given $\mathbf{J}_0 \in \mathcal{U}$, find fields or vectors $\mathbf{E}_0 \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ such that (2.19) is satisfied. If there is a unique solution for these fields for all $\mathbf{J}_0 \in \mathcal{U}$ then since $\mathbf{E}_0$ depends linearly on $\mathbf{J}_0$ we can write

$$\mathbf{E}_0 = \mathbf{L}_*^{-1}\mathbf{J}_0, \tag{2.21}$$

which defines the inverse effective operator. This formulation defining effective tensors is central to the theory of periodic composites, where $\mathcal{H}$ consists of periodic fields of some sort (which may have elements which are tensors, vectors, or scalars), that are square integrable over the unit cell of periodicity, $\mathcal{U}$ consists of constant fields, and the subspaces $\mathcal{E}$ and $\mathcal{J}$ have meanings according to the problem one is considering, with the projections onto these subspaces being local in Fourier space. The simplest example is electrical conductivity, where $\mathcal{H}$ consists of vector fields with the same periodicity as the composite that are square integrable over the unit cell, $\mathcal{U}$ consists of constant vector fields, $\mathcal{E}$ consists of electrical fields $\mathbf{E}(x)$ which are minus the gradient of a potential with the same periodicity as the composite, and $\mathcal{J}$ consists of current fields $\mathbf{J}(\mathbf{x})$ with the same periodicity as the composite, which have zero divergence and which have zero average value. The operator $\mathbf{L}$ is local in real space and can be identified with the conductivity tensor $\boldsymbol{\sigma}(\mathbf{x})$. Given an applied electric field

$\mathbf{E}_0 \in \mathcal{U}$, representing the average of the total electric in the composite, the action of $\mathbf{L} = \boldsymbol{\sigma}$ is to multiply the total electric field $\mathbf{E}_0 + \mathbf{E}(\mathbf{x})$ by $\boldsymbol{\sigma}(\mathbf{x})$ to produce the total current field

$$\mathbf{J}_0 + \mathbf{J}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})[\mathbf{E}_0 + \mathbf{E}(\mathbf{x})]. \tag{2.22}$$

Here $\mathbf{J}_0 \in \mathcal{U}$ is the average of the total current field, and the linear relation between $\mathbf{J}_0$ and $\mathbf{E}_0$ defines the effective conductivity tensor $\mathbf{L}_* = \boldsymbol{\sigma}_*$, i.e, $\mathbf{J}_0 = \boldsymbol{\sigma}_* \mathbf{E}_0$.

2.4 Solving for the effective tensor and fields

To solve for the effective tensor and fields, we introduce the projections $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Gamma}_2$ onto the spaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$. In the theory of composites the action of these projections are readily calculated in Fourier space, but here we will just assume they are projection operators. Applying $\boldsymbol{\Gamma}_1$ to both sides of the relation (2.19), gives

$$0 = \boldsymbol{\Gamma}_1 \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) = \boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_0 \mathbf{E}_0 + \boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1 \mathbf{E}, \tag{2.23}$$

which implies

$$\mathbf{J}_0 = \boldsymbol{\Gamma}_0 \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) = \boldsymbol{\Gamma}_0 \mathbf{L}[\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1 (\boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1)^{-1} \boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_0] \mathbf{E}_0, \tag{2.24}$$

where the inverse of $\boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1$, if it exists, is to be taken on the subspace $\mathcal{E}$. This gives the formula

$$\mathbf{L}_* = \boldsymbol{\Gamma}_0 \mathbf{L} \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_0 \mathbf{L} \boldsymbol{\Gamma}_1 (\boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1)^{-1} \boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_0, \tag{2.25}$$

for the effective operator $\mathbf{L}_*$. If $\mathcal{E}$ is a finite-dimensional vector space then inverting $\boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1$ will be inverting a matrix, which will generally be no problem. If however $\mathcal{E}$ has infinite dimension, then to ensure invertibility of the operator $\boldsymbol{\Gamma}_1 \mathbf{L} \boldsymbol{\Gamma}_1$, we should make additional assumptions, for instance that $\mathbf{L}$ is bounded on the subspace $\mathcal{E}$, i.e., there exists $\beta > 0$ such that

$$\beta > \sup_{\substack{\mathbf{E} \in \mathcal{E} \\ |\mathbf{E}|=1}} |\mathbf{L}\mathbf{E}|, \tag{2.26}$$

and that $\mathbf{L}$ has the following coercivity property on the subspace $\mathcal{E}$: there exists $\alpha > 0$ such that

$$\mathrm{Re}(\mathbf{E}, \mathbf{L}\mathbf{E}) \geq \alpha |\mathbf{E}|^2 \quad \text{for all } \mathbf{E} \in \mathcal{E}. \tag{2.27}$$

To obtain an alternative formula for the effective tensor $\mathbf{L}_*$ we introduce the "polarization field" (or polarization vector if $\mathcal{H}$ has finite dimension)

$$\mathbf{P} = (\mathbf{L} - \sigma_0 \mathbf{I})[\mathbf{E}_0 + \mathbf{E}] = \mathbf{J}_0 + \mathbf{J} - \sigma_0 [\mathbf{E}_0 + \mathbf{E}], \tag{2.28}$$

where we are free to appropriately choose the constant σ_0, which may possibly be complex. Applying the projection operator $\boldsymbol{\Gamma}_1$ onto to the space $\mathcal{E}$ to $\mathbf{P}$ gives

$$\boldsymbol{\Gamma}_1 \mathbf{P} = -\sigma_0 \mathbf{E}, \tag{2.29}$$

and hence we have

$$[\mathbf{I} + \boldsymbol{\Gamma}_1 (\mathbf{L}/\sigma_0 - \mathbf{I})](\mathbf{E}_0 + \mathbf{E}) = \mathbf{E}_0 + \mathbf{E} + \boldsymbol{\Gamma}_1 \mathbf{P}/\sigma_0 = \mathbf{E}_0, \tag{2.30}$$

which gives

$$\mathbf{J}_0 + \mathbf{J} = \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) = \mathbf{L}[\mathbf{I} + \boldsymbol{\Gamma}_1 (\mathbf{L}/\sigma_0 - \mathbf{I})]^{-1} \mathbf{E}_0. \tag{2.31}$$

Applying the projection $\mathbf{\Gamma}_0$ onto the space $\mathcal{U}$ to (2.31) we obtain the desired formula for the field

$$\mathbf{J}_0 = \mathbf{\Gamma}_0\mathbf{L}[\mathbf{I} + \mathbf{\Gamma}_1(\mathbf{L}/\sigma_0 - \mathbf{I})]^{-1}\mathbf{E}_0, \tag{2.32}$$

and hence for the effective operator,

$$\mathbf{L}_* = \mathbf{\Gamma}_0\mathbf{L}[\mathbf{I} + \mathbf{\Gamma}_1(\mathbf{L}/\sigma_0 - \mathbf{I})]^{-1}\mathbf{\Gamma}_0. \tag{2.33}$$

From (2.30) and (2.31) we obtain formulas for the fields $\mathbf{E}$ and $\mathbf{J}$:

$$\begin{aligned} \mathbf{E} &= \mathbf{\Gamma}_1[\mathbf{I} + \mathbf{\Gamma}_1(\mathbf{L}/\sigma_0 - \mathbf{I})]^{-1}\mathbf{E}_0, \\ \mathbf{J} &= \mathbf{\Gamma}_2\mathbf{L}[\mathbf{I} + \mathbf{\Gamma}_1(\mathbf{L}/\sigma_0 - \mathbf{I})]^{-1}\mathbf{E}_0. \end{aligned} \tag{2.34}$$

Expanding each inverse gives the associated series expansions

$$\begin{aligned} \mathbf{L}_* &= \sum_{j=0}^{\infty} \mathbf{\Gamma}_0\mathbf{L}[\mathbf{\Gamma}_1(\mathbf{I} - \mathbf{L}/\sigma_0)]^j\mathbf{\Gamma}_0, \\ \mathbf{J}_0 &= \sum_{j=0}^{\infty} \mathbf{\Gamma}_0\mathbf{L}[\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L}/\sigma_0)]^j\mathbf{E}_0, \\ \mathbf{E} &= \sum_{j=0}^{\infty} [\mathbf{\Gamma}_1(\mathbf{I} - \mathbf{L}/\sigma_0)]^j\mathbf{E}_0, \\ \mathbf{J} &= \sum_{j=0}^{\infty} \mathbf{\Gamma}_2\mathbf{L}[\mathbf{\Gamma}_1(\mathbf{I} - \mathbf{L}/\sigma_0)]^j\mathbf{E}_0. \end{aligned} \tag{2.35}$$

If these series expansions converge for a given $\mathbf{E}_0 \in \mathcal{U}$ it is quite clear that the fields lie in the right subspaces $\mathbf{J}_0 \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$, and $\mathbf{J} \in \mathcal{J}$, and that the relation (2.19) is satisfied. This is one way to show that a solution to the original equations exists. Although the converged fields $\mathbf{J}_0$, $\mathbf{E}$ and $\mathbf{J}$, and the effective tensor $\mathbf{L}_*$ appear to depend on the choice of σ_0, this is not the case: the tensor $\sigma_0\mathbf{I}$ just serves as the expansion point for these series as they converge quickest when $\mathbf{L} - \sigma_0\mathbf{I}$ is small. To see that the converged fields do not depend on σ_0, note that if we vary σ_0 in (2.30) the left hand side changes by an amount proportional to $\mathbf{\Gamma}_1\mathbf{L}(\mathbf{E}_0 + \mathbf{E})$, which from (2.23) vanishes. While the effective tensor $\mathbf{L}_*$ is of primary interest, it can also be important in composites to know the local field $\mathbf{E}_0 + \mathbf{E}(\mathbf{x})$, or equivalently $\mathbf{J}_0 + \mathbf{J}(\mathbf{x})$, to see for example if at any point in the material the field is close to (or has exceeded) a critical value which would cause the material to break down (the "yield surface" for plastic yielding or the critical electric field strength for dielectric breakdown) or to cause the onset of other nonlinearities which would render the analysis invalid. Knowing the local field $\mathbf{E}_0 + \mathbf{E}(\mathbf{x})$ is also useful if we are interested in knowing how perturbations of $\mathbf{L}(\mathbf{x})$ effect the effective tensor $\mathbf{L}_*$. [See **Chapter 16** of Milton (2002) and references therein.]

For composites these expansions are useful in Fast Fourier Transform methods for computing effective tensors and fields (1994, 1998). Variants of the expansion are useful for accelerated Fourier transform methods for computing effective tensors and fields (Eyre and Milton 1999; see also the generalization in Section 14.9 of Milton 2002, and in particular equation (14.38)). They are also useful for deriving algebraic conditions for exact microstructure independent relations satisfied by effective tensors (Grabovsky, Milton, and Sage 2000; see also **Chapter 17** of Milton 2002).

In multicomponent media the series expansions (2.35) are useful for proving that the fields and effective tensor are analytic functions of the component moduli (provided those component moduli are such that the series converges): see Bruno (1991b), Bruno and Leo (1992), page 372 of Milton (2002) [though there I should have referenced Section 14.7 page 300 since if $\mathbf{L}(\mathbf{x})$ is complex it *is not self-adjoint*], and also Section 3.5 of the current book.

To show convergence we recall the definition of the norm of an operator $\mathbf{A}$:

$$\|\mathbf{A}\| = \sup_{\substack{\mathbf{P}(\mathbf{x}) \\ |\mathbf{P}|=1}} |\mathbf{AP}|, \tag{2.36}$$

where

$$|\mathbf{Q}| = (\mathbf{Q}, \mathbf{Q})^{1/2}. \tag{2.37}$$

The operator norm has the properties

$$\|\mathbf{AB}\| \le \|\mathbf{A}\|\,\|\mathbf{B}\|, \quad \|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\|+\|\mathbf{B}\|, \tag{2.38}$$

and $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ being projections have norm 1. Hence all the series in (2.35) will certainly converge if

$$\|\mathbf{I} - \mathbf{L}/\sigma_0\| < 1. \tag{2.39}$$

To establish that $\mathbf{I} - \mathbf{L}/\sigma_0$ has norm less than 1, for an appropriate real value of σ_0, when $\mathbf{L}$ is bounded, i.e., there is some $\beta > 0$ such that

$$\beta > \sup_{\substack{\mathbf{P}\in\mathcal{H} \\ |\mathbf{P}|=1}} |\mathbf{LP}|, \tag{2.40}$$

and $\mathbf{L}$ is coercive in the sense that there is some $\alpha > 0$ such that

$$\mathrm{Re}(\mathbf{P}, \mathbf{LP}) \ge \alpha |\mathbf{P}|^2 \quad \text{for all } \mathbf{P} \in \mathcal{H}, \tag{2.41}$$

we follow (with some corrections) Section 14.7 of Milton (2002). For any field $\mathbf{Q} \in \mathcal{H}$, introduce $\mathbf{Q}' = (\mathbf{I} - \mathbf{L}/\sigma_0)\mathbf{Q}$, so that $\mathbf{LQ} = \sigma_0(\mathbf{Q} - \mathbf{Q}')$. Then the coercivity property (2.41) with $\mathbf{P} = \mathbf{Q}$ implies

$$\sigma_0[|\mathbf{Q}|^2 - \mathrm{Re}(\mathbf{Q}, \mathbf{Q}')] \ge \alpha |\mathbf{Q}|^2, \tag{2.42}$$

whereas the boundedness of $\mathbf{L}$ (2.40) implies

$$\beta^2 |\mathbf{Q}|^2 \ge \sigma_0^2 |\mathbf{Q} - \mathbf{Q}'|^2, \tag{2.43}$$

which, when expanded out, gives

$$2\,\mathrm{Re}(\mathbf{Q}, \mathbf{Q}') \ge [1 - (\beta/\sigma_0)^2]|\mathbf{Q}|^2 + |\mathbf{Q}'|^2. \tag{2.44}$$

Combining (2.42) and (2.44) gives

$$[1 + (\beta/\sigma_0)^2 - 2(\alpha/\sigma_0)]|\mathbf{Q}|^2 \ge |\mathbf{Q}'|^2, \tag{2.45}$$

which implies

$$\|\mathbf{I} - \mathbf{L}/\sigma_0\| \le [1 + (\beta/\sigma_0)^2 - 2(\alpha/\sigma_0)]^{1/2}. \tag{2.46}$$

Thus by choosing $\sigma_0 = \beta^2/\alpha$ we obtain the bound

$$\|\mathbf{I} - \mathbf{L}/\sigma_0\| \leq \sqrt{1-(\alpha/\beta)^2} < 1, \tag{2.47}$$

which ensures convergence of the series.

When the coercivity property (2.41) is satisfied, then solution to the equations is unique. To see this, suppose there is another solution $\mathbf{J}_0' \in \mathcal{U}$, $\mathbf{E}' \in \mathcal{E}$, and $\mathbf{J}' \in \mathcal{J}$ to (2.19) with the same value of $\mathbf{E}_0 \in \mathcal{U}$. Subtracting solutions we get

$$(\mathbf{J}_0' - \mathbf{J}_0 + \mathbf{J}' - \mathbf{J}) = \mathbf{L}(\mathbf{E}' - \mathbf{E}). \tag{2.48}$$

Using the coercivity condition (2.41) with $\mathbf{P} = \mathbf{E}' - \mathbf{E}$, and using the orthogonality of the subspaces, implies

$$0 = \mathrm{Re}(\mathbf{E}' - \mathbf{E}, (\mathbf{J}_0' - \mathbf{J}_0 + \mathbf{J}' - \mathbf{J})) = \mathrm{Re}(\mathbf{E}' - \mathbf{E}, \mathbf{L}(\mathbf{E}' - \mathbf{E})) \geq \alpha |\mathbf{E}' - \mathbf{E}|^2. \tag{2.49}$$

From this we deduce that $\mathbf{E}' = \mathbf{E}$, and then projecting (2.48) onto the subspaces $\mathcal{U}$ and $\mathcal{J}$ gives $\mathbf{J}_0' = \mathbf{J}_0$ and $\mathbf{J}' = \mathbf{J}$. This uniqueness provides an alternative proof that the solutions (2.35) cannot depend on σ_0.

Note that the norm $\|\mathbf{I} - \mathbf{L}/\sigma_0\|$ does not change if we multiply σ_0 by a phase factor $e^{i\theta}$ and at the same time multiply $\mathbf{L}$ by this same factor. Hence we obtain convergence when $\mathbf{L}$ is bounded and $\mathbf{L}$ is coercive in the sense that there is some $\alpha > 0$ and angle θ such that

$$\mathrm{Re}(\mathbf{P}, e^{i\theta}\mathbf{L}\mathbf{P}) \geq \alpha |\mathbf{P}|^2 \quad \text{for all } \mathbf{P} \in \mathcal{H}. \tag{2.50}$$

2.5 An example: a subspace collection associated with a function that takes positive semidefinite matrix values

Consider a positive semidefinite Hermitian $n \times n$ matrix valued function $\mathbf{m}(y)$ defined on the interval $y \in [0,1]$, with bounded integral over this interval. Now given two $2n$-dimensional vector fields $\mathbf{P}_1(y)$ and $\mathbf{P}_2(y)$ defined on the interval $y \in [0,1]$ of the form

$$\mathbf{P}_1(y) = \begin{pmatrix} \mathbf{A}_1(y) \\ \mathbf{B}_1(y) \end{pmatrix}, \quad \mathbf{P}_2(y) = \begin{pmatrix} \mathbf{A}_2(y) \\ \mathbf{B}_2(y) \end{pmatrix}, \tag{2.51}$$

we define their inner product to be

$$(\mathbf{P}_1, \mathbf{P}_2) = \int_0^1 \overline{\mathbf{A}_1(y)} \cdot \mathbf{A}_2(y) + \overline{\mathbf{B}_1(y)} \cdot \mathbf{B}_2(y)\, dy, \tag{2.52}$$

and the associated norm is then $|\mathbf{P}| = (\mathbf{P}, \mathbf{P})^{1/2}$. We take our Hilbert space $\mathcal{H}$ to consist of such $2n$-dimensional vector fields $\mathbf{P}(y)$, with $y \in [0,1]$ having finite norm. The subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ are defined to consist of fields that can be expressed in the form

$$\mathbf{P}_1(y) = \begin{pmatrix} \sqrt{y}\,\mathbf{p}_1(y) \\ \sqrt{(1-y)}\,\mathbf{p}_1(y) \end{pmatrix}, \quad \mathbf{P}_2(y) = \begin{pmatrix} \sqrt{(1-y)}\,\mathbf{p}_2(y) \\ -\sqrt{y}\,\mathbf{p}_2(y) \end{pmatrix}, \tag{2.53}$$

respectively for some choice of n-dimensional vector fields $\mathbf{p}_1(y)$ and $\mathbf{p}_2(y)$ defined on the interval $y \in [0,1]$. They are clearly orthogonal spaces, with respect to the inner product (2.52), and the associated projection operators are

$$\chi_1 = \begin{pmatrix} y\,\mathbf{I}_n & \sqrt{y(1-y)}\,\mathbf{I}_n \\ \sqrt{y(1-y)}\,\mathbf{I}_n & (1-y)\,\mathbf{I}_n \end{pmatrix}, \quad \chi_2 = \begin{pmatrix} (1-y)\,\mathbf{I}_n & -\sqrt{y(1-y)}\,\mathbf{I}_n \\ -\sqrt{y(1-y)}\,\mathbf{I}_n & y\,\mathbf{I}_n \end{pmatrix}, \tag{2.54}$$

in which $\mathbf{I}_n$ is the $n \times n$ identity matrix. Clearly $\chi_1 + \chi_2$ is the $2n \times 2n$ identity matrix $\mathbf{I}_{2n}$. We define $\mathcal{U}$ and $\mathcal{E}$ to consist of all fields that can be expressed in the form

$$\mathbf{U} = \begin{pmatrix} \sqrt{\mathbf{m}(y)}\,\mathbf{u} \\ 0 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} 0 \\ \mathbf{e}(y) \end{pmatrix}, \tag{2.55}$$

for some choice of n-dimensional vector $\mathbf{u}$, and for some choice of n-dimensional vector field $\mathbf{e}(y)$ defined on the interval $y \in [0,1]$. Here the square root of the matrix $\mathbf{m}(y)$ is defined in the usual way: $\sqrt{\mathbf{m}(y)}$ is Hermitian and positive semidefinite and has the same eigenvectors as $\mathbf{m}(y)$, but eigenvalues which are the square root of those of $\mathbf{m}(y)$. To obtain an orthonormal basis for $\mathcal{U}$, we consider a set of n fields of the form

$$\mathbf{U}_j = \begin{pmatrix} \sqrt{\mathbf{m}(y)}\mathbf{u}_j \\ 0 \end{pmatrix}, \tag{2.56}$$

where the n-dimensional vectors $\mathbf{u}_j$, $j = 1, 2, \ldots, n$, are real. These have inner products

$$(\mathbf{U}_j, \mathbf{U}_k) = \overline{\mathbf{u}_j} \cdot \mathbf{W}\mathbf{u}_k, \quad \text{where } \mathbf{W} = \int_0^1 \mathbf{m}(y)\, dy. \tag{2.57}$$

Clearly the matrix $\mathbf{W}$ is Hermitian and positive semidefinite. If we assume it is in fact strictly positive definite, then we can find an orthogonal set of n vectors $\mathbf{u}_j$ such that

$$\mathbf{W} = \sum_{i=1}^{n} \mathbf{u}_j \overline{\mathbf{u}_j}^T / |\mathbf{u}_j|^4, \tag{2.58}$$

i.e., the $\mathbf{u}_j$ should be chosen as the eigenvectors of $\mathbf{W}$ with lengths $|\mathbf{u}_j| = (\overline{\mathbf{u}_j}^T \cdot \mathbf{u}_j)^{1/2}$ chosen so $1/|\mathbf{u}_j|^2$ is the corresponding eigenvalue. Then from (2.57) we see that the fields $\mathbf{U}_j$, $j = 1, 2, \ldots, n$, form an orthonormal basis for $\mathcal{U}$, satisfying $(\mathbf{U}_j, \mathbf{U}_k) = \delta_{jk}$. The projections onto $\mathcal{U}$ and $\mathcal{E}$ are then given respectively by

$$\mathbf{\Gamma}_0 \mathbf{P} = \sum_{j=1}^{n} \mathbf{U}_j (\mathbf{U}_j, \mathbf{P}), \quad \mathbf{\Gamma}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{I}_n \end{pmatrix}. \tag{2.59}$$

We define $\mathcal{J}$ to be the orthogonal complement of $\mathcal{U} \oplus \mathcal{E}$ in the Hilbert space $\mathcal{H}$. Let us suppose the operator $\mathbf{L}$ acting on fields in $\mathcal{H}$ has the form

$$\mathbf{L} = \sigma_1 \chi_1 + \sigma_2 \chi_2, \tag{2.60}$$

where σ_1 and σ_2 are (possibly complex) scalars. Then, since

$$\mathbf{\Gamma}_1 (\mathbf{\Gamma}_1 \mathbf{L} \mathbf{\Gamma}_1)^{-1} \mathbf{\Gamma}_1 = \begin{pmatrix} [0 & 0 \\ 0 & [y\sigma_2 + (1-y)\sigma_1]^{-1} \mathbf{I}_n \end{pmatrix}, \tag{2.61}$$

we deduce that

$$\begin{aligned} (\mathbf{U}_j, \mathbf{L}\mathbf{\Gamma}_1 (\mathbf{\Gamma}_1 \mathbf{L} \mathbf{\Gamma}_1)^{-1} \mathbf{\Gamma}_1 \mathbf{L} \mathbf{U}_k) &= \int_0^1 \frac{y(1-y)(\sigma_1 - \sigma_2)^2 \overline{\mathbf{u}_j} \cdot \mathbf{m}(y)\mathbf{u}_k}{y\sigma_2 + (1-y)\sigma_1}\, dy, \\ (\mathbf{U}_j, \mathbf{L}\mathbf{U}_k) &= \int_0^1 (y\sigma_1 + (1-y)\sigma_2) \overline{\mathbf{u}_j} \cdot \mathbf{m}(y)\mathbf{u}_k\, dy, \end{aligned} \tag{2.62}$$

implying, via (2.25), that the effective operator $\mathbf{L}_*$ using this basis $\mathbf{U}_j$, $j = 1, 2, \ldots, n$, of $\mathcal{U}$ is represented by the matrix with elements

$$\{\mathbf{L}_*\}_{jk} = \int_0^1 \frac{\overline{\mathbf{u}_j} \cdot \mathbf{m}(y)\mathbf{u}_k}{y/\sigma_1 + (1-y)/\sigma_2}\, dy. \tag{2.63}$$

Introducing the matrix valued measure $d\mathbf{M}(y)$ with elements $\{d\mathbf{M}(y)\}_{jk} = \overline{\mathbf{u}_j} \cdot \mathbf{m}(y)\mathbf{u}_k\, dy$, we can rewrite this as

$$\{\mathbf{L}_*\}_{jk} = \int_0^1 \frac{\{d\mathbf{M}(y)\}_{jk}}{y/\sigma_1 + (1-y)/\sigma_2}, \quad \text{where} \quad \int_0^1 \{d\mathbf{M}(y)\}_{jk} = \delta_{jk}, \tag{2.64}$$

in which the last identity follows by integrating $\{d\mathbf{M}(y)\}_{jk}$ over $y \in [0, 1]$, giving $\overline{\mathbf{u}_j} \cdot \mathbf{W}\mathbf{u}_k = \delta_{jk}$.

In fact, the integral representation formula (2.64) for the effective operator $\mathbf{L}_*$ as a function of σ_1 and σ_2 holds true, not just for our example, but more generally whenever the three subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are mutually orthogonal, and $\mathbf{L}$ has the form (2.60) where χ_1 and χ_2 project onto subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ that are orthogonal complements in the space $\mathcal{H}$. This representation formula follows from the analytic properties of the function $\mathbf{L}_*(\sigma_1, \sigma_2)$ (Bergman 1978; Milton 1981a; Golden and Papanicolaou 1983; see also **Chapter 2** of Milton 2002). The measure $d\mathbf{M}(y)$ can be recovered from the values that $\mathbf{L}_*(\sigma_1, 1)$ takes as σ_1 ranges just above the negative real axis. What the above example shows is that any function $\mathbf{L}_*(\sigma_1, \sigma_2)$ of the form (2.64), where the $n \times n$ matrix valued measure $d\mathbf{M}(y)$ takes positive semidefinite values, can be associated approximately with the effective tensor of a subspace collection, by replacing $d\mathbf{M}(y)$ with a continuous measure $\mathbf{m}(y)\, dy$. (Then $\mathbf{W} = \mathbf{I}$ and the vectors $\mathbf{u}_j$, $j = 1, 2, \ldots, n$ can be chosen to be orthonormal.) For conducting composites, in two or more dimensions, one can find periodic microstructures that realize (to an arbitrarily high degree of approximation) as a diagonal element of their effective conductivity tensor $\boldsymbol{\sigma}_*$ any scalar valued function $\{\boldsymbol{\sigma}_*\}_{11}(\sigma_1, \sigma_2)$ of the component conductivities σ_1 and σ_2 having this integral representation (Milton 1981c; see also Section 18.4 of Milton 2002): y has the physical interpretation of the volume fraction of phase 1 in a laminate of both phases with layer surfaces perpendicular to the x_1-axis (so the effective conductivity of that laminate in the x_1-direction is the harmonic mean $1/(y/\sigma_1 + (1-y)/\sigma_2)$) and at a much larger length scale these laminates are layered together in an orthogonal direction (so one gets an arithmetic average of their effective conductivities in the x_1-direction), with the measure $\{d\mathbf{M}(y)\}_{11}$ giving the proportions of these laminates in the final microstructure.

2.6 Some properties of the effective tensor

Not only is the norm $\|\mathbf{I} - \mathbf{L}/\sigma_0\|$ invariant when multiply $\mathbf{L}$ and σ_0 by a phase factor, but more generally if $\mathbf{E}_0, \mathbf{J}_0 \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ solve (2.19) then for any complex number λ, $\mathbf{E}_0, \lambda\mathbf{J}_0 \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$ and $\lambda\mathbf{J} \in \mathcal{J}$, will solve the equations when $\mathbf{L}$ is replaced by $\lambda\mathbf{L}$. In this way we see that if $\mathbf{L}$ is multiplied by a constant λ so will be $\mathbf{L}_*$.

The effective tensor has the property that when we replace the operator $\mathbf{L}$ by its adjoint, then the effective tensor is replaced by its adjoint. To see this, we follow Section 12.10 of Milton (2002), and suppose that we are given fields $\mathbf{E}_0, \mathbf{E}_0' \in \mathcal{U}$. Let $\mathbf{J}_0 \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$, and $\mathbf{J} \in \mathcal{J}$ be the associated fields which solve (2.19) and let $\mathbf{J}_0' \in \mathcal{U}$, $\mathbf{E}' \in \mathcal{E}$, and $\mathbf{J}' \in \mathcal{J}$ be fields which solve the adjoint problem,

$$\mathbf{J}_0' + \mathbf{J}' = \mathbf{L}^\dagger(\mathbf{E}_0' + \mathbf{E}'), \tag{2.65}$$

where $\mathbf{L}^\dagger$ is the adjoint of $\mathbf{L}$, meaning that

$$(\mathbf{P}', \mathbf{L}\mathbf{P}) = (\mathbf{L}^\dagger\mathbf{P}', \mathbf{P}), \quad \text{for all } \mathbf{P}', \mathbf{P} \in \mathcal{H}. \tag{2.66}$$

Now the orthogonality of the subspaces implies the string of identities

$$\begin{aligned}(\mathbf{E}_0', \mathbf{L}_*\mathbf{E}_0) &= (\mathbf{E}_0', \mathbf{J}_0) = (\mathbf{E}_0' + \mathbf{E}', \mathbf{J}_0 + \mathbf{J}) = (\mathbf{E}_0' + \mathbf{E}', \mathbf{L}(\mathbf{E}_0 + \mathbf{E})) \\ &= (\mathbf{L}^\dagger(\mathbf{E}_0' + \mathbf{E}'), \mathbf{E}_0 + \mathbf{E}) = (\mathbf{J}_0' + \mathbf{J}', \mathbf{E}_0 + \mathbf{E}') = (\mathbf{J}_0', \mathbf{E}_0) \\ &= (\{\mathbf{L}^\dagger\}_*\mathbf{E}_0', \mathbf{E}_0). \end{aligned} \tag{2.67}$$

We conclude that the problem with tensor $\mathbf{L}^\dagger$ has an effective tensor $\{\mathbf{L}^\dagger\}_*$ which is the adjoint of $\mathbf{L}_*$. A corollary is if $\mathbf{L}$ is self-adjoint (meaning $\mathbf{L}^\dagger = \mathbf{L}$) so too will be the effective tensor $\mathbf{L}_*$.

We also have the identity that

$$\begin{aligned}(\mathbf{E}_0, (\mathbf{L}_* + \{\mathbf{L}^\dagger\}_*)\mathbf{E}_0) &= (\mathbf{E}_0, \mathbf{J}_0) + (\mathbf{J}_0, \mathbf{E}_0) = (\mathbf{E}_0 + \mathbf{E}, \mathbf{J}_0 + \mathbf{J}) + (\mathbf{J}_0 + \mathbf{J}, \mathbf{E}_0 + \mathbf{E}) \\ &= (\mathbf{E}_0 + \mathbf{E}, (\mathbf{L} + \mathbf{L}^\dagger)(\mathbf{E}_0 + \mathbf{E})). \end{aligned} \tag{2.68}$$

Defining

$$\mathbf{L}_s^* = (\mathbf{L}_* + \{\mathbf{L}^\dagger\}_*)/2, \quad \mathbf{L}_s = (\mathbf{L} + \mathbf{L}^\dagger)/2, \tag{2.69}$$

as the self-adjoint parts of $\mathbf{L}_*$ and $\mathbf{L}$, we see that $\mathbf{L}_s^*$ is positive semidefinite on $\mathcal{U}$ whenever $\mathbf{L}_s$ is positive semidefinite on $\mathcal{H}$. More generally using the invariance discussed at the beginning of this section, if for any value of θ, $(e^{i\theta}\mathbf{L} + e^{-i\theta}\mathbf{L}^\dagger)/2$ is positive semidefinite, so too will $(e^{i\theta}\mathbf{L}_* + e^{-i\theta}\{\mathbf{L}^\dagger\}_*)/2$ be positive semidefinite for that value of θ. In composites, with a symmetric tensor $\mathbf{L}(\mathbf{x})$, and with $e^{i\theta} = -i$ this means $\mathbf{L}_*$ will have an imaginary part which is positive semidefinite if the imaginary part of $\mathbf{L}(\mathbf{x})$ is positive semidefinite.

2.7 Variational principles and elementary bounds

When the operator $\mathbf{L}$ is self-adjoint and positive definite, then one has a variational principle for the effective tensor $\mathbf{L}_*$:

$$(\mathbf{E}_0, \mathbf{L}_*\mathbf{E}_0) = \inf_{\underline{\mathbf{E}} \in \mathcal{E}} (\mathbf{E}_0 + \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E}_0 + \underline{\mathbf{E}})). \tag{2.70}$$

For conductivity this is a corollary of the well-known Dirichlet variational principle. It is easily established in the abstract case and follows from the orthogonality of the subspaces. Following Section 13.1 in Milton (2002), given $\mathbf{E} \in \mathcal{E}$, $\mathbf{J} \in \mathcal{J}$ and $\mathbf{J}_0 \in \mathcal{U}$ solving (2.19) for some $\mathbf{E}_0 \in \mathcal{U}$ we have the chain of inequalities

$$\begin{aligned} 0 &\le (\mathbf{E} - \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E} - \underline{\mathbf{E}})) = (\mathbf{E}_0 + \mathbf{E} - \mathbf{E}_0 - \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E}_0 - \mathbf{E} - \mathbf{E}_0\underline{\mathbf{E}})) \\ &= (\mathbf{E}_0 + \mathbf{E}, \mathbf{J}_0 + \mathbf{J}) - (\mathbf{J}_0 + \mathbf{J}, \mathbf{E}_0 + \underline{\mathbf{E}}) - (\mathbf{E}_0 + \mathbf{E}, \mathbf{J}_0 + \mathbf{J}) + (\mathbf{E}_0 + \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E}_0 + \underline{\mathbf{E}})) \\ &= (\mathbf{E}_0 + \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E}_0 + \underline{\mathbf{E}})) - (\mathbf{J}_0, \mathbf{E}_0) \\ &= (\mathbf{E}_0 + \underline{\mathbf{E}}, \mathbf{L}(\mathbf{E}_0 + \underline{\mathbf{E}}) - (\mathbf{E}_0, \mathbf{L}_*\mathbf{E}_0), \end{aligned} \tag{2.71}$$

where we have used the fact that $\mathbf{L}$ and hence $\mathbf{L}_*$ is self-adjoint. Since the left hand side is zero when $\underline{\mathbf{E}} = \mathbf{E}$ we immediately have the variational principle (2.70). The dual variational principle, (also called the Thompson variational principle)

$$(\mathbf{J}_0, \mathbf{L}_*^{-1}\mathbf{J}_0) = \inf_{\underline{\mathbf{J}} \in \mathcal{J}} (\mathbf{J}_0 + \underline{\mathbf{J}}, \mathbf{L}^{-1}(\mathbf{J}_0 + \underline{\mathbf{J}})), \tag{2.72}$$

also immediately follows by switching the roles of the subspaces $\mathcal{E}$ and $\mathcal{J}$, and switching $\mathbf{L}_*$ and $\mathbf{L}$ with their inverses $\mathbf{L}_*^{-1}$ and $\mathbf{L}^{-1}$.

From these variational principles we easily obtain the elementary bounds:

$$\mathbf{L}_* \geq 0, \quad \mathbf{L}_* \leq \boldsymbol{\Gamma}_0 \mathbf{L} \boldsymbol{\Gamma}_0, \quad \mathbf{L}_*^{-1} \leq \boldsymbol{\Gamma}_0 \mathbf{L}^{-1} \boldsymbol{\Gamma}_0, \tag{2.73}$$

where $\boldsymbol{\Gamma}_0$ is the projection onto the space $\mathcal{U}$. The first bound follows directly from (2.70) because the right hand side is nonnegative. The other two bounds follow from (2.70) and (2.72), respectively, by taking the simplest possible choice of trial fields namely $\underline{\mathbf{E}} = 0$ and $\underline{\mathbf{J}} = 0$.

In the simplest example of a periodic conducting composite with a local real conductivity $\boldsymbol{\sigma}(\mathbf{x}) > 0$ and effective conductivity $\boldsymbol{\sigma}_*$ the bounds (2.73) become

$$\boldsymbol{\sigma}_* \geq 0, \quad \boldsymbol{\sigma}_* \leq \langle \boldsymbol{\sigma} \rangle, \quad \boldsymbol{\sigma}_*^{-1} \leq \langle \boldsymbol{\sigma}^{-1} \rangle, \tag{2.74}$$

where the angular brackets denote a volume average over the period cell. The latter arithmetic/harmonic mean bounds are known as the Wiener (1912) bounds. More generally, for many other equations in periodic composites where the local real self-adjoint tensor is $\mathbf{L}(\mathbf{x}) > 0$ and the effective tensor is $\mathbf{L}_*$ one has the classical bounds

$$\mathbf{L}_* \geq 0, \quad \mathbf{L}_* \leq \langle \mathbf{L} \rangle, \quad \mathbf{L}_*^{-1} \leq \langle \mathbf{L}^{-1} \rangle. \tag{2.75}$$

For example, for elasticity, the latter two bounds were derived by Hill (1952) and are known as the Voigt–Reuss–Hill bounds. [Voigt (1889, 1910) and Reuss (1929) had suggested the arithmetic/harmonic mean averages as approximations, but in fact did not prove they were bounds].

2.8 The abstract setting for defining Y-tensors

Here we review the general setting of the Y-tensor problem. In the subsequent sections of the chapter we will see how it is related to the effective tensor, and how it plays an important role in the theory of composites and in characterizing the response of multiterminal impedance networks. The setting is now a Hilbert space or a finite-dimensional vector space $\mathcal{K}$ (with an inner product satisfying (2.16)) that has the decomposition

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{H}, \tag{2.76}$$

where the spaces $\mathcal{E}$ and $\mathcal{J}$ are orthogonal complements, as are the spaces $\mathcal{V}$ and $\mathcal{H}$. (For the moment these are not to be confused with the spaces $\mathcal{E}$, $\mathcal{J}$ and $\mathcal{H}$ associated with effective tensors). We let $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ denote the projections onto the spaces $\mathcal{E}$ and $\mathcal{J}$, while we let $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ denote the projections onto the spaces $\mathcal{V}$ and $\mathcal{H}$. Given a linear operator $\mathbf{L}$ which maps $\mathcal{H}$ to $\mathcal{H}$, the Y-tensor problem is to find for each given field (or vector) $\mathbf{E}_1 \in \mathcal{V}$ the associated fields

$$\mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}, \quad \mathbf{J}_1 \in \mathcal{V}, \quad \text{with } \mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}, \quad \mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}, \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2. \tag{2.77}$$

At first sight the Y-tensor problem looks almost the same as the effective tensor problem, but what is important is the different partitioning of the Hilbert space ((2.76) compared to (2.17)) and the different association of fields ((2.77) compared to (2.18) and (2.19)). Thus the action of $\mathbf{L}$ needs only to be defined on the subspace $\mathcal{H}$ and its action on fields or vectors in $\mathcal{V}$ may not even be defined. Supposing that a unique solution exists for each $\mathbf{E}_1 \in \mathcal{V}$, the associated field (or vector) $\mathbf{J}_1$ must depend linearly of $\mathbf{E}_1$ and this linear relation,

$$\mathbf{J}_1 = -\mathbf{Y}_* \mathbf{E}_1, \tag{2.78}$$

defines the linear operator $\mathbf{Y}_*$ which maps $\mathcal{V}$ to $\mathcal{V}$, or to a subspace of $\mathcal{V}$. The dual problem consists of finding for each given field (or vector) $\mathbf{J}_1 \in \mathcal{V}$ the associated fields

$$\mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}, \quad \mathbf{E}_1 \in \mathcal{V}, \quad \text{with } \mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}, \quad \mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}, \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2. \tag{2.79}$$

Then supposing that a unique solution exists for each $\mathbf{J}_1 \in \mathcal{V}$, the linear relation,

$$\mathbf{E}_1 = -\mathbf{Y}_*^{-1}\mathbf{J}_1, \tag{2.80}$$

defines the inverse operator $\mathbf{Y}_*^{-1}$. The appearance of the minus sign in front of (2.78) and (2.80) looks a little strange but is motivated by the fact that defined in this way $\mathbf{Y}_*$ is positive semidefinite, when $\mathbf{L}$ is Hermitian and positive semidefinite. To see this we follow equation (19.12) in Milton (2002) and note that the orthogonality of the various subspaces implies

$$0 = (\mathbf{E}_1 + \mathbf{E}_2, \mathbf{J}_1 + \mathbf{J}_2) = (\mathbf{E}_1, \mathbf{J}_1) + (\mathbf{E}_2, \mathbf{J}_2) = -(\mathbf{E}_1, \mathbf{Y}_*\mathbf{E}_1) + (\mathbf{E}_2, \mathbf{L}\mathbf{E}_2), \tag{2.81}$$

or equivalently that $(\mathbf{E}_1, \mathbf{Y}_*\mathbf{E}_1) = (\mathbf{E}_2, \mathbf{L}\mathbf{E}_2)$ where the latter is nonnegative if $\mathbf{L}$ is Hermitian and positive semidefinite.

It is easy to check, following a string of identities like (2.67), that if an operator $\mathbf{L}$ has a Y-tensor $\mathbf{Y}_*$ then the adjoint operator $\mathbf{L}^\dagger$ will have a Y-tensor $\mathbf{Y}_*^\dagger$ which is the adjoint of $\mathbf{Y}_*$. Hence adding (2.81) to its complex conjugate we obtain

$$(\mathbf{E}_1, (\mathbf{Y}_* + \mathbf{Y}_*^\dagger)\mathbf{E}_1) = (\mathbf{E}_2, (\mathbf{L} + \mathbf{L}^\dagger)\mathbf{E}_2). \tag{2.82}$$

More generally, noting that if we multiply $\mathbf{L}$ by $e^{i\theta}$ then $\mathbf{Y}_*$ will be multiplied by $e^{i\theta}$ we conclude that if $e^{i\theta}\mathbf{L} + e^{-i\theta}\mathbf{L}^\dagger$ is positive definite so to will be $e^{i\theta}\mathbf{Y}_* + e^{-i\theta}\mathbf{Y}_*^\dagger$. Taking $e^{i\theta} = -i$ we conclude that if $\mathbf{L}$ is symmetric with positive semi-definite imaginary part, then $\mathbf{Y}_*$ will be symmetric with positive semi-definite imaginary part.

To solve for the tensor $\mathbf{Y}_*$ one notes that since $\mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}$ it follows that

$$\boldsymbol{\Gamma}_2\mathbf{E}_1 = -\boldsymbol{\Gamma}_2\mathbf{E}_2 = -\boldsymbol{\Gamma}_2\mathbf{L}^{-1}\mathbf{J}_2 = -\boldsymbol{\Gamma}_2\mathbf{L}^{-1}\boldsymbol{\Pi}_2\boldsymbol{\Gamma}_2\mathbf{J}, \tag{2.83}$$

which, provided the operator $\boldsymbol{\Gamma}_2\mathbf{L}^{-1}\boldsymbol{\Pi}_2\boldsymbol{\Gamma}_2$ is nonsingular on the space $\mathcal{J}$, implies

$$\mathbf{J}_1 = \boldsymbol{\Pi}_1\mathbf{J} = -\boldsymbol{\Pi}_1\boldsymbol{\Gamma}_2[\boldsymbol{\Gamma}_2\mathbf{L}^{-1}\boldsymbol{\Pi}_2\boldsymbol{\Gamma}_2]^{-1}\boldsymbol{\Gamma}_2\mathbf{E}_1, \tag{2.84}$$

and this gives the formula

$$\mathbf{Y}_* = \boldsymbol{\Pi}_1\boldsymbol{\Gamma}_2[\boldsymbol{\Gamma}_2\mathbf{L}^{-1}\boldsymbol{\Pi}_2\boldsymbol{\Gamma}_2]^{-1}\boldsymbol{\Gamma}_2\boldsymbol{\Pi}_1, \tag{2.85}$$

where the inverse is to be taken on the subspace $\mathcal{J}$. A necessary constraint for this inverse to be nonsingular is that there is no field (or vector) in $\mathcal{J}$ which is in the null-space of $\boldsymbol{\Gamma}_2$, i.e

$$\mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}. \tag{2.86}$$

Indeed if there is such a field $\mathbf{W}$ in both spaces, then the equations are still satisfied if we add $\mathbf{W}$ to both $\mathbf{J}_1$ and $\mathbf{J}$.

2.9 Example of a Y-tensor problem in two-phase composites

Y-tensor problems most naturally arise in the context of electrical circuits. However, another context in which they have a natural significance is in conduction in two-phase periodic composites where the conductivity tensor $\boldsymbol{\sigma}$ takes the form

$$\boldsymbol{\sigma}(\mathbf{x}) = \boldsymbol{\sigma}_1\chi_1(\mathbf{x}) + \boldsymbol{\sigma}_2\chi_2(\mathbf{x}), \tag{2.87}$$

where the indicator function $\chi_i(\mathbf{x})$ takes the value 1 in phase i, and is zero in the other phase, and the conductivity tensors $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ of the two phases may be anisotropic (and not necessarily commute). The space $\mathcal{K}$ is taken to be all fields which have zero average value, $\mathcal{H}$ those fields which average to zero in each phase separately, $\mathcal{E}$ is taken to be the space of electric fields that are minus the gradient of a periodic potential, and $\mathcal{J}$ the space of divergence free fields that have zero average value (though not necessarily zero average value taken over each phase separately). The operator $\mathbf{L}$ is to be identified with the conductivity tensor $\boldsymbol{\sigma}(\mathbf{x})$, with its action restricted to $\mathcal{H}$ (as $\mathcal{K}$ is not closed under the action of $\boldsymbol{\sigma}(\mathbf{x})$). Note that $\mathcal{H}$ is closed under the action of $\boldsymbol{\sigma}(\mathbf{x})$. The space $\mathcal{V}$ consists of fields of the form

$$\mathbf{V}(\mathbf{x}) = [f_2\chi_1(\mathbf{x}) - f_1\chi_2(\mathbf{x})]\mathbf{v} = [\chi_1(\mathbf{x}) - f_1]\mathbf{v}, \tag{2.88}$$

that are constant in each phase, with zero overall average, where to obtain the last expression in (2.88) we have used the fact that $\chi_2(\mathbf{x}) = 1 - \chi_1(\mathbf{x})$. Due to the simple form of these fields the operator $\mathbf{Y}_*$ can be represented by a local constant operator (this is not true for composites of more than two phases). Thus we have the relations

$$\mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}, \quad \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}, \quad \mathbf{J}_2(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x})\mathbf{E}_2(\mathbf{x}), \quad \mathbf{J}_1(\mathbf{x}) = -\mathbf{Y}_*\mathbf{E}_1(\mathbf{x}). \tag{2.89}$$

Given a solution to these equations, we can now consider the larger space

$$\mathcal{H}' = \mathcal{U} \oplus \mathcal{K} = \mathcal{U} \oplus \mathcal{V} \oplus \mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}, \tag{2.90}$$

where $\mathcal{U}$ is the subspace of uniform fields and $\mathcal{H}'$ is now the space of periodic square integrable fields. We postulate that the solution to the usual conductivity equations can be expressed in terms of the solution to the Y-tensor problem, and takes the form

$$\mathbf{J}_0 + \mathbf{J} = \boldsymbol{\sigma}(\mathbf{E}_0 + \mathbf{E}), \quad \text{with } \mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2, \quad \mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2, \tag{2.91}$$

with $\mathbf{E}_0, \mathbf{J}_0 \in \mathcal{U}$, $\mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}$, and $\mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}$. Since $\mathbf{J}_2 = \mathbf{L}\mathbf{E}_2$ this reduces to a simpler problem,

$$\mathbf{J}_0 + \mathbf{J}_1 = \boldsymbol{\sigma}(\mathbf{E}_0 + \mathbf{E}_1), \quad \mathbf{J}_1 = -\mathbf{Y}_*\mathbf{E}_1, \tag{2.92}$$

that only involves piecewise constant fields. Writing $\mathbf{E}_1 = [\chi_1(\mathbf{x}) - f_1]\mathbf{v}$ the equation becomes

$$\begin{aligned} \mathbf{J}_0 - [\chi_1(\mathbf{x}) - f_1]\mathbf{Y}_*\mathbf{v} &= [(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)\chi_1(\mathbf{x}) + \boldsymbol{\sigma}_2]\{\mathbf{E}_0 + [\chi_1(\mathbf{x}) - f_1]\mathbf{v}\} \\ &= \{(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)[\mathbf{E}_0 + f_2\mathbf{v}] + \boldsymbol{\sigma}_2\mathbf{v}\}\chi_1(\mathbf{x}) + \boldsymbol{\sigma}_2(\mathbf{E}_0 - f_1\mathbf{v}), \end{aligned} \tag{2.93}$$

which separates into

$$\mathbf{Y}_*\mathbf{v} = -\{(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)[\mathbf{E}_0 + f_2\mathbf{v}] + \sigma_2\mathbf{v}\}, \quad \mathbf{J}_0 + f_1\mathbf{Y}_*\mathbf{v} = \boldsymbol{\sigma}_2(\mathbf{E}_0 - f_1\mathbf{v}). \tag{2.94}$$

From the first equation we get

$$\mathbf{v} = -[\mathbf{Y}_* + f_2\boldsymbol{\sigma}_1 + f_1\boldsymbol{\sigma}_2]^{-1}(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)\mathbf{E}_0, \tag{2.95}$$

and by substituting the first into the second we get

$$\mathbf{J}_0 = (f_1\boldsymbol{\sigma}_1 + f_2\boldsymbol{\sigma}_2)\mathbf{E}_0 + f_1f_2(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)\mathbf{v}, \tag{2.96}$$

and taken together they imply a formula for the effective conductivity tensor $\mathbf{L}_* = \boldsymbol{\sigma}_*$ in terms of the Y-tensor:

$$\boldsymbol{\sigma}_* = f_1\boldsymbol{\sigma}_1 + f_2\boldsymbol{\sigma}_2 - f_1f_2(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)[\mathbf{Y}_* + f_2\boldsymbol{\sigma}_1 + f_1\boldsymbol{\sigma}_2]^{-1}(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2). \tag{2.97}$$

Using the last equation in (2.89) we also see that, for either $i = 1$ or $i = 2$,

$$\mathbf{Y}_*\langle\chi_i\mathbf{E}\rangle = \mathbf{Y}_*\langle\chi_i\mathbf{E}_1\rangle = \langle\chi_i\mathbf{Y}_*\mathbf{E}_1\rangle = \langle\chi_i\mathbf{J}_1\rangle = \langle\chi_i\mathbf{J}\rangle, \tag{2.98}$$

which gives a direct physical meaning to the tensor $\mathbf{Y}_*$ in two-phase composites (Gibiansky and Milton 1993).

One advantage of introducing the tensor $\mathbf{Y}_*$ is that well-known bounds take a simpler form. As follows from the observations of Torquato (1980), Milton (1981b, 1982) and Berryman (1982), when the components and effective tensor are isotropic, i.e., $\boldsymbol{\sigma}_i = \sigma_i\mathbf{I}$ for $i = 1, 2, *$, and $\mathbf{Y}_* = y_*\mathbf{I}$, the well-known bounds of Hashin and Shtrikman (1962) on the effective conductivity σ_* of a three-dimensional composite, when $\sigma_1 \geq \sigma_2$, reduce to

$$\sigma_1 \geq y_*/2 \geq \sigma_2, \tag{2.99}$$

while the bounds of Beran (1965) reduce to

$$\zeta_1\sigma_1 + \zeta_2\sigma_2 \geq y_*/2 \geq (\zeta_1/\sigma_1 + \zeta_2/\sigma_2)^{-1}, \tag{2.100}$$

where ζ_1 and $\zeta_2 = 1 - \zeta_1$ are nonnegative weights, that can be determined from three-point correlation functions:

$$\zeta_1 = 1 - \zeta_2 = \frac{9}{2f_1f_2}\int_0^\infty dr\int_0^\infty ds\int_{-1}^{+1} du\frac{f_{111}(r,s,u)}{rs}P_2(u), \tag{2.101}$$

in which $P_2(u) = (3u^2 - 1)/2$ is a Legendre polynomial, and $f_{111}(r, s, u)$ is the probability that all three vertices of a triangle having side lengths r and s, and included angle $\cos^{-1} u$ land in phase 1 when thrown randomly in the composite.

Curiously, Berryman (1982) found that the self-consistent equation

$$y_*/2 = \sigma_*, \tag{2.102}$$

corresponded to the well-known effective medium approximation of Bruggeman (1935) that is realized by a hierarchical model consisting of spheres of the two-components with a very wide range of sizes, distributed so spheres of similar size are well-separated (Milton 1984, 1985). This result of Berryman extends to multi-component media too (Milton 1987a).

2.10 Analytic properties preserving feature of the Y-transformation in two-phase composites

The transformation from σ_* to y_* is also important (Milton and Golden 1985), for simplifying and deriving bounds when the conductivities σ_1 and σ_2 are complex. When σ_*, for example, represents a diagonal elements of the effective conductivity tensor $\boldsymbol{\sigma}_*$, then bounds on $\boldsymbol{\sigma}_*$ have been obtained using the analytic properties of the function $\sigma_*(\sigma_1, \sigma_2)$ established by Bergman (1978), Milton (1981a) and Golden and Papanicolaou (1983). These properties are that the function satisfies the:

1. **The Homogeneity property**: $\sigma_*(\lambda\sigma_1, \lambda\sigma_2) = \lambda\sigma_*(\sigma_1, \sigma_2)$ for all real or complex λ;
2. **The Analyticity property**: $\sigma_*(\sigma_1, \sigma_2)$ is an analytic function of the complex variables σ_1 and σ_2 except possibly when σ_1/σ_2 is real, and zero or negative;
3. **The Herglotz property**: $\mathrm{Im}(\sigma_*(\sigma_1, \sigma_2)) > 0$ when $\mathrm{Im}(\sigma_1) > 0$ and $\mathrm{Im}(\sigma_2) > 0$;
4. **The Normalization property**: $\sigma_*(1, 1) = 1$.

From a physical viewpoint it is more natural to require $\mathrm{Re}(\sigma_*) > 0$ when $\mathrm{Re}(\sigma_1) > 0$ and $\mathrm{Re}(\sigma_2) > 0$ rather than the Herglotz property (unless σ is representing the dielectric constant) but both are equivalent as follows from the homogeneity property with $\lambda = i$. If we ignore the normalization property 4, and know nothing else about the function $\sigma_*(\lambda\sigma_1, \lambda\sigma_1)$ other than properties 1,2 and 3, then the most we can say is that σ_* satisfies **the wedge bounds**: σ_* lies in the wedge W in the complex plane bounded by the two straight lines which both pass through the origin, and one through σ_1 and the other through σ_2 (these wedge bounds follow easily from the Herglotz property and the fact that one can rotate the complex plane due to the homogeneity property with $\lambda = e^{i\theta}$).

Tighter bounds when more information is known were obtained by exploiting the analytic properties of the function $\sigma_*(\lambda\sigma_1, \lambda\sigma_1)$ by Milton(1980, 1981a, 1981b), Bergman(1980, 1982) and Clark and Milton (1995): see also Milton (1979), Golden and Papanicolaou (1983) and Bergman (1993). In a wider mathematical context most of these bounds follow from bounds of Stieltjes functions, see Milton (1986a), the discussion in the Introduction of Milton 1987b and references therein, and **Chapter V** in Kreĭn and Nudel'man (1974). Many of them can alternatively be derived from variational principles (Milton and McPhedran 1982; Cherkaev and Gibiansky 1994; Milton 1990).

As brought recently to my attention by Mihai Putinar, there is also a close connection to the Nevanlinna–Pick interpolation problem, solved by Nevanlinna (1919, 1929) and Pick (1915), of obtaining sharp bounds which correlate the values a Herglotz-function takes at a set of points in the upper half of the complex plane (a Herglotz function is a function which is analytic in the upper half of the complex plane, and has positive imaginary part there). Generalizations of the Nevanlinna–Pick interpolation problem, and different ways to obtain these generalizations, have been the subject of much research [see Ball and Trent (1998), the book of Agler and McCarthy (2002), the appendix of Charina, Putinar, Scheiderer, and Stöckler (2015), and references therein]. Of particular relevance to the theory of composites (as for anisotropic materials the effective tensor is represented by a matrix) is that the Nevanlinna–Pick interpolation problem has been solved for matrix-valued Herglotz functions (Delsarte, Genin, and Kamp 1979) and that algorithms are available for computing interpolations (Chen and Koç 1994, 1995). Nevanlinna–Pick interpolation and its generalizations to multivariate functions are also important in circuit and system theory, network synthesis, and control theory (Delsarte, Genin, and Kamp 1979; Kummert 1989; Ball and ter Horst 2010).

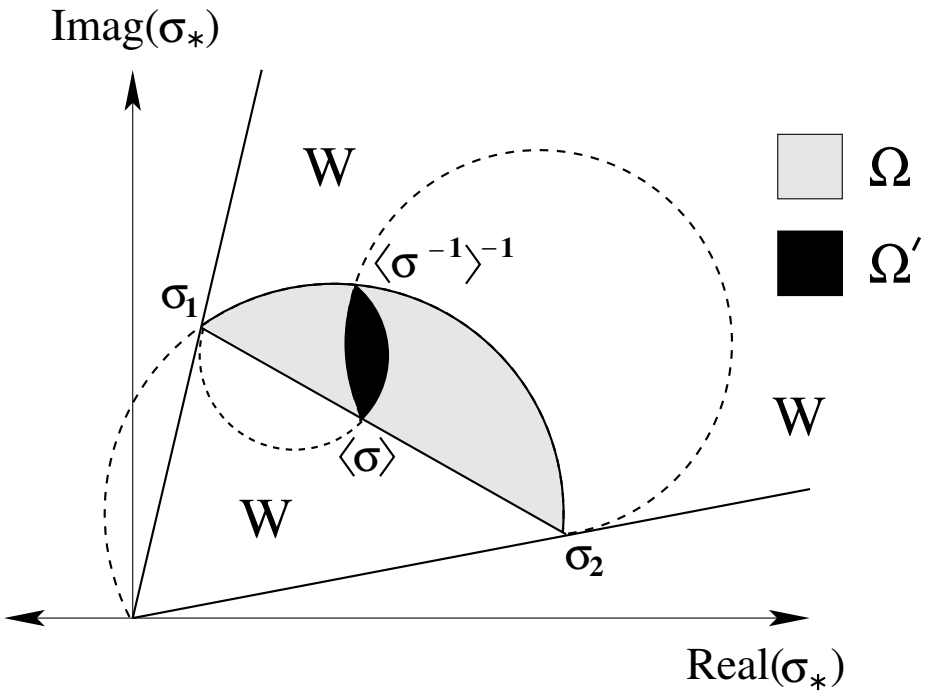


Figure 2.5: Bounds on a diagonal element σ_* of the complex effective tensor $\boldsymbol{\sigma}_*$. If one ignores the normalization property that $\sigma_*(1,1) = 1$, σ_* is only confined to the wedge W. If the normalization property is taken into account, σ_* is confined to Ω, and if the volume fractions are known, σ_* is confined to Ω'. The Y-transformation (2.105) maps Ω' back to the wedge.

When we include the normalization property, σ_* is confined to the lens–shaped shaded region Ω in **Figure 2.5**, bounded on one side by the straight line joining σ_1 and σ_2, and on the other side by the circular arc joining these two points that when extended passes through the origin. When the volume fractions of the phases are known, this information translates to knowledge of the first derivative,

$$f_1 = \left.\frac{d\sigma_*(\sigma_1, 1)}{d\sigma_1}\right|_{\sigma_1=1}, \tag{2.103}$$

and σ_* is confined to the lens–shaped black region Ω' in **Figure 2.5**, bounded by two circular arcs, both passing through

$$\langle\sigma\rangle = f_1\sigma_1 + f_2\sigma_2, \quad \text{and } \langle\sigma^{-1}\rangle^{-1} = 1/(f_1/\sigma_1 + f_2/\sigma_2), \tag{2.104}$$

one of which, when extended, passes through σ_1 while the other, when extended, passes through σ_2.

Now lets look for a transformation which maps the more complicated bound, represented by the black region Ω' back to the most elementary bounds, represented by the wedge W. Since the boundaries of these regions are straight lines or circular arcs it makes sense to look in the class of fractional linear transformations since these map circular arcs or straight lines to straight lines or circular arcs. The point $\langle\sigma^{-1}\rangle^{-1}$ since it is at the intersection of the two circular arcs bounding Ω', should get mapped to the origin (or the point at infinity)

being the intersection of the straight lines forming the wedge. Similarly the point $\langle\sigma\rangle$, should get mapped to infinity (or the point at the origin) being the other point of intersection of the straight lines forming the wedge. The transformation from σ_* to y_* does precisely this:

$$\begin{aligned} y_* \equiv Y(\sigma_*) &= -f_2\sigma_1 - f_1\sigma_2 + f_1f_2(\sigma_1-\sigma_2)^2[f_1\sigma_1 + f_2\sigma_2 - \sigma_*]^{-1} \\ &= \sigma_1\sigma_2(\sigma_*\langle\sigma^{-1}\rangle - 1)/(\langle\sigma\rangle - \sigma_*), \end{aligned} \tag{2.105}$$

as follows from the fact that

$$Y(\langle\sigma^{-1}\rangle^{-1}) = 0, \quad Y(\langle\sigma\rangle) = \infty, \quad Y(\sigma_1) = -\sigma_1, \quad Y(\sigma_1) = -\sigma_1. \tag{2.106}$$

Thus the circular arcs forming the boundary of Ω' that pass through the points $\langle\sigma\rangle$ and $\langle\sigma^{-1}\rangle^{-1}$ which when extended pass through σ_1 and σ_2 map to rays from the origin to infinity, that when extended in the opposite direction pass through $-\sigma_1$ and $-\sigma_2$. These are the rays that bound W.

The wedge bounds immediately imply $y_*(\sigma_1,\sigma_2)$ satisfies the Herglotz property, and hence shares with $\sigma_*(\sigma_1,\sigma_2)$ the analytic properties 1, 2, and 3. (Note that $y_*(\sigma_1,\sigma_2)$ cannot have a zero or pole except when σ_1/σ_2 is real and negative, since in the vicinity of that pole or zero, the wedge bounds would be violated. Also $y_*(\sigma_1,\sigma_2)$ cannot have other singularities, except when σ_1/σ_2 is real and negative, as these would transfer to $\sigma_*(\sigma_1,\sigma_2)$). A more direct proof that the transformation (2.105), or in fact its matrix analog, preserves analytic properties is given in Section 28.4 of Milton (2002).

Now in a three-dimensional isotropic composite (2.99) (or the series expansion for $\sigma_*(\sigma_1,\sigma_2)$) implies $y_*(1,1) = 2$. Making a normalization transformation of $y_*(\sigma_1,\sigma_1)$ by this factor $n = y_*(1,1) = 2$ we obtain a function

$$\sigma_*^{(1)}(\sigma_1,\sigma_2) = y_*(\sigma_1,\sigma_1)/2, \tag{2.107}$$

satisfying all four analytic properties. In particular $\sigma_*^{(1)}$ is also confined the lens–shaped shaded region Ω in **Figure 2.5**, bounded on one side by the straight line joining σ_1 and σ_2, and on the other side by the circular arc joining these two points that when extended passes through the origin. When mapped to the σ_* plane we see that σ_* is confined to a lens–shaped region Ω'' inside Ω'. Furthermore knowledge of the parameter $\zeta_1 = 1-\zeta_2$ given by (2.101), implies through (2.100) (or through the series expansion for $\sigma_*(\sigma_1,\sigma_2)$), knowledge of

$$\zeta_1 = \left.\frac{d\sigma_*^{(1)}(\sigma_1,1)}{d\sigma_1}\right|_{\sigma_1=1}. \tag{2.108}$$

Hence we can define

$$y_*^{(1)} = -\zeta_2\sigma_1 - \zeta_1\sigma_2 + \zeta_1\zeta_2(\sigma_1-\sigma_2)^2[\zeta_1\sigma_1 + \zeta_2\sigma_2 - \sigma_*^{(1)}]^{-1}, \tag{2.109}$$

and the wedge bounds on $y_*^{(1)}$ provide even tighter bounds on $\sigma_*^{(1)}$ and hence on σ_*. As more and more series expansion coefficients are incorporated in the bounds one can introduce a hierarchy of functions $\sigma_*^{(n)}(\sigma_1,\sigma_1)$ and $y_*^{(n)}(\sigma_1,\sigma_1)$, that are linked by fractional linear transformations and normalization transformations. In this way, one obtains a nested sequence of lens–shaped regions in the complex plane which bound σ_*. The effect of the factor of $(\sigma_1-\sigma_2)^2$ entering the Y-transformation (2.105), is to shift to order m information that is contained in the series expansion of $\sigma_*(\sigma_1,\sigma_2)$ at order $m+2$.

For other problems in composites, where one has tensors $\mathbf{L}_1$ and $\mathbf{L}_2$ in phases 1 and 2 (which may for example, be elasticity tensors) and an effective tensor $\mathbf{L}_*$, (2.97) generalizes to

$$\mathbf{L}_* = f_1\mathbf{L}_1 + f_2\mathbf{L}_2 - f_1f_2(\mathbf{L}_1-\mathbf{L}_2)[\mathbf{Y}_* + f_2\mathbf{L}_1 + f_1\mathbf{L}_2]^{-1}(\mathbf{L}_1-\mathbf{L}_2). \tag{2.110}$$

which is the inverse of the relation (2.15).

2.11 Bounds on the Y-tensor in two-phase composites using the translation method

As an example of the usefulness of the Y-tensor in two phase composites let us derive the associated translation bounds. Translation bounds on the Y-tensor were first derived using algebraic manipulations of the translation bounds on $\mathbf{L}_*$ by Cherkaev and Gibiansky (1992) assuming $\mathbf{L}_1$ and $\mathbf{L}_2$ commute. This restriction was subsequently removed (Milton 1991). The following simple derivation follows Section 24.10 in Milton (2002).

One looks for constant self-adjoint tensors $\mathbf{T}$ (translations) whose associated quadratic forms are quasiconvex on the space $\mathcal{E}$, i.e, such that

$$(\mathbf{E}, \mathbf{T}\mathbf{E}) \geq 0, \quad \text{for all } \mathbf{E} \in \mathcal{E}. \tag{2.111}$$

By taking Fourier transforms this condition reduces to an algebraic condition on $\mathbf{T}$: for example if $\mathbf{E} = \nabla \mathbf{u}$, then the Fourier components of $\mathbf{E}$ are rank-one matrices, and so $\underline{\mathbf{A}} \cdot \mathbf{T}\mathbf{A}$ must be nonnegative for all rank-one matrices $\mathbf{A}$. Therefore it is quite easy to see if the quadratic form associated with a tensor $\mathbf{T}$ is quasiconvex or not (but it is less easy to see which are the best translations to take).

Now $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}$ and $\mathbf{T}\mathbf{E}_1$ is piecewise constant and hence orthogonal to $\mathbf{E}_2$. So we deduce that

$$0 \leq (\mathbf{E}_1 + \mathbf{E}_2, \mathbf{T}(\mathbf{E}_1 + \mathbf{E}_2)) = (\mathbf{E}_1 \mathbf{T}\mathbf{E}_1) + (\mathbf{E}_2, \mathbf{T}\mathbf{E}_2). \tag{2.112}$$

By the orthogonality of the subspaces we also have the identity

$$0 = (\mathbf{E}_1 + \mathbf{E}_2, \mathbf{J}_1 + \mathbf{J}_2) = (\mathbf{E}_1, \mathbf{J}_1) + (\mathbf{E}_2, \mathbf{J}_2) = (\mathbf{E}_2, \mathbf{L}\mathbf{E}_2) - (\mathbf{E}_1, \mathbf{Y}_*\mathbf{E}_1). \tag{2.113}$$

Now suppose $\mathbf{T}$ is such that $\mathbf{L} - \mathbf{T} \geq 0$, i.e., $\mathbf{L}_1 - \mathbf{T}$ and $\mathbf{L}_2 - \mathbf{T}$ are positive semidefinite matrices. Then it follows from the above two equations that

$$(\mathbf{E}_1, \mathbf{Y}_*\mathbf{E}_1) = (\mathbf{E}_2, \mathbf{L}\mathbf{E}_2) \geq (\mathbf{E}_2, \mathbf{T}\mathbf{E}_2) \geq -(\mathbf{E}_1, \mathbf{T}\mathbf{E}_1). \tag{2.114}$$

So we have the translation bound that

$$\mathbf{Y}_* + \mathbf{T} \geq 0 \quad \text{if } \mathbf{L}_1 - \mathbf{T} \geq 0 \text{ and } \mathbf{L}_1 - \mathbf{T} \geq 0. \tag{2.115}$$

By contrast the corresponding bound on $\mathbf{L}_*$ is more complicated:

$$\mathbf{L}_* \geq \mathbf{T} + [f_1(\mathbf{L}_1 - \mathbf{T})^{-1} + f_1(\mathbf{L}_2 - \mathbf{T})^{-1}]^{-1}. \tag{2.116}$$

The form (2.115) of the translation bounds makes it clear that if $\mathbf{T} = \mathbf{T}' + \mathbf{A}$ where $\mathbf{A}$ is positive semidefinite and $\mathbf{T}'$ also satisfies (2.111) then the bounds using $\mathbf{T}'$ will be as least as good as those using $\mathbf{T}$. Thus the best translations to use are the extremal ones, which are such that the associated quasiconvex quadratic form loses its quasiconvexity whenever a nonzero positive-definite quadratic form is subtracted from it (Milton 1990). An algorithm for constructing extremal translations was given in Milton (2013b), and an explicit example of a (nontrivial) extremal translation was presented by Harutyunyan and Milton (2015a). Curiously there seems to be a connection between extremal translations and extremal polynomials (Harutyunyan and Milton 2015b).

2.12 Introducing the Y-tensor in multiphase composites using variational principles

In composite materials, when the local tensor $\mathbf{L}(\mathbf{x})$ is real, symmetric, and positive definite, the effective tensor is given by the variational principle

$$\mathbf{E}_0 \cdot \mathbf{L}_* \mathbf{E}_0 = \min_{\mathbf{E} \in \mathcal{E}} \langle (\mathbf{E}_0 + \mathbf{E}) \cdot \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) \rangle, \tag{2.117}$$

as proved in Section 2.7. When the material has n-phases (each with constant orientation) we can introduce the space $\mathcal{V}$ consisting of those fields $\mathbf{V}$ which are constant in each phases, yet average to zero so they are orthogonal to the space $\mathcal{U}$:

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{v}_i \chi_i(\mathbf{x}), \quad \text{where} \quad \sum_{i=1}^{n} f_i \mathbf{v}_i = 0, \tag{2.118}$$

in which $\chi_i(\mathbf{x})$ is the indicator function taking the value 1 in phase i and zero elsewhere, while $f_i = \langle \chi_i \rangle$ is the volume fraction of component i. Also let us introduce the space $\mathcal{H}^{(1)}$ consisting of those fields $\mathbf{P}(\mathbf{x})$ that have zero average value over each phase: i.e., $\mathbf{P}(\mathbf{x})\chi_i(\mathbf{x}) = 0$ for all i. Now the trial field $\mathbf{E}$ can be decomposed into a sum

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \quad \text{where } \mathbf{E}_1 \in \mathcal{V}, \quad \mathbf{E}_2 \in \mathcal{H}^{(1)}, \tag{2.119}$$

and $\mathbf{L}(\mathbf{x})$, by assumption, takes the form

$$\mathbf{L}(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{L}_i \chi_i(\mathbf{x}). \tag{2.120}$$

Since the spaces $\mathcal{U} \oplus \mathcal{V}$ and $\mathcal{H}^{(1)}$ are each closed under the action of $\mathbf{L}$ and are orthogonal to each other it follows that

$$\langle (\mathbf{E}_0 + \mathbf{E}) \cdot \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) \rangle = \langle (\mathbf{E}_0 + \mathbf{E}_1) \cdot \mathbf{L}(\mathbf{E}_0 + \mathbf{E}_1) \rangle + \langle \mathbf{E}_2 \cdot \mathbf{L}\mathbf{E}_2 \rangle. \tag{2.121}$$

Consequently the minimization in the variational principle for $\mathbf{L}_*$ can be done in two steps: first the computation of the quadratic form,

$$(\mathbf{E}_1, \mathbf{Y}_* \mathbf{E}_1) = \min_{\substack{\mathbf{E}_2 \in \mathcal{H}^{(1)} \\ \mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}}} \langle \mathbf{E}_2 \cdot \mathbf{L}\mathbf{E}_2 \rangle, \tag{2.122}$$

which defines the Y-tensor $\mathbf{Y}_*$, mapping $\mathcal{V}$ to $\mathcal{V}$, and then the computation of

$$\mathbf{E}_0 \cdot \mathbf{L}_* \mathbf{E}_0 = \min_{\mathbf{E}_1 \in \mathcal{V}} \langle (\mathbf{E}_0 + \mathbf{E}_1) \cdot \mathbf{L}(\mathbf{E}_0 + \mathbf{E}_1) \rangle + (\mathbf{E}_1, \mathbf{Y}_* \mathbf{E}_1), \tag{2.123}$$

which gives the relation between the effective tensor $\mathbf{L}_*$ and the tensor $\mathbf{Y}_*$.

If in the variational principle (2.122), $\mathbf{E}_2(\mathbf{x})$ is varied by a small perturbing field $\delta\mathbf{E}_2(\mathbf{x}) \in \mathcal{E} \cap \mathcal{H}^{(1)}$ then the first order variation in the integrand will be $2\langle \delta\mathbf{E}_2 \cdot \mathbf{L}\mathbf{E}_2 \rangle$ and so will vanish if $\mathbf{J}_2 \equiv \mathbf{L}\mathbf{E}_2 \in \mathcal{J} \oplus \mathcal{V}$. Since $\mathcal{H}^{(1)}$ is closed under the action of $\mathbf{L}$, it follows that $\mathbf{J}_2 \in \mathcal{H}^{(1)}$ and

$$\mathbf{J}_2 = \mathbf{J} - \mathbf{J}_1, \quad \text{with } \mathbf{J} \in \mathcal{J},\ \mathbf{J}_1 \in \mathcal{V}. \tag{2.124}$$

The value of the minimum in (2.122) is therefore

$$\langle (\mathbf{E} - \mathbf{E}_1) \cdot (\mathbf{J} - \mathbf{J}_1) \rangle = \langle \mathbf{E}_1 \cdot \mathbf{J}_1 - (\mathbf{E}_1 + \mathbf{E}_2) \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot (\mathbf{J}_1 + \mathbf{J}_2) \rangle = -\langle \mathbf{E}_1 \cdot \mathbf{J}_1 \rangle. \tag{2.125}$$

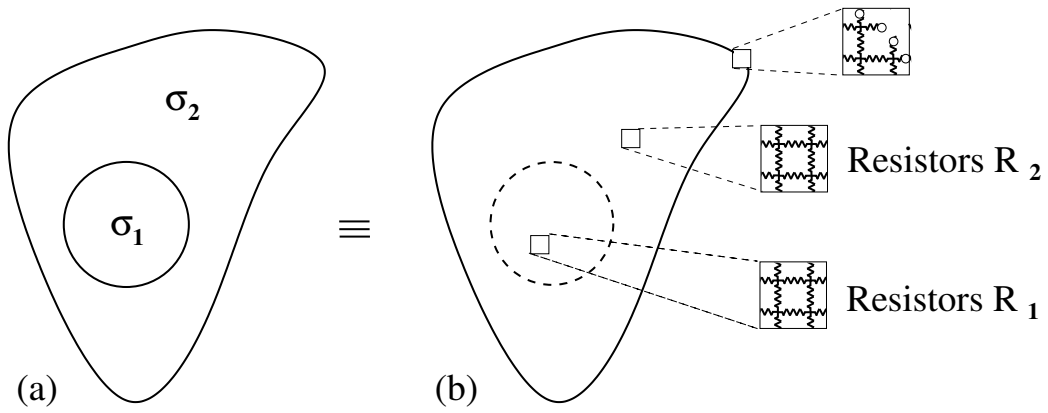


Figure 2.6: A conducting body may be approximated by a discrete network.

Comparing this with the left of (2.122) we see that $\mathbf{J}_1 = -\mathbf{Y}_*\mathbf{E}_1$. This establishes the minimization variational principle (2.122) for $\mathbf{Y}_*$.

From (2.123) one can deduce (see **Section 19.2** in Milton 2002) that the operator $\mathbf{L}_*$ and the operator $\mathbf{Y}_*$ are linked via the relation

$$\mathbf{L}_* = \langle \mathbf{L} \rangle - \mathbf{\Gamma}_0 \mathbf{L} \mathbf{\Pi}_1 [\mathbf{\Pi}_1 \mathbf{L} \mathbf{\Pi}_1 + \mathbf{Y}_*]^{-1} \mathbf{\Pi}_1 \mathbf{L} \mathbf{\Gamma}_0, \tag{2.126}$$

in which $\mathbf{\Pi}_1$ is the projection onto $\mathcal{V}$ and the inverse is to be done on this space. The relation between the matrix representing $\mathbf{Y}_*$ and the matrix representing $\mathbf{L}_*$ is given by (2.126) but depends on what basis one uses for the subspaces $\mathcal{U}$ and $\mathcal{V}$. Unlike the case when there are two-phases, one cannot for a fixed value of $\mathbf{L}$ recover $\mathbf{Y}_*$ from $\mathbf{L}_*$. However, if one looks at the equation with source terms (constant in each phase), then one can recover $\mathbf{Y}_*$ from the macroscopic response: see Section 19.3 in Milton (2002).

2.13 Effective tensors and Y-tensors for discrete electrical circuits

The connection between Dirichlet-to-Neumann maps and the theory of composites established in the next chapter is clearer if we replace the continuous body by a discrete electrical network. For an entertaining introduction to electrical circuits and their connection with random walks see Doyle and Snell (1984). The approach we follow, using incidence matrices, is nicely presented by Strang (1986), pages 87–95.

Suppose, for simplicity, we were considering electrical conductivity and the body was composed of two conducting materials with conductivities σ_1 and σ_2. Then we could replace the body by a discrete resistor network as shown in **Figure 2.6**, where applied electrical potentials and the resulting fluxes of current through the boundary are replaced by potentials and fluxes of current through electrical terminals placed at the former boundary of the body. That effective tensors and Y-tensors may be associated with such resistor networks is discussed in depth in **Chapter 20** of Milton (2002). Briefly, the finite-dimensional vector space $\mathcal{H}$ associated with the network consists of fields which are constant in each bond, but which have a direction. The bonds have an arrow associated with them, and the field component in that bond is positive if the field is directed in

the direction of the arrow, and is negative if the field is directed opposite to the arrow. It is helpful to introduce the incidence matrix $\mathbf{M}$, with $M_{ij} = 1$ if the arrow of bond i points towards node j, $M_{ij} = -1$ if the arrow of bond i points away from node j, and $M_{ij} = 0$ if bond i and node j are not connected. Let $z_0 = 1$ be our reference admittance. We take the space $\mathcal{U}$ to consist of fields $\mathbf{U}$ satisfying

$$\mathbf{U}_0 = -\mathbf{M}\boldsymbol{\varphi}, \quad \mathbf{M}^T\mathbf{U}_0 = \mathbf{I}_0, \tag{2.127}$$

where $\boldsymbol{\varphi}$ represents the potentials at the nodes taking the value $\boldsymbol{\varphi} = \boldsymbol{\varphi}_0$ at the terminal nodes, and $\mathbf{I}_0$ represents the net current flowing out of each node (taking a negative value if there is a net current flowing into that node) which we restrict to be zero for all nodes but the terminal nodes. We take $\mathcal{E}$ to consist of fields expressible as $\mathbf{E} = -\mathbf{M}\boldsymbol{\varphi}$ where $\boldsymbol{\varphi}$ is zero at the terminal nodes, and we take $\mathcal{J}$ to consist of fields $\mathbf{J}$ such that $\mathbf{M}^T\mathbf{J} = 0$. It is easy to check that the spaces $\mathcal{U}$, $\mathcal{E}$, and $\mathcal{J}$ are mutually orthogonal using the fact that the null space of $\mathbf{M}^T$ is orthogonal to the range of $\mathbf{M}$. Solving the conductivity equations in the network is then equivalent to finding fields $\mathbf{E}_0$ and $\mathbf{J}_0$ in $\mathcal{U}$, $\mathbf{E} \in \mathcal{E}$, and $\mathbf{J} \in \mathcal{J}$ such that

$$(\mathbf{J}_0 + \mathbf{J}) = \mathbf{L}(\mathbf{E}_0 + \mathbf{E}), \tag{2.128}$$

where

$$\mathbf{L} = \sum_{i=1}^{n} z_i \boldsymbol{\chi}_i, \tag{2.129}$$

and the indicator function $\boldsymbol{\chi}_i$ is 1 in those bonds having admittance z_i, and zero otherwise. (The admittance is the inverse of the resistance). The effective tensor $\mathbf{L}_*$ by definition governs the relation between $\mathbf{J}_0$ and $\mathbf{E}_0$,

$$\mathbf{J}_0 = \mathbf{L}_*\mathbf{E}_0, \tag{2.130}$$

and measures the overall response of the network, relative to its response when $z = z_0$ in every bond in the network. To see this take an orthonormal basis $\mathbf{U}_\beta$ of $\mathcal{U}$ indexed by the integer β, and resolve $\mathbf{E}_0$ into its components

$$E^0_\beta = \mathbf{U}_\beta \cdot \mathbf{E}_0 = -(\mathbf{M}^T\mathbf{U}_\beta) \cdot \boldsymbol{\varphi}_0 = -\mathbf{I}_\beta \cdot \boldsymbol{\varphi}_0, \tag{2.131}$$

where $\boldsymbol{\varphi}_0$ are the potentials we apply at the terminals. Similarly we resolve $\mathbf{J}_0$ into its components

$$J^0_\gamma = \mathbf{U}_\gamma \cdot \mathbf{J}_0 = -(\mathbf{M}^T\mathbf{J}_0) \cdot \boldsymbol{\varphi}_{(\gamma)} = -\mathbf{I}_0 \cdot \boldsymbol{\varphi}_{(\gamma)}, \tag{2.132}$$

where $\mathbf{I}_0$ represents the current fluxes we measure at the terminals.

In this basis the relation (2.130) takes the form

$$J^0_\gamma = L_{*\gamma\beta}E^0_\beta, \tag{2.133}$$

implying

$$\mathbf{I}_0 \cdot \boldsymbol{\varphi}_{(\gamma)} = L_{*\gamma\beta}\mathbf{I}_\beta \cdot \boldsymbol{\varphi}_0. \tag{2.134}$$

Writing this relation out using indices we have

$$I^0_i \cdot \varphi_{i\gamma} = L_{*\gamma\beta}\mathbf{I}_\beta \cdot \boldsymbol{\varphi}_0, \tag{2.135}$$

giving

$$I^0_m = K_{m\gamma}L_{*\gamma\beta}\mathbf{I}_\beta \cdot \boldsymbol{\varphi}_0, \tag{2.136}$$

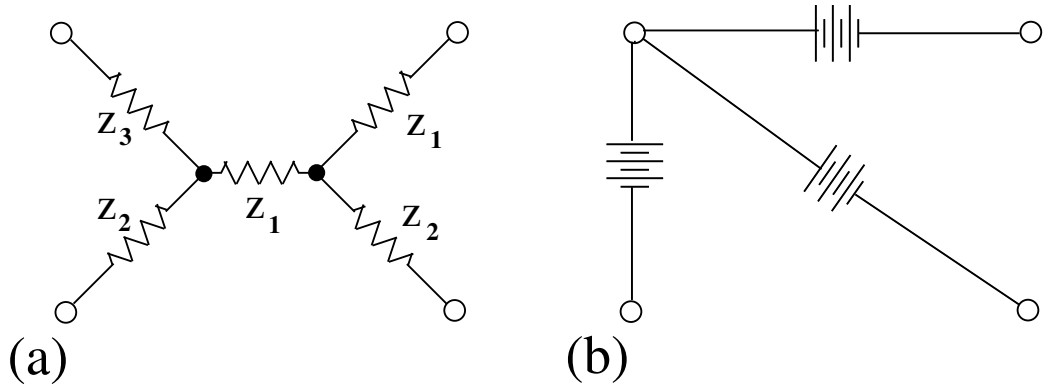


Figure 2.7: The resistor network is on one side (a) of the circuit board. On the other side (b) is a collection of batteries which power the network, and the relation between the potentials across each battery and the currents going through them define the Y_*-tensor.

where $\mathbf{K}$ is the inverse matrix satisfying

$$\varphi_{i\gamma} K_{m\gamma} = \delta_{im}, \tag{2.137}$$

assuming it exists. We rewrite (2.136) as

$$-I^0_m = D_{mn}\varphi^0_n, \quad \text{where } D_{mn} = -K_{m\gamma} L_{*\gamma\beta} I_{\beta n}, \tag{2.138}$$

which expresses the discrete Dirichlet-to-Neumann map $\mathbf{D}$ in terms of $\mathbf{L}_*$, where $I_{\beta n}$ is the value of $\mathbf{I}_\beta$ at terminal n. Given prescribed potentials φ^0_n at the terminals, this map gives the current $-I^0_m$ flowing into terminal m.

One can also complete the network by adding the part of the circuit associated with the batteries that power it. (Some of these "batteries" could in fact be resistors depending on whether they produce or absorb power). For instance, consider the resistor network in **Figure 2.7**(a). We drill holes in the circuit board and attach batteries to the other side of the circuit board. The relation between the currents through these batteries and the voltages across them, which is measured by the Y-tensor, provides an alternative description of the response of the network on the front side of the circuit board.

Now the appropriate finite-dimensional vector space $\mathcal{K}$ is the direct sum of $\mathcal{H}$ and the space $\mathcal{V}$ representing those directed bonds on the reverse side of the circuit board associated with the network of batteries. We now have a new incidence matrix $\widetilde{\mathbf{M}}$, associated with the full circuit. We define $\widetilde{\mathcal{E}}$ to consist of those fields $\widetilde{\mathbf{E}}$ such that $\widetilde{\mathbf{E}} = -\widetilde{\mathbf{M}}\boldsymbol{\varphi}$ for some potential $\boldsymbol{\varphi}$ defined on the nodes (that needs not be zero on the terminal nodes), and we define $\widetilde{\mathcal{J}}$ to consist of those fields $\widetilde{\mathbf{J}}$ such that $\widetilde{\mathbf{M}}^T\widetilde{\mathbf{J}} = 0$. Since the null space of $\widetilde{\mathbf{M}}^T$ is the orthogonal complement of the range of $\mathbf{M}$ the spaces $\widetilde{\mathcal{J}}$ and $\widetilde{\mathcal{E}}$ are orthogonal and span $\mathcal{K}$, so we have

$$\mathcal{K} = \widetilde{\mathcal{E}} \oplus \widetilde{\mathcal{J}} = \mathcal{V} \oplus \mathcal{H}. \tag{2.139}$$

To find the $\mathbf{Y}_*$ tensor one looks for fields $\widetilde{\mathbf{E}} \in \widetilde{\mathcal{E}}$ and $\widetilde{\mathbf{J}} \in \widetilde{\mathcal{J}}$ which have a decomposition

$$\widetilde{\mathbf{E}} = \mathbf{E}_1 + \mathbf{E}_2, \quad \widetilde{\mathbf{J}} = \mathbf{J}_1 + \mathbf{J}_2, \tag{2.140}$$

with

$$\mathbf{E}_1, \mathbf{J}_1 \in \mathcal{V}, \quad \mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}, \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2. \tag{2.141}$$

Then the relation between $\mathbf{J}_1$ and $\mathbf{E}_1$ determines $\mathbf{Y}_*$:

$$\mathbf{J}_1 = -\mathbf{Y}_*\mathbf{E}_1. \tag{2.142}$$

If no restrictions are placed on $\mathbf{Y}_*$ then a necessary constraint for $\mathbf{Y}_*$ to be uniquely determined is that

$$\mathcal{V} \cap \widetilde{\mathcal{J}} = \{\mathbf{0}\}, \tag{2.143}$$

since otherwise if there is a field $\mathbf{W}$ in this intersection (2.139) and (2.140) are still satisfied if we add $\mathbf{W}$ to both $\widetilde{\mathbf{J}}$ and $\mathbf{J}_1$. The restriction (2.143) says there are no closed loops in the network $\mathcal{V}$ on the reverse side of the circuit board, as in Figure 2.7(b)

3 A new perspective on boundary-value problems in mathematics, physics, and engineering

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

We show that the problem of determining the Dirichlet-to-Neumann map can be reformulated as the problem of determining an effective operator, associated with exactly the same sort of abstract problem for determining effective tensors in the theory of composites. In particular this implies that for a body containing n, possibly polycrystalline phases, the effective operator, and hence the Dirichlet-to-Neumann map is an analytic function of the elements of the component tensors $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ in the domain consisting of the union over θ of the region where all the tensors $e^{i\theta}\mathbf{L}_j$ have positive definite self-adjoint part.

3.1 Introduction

Generally the response of a body is determined by the "Dirichlet-to-Neumann" map which is a generic term: in electromagnetism it could measure current fluxes which result when one applies potentials to the body boundary, or it could measure the tangential components of the magnetic field that result when tangential components of the electric field are applied, or for linear elasticity it could measure the tractions (forces) at the boundary of the body when displacements are prescribed at the boundary. One can sort of think of the Dirichlet-to-Neumann map as a matrix: really it's a linear operator but one can approximate it by a matrix.

Here we show that the Dirichlet-to-Neumann map, when appropriately defined, is mathematically speaking the exact analog of an effective tensor in a composite material. This link between the Dirichlet-to-Neumann map and an effective tensor in the abstract theory of composites is the analog at the continuum level of the way effective tensors have been obtained from the response matrix of discrete electrical networks: see Sections 20.4, 20.5, and 20.6 in Milton (2002) and also Section 2.13 of the previous chapter. Roughly speaking, the Dirichlet-to-Neumann map gets replaced by a map (the effective operator in the abstract theory of composites) which acts on the space of fields that (modulo multiplication by an appropriate matrix) solve the equations when the body is filled with a homogeneous "reference medium": thus boundary conditions are removed from the problem.

When suitably formulated, the equations defining the Dirichlet-to-Neumann map are exactly those of effective tensors in the abstract theory of composites. Consequently, many of the tools that have been developed in the theory composite materials essentially carry over directly to Dirichlet-to-Neumann maps. In particular, variational principles for composites map over to variational principles for boundary-value problems, the theory of bounds on effective tensors carries over to an analogous theory of bounds on Dirichlet-to-Neumann maps, and the analyticity properties of effective tensors as functions of the component moduli map over to analyticity properties of the Dirichlet-to-Neumann map as functions of the component tensors within the body, assuming the body contains a multiphase mixture. Some results do not directly carry over, such as the general theory of exact relations (Grabovsky 1998; Grabovsky and Sage 1998; Grabovsky, Milton, and Sage 2000; see also **Chapter 17** of Milton 2002 and Grabovsky 2004) and Fast Fourier Transform methods for computing fields that solve the equations (Moulinec and Suquet 1994, 1998; Eyre and Milton 1999; Moulinec and Silva 2014; Willot, Abdallah, and Pellegrini (2014); Willot 2015). These rely heavily on the simple form of the operator $\mathbf{\Gamma}_1$ (the projection onto the subspace $\mathcal{E}$) in Fourier space that is dictated by the differential constraints on the fields.

This is not the first time that results from the theory of composites have been carried over to the response of bodies. Huet (1990) obtained elementary bounds on the response of bodies to special boundary conditions (such as affine boundary conditions) that were the analog of classic arithmetic average–harmonic average bounds of composites. [The same observation was made by Willis in a 1989 private communication to Nemat-Nasser and Hori (1993).] Milgrom (1990) found that exact relations satisfied by the effective moduli of composites in coupled field problems carried over to exact relation satisfied by the response of bodies (see also **Chapter 9** in this book). For ellipsoidal bodies Nemat-Nasser and Hori (1993) and Hori and Nemat–Nasser (1995,1998) obtained bounds that were the analog of the famous Hashin–Shtrikman bounds on the effective moduli of composites. Subsequently, Milton (2012) removed the restriction that the bodies had to be ellipsoidal. The Hashin–Shtrikman method was also used by Capdeboscq and Vogelius (2003, 2004) to asymptotically bound the volume of a dilute suspension of inclusions in a body. Variational minimization principles that had been primarily developed for bounding the quasistatic effective moduli of composites (Cherkaev and Gibiansky 1994), led to variational minimization principles for the full time-harmonic wave equations (of acoustics, elastodynamics, and electromagnetism) in lossy inhomogeneous bodies (Milton, Seppecher, and Bouchitté 2009, Milton and Willis 2010). The translation method for bounding the effective tensors of composites was extended to bodies, and used in an inverse way to bound the volume fraction of an inclusion in a two-phase body (Kang, Kim, and Milton 2012; Kang, Milton, and Wang 2014; Milton and Nguyen 2012; Kang, Milton, and Wang 2014; Kang and Milton 2013; Kang, Kim, Lee, Li, and Milton 2014). It allowed for more general boundary conditions, and led to the related splitting method for bounding the volume fraction of an inclusion in a body (Milton and Nguyen 2012; Thaler and Milton 2015).

What sets this work apart is establishing a direct mathematical isomorphism between the theory of effective tensors of composites and the theory of the Dirichlet-to-Neumann map for bodies. The chapter assumes the reader is familiar with the contents of **Chapters 1** and **2**.

3.2 General theory

The general problem we address in this chapter is the response of an inhomogeneous body Ω to fields, which may be waves or static fields. The starting point is the section "General Theory" in the paper Milton, Seppecher, and Bouchitté (2009) which provides a framework for treating the time-harmonic equations of acoustics and elastodynamics on one footing. Also the framework encompasses the time-harmonic

Schrödinger equation since for time-harmonic fields it is directly analogous to the acoustics equation.

Let's begin by assuming there are no sources in the body (i.e., $\mathbf{h} = 0$ in the terminology of the paper of Milton, Seppecher, and Bouchitté (2009)). First let us define the operators $\sqcap$ and $\sqcup$: given a vector (or scalar) "potential" $\mathbf{u}$ we define

$$\sqcap \mathbf{u} \equiv \begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix} \tag{3.1}$$

and given a field

$$\mathcal{G} \equiv \begin{pmatrix} \mathbf{G} \\ \mathbf{g} \end{pmatrix} \tag{3.2}$$

where the "current" $\mathbf{G}$ is a second order tensor field and $\mathbf{g}$ is a vector field (or $\mathbf{G}$ is a vector field and $\mathbf{g}$ is a scalar field when $\mathbf{u}$ is scalar) we define

$$\sqcup \mathcal{G} \equiv -\nabla \cdot \mathbf{G} + \mathbf{g}. \tag{3.3}$$

One can think of these operators $\sqcap$ and $\sqcup$ as being a little analogous to the gradient ∇, and divergence $\nabla\cdot$ that one is familiar with in electrical conduction.

Now inside a body Ω we consider fields $\mathcal{F}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ that satisfy the constitutive relation

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\mathcal{F}(\mathbf{x}), \tag{3.4}$$

and are subject to the differential constraints that

$$\sqcup \mathcal{G} = 0, \quad \mathcal{F} \equiv \begin{pmatrix} \mathbf{F} \\ \mathbf{f} \end{pmatrix} = \sqcap \mathbf{u}, \tag{3.5}$$

where $\mathbf{F}$ is a second order tensor field and $\mathbf{f}$ is a vector field (or $\mathbf{F}$ is a vector field and $\mathbf{f}$ is a scalar field when $\mathbf{u}$ is scalar). The tensor field $\mathbf{Z}(\mathbf{x})$ could be complex-valued, but is typically such that the self adjoint part of $\mathbf{Z}(\mathbf{x})/i$ is positive semidefinite, or more generally such that the self adjoint part of $e^{i\theta}\mathbf{Z}(\mathbf{x})$ is positive semidefinite for at least one real value of θ.

Equivalently, these differential constraints imply that $\mathcal{F}$ and $\mathcal{G}$ take the form

$$\mathcal{F} = \begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix}, \quad \mathcal{G} = \begin{pmatrix} \mathbf{G} \\ \nabla \cdot \mathbf{G} \end{pmatrix}. \tag{3.6}$$

The two simplest examples are those given in Section 1.7: time-harmonic acoustics,

$$\underbrace{\begin{pmatrix} -i\mathbf{v} \\ -i\nabla \cdot \mathbf{v} \end{pmatrix}}_{\mathcal{G}(\mathbf{x})} = \underbrace{\begin{pmatrix} -(\omega \boldsymbol{\rho})^{-1} & 0 \\ 0 & \omega/\kappa \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \underbrace{\begin{pmatrix} \nabla P \\ P \end{pmatrix}}_{\mathcal{F}(\mathbf{x})}, \tag{3.7}$$

where $P(\mathbf{x})$ is the pressure, and $\mathbf{v}(\mathbf{x})$ the velocity; and time-harmonic elastodynamics

$$\underbrace{\begin{pmatrix} -\boldsymbol{\sigma}/\omega \\ i\mathbf{p} \end{pmatrix}}_{\mathcal{G}(\mathbf{x})} = \underbrace{\begin{pmatrix} -\boldsymbol{\mathcal{C}}/\omega & 0 \\ 0 & \omega \boldsymbol{\rho} \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \underbrace{\begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix}}_{\mathcal{F}(\mathbf{x})}, \tag{3.8}$$

where $\mathbf{u}(\mathbf{x})$ is the displacement, $\boldsymbol{\sigma}(\mathbf{x})$ is the stress, and $\mathbf{p}(\mathbf{x}) = i\nabla \cdot \boldsymbol{\sigma}(\mathbf{x})/\omega$ is the momentum.

In the case where $\mathbf{u}$ is a vector field we have the key identity

$$\int_\Omega \mathcal{G}\cdot\mathcal{F} = \int_\Omega \mathbf{G} : \nabla\mathbf{u} + (\nabla\cdot\mathbf{G})\cdot\mathbf{u} = \int_\Omega \nabla\cdot(\mathbf{G}\cdot\mathbf{u}) = \int_{\partial\Omega} \mathbf{n}\cdot\mathbf{G}\mathbf{u}, \tag{3.9}$$

where $\mathbf{n}$ is normal to the boundary $\partial\Omega$. [In fact, as pointed out in Section 1.7, there are a multitude of key identities parameterized by the second order tensor $\mathbf{M}$ entering (1.97), but we will ignore these additional key identities in this chapter.]

In the case where $\mathbf{u}$ is a scalar P and $\mathbf{G}$ is a vector, as for acoustics, the key identity becomes

$$\int_\Omega \mathcal{G}\cdot\mathcal{F} = \int_\Omega \mathbf{G}\cdot\nabla P + (\nabla\cdot\mathbf{G})\cdot\mathbf{u} = \int_\Omega \nabla\cdot(P\mathbf{G}) = \int_{\partial\Omega} P\mathbf{n}\cdot\mathbf{G}. \tag{3.10}$$

Choose a *constant* comparison medium or "reference medium" $\mathbf{Z}(\mathbf{x}) = \mathbf{Z}_0$, where $\mathbf{Z}_0$ is a positive definite Hermitian (or real) tensor which is independent of $\mathbf{x}$. Consider, as the boundary fields range over all possible data, the set of all solutions to the equation

$$\mathcal{G}_0(\mathbf{x}) = \mathbf{Z}_0\mathcal{F}_0(\mathbf{x}), \tag{3.11}$$

subject to the differential constraints that

$$\sqcup\mathcal{G}_0(\mathbf{x}) = 0, \quad \mathcal{F}_0(\mathbf{x}) = \sqcap\mathbf{u}_0 \tag{3.12}$$

for some potential $\mathbf{u}_0$.

Suppose some real boundary potential $\mathbf{u}_0(\mathbf{x}) = \mathbf{v}(\mathbf{x})$ is prescribed for $\mathbf{x}\in\partial\Omega$. Consider the variational formula

$$\min_{\substack{\mathbf{u}\\ \mathbf{u}=\mathbf{v}\text{ on }\partial\Omega}} \int_\Omega (\sqcap\mathbf{u})\cdot\mathbf{Z}_0(\sqcap\mathbf{u}). \tag{3.13}$$

Provided $\mathbf{v}$ is regular enough (in the space $H^{1/2}$) the minimum exists, is unique and at the minimum $\mathbf{u}$ takes a value $\mathbf{u}_0$ such that for all variations $\delta\mathbf{u}$ vanishing on $\partial\Omega$ we have

$$\int_\Omega (\nabla\delta\mathbf{u})\cdot\mathbf{G}_0 + \delta\mathbf{u}\cdot\mathbf{g}_0 = 0, \quad \text{where } \begin{pmatrix}\mathbf{G}_0\\ \mathbf{g}_0\end{pmatrix} = \mathbf{Z}_0\cdot(\sqcap\mathbf{u}_0). \tag{3.14}$$

By integrating by parts the first term in the integral we see that a necessary condition for this to hold is that $\mathbf{g}_0 = \nabla\cdot\mathbf{G}_0$, i.e., that the Euler–Lagrange equation $\sqcup\mathbf{Z}_0\sqcap\mathbf{u}_0 = 0$ is satisfied. So for any real regular boundary potential $\mathbf{u} = \mathbf{v}$ there exists a unique field $\mathcal{F}_0(\mathbf{x}) = \sqcap\mathbf{u}_0$ which corresponds to it.

Given a complex boundary potential $\mathbf{v} = \mathbf{v}' + i\mathbf{v}''$ then we find the fields $\mathcal{F}_0{}'(\mathbf{x})$ and $\mathcal{F}_0{}''(\mathbf{x})$ associated with the real and imaginary parts of the potential $\mathbf{v}'$ and $\mathbf{v}''$, to generate the field $\mathcal{F}_0(\mathbf{x}) = \mathcal{F}_0{}'(\mathbf{x}) + i\mathcal{F}_0{}''(\mathbf{x}) = \sqcap\mathbf{u}_0$ associated with the complex potential $\mathbf{u}_0 = \mathbf{u}_0{}' + i\mathbf{u}_0{}''$ that has the boundary-value $\mathbf{v} = \mathbf{v}' + i\mathbf{v}''$.

Similarly if for given real values of the flux $\mathbf{t}(\mathbf{x}) = \mathbf{n}\cdot\mathbf{G}(\mathbf{x})$ we consider

$$\min_{\substack{\mathbf{G}(\mathbf{x})\\ \mathbf{n}\cdot\mathbf{G}(\mathbf{x})=\mathbf{t}(\mathbf{x})\text{ on }\partial\Omega}} \int_\Omega \begin{pmatrix}\mathbf{G}\\ \nabla\cdot\mathbf{G}\end{pmatrix}\cdot\mathbf{Z}_0^{-1}\begin{pmatrix}\mathbf{G}\\ \nabla\cdot\mathbf{G}\end{pmatrix}, \tag{3.15}$$

then if $\mathbf{t}$ is regular enough (in the space $H^{-1/2}$), the minimum exists and is unique. If we let

$$\mathbf{Z}_0^{-1} = \begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^T & \mathbf{C}\end{pmatrix}, \tag{3.16}$$

then at the minimum $\mathbf{G}(\mathbf{x})$ takes a value $\mathbf{G}_0(\mathbf{x})$ such that

$$\int_\Omega \delta\mathbf{G}[\mathbf{A}\mathbf{G}_0 + \mathbf{B}\nabla\cdot\mathbf{G}_0] + \nabla\cdot\delta\mathbf{G}[\mathbf{B}^T\mathbf{G}_0 + \mathbf{C}\nabla\cdot\mathbf{G}_0] = 0, \tag{3.17}$$

for all variations $\delta\mathbf{G}$ meeting the boundary condition that $\mathbf{n}\cdot\delta\mathbf{G} = 0$ on $\partial\Omega$. By integrating by parts this last term we see that $\mathbf{G}_0(\mathbf{x})$ must satisfy the Euler–Lagrange equation

$$[\mathbf{A}\mathbf{G}_0 + \mathbf{B}\nabla\cdot\mathbf{G}_0] = \nabla\mathbf{u}_0, \quad \text{where } \mathbf{u}_0 = \mathbf{B}^T\mathbf{G}_0 + \mathbf{C}\nabla\cdot\mathbf{G}_0, \tag{3.18}$$

implying

$$\mathbf{Z}_0^{-1}\begin{pmatrix}\mathbf{G}_0\\ \nabla\cdot\mathbf{G}_0\end{pmatrix} = \begin{pmatrix}\nabla\mathbf{u}_0\\ \mathbf{u}_0\end{pmatrix}. \tag{3.19}$$

So for any real regular boundary condition $\mathbf{n}\cdot\mathbf{G}(\mathbf{x}) = \mathbf{t}(\mathbf{x})$ there exists a unique field $\mathcal{G}_0(\mathbf{x})$ which corresponds to it. Given a complex-valued boundary flux $\mathbf{t} = \mathbf{t}' + i\mathbf{t}''$ then we find the fields $\mathcal{G}_0{}'(\mathbf{x})$ and $\mathcal{G}_0{}''(\mathbf{x})$ associated with the real and imaginary parts of the potential $\mathbf{t}'$ and $\mathbf{t}''$, to generate the field $\mathcal{G}_0(\mathbf{x}) = \mathcal{G}_0{}'(\mathbf{x}) + i\mathcal{G}_0{}''(\mathbf{x})$ associated with the complex flux $\mathbf{t} = \mathbf{t}' + i\mathbf{t}''$.

Thus, for either prescribed potentials or prescribed fluxes, we have

$$\mathbf{Z}_0^{-1/2}\mathcal{G}_0(\mathbf{x}) = \mathbf{Z}_0^{1/2}\mathcal{F}_0(\mathbf{x}) \equiv \mathbf{U}_0(\mathbf{x}), \tag{3.20}$$

and we let $\mathcal{U}$ be the space of all these fields $\mathbf{U}_0(\mathbf{x})$ as the boundary data varies (over real and complex values). Similar in some respects to the trick of introducing $\mathbf{Z}_0^{1/2}$ are the reference transformations discussed in Section 9.7 of Milton (2002). Reference transformations multiply the fields on the left side of the constitutive law by a constant tensor $\mathbf{L}_0^{-1/2}$ and the fields on the right hand side of the constitutive law by $\mathbf{L}_0^{+1/2}$, thus preserving the key identity and the orthogonality of the subspaces when the action of $\mathbf{L}_0^{+1/2}$ or $\mathbf{L}_0^{-1/2}$ on $\mathcal{U}$ leaves it invariant: under a reference transformation in the abstract theory of composites the three mutually orthogonal subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ get replaced by the three mutually orthogonal subspaces $\mathcal{U}$, $\mathbf{L}_0^{+1/2}\mathcal{E}$ and $\mathbf{L}_0^{-1/2}\mathcal{J}$.

Now let $\mathcal{E}$ be the space of all fields $\mathbf{E}(\mathbf{x})$ that are square integrable in Ω (don't confuse them with electric fields) such that

$$\mathbf{E}(\mathbf{x}) = \mathbf{Z}_0^{1/2}\sqcap\mathbf{u}, \tag{3.21}$$

for some potential $\mathbf{u}$ with $\mathbf{u} = 0$ on $\partial\Omega$. This last boundary condition ensures that $\mathbf{E}$ is perpendicular to every field $\mathbf{U}_0$ in $\mathcal{U}$, where given square integrable fields $\mathbf{P}_1(\mathbf{x})$ and $\mathbf{P}_2(\mathbf{x})$ we take the usual inner product

$$(\mathbf{P}_1, \mathbf{P}_2) = \int_\Omega \overline{\mathbf{P}}_1\cdot\mathbf{P}_2. \tag{3.22}$$

The orthogonality condition can then be written as

$$0 = (\mathbf{E}, \mathbf{U}_0) = \int_\Omega \mathbf{E}(\mathbf{x})\cdot\mathbf{U}_0(\mathbf{x}) = \int_\Omega [\sqcap\mathbf{u}]\cdot\mathcal{G}_0(\mathbf{x}) = \int_{\partial\Omega}(\mathbf{G}_0\cdot\mathbf{n})\cdot\mathbf{u}, \tag{3.23}$$

which vanishes because $\mathbf{u} = 0$ on $\partial\Omega$.

We let $\mathcal{J}$ be the space of all fields $\mathbf{J}(\mathbf{x})$ that are square integrable in Ω (don't confuse them with current fields) such that for some $\mathbf{G}(\mathbf{x})$, with $\mathbf{n}\cdot\mathbf{G}(\mathbf{x}) = 0$ on $\partial\Omega$,

$$\mathbf{J}(\mathbf{x}) = \mathbf{Z}_0^{-1/2}\begin{pmatrix}\mathbf{G}\\ \nabla\cdot\mathbf{G}\end{pmatrix}. \tag{3.24}$$

The flux condition $\mathbf{n} \cdot \mathbf{G}(\mathbf{x}) = 0$ on $\partial\Omega$ ensures that $\mathbf{J}(\mathbf{x})$ is perpendicular to every field $\mathbf{U}_0$ in $\mathcal{U}$:

$$0 = (\mathbf{J}, \mathbf{U}_0) = \int_\Omega \mathbf{J}(\mathbf{x}) \cdot \mathbf{U}_0(\mathbf{x}) = \int_\Omega \begin{pmatrix} \mathbf{G} \\ \nabla \cdot \mathbf{G} \end{pmatrix} \cdot \mathcal{F}_0(\mathbf{x}) = \int_{\partial\Omega} (\mathbf{G} \cdot \mathbf{n}) \cdot \mathbf{U}_0, \tag{3.25}$$

which vanishes because $\mathbf{G} \cdot \mathbf{n} = 0$ on $\partial\Omega$.

Finally we show that $\mathcal{E}$ and $\mathcal{J}$ are orthogonal: the inner product

$$\int_\Omega \mathbf{E}(\mathbf{x}) \cdot \mathbf{J}(\mathbf{x}) = \int_\Omega \begin{pmatrix} \mathbf{G} \\ \nabla \cdot \mathbf{G} \end{pmatrix} \cdot \sqcap \mathbf{u} = \int_{\partial\Omega} (\mathbf{G} \cdot \mathbf{n}) \cdot \mathbf{u} = 0, \tag{3.26}$$

vanishes because we have shown for fields in $\mathcal{E}$ the potential $\mathbf{u}$ is zero at the boundary, and for fields in $\mathcal{J}$ the flux $\mathbf{G} \cdot \mathbf{n}$ is zero.

Now we can consider equations

$$\mathbf{J}_0(\mathbf{x}) + \mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})(\mathbf{E}_0(\mathbf{x}) + \mathbf{E}(\mathbf{x})), \tag{3.27}$$

with $\mathbf{J}_0(\mathbf{x}),\ \mathbf{E}_0(\mathbf{x}) \in \mathcal{U}$, $\mathbf{J}(\mathbf{x}) \in \mathcal{J}$, $\mathbf{E}(\mathbf{x}) \in \mathcal{E}$, and

$$\mathbf{L}(\mathbf{x}) = \mathbf{Z}_0^{-1/2} \mathbf{Z}(\mathbf{x}) \mathbf{Z}_0^{-1/2}. \tag{3.28}$$

This is exactly the abstract formulation associated with the theory of composite materials (see Section 2.2 of the review **Chapter 2**, or Section 12.7 of Milton 2002). For example, in the concrete setting of the conductivity equations in a periodic medium, $\mathcal{U}$ consists of the space of constant fields, $\mathcal{E}$ consists of electric fields that are gradients of periodic potentials, $\mathbf{E}(\mathbf{x}) = -\nabla V(\mathbf{x})$, and $\mathcal{J}$ consists of divergence free fields $\mathbf{J}(\mathbf{x})$ that have zero average value (i.e., $\nabla \cdot \mathbf{J} = 0$ and $\langle \mathbf{J} \rangle = 0$ where the angular brackets denote a volume average).

Therefore many results about composites carry over directly to this new setting. In particular, in Section 3.5, we will see that the analyticity properties of the effective tensor $\mathbf{L}_*$ as a function of the component moduli carry over. One distinction between solving (3.27) in a body, rather than in a composite, is that in a composite the operators $\mathbf{\Gamma}_0$, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ defined as the projections onto the spaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are local operators in Fourier space, which is *not* the case when we are considering a body, although being projections they still have norm 1.

The inverse problem we discuss in **Chapter 5** is recovering $\mathbf{L}(\mathbf{x})$ from knowledge of the effective operator $\mathbf{L}_*$ using tools from the abstract theory of composites (such as bounds obtained from variational principles and the translation method, and using analyticity and integral representations of the relevant analytic functions). As shown in the next section, this effective operator $\mathbf{L}_*$ can be obtained from the Dirichlet-to-Neumann map. Once one has information about $\mathbf{L}(\mathbf{x})$ one can of course transfer it to information about $\mathbf{Z}(\mathbf{x})$, using (3.28). This, of course, is the real goal in imaging: obtaining information about $\mathbf{Z}(\mathbf{x})$ from information (which may be complete or partial) about the Dirichlet-to-Neumann map, or equivalently about $\mathbf{L}_*$.

3.3 Relating the effective operator to the Dirichlet-to-Neumann map

Given a potential $\mathbf{u}_0(\mathbf{x})$ at the boundary $\partial\Omega$ one may solve the minimization problem (3.13) to find the associated field $\mathbf{E}_0(\mathbf{x})$. Then one can solve the equations (3.27) for the remaining fields $\mathbf{E}(\mathbf{x})$, $\mathbf{J}_0(\mathbf{x})$ and $\mathbf{J}(\mathbf{x})$. This is exactly equivalent to solving the equations (3.4) subject to the differential constraints (3.6) with $\mathbf{u}_0(\mathbf{x})$ prescribed at the boundary $\partial\Omega$. Since $\mathbf{J}_0$ depends linearly on $\mathbf{E}_0$ we may write

$$\mathbf{J}_0 = \mathbf{L}_* \mathbf{E}_0, \tag{3.29}$$

which defines the effective operator $\mathbf{L}_*$ which is a linear map from $\mathcal{U}$ to $\mathcal{U}$. It is to be stressed that $\mathbf{L}_*$ is an operator and that (3.29) *does not* imply $\mathbf{J}_0(\mathbf{x}) = \mathbf{L}_*\mathbf{E}_0(\mathbf{x})$ for some matrix $\mathbf{L}_*$.

To see the meaning of this take an orthonormal basis $\mathbf{U}_\beta$ of $\mathcal{U}$ indexed by β. For simplicity we will assume that our basis is a countable set so that β ranges over a discrete set $\aleph$. For example, in two-dimensions if Ω was the unit disk and $\mathbf{u}$ was an m-component vector then we could take $\aleph$ as the set of m-tuples $\beta = (k_1, k_2, \ldots, k_m)$, where the k_i are integers, and the field $\mathbf{U}_\beta = \mathbf{Z}_0^{1/2} \sqcap \mathbf{u}_\beta$ could be that associated with solving the equation $\sqcup\mathbf{Z}_0 \sqcap \mathbf{u}_\beta = 0$ with boundary-values

$$\begin{aligned} \{\mathbf{u}_\beta\}_j(\mathbf{x}) &= \cos(|k_j|\phi) \quad \text{if } k_j \geq 0 \\ &= \sin(|k_j|\phi) \quad \text{if } k_j < 0, \end{aligned} \tag{3.30}$$

at points $\mathbf{x} = (\cos\phi, \sin\phi)$ on the boundary $\partial\Omega$, where $\{\mathbf{u}_\beta\}_j(\mathbf{x})$ denotes the j-th component of the vector $\mathbf{u}_\beta(\mathbf{x})$, $j = 1, 2, \ldots, m$. If we were in three dimensions and Ω was a sphere, we could prescribe the boundary-value of the component $\{\mathbf{u}_\beta\}_j(\mathbf{x})$ to be a spherical harmonic. If we had some other smoothly shaped domain we could choose boundary-values that were the image of spherical harmonics under a map that takes the boundary of the sphere to the boundary of this smoothly shaped domain.

We now resolve our "applied field" $\mathbf{E}_0 = \mathbf{Z}_0^{1/2} \sqcap \mathbf{u}_0$ into its components

$$E_{0\beta} = \int_\Omega \mathbf{U}_\beta \cdot \mathbf{E}_0 = \int_\Omega \mathcal{G}_\beta(\mathbf{x}) \cdot \sqcap\mathbf{u}_0 = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{G}_\beta \mathbf{u}_0, \tag{3.31}$$

where the boundary-value of $\mathbf{u}_0$ is the potential which we impose at $\partial\Omega$. We assume the field $\mathbf{U}_\beta$ has been calculated (this probably has to be done numerically, rather than analytically, if Ω is not a circular disk or sphere) and thus the boundary-value of $\mathbf{n} \cdot \mathbf{G}_\beta$ and the coefficients $E_{0\beta}$ can be obtained. We similarly resolve our "response field" $\mathbf{J}_0 = \mathbf{Z}_0^{-1/2}\mathcal{G}_0$ into its components as well:

$$J_{0\gamma} = \int_\Omega \mathbf{U}_\gamma \cdot \mathbf{J}_0 = \int_\Omega (\sqcap\mathbf{u}_\gamma) \cdot \mathcal{G}_0(\mathbf{x}) = \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{G}_0 \mathbf{u}_\gamma, \tag{3.32}$$

where $\mathcal{G}_0$ is associated with $\mathbf{J}_0$, i.e., $\mathcal{G}_0 = \mathbf{Z}^{-1/2}\mathbf{J}_0$, and $\mathbf{n} \cdot \mathbf{G}_0$ is the "surface flux" associated with it. Again the coefficients can be obtained from the prescribed value of $\mathbf{u}_\gamma$ on $\partial\Omega$ (see, for example, (3.30)) and the measured value of $\mathbf{n} \cdot \mathbf{G}_0$.

In this basis the relation (3.29) takes the form

$$J_{0\gamma} = \sum_{\beta\in\aleph} L_{*\gamma\beta} E_{0\beta}. \tag{3.33}$$

where the $L_{*\gamma\beta}$ are the elements of the matrix representing the operator $\mathbf{L}_* : \mathcal{U} \to \mathcal{U}$ in this basis. Since the basis of $\mathcal{U}$ contains infinitely many fields one would in practice want to truncate this matrix: for example, by ignoring those basis fields that are generated from potentials that oscillate rapidly around the boundary, such as the high order Fourier modes in (3.30) with large k_j for some j. Note that from the definition of the effective tensor, there are, according to (3.27), fields $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ such that

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}_0^{1/2}[\mathbf{J}_0(\mathbf{x}) + \mathbf{J}(\mathbf{x})], \quad \mathcal{F}(\mathbf{x}) = \mathbf{Z}_0^{-1/2}[\mathbf{E}_0(\mathbf{x}) + \mathbf{E}(\mathbf{x})] \tag{3.34}$$

satisfy

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\mathcal{F}(\mathbf{x}), \quad \sqcup\mathcal{G} = 0, \quad \mathcal{F} = \sqcap\mathbf{u}, \tag{3.35}$$

where on the boundary $\partial\Omega$, $\mathbf{n}\cdot\mathbf{G}(\mathbf{x}) = \mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$ and $\mathbf{u}(\mathbf{x}) = \mathbf{u}_0(\mathbf{x})$. These boundary fields $\mathbf{u}_0(\mathbf{x})$ and $\mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$ are thus related via the Dirichlet-to-Neumann map.

Now (3.31), (3.32) and (3.33) imply

$$\int_{\partial\Omega} \mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})\mathbf{u}_\gamma(\mathbf{x})\,dS = \sum_{\beta\in\aleph} L_{*\gamma\beta}\int_{\partial\Omega}\mathbf{n}\cdot\mathbf{G}_\beta(\mathbf{x})\mathbf{u}_0(\mathbf{x})dS. \tag{3.36}$$

Knowing the left-hand side for all $\gamma\in\aleph$ gives us the components of $\mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$ in the basis of surface fields $\mathbf{u}_\gamma(\mathbf{x})$, from which we can recover $\mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$. Thus, from the effective tensor $\mathbf{L}_*$ we obtain the Dirichlet-to-Neumann map Λ mapping the potential $\mathbf{u}_0$ on $\partial\Omega$ to the flux $\mathbf{t}_0 = \mathbf{n}\cdot\mathbf{G}_0$ on $\partial\Omega$. In the example where Ω was the unit disk, the choice of basis fields (3.30) gives us the Fourier components of $\mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$, which, by taking the Fourier transform, allows us to recover $\mathbf{n}\cdot\mathbf{G}_0(\mathbf{x})$. The effective tensor $\mathbf{L}_*$ gives us an alternate representation of the Dirichlet-to-Neumann map. Of course we need to know the fields in the space $\mathcal{U}$, and in general this requires some numerical computation.

Alternatively, we could prescribe the flux $\mathbf{t}(\mathbf{x}) = \mathbf{n}\cdot\mathbf{G}(\mathbf{x})$ at the boundary $\partial\Omega$ and solve the minimization problem (3.15) for the associated field $\mathbf{J}_0(\mathbf{x})$. Then we could solve (3.27) for the remaining fields $\mathbf{J}(\mathbf{x})$, $\mathbf{E}_0(\mathbf{x})$, and $\mathbf{E}(\mathbf{x})$. Since the relation between $\mathbf{E}_0$ and $\mathbf{J}_0$ is linear we can write

$$\mathbf{E}_0 = \mathbf{L}_*^{-1}\mathbf{J}_0, \tag{3.37}$$

which defines the inverse effective operator $\mathbf{L}_*^{-1}$ which is a map from $\mathcal{U}$ to $\mathcal{U}$, and which can be connected with the inverse of the Dirichlet-to-Neumann map.

A remark is that everything extends more or less directly to coupled field problems where the constitutive relation looks like

$$\begin{pmatrix} \mathcal{G}^{(1)} \\ \mathcal{G}^{(2)} \\ \vdots \\ \mathcal{G}^{(n)} \\ \widetilde{\mathcal{F}}^{(1)} \\ \widetilde{\mathcal{F}}^{(2)} \\ \vdots \\ \widetilde{\mathcal{F}}^{(n)} \end{pmatrix} = \mathbf{M}(\mathbf{x}) \begin{pmatrix} \mathcal{F}^{(1)} \\ \mathcal{F}^{(2)} \\ \vdots \\ \mathcal{F}^{(n)} \\ \widetilde{\mathcal{G}}^{(1)} \\ \widetilde{\mathcal{G}}^{(2)} \\ \vdots \\ \widetilde{\mathcal{G}}^{(n)} \end{pmatrix}, \tag{3.38}$$

where

$$\mathcal{G}^{(i)} = \begin{pmatrix} \mathbf{G}^{(i)} \\ \nabla\cdot\mathbf{G}^{(i)} \end{pmatrix}, \quad \mathcal{F}^{(i)} = \sqcap\mathbf{u}^{(i)}, \quad \widetilde{\mathcal{F}}^{(i)} = \sqcap\widetilde{\mathbf{u}}^{(i)}, \quad \widetilde{\mathcal{G}}^{(i)} = \begin{pmatrix} \widetilde{\mathbf{G}}^{(i)} \\ \nabla\cdot\widetilde{\mathbf{G}}^{(i)} \end{pmatrix}. \tag{3.39}$$

3.4 Quadratic forms

The connection between the Dirichlet-to-Neumann map and the effective operator can also be made through the quadratic forms associated with them. Given two possibly complex-valued vector fields $\mathbf{p}_1$ and $\mathbf{p}_2$ (at least one of which is infinitely differentiable, in C^∞) defined on the surface $\partial\Omega$ let us define their "$\circ$" product to be

$$\mathbf{p}_1\circ\mathbf{p}_2 = \int_{\partial\Omega}\mathbf{p}_1(\mathbf{x})\cdot\mathbf{p}_2(\mathbf{x})\,dS. \tag{3.40}$$

Note that there is no complex conjugation here. Assuming $\mathbf{Z}(\mathbf{x})$ is a symmetric (possibly complex-valued) matrix, the Dirichlet-to-Neumann map Λ is a symmetric operator in the sense that

$$\mathbf{u}_1 \circ (\Lambda \mathbf{u}_2) = (\Lambda \mathbf{u}_1) \circ \mathbf{u}_2 \quad \text{for all } \mathbf{u}_1,\ \mathbf{u}_2 \in C^\infty. \tag{3.41}$$

To see this, let $\mathbf{u}_1(\mathbf{x})$, $\mathbf{G}_1(\mathbf{x})$, $\mathcal{F}_1(\mathbf{x})$, $\mathcal{G}_1(\mathbf{x})$ and $\mathbf{u}_2(\mathbf{x})$, $\mathbf{G}_2(\mathbf{x})$, $\mathcal{F}_2(\mathbf{x})$, $\mathcal{G}_2(\mathbf{x})$ be the fields inside the body, respectively, associated with the given boundary potentials $\mathbf{u}_1(\mathbf{x})$ and $\mathbf{u}_2(\mathbf{x})$. Then, using the key identity and the symmetry of $\mathbf{Z}(\mathbf{x})$ [that $\mathbf{Z}^T(\mathbf{x}) = \mathbf{Z}(\mathbf{x})$], we have

$$\begin{aligned} \mathbf{u}_1 \circ (\Lambda \mathbf{u}_2) &= \int_{\partial\Omega} \mathbf{u}_1 \cdot (\mathbf{n} \cdot \mathbf{G}_2) \\ &= \int_\Omega \mathcal{F}_1 \cdot \mathcal{G}_2 = \int_\Omega \mathcal{F}_1 \cdot [\mathbf{Z}\mathcal{F}_2] = \int_\Omega [\mathbf{Z}\mathcal{F}_1] \cdot \mathcal{F}_2 = \int_\Omega \mathcal{G}_1 \cdot \mathcal{F}_2 \\ &= \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{G}_1 \mathbf{u}_2 = (\Lambda \mathbf{u}_1) \circ \mathbf{u}_2, \end{aligned} \tag{3.42}$$

which establishes (3.42). Thus the form in (3.41) is a symmetric bilinear form of $\mathbf{u}_1$ and $\mathbf{u}_2$, and by the "polarization identity"

$$4\mathbf{u}_1 \circ (\Lambda \mathbf{u}_2) = (\mathbf{u}_1 + \mathbf{u}_2) \circ [\Lambda(\mathbf{u}_1 + \mathbf{u}_2)] - (\mathbf{u}_1 - \mathbf{u}_2) \circ [\Lambda(\mathbf{u}_1 - \mathbf{u}_2)], \tag{3.43}$$

we see that we can recover Λ from knowledge of the quadratic form

$$f(\mathbf{u}_0) = \mathbf{u}_0 \circ (\Lambda \mathbf{u}_0) = \int_\Omega \mathcal{F} \cdot \mathcal{G}, \tag{3.44}$$

where $\mathcal{F}(\mathbf{x})$ and $\mathcal{G}(\mathbf{x})$ are the fields inside the body, satisfying $\mathcal{F} = \sqcap \mathbf{u}$, $\sqcup \mathcal{G} = 0$ where $\mathbf{u}(\mathbf{x}) = \mathbf{u}_0$ on $\partial\Omega$. Making the substitution (3.34) and using the orthogonality of the spaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ we see that $f(\mathbf{u})$ can be determined from $\mathbf{L}_*$ and from the field $\mathbf{E}_0$ associated with the boundary-value of $\mathbf{u}$:

$$f(\mathbf{u}_0) = \int_\Omega \mathbf{E}_0(\mathbf{x}) \cdot \mathbf{J}_0(\mathbf{x}), \tag{3.45}$$

in which $\mathbf{J}_0 = \mathbf{L}_* \mathbf{E}_0$. [Again, to emphasize the point, we cannot write the integrand as $\mathbf{E}_0(\mathbf{x}) \cdot \mathbf{L}_* \mathbf{J}_0(\mathbf{x})$.]

By following steps similar to that in (3.42) we see that

$$\begin{aligned} \overline{\mathbf{u}_0} \circ (\Lambda \mathbf{u}_0) &= \int_\Omega \overline{\mathcal{F}} \cdot [\mathbf{Z}\mathcal{F}] \\ &= \int_\Omega (\mathcal{F}' - i\mathcal{F}'') \cdot [\mathbf{Z}(\mathcal{F}' + i\mathcal{F}'')] \\ &= \int_\Omega \mathcal{F}' \cdot [\mathbf{Z}\mathcal{F}'] + \mathcal{F}'' \cdot [\mathbf{Z}\mathcal{F}''], \end{aligned} \tag{3.46}$$

where we have used the symmetry of $\mathbf{Z}$ to cancel the cross terms, and $\mathcal{F}'$ and $\mathcal{F}''$ are the real and imaginary parts of $\mathbf{F}$. Taking imaginary parts of both sides gives

$$\mathrm{Im}[\overline{\mathbf{u}_0} \circ (\Lambda \mathbf{u}_0)] = \int_\Omega \mathcal{F}' \cdot [\mathrm{Im}(\mathbf{Z})\mathcal{F}'] + \mathcal{F}'' \cdot [\mathrm{Im}(\mathbf{Z})\mathcal{F}'']. \tag{3.47}$$

So the quadratic form associated with the Dirichlet-to-Neumann map Λ has a positive semi-definite imaginary part when the imaginary part of $\mathbf{Z}(\mathbf{x})$ is positive semi-definite for all $\mathbf{x}$.

3.5 Analyticity properties of effective tensors

Suppose the body has n-phases, each having varying orientations, so that the tensor field $\mathbf{L}$ takes the form

$$\mathbf{L}(\mathbf{x}) = \mathbf{Q}(\mathbf{R}(\mathbf{x}))\left[\sum_{i=1}^{n}\chi_i(\mathbf{x})\mathbf{L}_i\right][\mathbf{Q}(\mathbf{R}(\mathbf{x}))]^T, \tag{3.48}$$

in which $\mathbf{Q}(\mathbf{R})$ is the orthogonal matrix (satisfying $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) associated with a rotation $\mathbf{R}$ acting on elements in the tensor space, $\mathbf{R}(\mathbf{x})$ is a field of rotation matrices giving the local orientation of each phase, and $\chi_i(\mathbf{x})$ represents the characteristic function that is 1 in phase i and zero elsewhere. For example, $\mathbf{L}(\mathbf{x})$ could represent the moduli in an n-phase polycrystalline material. In that case the matrices $\mathbf{L}_1, \ldots, \mathbf{L}_n$ represent the moduli of the pure crystalline phases, while $\mathbf{R}(\mathbf{x})$ represents the field of rotation matrices required to account for the different crystal orientations throughout space.

The proof of the analytic properties of $\mathbf{L}_*$ as a function of the component tensors $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ carries over directly from the theory of composites, and is a straightforward extension of an argument of Bruno (1991b) (see also Bruno and Leo 1992). For composites, the extension is given on page 372 of Milton (2002), although there I should have referenced Section 14.7 page 300, rather than Section 14.6 on page 298, since if $\mathbf{L}(\mathbf{x})$ is complex it *is not self-adjoint*). The argument given in Section 14.7 itself needs a minor correction, as the inner product $(\mathbf{P}, \mathbf{L}\mathbf{P})$ needs to be replaced by $\mathrm{Re}(\mathbf{P}, \mathbf{L}\mathbf{P})$: the correct analysis is given here in Section 2.4.

In Section 2.4 we obtained the formula (2.35) for the effective tensor as a series expansion

$$\mathbf{L}_* = \sum_{j=0}^{\infty}\boldsymbol{\Gamma}_0\mathbf{L}[\boldsymbol{\Gamma}_1(\mathbf{I} - \mathbf{L}/\sigma_0)^j]\boldsymbol{\Gamma}_0. \tag{3.49}$$

Note that each term in this series expansion is a polynomial in the elements of the matrices $\mathbf{L}_1, \ldots, \mathbf{L}_n$. As a sequence of analytic functions that converge uniformly on any compact subset of a domain is analytic in that domain [see theorem 10.28 Rudin (1987)], it follows that if this series converges, then it will be an analytic function of all the elements of $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ in this region of convergence.

In Section 2.4 we established the series converges when $\mathbf{L}$ is bounded, i.e., there is some $\beta > 0$ such that

$$\beta > \sup_{\substack{\mathbf{P}\in\mathcal{H}\\ |\mathbf{P}|=1}} |\mathbf{L}\mathbf{P}|, \tag{3.50}$$

and $\mathbf{L}$ is coercive in the sense that there is some $\alpha > 0$ with

$$\mathrm{Re}(\mathbf{P}, \mathbf{L}\mathbf{P}) \geq \alpha|\mathbf{P}|^2 \quad \text{for all } \mathbf{P} \in \mathcal{H}. \tag{3.51}$$

With the natural choice (3.22) of inner product these conditions are satisfied if the tensors $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ are all bounded, and coercive in the sense that (2.50) holds, i.e., if for some angle θ (independent of j), the self adjoint part of $e^{i\theta}\mathbf{L}_j$ is strictly positive definite for all j. This condition holds, for example, when the tensors $\mathbf{L}_j$ are all symmetric with positive definite imaginary parts. Hence $\mathbf{L}_*$ is an analytic function of $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ in the domain where for some angle θ (independent of j), the self adjoint part of $e^{i\theta}\mathbf{L}_j$ is strictly positive definite for all j.

For time-harmonic problems each of the tensors $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ are often analytic functions of the frequency ω in the upper half plane $\mathrm{Im}\,\omega > 0$, having positive definite imaginary parts, $\mathrm{Im}\,\mathbf{L}_i > 0$ for all i.

Since an analytic function of analytic function is an analytic function, it follows that in this case $\mathbf{L}_*$ will be an analytic function of frequency in the upper half plane $\mathrm{Im}\,\omega > 0$.

Note that, since in (3.36) the basis fields $\mathbf{U}_\beta(\mathbf{x})$ (and hence $\mathbf{u}_\beta(\mathbf{x})$ and $\mathbf{G}_\beta(\mathbf{x})$) only depend on $\mathbf{Z}_0$ and not on the component moduli, it follows that the Dirichlet-to-Neumann map Λ will inherit the analytic properties of the effective operator $\mathbf{L}_*$ as a function of the component moduli, or component tensors. The Herglotz properties of Λ as a function of the component moduli, or component tensors, then follow immediately from (3.47).

We also remark that there are results (Alessandrini and Vessella 2005; Beretta and Francini 2011; Beretta, de Hoop, and Qiu 2013; Beretta, Francini, and Vessella 2014) on Lipschitz stability of the Dirichlet-to-Neumann map, when the moduli are piecewise constant. These results are not only for stability when the moduli change with fixed geometry (which could be derived using the analyticity properties established here) but also, and more importantly, for stability under changes in the geometry.

3.6 Partial data

The structure extends to the case when there is partial data. Again there is a modified subspace collection. This observation is basically made (in the general setting) in (29.1) and below on page 620 of Milton (2002). One considers the restricted subspace $\mathcal{U}'$ to be that subspace of $\mathcal{U}$ associated with boundary data where $\mathbf{u}$ can only be nonzero on a portion $\partial\Omega'$ of $\partial\Omega$. Then we suppose $\mathcal{U}'_\perp$ is the orthogonal complement of $\mathcal{U}'$ in the space $\mathcal{U}$. Fields in $\mathcal{U}'_\perp$ have a flux $\mathbf{G}\cdot\mathbf{n}$ which is zero on $\partial\Omega'$ (as the inner product of two fields in $\mathcal{U}$ involves at the boundary the flux of one, and the potential of the other). Keep the Hilbert space the same as before, but redefine the subspaces

$$\mathcal{J}' = \mathcal{U}'_\perp \oplus \mathcal{J}, \quad \mathcal{E}' = \mathcal{E}. \tag{3.52}$$

Then everything goes through. The solutions are exactly the same as before but now the field $\mathbf{J}_0 + \mathbf{J}$ gets re-expressed as $\mathbf{J}_0{}' + \mathbf{J}'$, where $\mathbf{J}_0{}' \in \mathcal{U}'$ and $\mathbf{J}' \in \mathcal{J}'$. There is an effective tensor $\mathbf{L}'_*$ which maps fields in $\mathcal{U}'$ to fields in $\mathcal{U}'$.

More generally, we can just do a set of M measurements where, say, $\mathbf{u}$ at the boundary takes values $\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_M$. Given $\mathbf{Z}_0$, associated with these are fields $\mathbf{E}_1$, $\mathbf{E}_2$, ..., $\mathbf{E}_M$ in $\mathcal{U}$. We then take $\mathcal{U}'$ to be the space spanned by these fields, and define $\mathcal{J}'$ and $\mathcal{E}'$ through (3.52). The associated effective operator $\mathbf{L}'_*$ can then be represented by an $M \times M$ matrix.

One can also consider partial data where the flux $\mathbf{G}\cdot\mathbf{n}$ is nonzero only on a portion $\partial\Omega''$ of $\partial\Omega$. Again there is a subspace $\mathcal{U}''$ of $\mathcal{U}$ associated with such boundary data. We suppose $\mathcal{U}''_\perp$ is the orthogonal complement of $\mathcal{U}''$ in the space $\mathcal{U}$. Fields in $\mathcal{U}''_\perp$ have a potential $\mathbf{u}$ which is zero on $\partial\Omega''$. Let

$$\mathcal{J}'' = \mathcal{J}, \quad \mathcal{E}'' = \mathcal{E} \oplus \mathcal{U}''_\perp, \tag{3.53}$$

and everything proceeds as before. There is an effective tensor $\mathbf{L}''_*$ which maps fields in $\mathcal{U}''$ to fields in $\mathcal{U}''$.

Again more generally, we can just do a set of M measurements where, say, the flux $\mathbf{G}\cdot\mathbf{n}$ at the boundary takes values $\mathbf{t}_1$, $\mathbf{t}_2$, ..., $\mathbf{t}_M$, and obtain an effective operator $\mathbf{L}'_*$ represented by an $M \times M$ matrix.

3.7 Mixed data

Now suppose we have mixed data where $\mathbf{u}$ is prescribed on a portion $\partial\Omega''$ of $\partial\Omega$, while the flux $\mathbf{G}\cdot\mathbf{n}$ is prescribed on the remaining portion $\partial\Omega'$ of $\partial\Omega$, where $\partial\Omega' \cup \partial\Omega'' = \partial\Omega$. We first consider two problems

separately: first the problem where the flux $\mathbf{G}\cdot\mathbf{n}$ is prescribed on the portion $\partial\Omega'$ of $\partial\Omega$, and $\mathbf{u}$ is zero on $\partial\Omega''$; and second where $\mathbf{u}$ is prescribed on a portion $\partial\Omega''$ of $\partial\Omega$, and the flux $\mathbf{G}\cdot\mathbf{n}$ is zero on $\partial\Omega'$. These are essentially the same problems considered in the previous section when we have partial data. Therefore let us define the subspaces $\mathcal{U}'$, $\mathcal{E}'$, $\mathcal{J}'$, and $\mathcal{U}''$, $\mathcal{E}''$, $\mathcal{J}''$ as in the previous section. For the first problem the appropriate effective tensor is $\mathbf{L}'_*$, since the field ${\mathbf{E}_0}'$ will have nonzero $\mathbf{u}$ on $\partial\Omega'$ and zero $\mathbf{u}$ on $\partial\Omega''$. For the second problem the appropriate effective tensor is $\mathbf{L}''_*$, since the field ${\mathbf{J}_0}''$ will have nonzero flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega''$ and a zero flux $\mathbf{G}\cdot\mathbf{n}=0$ on the remaining part $\partial\Omega'$ of the boundary.

In the first problem we apply a field ${\mathbf{J}_0}'$ associated with the prescribed flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega'$, then ${\mathbf{E}_0}' = (\mathbf{L}'_*)^{-1}{\mathbf{J}_0}'$ gives us the potential $\mathbf{u}$ on $\partial\Omega$, which will be zero on the portion $\partial\Omega''$. Applying $\mathbf{L}_*$ to ${\mathbf{E}_0}'$ gives

$$\mathbf{J}_0 = \mathbf{L}_*(\mathbf{L}'_*)^{-1}{\mathbf{J}_0}', \tag{3.54}$$

and from this field we can obtain the flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega$, which will coincide with the prescribed flux on $\partial\Omega'$. Thus we obtain the response, which is the potential $\mathbf{u}$ on $\partial\Omega'$ and the flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega''$.

In the second problem we apply a field ${\mathbf{E}_0}'$ associated with the prescribed potential $\mathbf{u}$ on $\partial\Omega''$, then ${\mathbf{J}_0}'' = \mathbf{L}''_*{\mathbf{E}_0}''$ gives us the flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega$, which will be zero on the portion $\partial\Omega'$. Applying $(\mathbf{L}_*)^{-1}$ to ${\mathbf{J}_0}''$ gives

$$\mathbf{E}_0 = (\mathbf{L}_*)^{-1}\mathbf{L}''_*{\mathbf{J}_0}'', \tag{3.55}$$

and from this field we can obtain the potential $\mathbf{u}$ on $\partial\Omega$, which will coincide with the prescribed potential on $\partial\Omega''$. Thus we obtain the response, which is the potential $\mathbf{u}$ on $\partial\Omega'$ and the flux $\mathbf{G}\cdot\mathbf{n}$ on $\partial\Omega''$.

Finally by summing the responses for the first and second problem, we obtain the response for the original mixed boundary conditions.

3.8 Applicability to the Schrödinger and heat conduction equations in the time domain

Recall from **Chapter 1**, the heat conduction (diffusion) equation (1.151) for the temperature (or particle concentration) T,

$$\underbrace{\begin{pmatrix}\mathbf{q}_x\\ q_t\\ \nabla\cdot\mathbf{q}_x+\frac{\partial q_t}{\partial t}\end{pmatrix}}_{\mathcal{G}} = \underbrace{\begin{pmatrix} i\mathbf{k}(\mathbf{x}) & 0 & 0\\ 0 & 0 & -\frac{i\alpha(\mathbf{x})}{2}\\ 0 & \frac{i\alpha(\mathbf{x})}{2} & 0\end{pmatrix}}_{\mathbf{Z}}\underbrace{\begin{pmatrix}\nabla T\\ \frac{\partial T}{\partial t}\\ T\end{pmatrix}}_{\mathcal{F}}, \tag{3.56}$$

and the Schrödinger equation (1.187) for the wavefunction ψ of an electron, or many electrons, in a potential V:

$$\underbrace{\begin{pmatrix}\mathbf{q}_x\\ q_t\\ \nabla\cdot\mathbf{q}_x+\frac{\partial q_t}{\partial t}\end{pmatrix}}_{\mathcal{G}} = \underbrace{\begin{pmatrix} -\mathbf{A} & 0 & 0\\ 0 & 0 & -\frac{i\hbar}{2}\\ 0 & \frac{i\hbar}{2} & -V\end{pmatrix}}_{\mathbf{Z}}\underbrace{\begin{pmatrix}\nabla\psi\\ \frac{\partial\psi}{\partial t}\\ \psi\end{pmatrix}}_{\mathcal{F}}, \tag{3.57}$$

where the meanings of the various quantities are explained in **Chapter 1**. These equations are of the form (3.6) and (3.4) with ∇ replaced by

$$\underline{\nabla} = \begin{pmatrix}\nabla\\ \frac{\partial}{\partial t}\end{pmatrix}. \tag{3.58}$$

Therefore, all the preceding analysis applies with ∇ replaced by $\underline{\nabla}$ and Ω replaced by a "space–time" body $\underset{\sim}{\Omega}$.

3.9 Adding source terms

These are easily handled in the theory, and follow the treatment in Section 1.19. With a source term $\mathbf{h}(\mathbf{x})$ weighted by a constant θ_0 the equations take the form

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\mathcal{F}(\mathbf{x}) + \theta_0\mathbf{h}(\mathbf{x}), \tag{3.59}$$

which we can reformulate as

$$\underbrace{\begin{pmatrix}\mathcal{G}(\mathbf{x})\\ \nabla\cdot\mathbf{r}(\mathbf{x})\end{pmatrix}}_{\mathcal{G}'(\mathbf{x})} = \underbrace{\begin{pmatrix}\mathbf{Z}(\mathbf{x}) & \mathbf{h}(\mathbf{x})\\ \mathbf{h}(\mathbf{x})^\dagger & \mathbf{d}(\mathbf{x})\end{pmatrix}}_{Z'(\mathbf{x})}\underbrace{\begin{pmatrix}\mathcal{F}(\mathbf{x})\\ \theta_0\end{pmatrix}}_{\mathcal{F}'(\mathbf{x})}, \tag{3.60}$$

where we are free to choose $\mathbf{d}(\mathbf{x})$ (often it is chosen so $\mathbf{Z}'(\mathbf{x})$ is positive definite). Here $\mathcal{F}'$ satisfies the differential constraint that θ_0 is constant, $\mathcal{G}(\mathbf{x}) = \begin{pmatrix}\mathbf{G}\\ \nabla\cdot\mathbf{G}\end{pmatrix}$ and the vector field $\mathbf{r}(\mathbf{x})$ is not subject to any differential constraints (and in general is not uniquely determined by the equations, only its divergence). We have the key identity:

$$\int_\Omega \mathcal{G}'\cdot\mathcal{F}' = \int_{\partial\Omega}\begin{pmatrix}\mathbf{n}\cdot\mathbf{G}\\ \mathbf{n}\cdot\mathbf{r}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{u}(\mathbf{x})\\ \theta_0\end{pmatrix}, \tag{3.61}$$

and all of the analysis applies, although we may not be able to physically measure $\mathbf{r}(\mathbf{x})$ at the boundary since it is $\nabla\cdot\mathbf{r}(\mathbf{x})$ which has the physical significance.

We start by choosing a positive definite reference tensor $\mathbf{Z}'_0$ and look at those fields satisfying (3.60) with $\mathbf{Z}'(\mathbf{x})$ replaced by $\mathbf{Z}'_0$: the associated fields $(\mathbf{Z}'_0)^{1/2}\mathcal{F}'(\mathbf{x}) = (\mathbf{Z}'_0)^{-1/2}\mathcal{G}'(\mathbf{x})$ span a subspace which we define to be $\mathcal{U}'$. We take $\mathcal{E}'$ to consist of fields $\mathbf{E}'(\mathbf{x})$ that are square integrable in Ω taking the form

$$\mathbf{E}'(\mathbf{x}) = (\mathbf{Z}'_0)^{1/2}\begin{pmatrix}\nabla\mathbf{u}(\mathbf{x})\\ \mathbf{u}(\mathbf{x})\\ 0\end{pmatrix}, \tag{3.62}$$

for some potential $\mathbf{u}(\mathbf{x})$ with $\mathbf{u}(\mathbf{x}) = 0$ when $\mathbf{x}\in\partial\Omega$. The space $\mathcal{J}$ consists of fields $\mathbf{J}'(\mathbf{x})$ that are square integrable in Ω of the form

$$\mathbf{J}'(\mathbf{x}) = (\mathbf{Z}'_0)^{-1/2}\begin{pmatrix}\mathbf{G}(\mathbf{x})\\ \nabla\cdot\mathbf{G}(\mathbf{x})\\ \nabla\cdot\mathbf{r}(\mathbf{x})\end{pmatrix}, \tag{3.63}$$

where $\mathbf{n}\cdot\mathbf{G} = \mathbf{n}\cdot\mathbf{r} = 0$ on $\partial\Omega$. Equivalently $\mathcal{J}$ consists of fields of the form

$$\mathbf{J}'(\mathbf{x}) = (\mathbf{Z}'_0)^{-1/2}\begin{pmatrix}\mathbf{G}(\mathbf{x})\\ \nabla\cdot\mathbf{G}(\mathbf{x})\\ S(\mathbf{x})\end{pmatrix}, \tag{3.64}$$

where $\mathbf{n}\cdot\mathbf{G} = 0$ on $\partial\Omega$, and the average of $S(\mathbf{x})$ over Ω, $\langle S\rangle$, is zero.

Defining $\mathbf{L}'(\mathbf{x}) = (\mathbf{Z}'_0)^{-1/2}\mathbf{Z}'(\mathbf{x})(\mathbf{Z}'_0)^{-1/2}$ the equations become

$$(\mathbf{J}_0{}'(\mathbf{x}) + \mathbf{J}'(\mathbf{x})) = \mathbf{L}'(\mathbf{x})(\mathbf{E}_0{}'(\mathbf{x}) + \mathbf{E}'(\mathbf{x})), \tag{3.65}$$

with $\mathbf{J}_0{}', \ \mathbf{E}_0{}' \in \mathcal{U}, \mathbf{J}'(\mathbf{x}) \in \mathcal{J}, \mathbf{E}'(\mathbf{x}) \in \mathcal{E}$. The effective operator $\mathbf{L}'_*$ defined by

$$\mathbf{J}_0{}' = \mathbf{L}'_* \mathbf{E}_0{}' \tag{3.66}$$

then governs how the field $\mathbf{n}\cdot\mathbf{G}$ at the boundary $\partial\Omega$ responds to applied potentials $\mathbf{u}(\mathbf{x})$ and θ_0 at the boundary. We skip the analysis which parallels (3.29) to (3.36).

3.10 Static and quasistatic equations

We can consider electrical and thermal conduction, elasticity, piezoelectricity, magnetostriction, Hall Effect conductivity, thermoelectricity, thermoelasticity, the steady-state Biot equations, magnetic permeability, fluid flow in porous media (with spatially varying permeability), the dielectric problem, plate equations, chemical diffusion, neutron diffusion, antiplane elasticity (see e.g., Section 2 of Milton 2002). These can all be formulated as

$$\underbrace{\begin{pmatrix} \mathbf{G}(\mathbf{x}) \\ \nabla\cdot\mathbf{R}(\mathbf{x}) \end{pmatrix}}_{\mathcal{G}'(\mathbf{x})} = \underbrace{\begin{pmatrix} \boldsymbol{\mathcal{A}}(\mathbf{x}) & \boldsymbol{\mathcal{B}}(\mathbf{x}) \\ \boldsymbol{\mathcal{C}}(\mathbf{x}) & \boldsymbol{\mathcal{D}}(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}'(\mathbf{x})} \underbrace{\begin{pmatrix} \nabla\mathbf{u}(\mathbf{x}) \\ \boldsymbol{\theta}_0 \end{pmatrix}}_{\mathcal{F}'(\mathbf{x})}, \tag{3.67}$$

where $\boldsymbol{\theta}_0$ is a constant ℓ-component vector (possibly $\ell = 0$ when the equations are simply $\mathbf{G} = \boldsymbol{\mathcal{A}}\nabla\mathbf{u}$, $\nabla\cdot\mathbf{G} = 0$), $\mathbf{u}(\mathbf{x})$ is an $m-$component vector field, $\mathbf{G}(\mathbf{x})$ is a $d\times m$ dimensional matrix satisfying $\nabla\cdot\mathbf{G}(\mathbf{x}) = 0$ and $\mathbf{R}(\mathbf{x})$ is a $d \times \ell$ component matrix-valued vector field. When the body is subject to time-harmonic fields varying at a frequency ω such that the wavelength is much bigger than the size of the body, then the quasistatic equations are often appropriate. These retain the same form, only the fields $\mathcal{G}'(\mathbf{x})$, $\mathcal{F}'(\mathbf{x})$ and tensor $\mathbf{Z}'(\mathbf{x})$ become complex: the physical fields are $\mathrm{Re}[e^{-i\omega t}\mathcal{G}'(\mathbf{x})]$ and $\mathrm{Re}[e^{-i\omega t}\mathcal{F}'(\mathbf{x})]$.

We have the key identity:

$$\int_\Omega \begin{pmatrix} \mathbf{G}(\mathbf{x}) \\ \nabla\cdot\mathbf{R}(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} \nabla\mathbf{u}(\mathbf{x}) \\ \boldsymbol{\theta}_0 \end{pmatrix} = \int_{\partial\Omega} \mathbf{n}\cdot\mathbf{G}(\mathbf{x})\mathbf{u}(\mathbf{x}) + \mathbf{n}\cdot\mathbf{R}(\mathbf{x})\boldsymbol{\theta}_0, \tag{3.68}$$

and all of the analysis applies, although again we may not be able to physically measure $\mathbf{R}(\mathbf{x})$ at the boundary since it is $\nabla\cdot\mathbf{R}(\mathbf{x})$ which has the physical significance.

We start by choosing a positive definite reference tensor $\mathbf{Z}'_0$ and look at those fields satisfying (3.60) with $\mathbf{Z}'(\mathbf{x})$ replaced by $\mathbf{Z}'_0$: the associated fields $(\mathbf{Z}'_0)^{1/2}\mathcal{F}'(\mathbf{x}) = (\mathbf{Z}'_0)^{-1/2}\mathcal{G}'(\mathbf{x})$ span a subspace which we define to be $\mathcal{U}'$. We take $\mathcal{E}'$ to consist of fields $\mathbf{E}'(\mathbf{x})$ that are square integrable in Ω taking the form

$$\mathbf{E}'(\mathbf{x}) = \begin{pmatrix} \nabla\mathbf{u}(\mathbf{x}) \\ 0 \end{pmatrix}, \tag{3.69}$$

for some potential $\mathbf{u}(\mathbf{x})$ with $\mathbf{u}(\mathbf{x}) = 0$ when $\mathbf{x} \in \partial\Omega$. The space $\mathcal{J}$ consists of fields $\mathbf{J}'(\mathbf{x})$ that are square integrable in Ω of the form

$$\mathbf{J}'(\mathbf{x}) = \begin{pmatrix} \mathbf{G}(\mathbf{x}) \\ \nabla\cdot\mathbf{R}(\mathbf{x}) \end{pmatrix}, \tag{3.70}$$

where $\nabla\cdot\mathbf{G} = 0$ inside Ω and $\mathbf{n}\cdot\mathbf{G} = \mathbf{n}\cdot\mathbf{R} = 0$ on $\partial\Omega$. Equivalently $\mathcal{J}$ consists of fields of the form

$$\mathbf{J}'(\mathbf{x}) = \begin{pmatrix} \mathbf{G}(\mathbf{x}) \\ \mathbf{S}(\mathbf{x}) \end{pmatrix}, \tag{3.71}$$

where $\nabla \cdot \mathbf{G} = 0$ inside Ω, $\mathbf{n} \cdot \mathbf{G} = 0$ on $\partial\Omega$, and the average of $\mathbf{S}(\mathbf{x})$ over Ω, $\langle \mathbf{S} \rangle$, is zero.

Defining $\mathbf{L}'(\mathbf{x}) = (\mathbf{Z}_0')^{-1/2}\mathbf{Z}'(\mathbf{x})(\mathbf{Z}_0')^{-1/2}$ the equations become

$$(\mathbf{J}_0{}'(\mathbf{x}) + \mathbf{J}'(\mathbf{x})) = \mathbf{L}'(\mathbf{x})(\mathbf{E}_0{}'(\mathbf{x}) + \mathbf{E}'(\mathbf{x})), \tag{3.72}$$

with $\mathbf{J}_0{}',\ \mathbf{E}_0{}' \in \mathcal{U}$, $\mathbf{J}'(\mathbf{x}) \in \mathcal{J}$, $\mathbf{E}'(\mathbf{x}) \in \mathcal{E}$. The effective operator $\mathbf{L}_*'$, defined by

$$\mathbf{J}_0{}' = \mathbf{L}_*'\mathbf{E}_0{}', \tag{3.73}$$

then governs how the field $\mathbf{n}\cdot\mathbf{G}$ at the boundary $\partial\Omega$ responds to applied potentials $\mathbf{u}(\mathbf{x})$ and $\boldsymbol{\theta}_0$ at the boundary. Again we skip the analysis which parallels (3.29) to (3.36).

3.11 Applicability to acoustics and elastodynamics in the time domain

Recall from **Chapter 1**, that with

$$\underset{\sim}{\nabla} \equiv \begin{pmatrix} \nabla \\ \frac{\partial}{\partial x_4} \end{pmatrix} = \begin{pmatrix} \nabla \\ -\frac{\partial}{\partial t} \end{pmatrix}, \tag{3.74}$$

we have the equation (1.118) for acoustics in the time domain,

$$\underset{\sim}{\nabla} \cdot \mathbf{j}(\underset{\sim}{\mathbf{x}}) = 0, \quad \mathbf{j}(\underset{\sim}{\mathbf{x}}) = \mathbf{Z}(\underset{\sim}{\mathbf{x}})\underset{\sim}{\nabla} P(\underset{\sim}{\mathbf{x}}), \tag{3.75}$$

where P is the pressure and $\underset{\sim}{\mathbf{x}} = (x_1, x_2, x_3, x_4)$, with $x_4 = -t$ where t is the time. Also recall from **Chapter 1**, the equation (1.126) for elastodynamics in the time domain,

$$\underset{\sim}{\nabla} \cdot \mathbf{J} = 0, \quad \mathbf{J} = \mathbf{Z}\underset{\sim}{\nabla}\mathbf{v}, \tag{3.76}$$

where $\mathbf{v}$ is the velocity. Both these equations are of the form (3.67) (with $\boldsymbol{\theta}_0 = \boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{D}} = 0$) with $\mathbf{x}$ replaced by $\underset{\sim}{\mathbf{x}}$ and ∇ replaced by $\underset{\sim}{\nabla}$. Therefore, with Ω replaced by a space-time body $\underset{\sim}{\Omega}$ all the results carry through and the problem can be reformulated in the language of the theory of composites.

3.12 The electromagnetic equations in the frequency domain

For electromagnetism at fixed frequency ω in a three-dimensional body Ω the governing equations for the electric field $\mathbf{e}(\mathbf{x})$ and magnetic field intensity $\mathbf{h}(\mathbf{x})$ are

$$\underbrace{\begin{pmatrix} -i\mathbf{h} \\ i\nabla \times \mathbf{h} \end{pmatrix}}_{\mathcal{G}} = \underbrace{\begin{pmatrix} -[\omega\boldsymbol{\mu}(\mathbf{x})]^{-1} & 0 \\ 0 & \omega\boldsymbol{\varepsilon}(\mathbf{x}) \end{pmatrix}}_{\mathbf{Z}} \underbrace{\begin{pmatrix} \nabla \times \mathbf{e} \\ \mathbf{e} \end{pmatrix}}_{\mathcal{F}}, \tag{3.77}$$

where $\boldsymbol{\mu}(\mathbf{x})$ is the magnetic permeability tensor and $\boldsymbol{\varepsilon}(\mathbf{x})$ is the electrical permittivity tensor. We have the key identity (Poynting's theorem):

$$\int_\Omega \mathcal{G}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}) = \int_{\partial\Omega} -i\mathbf{n} \cdot (\mathbf{e} \times \mathbf{h}) = \int_{\partial\Omega} i(\mathbf{n} \times \mathbf{h}) \cdot \{[\mathbf{I} - \mathbf{n}\mathbf{n}^T]\mathbf{e}\}, \tag{3.78}$$

in which $[\mathbf{I}-\mathbf{n}\mathbf{n}^T]\mathbf{e}=\mathbf{n}\times\mathbf{e}\times\mathbf{n}$ is nothing but the tangential component of the electric field, $\mathbf{I}-\mathbf{n}\mathbf{n}^T$ being the projection operator onto this tangential component.

The mapping from the field $[\mathbf{I}-\mathbf{n}\mathbf{n}^T]\mathbf{e}$ at the boundary $\partial\Omega$ to the field $i(\mathbf{n}\times\mathbf{h})$ is what we define as our Dirichlet-to-Neumann map. The appearance of $\mathbf{n}(\mathbf{x})\times\mathbf{h}(\mathbf{x})$ on the left of this relation is quite natural. In the special relativity form of Maxwell's equations in Section 1.15 the electric field $\mathbf{e}(\mathbf{x})$ enters $\mathbf{F}$ as a vector with 3 spatial components while the magnetizing field $\mathbf{h}$ enters the field $\mathbf{G}$ (1.167) as an antisymmetric matrix:

$$\underline{\mathbf{h}}=\begin{pmatrix}0&-h_3&h_2\\h_3&0&-h_1\\-h_2&h_1&0\end{pmatrix}. \tag{3.79}$$

The associated flux is $\mathbf{n}\cdot\underline{\mathbf{h}}=\mathbf{n}\times\mathbf{h}$. In the next section, an alternative Dirichlet-to-Neumann map from $\mathbf{e}\times\mathbf{n}$ to $\mathbf{n}\times\mathbf{h}\times\mathbf{n}$ is used. If we use the invariance of the time-harmonic Maxwell's equations under the interchanges

$$\mathbf{e}\Leftrightarrow\mathbf{h},\quad \mathbf{d}\Leftrightarrow\mathbf{b},\quad i\Leftrightarrow-i,\quad \boldsymbol{\varepsilon}\Leftrightarrow\boldsymbol{\mu} \tag{3.80}$$

then this map transforms to a map from $\mathbf{h}\times\mathbf{n}$ to $\mathbf{n}\times\mathbf{e}\times\mathbf{n}$, which is the inverse of the map considered here. Both maps differ from the map of Uhlmann and Zhou (2015) which maps $\mathbf{n}\times\mathbf{e}$ to $\mathbf{n}\times\mathbf{h}$.

We take a positive definite real tensor $\mathbf{Z}_0$ and consider the solutions to

$$\underbrace{\begin{pmatrix}-i\mathbf{h}\\i\nabla\times\mathbf{h}\end{pmatrix}}_{\mathcal{G}_0}=\mathbf{Z}_0\underbrace{\begin{pmatrix}\nabla\times\mathbf{e}\\\mathbf{e}\end{pmatrix}}_{\mathcal{F}_0}, \tag{3.81}$$

as the boundary fields range over all possible data. For any prescribed real value of the tangential field $\mathbf{e}_T=(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{e}$ we calculate

$$\min_{\substack{\mathbf{e}\\(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{e}=\mathbf{e}_T\text{ on }\partial\Omega}}\int_\Omega\begin{pmatrix}\nabla\times\mathbf{e}\\\mathbf{e}\end{pmatrix}\cdot\mathbf{Z}_0\begin{pmatrix}\nabla\times\mathbf{e}\\\mathbf{e}\end{pmatrix}. \tag{3.82}$$

Provided $\mathbf{e}_T$ has sufficient regularity the minimum exists, is unique and at the minimum

$$0=\int_\Omega\begin{pmatrix}\nabla\times\delta\mathbf{e}\\\delta\mathbf{e}\end{pmatrix}\cdot\begin{pmatrix}-i\mathbf{h}\\\omega\mathbf{d}\end{pmatrix}=-i(\nabla\times\delta\mathbf{e})\cdot\mathbf{h}+\delta\mathbf{e}\cdot\omega\mathbf{d}, \tag{3.83}$$

for all $\delta\mathbf{e}$ with zero tangential value at the boundary. By integrating this last expression by parts we see that $\omega\mathbf{d}=i\nabla\times\mathbf{h}$. So for any sufficiently regular real-valued boundary condition $(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{e}=\mathbf{e}_T$ there exists a unique field $\mathcal{F}_0(\mathbf{x})$ which corresponds to it. For a complex prescribed field $\mathbf{e}_T=\mathbf{e}'_T+i\mathbf{e}''_T$ we find the fields $\mathcal{F}'_0(\mathbf{x})$ and $\mathcal{F}''_0(\mathbf{x})$ associated with $\mathbf{e}'_T$ and $\mathbf{e}''_T$ to get the field $\mathcal{F}_0=\mathcal{F}'_0+\mathcal{F}''_0$ associated with $\mathbf{e}_T=\mathbf{e}'_T+i\mathbf{e}''_T$.

We define $\mathcal{U}$ as the space of values of $\mathbf{Z}_0^{1/2}\mathcal{F}_0(\mathbf{x})$ as the tangential field $\mathbf{e}_T=(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{e}$ is varied. The space $\mathcal{E}$ consists of fields of the form

$$\mathbf{E}=\mathbf{Z}_0^{1/2}\begin{pmatrix}\nabla\times\mathbf{e}\\\mathbf{e}\end{pmatrix}, \tag{3.84}$$

as $\mathbf{e}$ is varied with $(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{e}=0$ at $\partial\Omega$. The space $\mathcal{J}$ consists of fields of the form

$$\mathbf{J}=\mathbf{Z}_0^{-1/2}\begin{pmatrix}-i\mathbf{h}\\i\nabla\times\mathbf{h}\end{pmatrix}, \tag{3.85}$$

as $\mathbf{h}$ is varied with $\mathbf{n} \times \mathbf{h} = 0$ on $\partial\Omega$. All the analysis goes through as before. In particular the effective tensor $\mathbf{L}_*$ and hence the Dirichlet-to-Neumann map, mapping the boundary field $\mathbf{e}_T = (\mathbf{I} - \mathbf{n}\mathbf{n}^T)\mathbf{e}$ to the boundary field $i(\mathbf{n} \times \mathbf{h})$ in an n-phase medium is an analytic function of the matrix elements of $\omega\boldsymbol{\varepsilon}_i$ and $\omega\boldsymbol{\mu}_i$, $i = 1, 2, \ldots, n$ when the imaginary parts of all these tensors are positive definite. A separate and rigorous proof of this fact is given in the next chapter.

3.13 The electromagnetic equations in the time domain

In the time domain, the desired form of the electromagnetic equations from (1.171), (1.172), and (1.173) is given by

$$\begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = \underbrace{\begin{pmatrix} -[\boldsymbol{\mu}(\mathbf{x})]^{-1} & 0 \\ 0 & \boldsymbol{\varepsilon}(\mathbf{x}) \end{pmatrix}}_{\mathbf{z}} \boldsymbol{\Theta} \begin{pmatrix} \boldsymbol{\Phi} \\ \phi \end{pmatrix}, \quad \boldsymbol{\Theta}^\dagger \begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = 0, \tag{3.86}$$

where

$$\boldsymbol{\Theta} = \begin{pmatrix} \nabla\times & 0 \\ -\frac{\partial}{\partial t} & -\nabla \end{pmatrix}, \quad \boldsymbol{\Theta}^\dagger = \begin{pmatrix} \nabla\times & \frac{\partial}{\partial t} \\ 0 & \nabla\cdot \end{pmatrix}. \tag{3.87}$$

We take a positive definite matrix $\mathbf{Z}_0$ and consider the solutions to the equations

$$\mathcal{G}_0(\mathbf{x}) = \mathbf{Z}_0\mathcal{F}_0(\mathbf{x}), \quad \mathcal{F}_0 = \boldsymbol{\Theta}\begin{pmatrix} \boldsymbol{\Phi}_0 \\ \phi_0 \end{pmatrix}, \quad \boldsymbol{\Theta}^\dagger\mathcal{G}_0(\mathbf{x}) = 0. \tag{3.88}$$

These are generated by taking some potentials $\boldsymbol{\Phi}_0(\mathbf{x})$ and ϕ_0 defined at the boundary of the space–time body $\underset{\sim}{\Omega}$ and considering

$$\min_{\substack{\boldsymbol{\Phi},\,\phi \\ \boldsymbol{\Phi}=\boldsymbol{\Phi}_0,\,\phi=\phi_0 \text{ on } \partial\Omega}} \int_{\underset{\sim}{\Omega}} \left[\boldsymbol{\Theta}\begin{pmatrix} \boldsymbol{\Phi} \\ \phi \end{pmatrix}\right] \cdot \mathbf{Z}_0 \left[\boldsymbol{\Theta}\begin{pmatrix} \boldsymbol{\Phi} \\ \phi \end{pmatrix}\right]. \tag{3.89}$$

The minimizing potentials, called $\boldsymbol{\Phi}_0$ and ϕ_0, give fields

$$\mathbf{U}_0 = \mathbf{Z}_0^{1/2}\boldsymbol{\Theta}\begin{pmatrix} \boldsymbol{\Phi}_0 \\ \phi_0 \end{pmatrix}, \tag{3.90}$$

which generate a space $\mathcal{U}$ as the boundary potentials are varied. The space $\mathcal{E}$ consists of fields of the form

$$\mathbf{E} = \mathbf{Z}_0^{1/2}\boldsymbol{\Theta}\begin{pmatrix} \boldsymbol{\Phi} \\ \phi \end{pmatrix}, \tag{3.91}$$

as $\boldsymbol{\Phi}$ and ϕ vary with $\boldsymbol{\Phi} = \phi = 0$ on $\partial\underset{\sim}{\Omega}$. The space $\mathcal{J}$ consists of fields of the form

$$\mathbf{J} = \mathbf{Z}_0^{-1/2}\begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix}, \quad \text{with } \boldsymbol{\Theta}^\dagger \begin{pmatrix} -\mathbf{h} \\ \mathbf{d} \end{pmatrix} = 0, \tag{3.92}$$

as $\mathbf{h}$ and $\mathbf{d}$ vary (subject to the last constraint in (3.92)) with $\mathbf{d} = 0$ and $\mathbf{n}_x \times \mathbf{h} = 0$ on $\partial\underset{\sim}{\Omega}$. All the analysis applies as before. However, from a practical viewpoint the vector potential $\boldsymbol{\Phi}$ cannot be directly measured: it is the magnetic field $\mathbf{b}(\mathbf{x})$ which has the physical significance. Nevertheless, if we have a n-phase body where the moduli $\boldsymbol{\mu}(\mathbf{x})$ and $\boldsymbol{\varepsilon}(\mathbf{x})$ are piecewise constant, then the Dirichlet-to-Neumann map will be a Herglotz function of the component tensors $\boldsymbol{\mu}_j$ and $\boldsymbol{\varepsilon}_j$, $j = 1, 2, \ldots, n$. Unlike in the time-harmonic case, complex values of $\boldsymbol{\mu}_i$ and $\boldsymbol{\varepsilon}_j$ do not have a physical significance.

Acknowledgments

G.W. Milton wishes to especially thank the National Science Foundation for support through grant DMS-1211359.

4 Analyticity of the Dirichlet-to-Neumann map for the time-harmonic Maxwell's equations

Maxence Cassier*, Aaron Welters and Graeme W. Milton***

*Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA
**Department of Mathematical Sciences, Florida Institute of Technology, Melbourne, FL 32901, USA
emails: cassier@math.utah.edu, awelters@fit.edu, milton@math.utah.edu

Abstract

In this chapter we derive the analyticity properties of the electromagnetic Dirichlet-to-Neumann map for the time-harmonic Maxwell's equations for passive linear multicomponent media. Moreover, we discuss the connection of this map to Herglotz functions for isotropic and anisotropic multicomponent composites.

Key words: multicomponent media, electromagnetic Dirichlet-to-Neumann map, analytic properties, Herglotz functions

4.1 Introduction

In this chapter, we study the analytic properties of the electromagnetic "Dirichlet-to-Neumann" (DtN) map for a composite material. Using passive linear multicomponent media, we will prove that this DtN map is an analytic function of the dielectric permittivities and magnetic permeabilities (multiplied by the frequency ω) which characterize each phase. More specifically, it belongs to a special class of functions known as Herglotz functions. In that sense, this chapter is highly connected to the previous one by Graeme Milton since both are proving analyticity properties on the DtN map, but with different methods. In **Chapter 3**, these analyticity properties are derived by using the theory of composite materials, whereas in this chapter they are proved via spectral theory in the usual functional framework associated with the time-harmonic Maxwell's equations. Maxwell's equations at fixed frequency ω involve the electric permittivity $\varepsilon(\mathbf{x}, \omega)$ (also called the dielectric constant if measured relative to the permittivity of the vacuum) and the magnetic permeability $\mu(\mathbf{x}, \omega)$. The approach taken in the current chapter has the important advantage of being applicable to bodies where the

moduli $\omega\varepsilon(\mathbf{x},\omega)$ and $\omega\mu(\mathbf{x},\omega)$ are not piecewise constant but instead vary smoothly (or not) with position. In this case we establish (in Subsection 4.3.4) the Herglotz properties of the Dirichlet-to-Neumann map, as a function of frequency, assuming the material is passive at each point $\mathbf{x}$, i.e., that $\omega\varepsilon(\mathbf{x},\omega)$ and $\omega\mu(\mathbf{x},\omega)$ are Herglotz functions of the frequency ω.

The use of theory of Herglotz functions in electromagnetism and in the theory of composites has many important impacts and consequences (Bergman 1978, 1980, 1982; Milton 1980, 1981a, 1981c, 2002; Golden and Papanicolaou 1983; Dell'Antonio, Figari, and Orlandi 1986; Bruno 1991a; Lipton 2000, 2001; Gustafsson and Sjöberg 2010; Bernland, Luger, and Gustafsson 2011; Liu, Guenneau, and Gralak 2013; Welters, Avniel, and Johnson 2014) especially in developing bounds on certain physical quantities. Based on this and the work of Golden and Papanicolaou (1985), Bergman (1986), Milton (1987a, 1987b) and Milton and Golden (1990) on developing bounds on effective tensors of composites containing more than two phases using analyticity of the effective tensors as a multivariable function of the moduli of the phases, we also establish that the DtN map is an analytic function of the permittivity and permeability tensors of each phase. Another potential application of these analytic properties is to derive information about the DtN map for real frequencies by using the theory of boundary-values of Herglotz functions (for instance, see Gesztesy and Tsekanovskii 2000 and Naboko 1996). Moreover, as the DtN map is usually used as data in electromagnetic inverse problems (see, for instance, Albanese and Monk 2006; Uhlmann and Zhou 2015, Ola, Päivärinta, and Somersalo 2012), we believe these analyticity properties and the connection to the theory of Herglotz functions will have important applications in this area of research (see **Chapter 5** of this book). The Herglotz properties might also be important to characterize the complete set of all possible Dirichlet-to-Neumann maps (either at fixed frequency or as a function of frequency) associated with multiphase bodies with frequency independent permittivity and permeability. Indeed such analyticity properties were a key ingredient to characterize the possible dynamic response functions of multiterminal mass-spring networks (Guevara Vasquez, Milton, and Onofrei 2011). These response functions are the discrete analogs of the Dirichlet-to-Neumann map in that problem. Additionally, analytic properties were a key ingredient in the theory of exact relations (Grabovsky 1998; Grabovsky and Sage 1998; Grabovsky, Milton, and Sage 2000: see also **Chapter 17** in Milton 2002 and Grabovsky 2004) satisfied by the effective tensors of composites, and for establishing links between effective tensors. These are generally nonlinear relations that are microstructure independent and thus, besides their intrinsic interest, are useful as benchmarks for numerical methods and approximations. They become linear (Grabovsky 1998) after a suitable fractional linear matrix transformation is made (which is nonunique and involves an arbitrary unit vector $\mathbf{n}$). After any such transformation is made and once certain algebraic relations are satisfied (for all unit vectors $\mathbf{n}$) it can be proved that all terms in the series expansion satisfy the exact relation, and then analyticity is needed to prove the relation holds (in the domain of analyticity) even if the series expansion does not converge (Grabovsky, Milton, and Sage 2000).

We split this chapter in three sections. In the first one, we consider the electromagnetic DtN map for a layered media. In this setting, the DtN map can be expressed explicitly in terms of the transfer matrix associated with the medium. This gives a good example in which one can see these analytic properties in the context of matrix perturbation theory (Baumgärtel 1985; Kato 1995; Welters 2011a). In the second section, we restrict ourselves to bounded media but with a large class of different geometries, more precisely, Lipschitz domains. In this case, using a variational reformulation of the time-harmonic Maxwell's equations (Cessenat 1996; Kirsch and Hettlich 2015; Monk 2003; Nedelec 2001), we prove both the well-posedness and the analyticity of the DtN map. Also we consider bodies where the moduli $\omega\varepsilon(\mathbf{x},\omega)$ and $\omega\mu(\mathbf{x},\omega)$ are not piecewise constant but instead vary with position, and at each point $\mathbf{x}$ are Herglotz functions of the frequency ω. In this case we establish the Herglotz properties of the Dirichlet-to-Neumann map, as a function of frequency. In both sections, the key step to prove the multivariable analyticity is Hartogs' Theorem from complex analysis which

essentially says that analyticity in each variable separately implies joint analyticity (see Theorem 4 below). Concerning the connection to Herglotz functions, an energy balance equation is derived (which is essentially Poynting's Theorem for complex frequencies) that allows us to prove that the imaginary part of the DtN map is positive definite, as a consequence of the positivity of the imaginary part of the material tensors. Nevertheless, in the case of anisotropic media, the connection to Herglotz functions has to be made more precise. Indeed, we leave here the usual framework of Herglotz functions of scalar variables since we are concerned with dielectric permittivity and magnetic permeability tensors as input variables. Thus, the purpose of the last section is to provide a rigorous definition of Herglotz functions in this general framework, that provides an alternative to the one developed in Section 18.8 of Milton (2002), by connecting this notion to the theory of holomorphic functions on tubular domains with nonnegative imaginary part as described in Vladimirov 2002 (see Sections 17–19). This new link is especially significant since this class of functions (like the Herglotz functions introduced in Section 18.8 of Milton (2002)) admits integral representations analogous to Herglotz functions of one complex variable (the representation in the one variable case as described in Theorem 3 below) and are deeply connected to the theory of multivariate passive linear systems (see Section 20 in Vladimirov 2002) with the notions of convolutions, passivity, causality, Laplace/Fourier transforms, and analyticity properties.

This chapter is essentially self-contained, and written in a rigorous mathematical style. Care has been taken to explain most technical definitions so that it should be accessible to non-mathematicians.

Before we proceed, let us introduce some notation, definitions and theorems used in this chapter. We denote:

- the complex upper-half plane by $\mathbb{C}^+ = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$,
- the Banach space of all $m \times n$ matrices with complex entries by $M_{m,n}(\mathbb{C})$ equipped with any norm, with the square matrices $M_{n,n}(\mathbb{C})$ denoted by $M_n(\mathbb{C})$, and we treat $\mathbb{C}^n$ as $M_{n,1}(\mathbb{C})$ (recall that a Banach space is a complete normed vector space: unlike a Hilbert space, it does not necessarily have an inner product defined on the space, just a norm.)
- by $\cdot$ the operation defined for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ via $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = u_i v_i$, where T denotes the transpose. Note that there is no complex conjugation in this definition, so $\mathbf{u} \cdot \mathbf{u}$ is not generally real.
- the open, connected, and convex subset of $M_n(\mathbb{C})$ of matrices with positive definite imaginary part by

$$M_n^+(\mathbb{C}) = \{\mathbf{M} \in M_n(\mathbb{C}) \mid \operatorname{Im} \mathbf{M} > 0\},$$

  where $\operatorname{Im} \mathbf{M} = (\mathbf{M} - \mathbf{M}^*)/(2i)$ with $\mathbf{M}^* = \overline{\mathbf{M}}^T$ the adjoint of $\mathbf{M}$, and the inequality $\mathbf{M} > 0$ holds in the sense of quadratic forms. We remark that this set is invariant by the operation: $\mathbf{M} \to -\mathbf{M}^{-1}$ since if $\mathbf{M} \in M_n^+(\mathbb{C})$ then $\mathbf{M}$ is invertible and

$$-\operatorname{Im}(\mathbf{M}^{-1}) = (\mathbf{M}^{-1})^* \operatorname{Im}(\mathbf{M}) \mathbf{M}^{-1} > 0$$

- by $L(E, F)$ the Banach space of all continuous linear operators from a Banach space E to a Banach space F equipped with the operator norm.

Definition 1. *(Analyticity) Let E and F be two complex Banach spaces and U be an open set of E. A function $f : U \to F$ is said to be a analytic if it is differentiable on $\mathcal{U}$.*

Definition 2. *(Herglotz functions) Let $m, n, N \in \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers (positive integers), and $\mathcal{T} = (\mathbb{C}^+)^n$ or $(M_N^+(\mathbb{C}))^n$. An analytic function $h : \mathcal{T} \to \mathbb{C}$ or $h : \mathcal{T} \to M_m(\mathbb{C})$ is called a Herglotz function (also called Pick or Nevanlinna function) if*

$$\mathrm{Im}(h(\mathbf{z})) \geq 0, \ \forall \mathbf{z} \in \mathcal{T}.$$

We note here that Definition 2 is the standard definition of a Herglotz function when $\mathcal{T} = \mathbb{C}^+$ (see Gesztesy and Tsekanovskii 2000, Berg 2008) and $\mathcal{T} = (\mathbb{C}^+)^n$ (in Agler, McCarthy, and Young 2012 it is called a Pick function), but not when $\mathcal{T} = (M_N^+(\mathbb{C}))^n$. Its justification in this last case is given in Section 4.4.

A particular and useful property of Herglotz functions defined on a scalar variable, which has been a key-tool to use analytic methods to derive bounds, is the following representation theorem.

Theorem 3. *A necessary and sufficient condition for a function $h : \mathbb{C}^+ \to \mathbb{C}$ to be a Herglotz function is that there exist $\alpha, \beta \in \mathbb{R}$ with $\alpha \geq 0$ and a positive regular Borel measure μ for which $\int_{\mathbb{R}} d\mu(\lambda)/(1+\lambda^2)$ is finite such that*

$$h(z) = \alpha\, z + \beta + \int_{\mathbb{R}} \left(\frac{1}{\lambda - z} - \frac{\lambda}{1+\lambda^2} \right) d\mu(\lambda), \ \textit{for } z \in \mathbb{C}^+. \tag{4.1}$$

For an extension of this representation theorem, for instance, in the case of matrix-valued Herglotz functions $h : \mathbb{C}^+ \to M_m(\mathbb{C})$, we refer to Gesztesy and Tsekanovskii (2000).

Theorem 4. *(Hartogs' Theorem) If $h : U \to E$ is a function with U an open subset of $\mathbb{C}^n$ and E is a Banach space then h is a multivariate analytic function (i.e., jointly analytic) if and only if it is an analytic function of each variable separately.*

A proof of Hartogs' Theorem when $E = \mathbb{C}$ can be found in Hörmander (1990) (see Section 2.2, p. 28, Theorem 2.2.8). For the general case, we refer the reader to Mujica (1986) (see Section 36, p. 265, Theorem 36.1).

Theorem 5. *Let E and F denote two Banach spaces and U an open subset of $\mathbb{C}^n$. If $h : U \to L(E, F)$ is an analytic function and for each $\mathbf{z} \in U$ the value $h(\mathbf{z})$ is an isomorphism, then the function $\mathbf{z} \to h(\mathbf{z})^{-1}$ is analytic from U into $L(F, E)$.*

For a proof of Theorem 5 when $n = 1$, we refer the reader to Kato (1995) (see **Chapter 7**, Section 1, pp. 365–366). The proof for an integer $n > 1$ is then obtained by using Hartogs' Theorem.

The next theorem, which is a rewriting of Theorem 3.12 of Kato (1995) shows that the notion of weak analyticity of a family of operators in $L(E, F)$ implies the analyticity of this family for the operator norm of $L(E, F)$. More precisely, we have the following result:

Theorem 6. *Let E and F be two Banach spaces, U an open subset of $\mathbb{C}$ and $h : U \to L(E, F)$. We denote by $\langle \cdot, \cdot \rangle$ the duality product of F and its dual F^*. If the function*

$$h_{\phi,\psi}(z) = \langle h(z)\phi, \psi \rangle \ , \forall z \in U,$$

is analytic on U for all ϕ in a dense subset of E and for all ψ in a dense subset of F^*, then h is analytic in U for the operator norm of $L(E, F)$.

The following is a theorem for taking the derivative under the integral of a function which depends analytically on a complex parameter (see Mattner 2001). It introduces the notion of a measure space that we briefly recall here. A measure space $(\Omega, \mathcal{F}, \mu)$ is roughly speaking a triple composed of a set Ω, a collection $\mathcal{F}$ of subsets of Ω that one wants to measure ($\mathcal{F}$ is called a $\sigma-$algebra) and a measure μ defined on $\mathcal{F}$.

Theorem 7. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let U be an open set of $\mathbb{C}$ and $f : \Omega \times U \to \mathbb{C}$ be a function subject to the following assumptions:*

- *$f(\cdot, z)$ is $\mathcal{F}$ measurable for all $z \in U$ and $f(\mathbf{x}, \cdot)$ is analytic for almost every $\mathbf{x}$ in Ω,*
- *$\int_\Omega |f(\mathbf{x}, \cdot)| \, \mathrm{d}\mu(\mathbf{x})$ is locally bounded, that is, for every $z_0 \in U$ there exists a $\delta > 0$ such that*

$$\sup_{z \in U \,||z - z_0| \leq \delta} \int_\Omega |f(\mathbf{x}, z)| \, \mathrm{d}\mu(\mathbf{x}) < \infty,$$

then the function $F : U \to \mathbb{C}$ defined by

$$F(z) = \int_\Omega f(\mathbf{x}, z) \, \mathrm{d}\mu(\mathbf{x}),$$

is analytic in U and one can take derivatives under the integral sign:

$$F^{(k)}(z) = \int_\Omega \frac{\partial^k f(\mathbf{x}, z)}{\partial z^k} \, \mathrm{d}\mu(\mathbf{x}), \ \forall k \in \mathbb{N}.$$

4.2 Analyticity of the DtN map for layered media

4.2.1 Formulation of the problem

We consider passive linear two-component layered media (material 1 with moduli $\boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1$; material 2 with moduli $\boldsymbol{\varepsilon}_2, \boldsymbol{\mu}_2$) with layers normal to the z-axis. The geometry of this problem, as illustrated in **Figure 4.1**, is as follows: First, a layered medium in the region $\Omega = \Omega_1 \cup \Omega_2 = [-d, d]$ consisting of a two-phase material lies between $z = -d$ and $z = d$. A homogeneous passive linear material lies between $-d_2 \leq z \leq d_2$ (denote this "inner" region by $\Omega_2 = [-d_2, d_2]$) with permittivity and permeability $\boldsymbol{\varepsilon}_2, \boldsymbol{\mu}_2$. Another homogeneous passive linear material lies between $-d \leq z < -d_2$, i.e., the region $\Omega_{1,-} = [-d, -d_2)$, and $d_2 < z \leq d$, i.e., the region $\Omega_{1,+} = (d_2, d]$ (denote "outer" region by $\Omega_1 = \Omega_{1,-} \cup \Omega_{1,+}$) with permittivity and permeability $\boldsymbol{\varepsilon}_1, \boldsymbol{\mu}_1$. The unit outward pointing normal vectors to the boundary surfaces of these regions are $\mathbf{n} \in \{\mathbf{e}_3, -\mathbf{e}_3\}$, where $\mathbf{e}_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.

The dielectric permittivity $\boldsymbol{\varepsilon}$ and magnetic permeability $\boldsymbol{\mu}$ are 3×3 matrices that depend on the frequency ω and the spatial variable z only (i.e., spatially homogeneous in each layer) which are defined by

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\omega, z) = \chi_{\Omega_1}(z)\boldsymbol{\varepsilon}_1(\omega) + \chi_{\Omega_2}(z)\boldsymbol{\varepsilon}_2(\omega), \quad z \in [-d, d], \ \omega \in \mathbb{C}^+, \tag{4.2}$$

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\omega, z) = \chi_{\Omega_1}(z)\boldsymbol{\mu}_1(\omega) + \chi_{\Omega_2}(z)\boldsymbol{\mu}_2(\omega), \quad z \in [-d, d], \ \omega \in \mathbb{C}^+. \tag{4.3}$$

Here χ_{Ω_j} denotes the indicator function of the region Ω_j, for $j = 1, 2$. Moreover, they have the passivity properties (see, for example, section 1.6)

$$\mathrm{Im}(\omega\boldsymbol{\varepsilon}(\omega, z)) > 0, \ \ \mathrm{Im}(\omega\boldsymbol{\mu}(\omega, z)) > 0, \ \text{for } \mathrm{Im}\,\omega > 0, \tag{4.4}$$

and $\boldsymbol{\varepsilon}, \boldsymbol{\mu}$ are analytic functions of ω in the complex upper-half plane for each fixed z, i.e.,

$$\omega\boldsymbol{\varepsilon}_j(\omega), \omega\boldsymbol{\mu}_j(\omega) : \mathbb{C}^+ \to M_3^+(\mathbb{C}) \text{ are Herglotz functions, for } j = 1, 2. \tag{4.5}$$

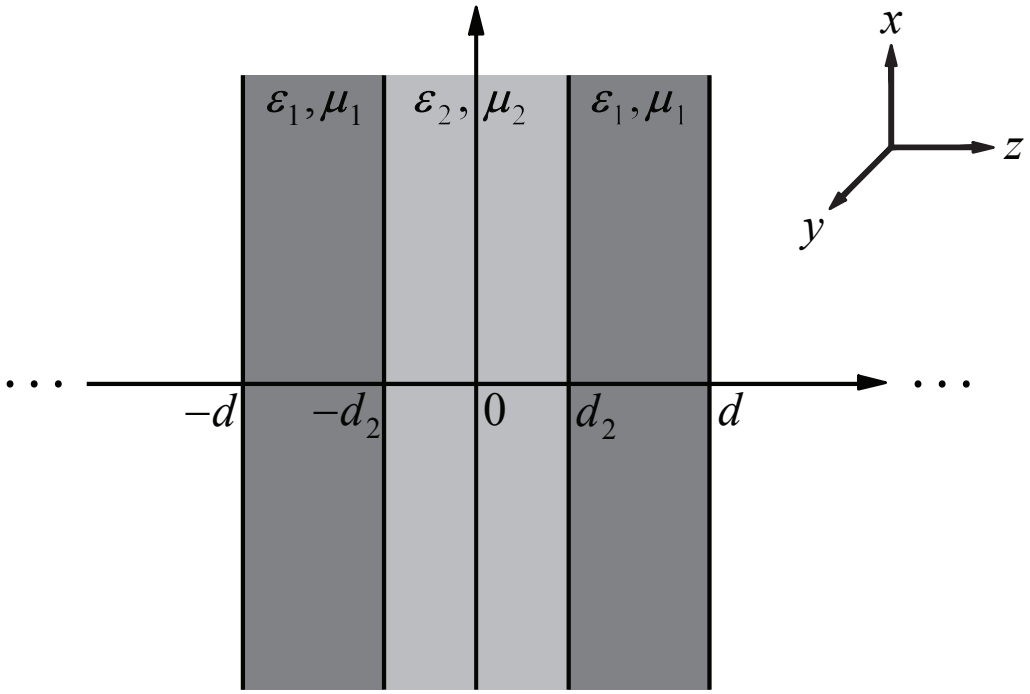


Figure 4.1: A plane-parallel, two-component layered medium Ω consisting of two phases, $\varepsilon_1, \boldsymbol{\mu}_1$ and $\varepsilon_2, \boldsymbol{\mu}_2$, of linear passive materials with layers normal to the z-axis. The core containing the homogeneous material 2 (with permittivity ε_2 and permeability $\boldsymbol{\mu}_2$) is sandwiched between the shell containing the homogeneous material 1 (with permittivity ε_1 and permeability $\boldsymbol{\mu}_1$). Moreover, the system is symmetric about the xy-plane.

The time-harmonic Maxwell's equations in Gaussian units without sources are

$$\operatorname{curl} \mathbf{E} = \frac{i\omega}{c}\mathbf{B}, \ \ \operatorname{curl} \mathbf{H} = -\frac{i\omega}{c}\mathbf{D}, \ \ \mathbf{D} = \boldsymbol{\varepsilon}\mathbf{E}, \ \ \mathbf{B} = \boldsymbol{\mu}\mathbf{H}, \tag{4.6}$$

where c denotes the speed of light in a vacuum.

Let us now introduce some classical functional spaces associated to the study of Maxwell's equations (4.6) in layered media:

- For a bounded interval $I \subseteq \mathbb{R}$, we denote by $L^1(I)$, the Lebesgue space of integrable functions on I. It is a Banach space with norm

$$||f||_1 = \int_I |f(z)|dz, \ f \in L^1(I). \tag{4.7}$$

- For a bounded interval $I \subseteq \mathbb{R}$, we denote by $AC(I)$, the Banach space of all absolutely continuous

functions equipped with the norm

$$||f||_{1,1}= \int_I |f(z)|dz + \int_I |f'(z)|dz,\ f \in AC(I). \tag{4.8}$$

Recall, that any $f \in AC(I)$ is continuous on I into $\mathbb{C}$, differentiable almost everywhere on I (i.e., except on a set of Lebesgue measure zero), and is given in terms of its derivative $f' = \frac{df}{dz}$ (which is integrable on I) by

$$f(z) = f(z_0) + \int_{z_0}^{z} f'(u)du,\ z_0, z \in I. \tag{4.9}$$

- Denote the Banach space of all $m \times n$ matrices with entries in the Banach space E with norm $||\cdot||$, where $(E, ||\cdot||) \in \{(L^1(I), ||\cdot||_1), (AC(I), ||\cdot||_{1,1}), (\mathbb{C}, |\cdot|)\}$, by $M_{m,n}(E)$ and equipped with norm

$$||\mathbf{M}||= \left(\sum_{i=1}^{m} \sum_{j=1}^{n} ||M_{ij}||^2 \right)^{\frac{1}{2}},\ \mathbf{M} = [M_{ij}] \in M_{m,n}(E), \tag{4.10}$$

  with $M_{n,n}(E)$ denoted by $M_n(E)$, and we treat E^n as $M_{n,1}(E)$.

- Similar to $AC(I)$, any $\mathbf{M} \in M_{m,n}(AC(I))$ is continuous on I, differentiable almost everywhere on I, and in terms of its derivative $\mathbf{M}' = \frac{d\mathbf{M}}{dz} = [M'_{ij}]$ is given by

$$\mathbf{M}(z) = \mathbf{M}(z_0) + \int_{z_0}^{z} \mathbf{M}'(u)du = \left[M_{ij}(z_0) + \int_{z_0}^{z} M'_{ij}(u)du \right],\ z_0, z \in I. \tag{4.11}$$

- Denote the standard inner product on $\mathbb{C}^n$ by $(\cdot,\cdot) : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$, where

$$(\psi_1, \psi_2) = \psi_1^T \overline{\psi_2},\ \ \psi_1, \psi_2 \in \mathbb{C}^n. \tag{4.12}$$

Now, because of the translation invariance of the layered media in the x, y coordinates, solutions of equation (4.6) are sought in the form

$$\begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{E}(z) \\ \mathbf{H}(z) \end{bmatrix} e^{i(k_1 x + k_2 y)},\ x, y \in \mathbb{R},\ z \in [-d, d],\ \boldsymbol{\kappa} = (k_1, k_2) \in \mathbb{C}^2,\ \omega \in \mathbb{C}^+, \tag{4.13}$$

in which $\boldsymbol{\kappa}$ is the tangential wavevector. Maxwell's equations (4.6) for this type of solution can be reduced [see the appendix in Shipman and Welters (2013) and also Berreman (1972) for more details] to an ordinary linear differential equation (ODE) for the vector of tangential electric and magnetic field components $\boldsymbol{\psi}$, where

$$\boldsymbol{\psi}(z) = \begin{bmatrix} E_1(z) & E_2(z) & H_1(z) & H_2(z) \end{bmatrix}^T, \tag{4.14}$$

$$-i\mathbf{J}\frac{d\boldsymbol{\psi}}{dz} = \mathbf{A}(z)\boldsymbol{\psi}(z),\ \ \boldsymbol{\psi} \in (AC([-d, d]))^4, \tag{4.15}$$

in which

$$\mathbf{J} = \begin{bmatrix} 0 & \boldsymbol{\rho} \\ \boldsymbol{\rho}^* & 0 \end{bmatrix},\ \ \boldsymbol{\rho} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},\ \ \mathbf{J}^* = \mathbf{J}^{-1} = \mathbf{J}, \tag{4.16}$$

$$\mathbf{A} = \mathbf{A}(z) = \mathbf{A}(z, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega)),\ z \in [-d, d],\ \boldsymbol{\kappa} \in \mathbb{C}^2,\ \omega \in \mathbb{C}^+, . \tag{4.17}$$

Here $\mathbf{A} = \mathbf{A}(z)$ is a piecewise constant function of z into $M_4(\mathbb{C})$ (for fixed $\boldsymbol{\kappa}, \omega$) with the following explicit representation in terms of the entries of the matrices $\boldsymbol{\varepsilon} = [\varepsilon_{ij}]$, $\boldsymbol{\mu} = [\mu_{ij}]$ in (4.2), (4.3):

$$\mathbf{A} = \mathbf{V}_{\|\|} - \mathbf{V}_{\|\perp} \left(\mathbf{V}_{\perp\perp}\right)^{-1} \mathbf{V}_{\perp\|}, \tag{4.18}$$

where

$$\mathbf{V}_{\perp\perp} = \tfrac{1}{c} \begin{bmatrix} \omega\varepsilon_{33} & 0 \\ 0 & \omega\mu_{33} \end{bmatrix}, \tag{4.19}$$

$$\mathbf{V}_{\|\|} = \tfrac{1}{c} \begin{bmatrix} \omega\varepsilon_{11} & \omega\varepsilon_{12} & 0 & 0 \\ \omega\varepsilon_{21} & \omega\varepsilon_{22} & 0 & 0 \\ 0 & 0 & \omega\mu_{11} & \omega\mu_{12} \\ 0 & 0 & \omega\mu_{21} & \omega\mu_{22} \end{bmatrix}, \tag{4.20}$$

$$\mathbf{V}_{\|\perp} = \tfrac{1}{c} \begin{bmatrix} \omega\varepsilon_{13} & 0 \\ \omega\varepsilon_{23} & 0 \\ 0 & \omega\mu_{13} \\ 0 & \omega\mu_{23} \end{bmatrix} + \begin{bmatrix} 0 & k_2 \\ 0 & -k_1 \\ -k_2 & 0 \\ k_1 & 0 \end{bmatrix}, \tag{4.21}$$

$$\mathbf{V}_{\perp\|} = \tfrac{1}{c} \begin{bmatrix} \omega\varepsilon_{31} & \omega\varepsilon_{32} & 0 & 0 \\ 0 & 0 & \omega\mu_{31} & \omega\mu_{32} \end{bmatrix} + \begin{bmatrix} 0 & 0 & -k_2 & k_1 \\ k_2 & -k_1 & 0 & 0 \end{bmatrix}. \tag{4.22}$$

From these matrices the normal electric and magnetic field components $\boldsymbol{\phi}$ are given in terms of their tangential components by

$$\boldsymbol{\phi} = \begin{bmatrix} E_3 & H_3 \end{bmatrix}^T = -(\mathbf{V}_{\perp\perp})^{-1} \mathbf{V}_{\perp\|} \boldsymbol{\psi}. \tag{4.23}$$

The fact that the matrix $\mathbf{V}_{\perp\perp}(z, \omega)$ is invertible follows immediately from the fact that the passivity properties (4.4) imply

$$\mathrm{Im}(\mathbf{V}_{\perp\perp}(z, \omega)) > 0. \tag{4.24}$$

We will now prove in the next proposition [using the methods developed in the appendix of Shipman and Welters (2013) and in the Ph.D. thesis of Welters (2011b)], some fundamental properties associated to the ODE (4.15). In particular, we will show that the solution of the initial-valued problem for the ODE (4.15) depends analytically on the phase moduli.

Proposition 8. *For each $z_0 \in [-d, d]$ (and for fixed $\boldsymbol{\kappa}, \omega$), the initial-value problem for the ODE (4.15), i.e.,*

$$-i\mathbf{J}\frac{d\boldsymbol{\psi}}{dz} = \mathbf{A}(z)\boldsymbol{\psi}(z), \quad \boldsymbol{\psi}(z_0) = \boldsymbol{\psi}_0, \tag{4.25}$$

has a unique solution $\boldsymbol{\psi}$ in $(AC([-d, d]))^4$ for each $\boldsymbol{\psi}_0 \in \mathbb{C}^4$ which is given by

$$\boldsymbol{\psi}(z) = \mathbf{T}(z_0, z)\boldsymbol{\psi}_0, \quad z \in [-d, d], \tag{4.26}$$

where the 4×4 matrix $\mathbf{T}(z_0, z)$ is called the transfer matrix. This transfer matrix $\mathbf{T}$ has the properties

$$\mathbf{T}(z_0, z) = \mathbf{T}(z_1, z)\mathbf{T}(z_0, z_1), \quad \mathbf{T}(z_0, z_1)^{-1} = \mathbf{T}(z_1, z_0), \quad \mathbf{T}(z_0, z_0) = \mathbf{I}, \tag{4.27}$$

for all $z_0, z_1, z \in [-d, d]$. Furthermore, the map

$$\begin{aligned}\mathbf{T} &= \mathbf{T}(z_0, z) = \mathbf{T}(z_0, z, \boldsymbol{\kappa}, \omega) \\ &= \mathbf{T}(z_0, z, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega)),\ z_0, z \in [-d, d],\ \boldsymbol{\kappa} \in \mathbb{C}^2,\ \omega \in \mathbb{C}^+,\end{aligned} \tag{4.28}$$

belongs to $M_4(AC([-d, d]))$ as a function of z (for fixed $z_0, \boldsymbol{\kappa}, \omega$) and it is an analytic function as a map of $(\boldsymbol{\kappa}, \omega)$ into $M_4(\mathbb{C})$ (for fixed z_0, z). More generally, the map

$$\mathbf{Z} \mapsto \mathbf{T}(z_0, z, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2) \tag{4.29}$$

is analytic as a function of $\mathbf{Z} = (\omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2) \in (M_3^+(\mathbb{C}))^4$ into $M_4(\mathbb{C})$ (for fixed $z_0, z, \boldsymbol{\kappa}$).

Proof. First, it follows from Hartogs' Theorem (see Theorem 4), the hypotheses (4.2), (4.3), (4.5), the formulas (4.17)–(4.22), and Theorem 5 that

$$(\boldsymbol{\kappa}, \omega) \mapsto \mathbf{A}(\cdot, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega)) \tag{4.30}$$

is analytic as a function into $M_4(L^1(I))$, where $I = [-d, d]$, from $\mathbb{C}^2 \times \mathbb{C}^+$. And, more generally, it follows from these theorems, hypotheses, and formulas that the map

$$(\boldsymbol{\kappa}, \mathbf{Z}) \mapsto \mathbf{A}(\cdot, \boldsymbol{\kappa}, \mathbf{Z}) \tag{4.31}$$

is analytic as a function of $(\boldsymbol{\kappa}, \mathbf{Z}) \in \mathbb{C}^2 \times (M_3^+(\mathbb{C}))^4$ into $M_4(L^1(I))$, where $\mathbf{Z}$ is the variable $\mathbf{Z} = (\omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2)$.

In particular, for either fixed variables $(\boldsymbol{\kappa}, \omega) \in \mathbb{C}^2 \times \mathbb{C}^+$ or $(\boldsymbol{\kappa}, \mathbf{Z}) \in \mathbb{C}^2 \times (M_3^+(\mathbb{C}))^4$, we have $\mathbf{A} = \mathbf{A}(z)$ from (4.17) is in $M_4(L^1(I))$. Fix a $z_0 \in I$. Then by the theory of linear ordinary differential equations [see, for instance, Theorem 1.2.1 in **Chapter 1** of Zettl (2005)], the initial-value problem (4.25) has a unique solution $\boldsymbol{\psi}$ in $(AC(I))^4$ for each $\boldsymbol{\psi}_0 \in \mathbb{C}^4$. Denote the standard orthonormal basis vectors of $\mathbb{R}^4$ by $\mathbf{w}_j$, for $j = 1, 2, 3, 4$. Let $\boldsymbol{\psi}_j \in (AC(I))^4$ denote the unique solution of the ODE (4.15) satisfying $\boldsymbol{\psi}_j(z_0) = \mathbf{w}_j$, for $j = 1, 2, 3, 4$. Now let $\mathbf{T}(z_0, z) = [\boldsymbol{\psi}_1(z)|\boldsymbol{\psi}_2(z)|\boldsymbol{\psi}_3(z)|\boldsymbol{\psi}_4(z)] \in M_4(\mathbb{C})$ denote the 4×4 matrix whose columns are $\mathbf{T}(z_0, z)\mathbf{w}_j = \boldsymbol{\psi}_j(z)$ for $j = 1, 2, 3, 4$ and $z \in I$. This matrix $\mathbf{T}(z_0, z)$ is known in the electrodynamics of layered media as the transfer matrix.

Now it follows immediately from the uniqueness of the solution to the initial-value problem (4.25) and the definition of the transfer matrix $\mathbf{T}(z_0, z)$, that $\mathbf{T} = \mathbf{T}(z_0, z)$ as a function of $z \in I$ belongs to $M_4(AC(I))$, it has the properties (4.27), and it is the unique matrix-valued function in $M_4(AC(I))$ satisfying: if $\boldsymbol{\psi}_0 \in \mathbb{C}^4$ then $\boldsymbol{\psi}(z) = \mathbf{T}(z_0, z)\boldsymbol{\psi}_0$ for all $z \in I$ is an $(AC(I))^4$ solution to the initial-value problem (4.25). From this uniqueness property of the transfer matrix $\mathbf{T}(z_0, z)$, it follows that $\mathbf{T}(z_0, z)$ is the unique solution to the initial-value problem:

$$\boldsymbol{\Psi}'(z) = i\mathbf{J}^{-1}\mathbf{A}(z)\boldsymbol{\Psi}(z),\ \boldsymbol{\Psi}(z_0) = \mathbf{I},\ \boldsymbol{\Psi} \in M_4(AC(I)), \tag{4.32}$$

where $\mathbf{I} \in M_4(\mathbb{C})$ is the identity matrix.

Now we wish to derive an explicit representation for $\mathbf{T}(z_0, z)$ in terms of $\mathbf{J}$ and $\mathbf{A}$. To do this we first introduce some results from the integral operator approach to the theory of linear ODEs. For fixed $\mathbf{M} \in M_4(L^1(I))$, define the linear map $\boldsymbol{\mathcal{I}}[\mathbf{M}, z_0] : M_4(AC(I)) \to M_4(AC(I))$ by

$$(\boldsymbol{\mathcal{I}}[\mathbf{M}, z_0]\mathbf{N})(z) = \int_{z_0}^{z} \mathbf{M}(u)\mathbf{N}(u)du,\ \mathbf{N} \in M_4(AC(I)),\ z \in I. \tag{4.33}$$

It follows that $\boldsymbol{\mathcal{I}}[\mathbf{M}, z_0]$ is a continuous linear operator on the Banach space $M_4(AC(I))$, i.e., it belongs to $L(M_4(AC(I)), M_4(AC(I)))$. Next, define the linear map $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0] : M_4(AC(I)) \to M_4(AC(I))$ by

$$\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0] = \mathbf{1} - \boldsymbol{\mathcal{I}}[\mathbf{M}, z_0], \tag{4.34}$$

where $\mathbf{1} \in L(M_4(AC(I)), M_4(AC(I)))$ denotes the identity operator on $M_4(AC(I))$. Then it follows that $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0] \in L(M_4(AC(I)), M_4(AC(I)))$ and, moreover, $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]$ is invertible with $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]^{-1} \in L(M_4(AC(I)), M_4(AC(I)))$, i.e., $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]$ is an isomorphism. The fact that $\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]$ is invertible follows immediately from the existence and uniqueness of the solution $\mathbf{Y} \in M_4(AC(I))$ for each $\mathbf{C} \in M_4(\mathbb{C})$, $\mathbf{F} \in M_4(L^1(I))$ to the inhomogeneous initial-value problem [see, for instance, Theorem 1.2.1 in **Chapter 1** of Zettl (2005)]:

$$\mathbf{Y}'(z) = \mathbf{M}(z)\mathbf{Y}(z) + \mathbf{F}(z), \ \mathbf{Y}(z_0) = \mathbf{C}. \tag{4.35}$$

In other words, $\mathbf{Y}$ is the unique solution in $M_4(AC(I))$ to the integral equation

$$\boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]\mathbf{Y} = \boldsymbol{\mathcal{I}}[\mathbf{M}, z_0]\mathbf{F} + \mathbf{C}. \tag{4.36}$$

Hence, the solution is given explicitly by

$$\mathbf{Y} = \boldsymbol{\mathcal{T}}[\mathbf{M}, z_0]^{-1}(\boldsymbol{\mathcal{I}}[\mathbf{M}, z_0]\mathbf{F} + \mathbf{C}). \tag{4.37}$$

In particular, it follows from this representation and the fact that the transfer matrix $\mathbf{T}(z_0, z)$ is the unique solution to the initial-value problem (4.32) that with $\mathbf{F} = \mathbf{0}$, $\mathbf{C} = \mathbf{I}$ in the notation above,

$$\mathbf{T}(z_0, \cdot) = \boldsymbol{\mathcal{T}}[i\mathbf{J}^{-1}\mathbf{A}, z_0]^{-1}(\mathbf{I}), \tag{4.38}$$

where $\mathbf{A} = \mathbf{A}(z)$ as you will recall belongs to $M_4(L^1(I))$ as a function of $z \in I$ (ignoring its dependence on the other variables) and hence so does $i\mathbf{J}^{-1}\mathbf{A}$.

Now since $i\mathbf{J}^{-1}\mathbf{A}$ is an analytic function of either of the variables $(\boldsymbol{\kappa}, \omega)$ or $(\boldsymbol{\kappa}, \mathbf{Z})$ into $M_4(L^1(I))$ as a function of $z \in I$, for fixed z_0, then it follows immediately from this, the representation (4.38), and Theorem 5 that $(\boldsymbol{\kappa}, \omega) \mapsto \mathbf{T}(z_0, z, \boldsymbol{\kappa}, \omega)$ and $(\boldsymbol{\kappa}, \mathbf{Z}) \mapsto \mathbf{T}(z_0, z, \boldsymbol{\kappa}, \mathbf{Z})$ are analytic functions into $M_4(AC(I))$ as a function of $z \in I$, for fixed z_0. Finally, the proof of the rest of this proposition now follows immediately from these facts and the fact that the Banach space $AC(I)$ can be continuously embedded into the Banach space $C(I)$ of continuous functions from I into $\mathbb{C}$ equipped with the sup norm $||f||_\infty = \sup_{z \in I}|f(z)|$, that is, the identity map $\iota : AC(I) \to C(I)$ between these two Banach spaces [i.e., $\iota(f) = f$ for $f \in AC(I)$] is a continuous (and hence bounded) linear map under their respective norms [i.e., $\iota \in L(AC(I), C(I))$]. □

Remark 9. *Using Proposition 8 and due to the simplicity of the layered media considered we can derive a simple explicit representation of the transfer matrix $\mathbf{T}(z_0, z)$ for all $z_0, z \in [-d, d]$. First, the transfer matrix $\mathbf{T}(-d, z)$, $z \in [-d, d]$ takes on the simple form*

$$\mathbf{T}(-d, z) = \begin{cases} e^{i\mathbf{J}\mathbf{A}_1(z+d)}, & -d \le z \le -d_2, \\ e^{i\mathbf{J}\mathbf{A}_1(d-d_2)}e^{i\mathbf{J}\mathbf{A}_2(z+d_2)}, & -d_2 \le z \le d_2, \\ e^{i\mathbf{J}\mathbf{A}_1(d-d_2)}e^{i\mathbf{J}\mathbf{A}_2(2d_2)}e^{i\mathbf{J}\mathbf{A}_1(z-d_2)}, & d_2 \le z \le d, \end{cases} \tag{4.39}$$

*where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the matrices (4.17) for a z-independent homogeneous medium filled with only material 1 (with permittivity and permeability $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\mu}_1$) and with only material 2 (with permittivity and permeability $\boldsymbol{\varepsilon}_2$ and $\boldsymbol{\mu}_2$), respectively (see **Figure 4.1**).*

Therefore, in terms of this explicit form for $\mathbf{T}(-d,z)$*, it follows from (4.27) that the transfer matrix* $\mathbf{T}(z_0,z)$*,* $z_0,z\in[-d,d]$ *is given explicitly in terms of (4.39) by*

$$\mathbf{T}(z_0,z)=\mathbf{T}(-d,z)\mathbf{T}(z_0,-d)=\mathbf{T}(-d,z)\mathbf{T}(-d,z_0)^{-1}. \tag{4.40}$$

4.2.2 Electromagnetic Dirichlet-to-Neumann Map

Now every solution to Maxwell's equations (4.6) of the form (4.13) has in terms of its tangential components (4.14) a corresponding solution of the ODE (4.15) with normal components given by (4.23). And conversely, every solution of the ODE (4.15) gives the tangential components of a unique solution to equations (4.6) of the form (4.13) with normal components expressed in terms of its tangential components by (4.23). We use this correspondence to now define the electromagnetic "Dirichlet-to-Neumann" (DtN) map in terms of the transfer matrix $\mathbf{T}$ whose properties are described in Proposition 8.

The DtN map is a function

$$\begin{gathered}\Lambda=\Lambda(z_0,z_1)=\Lambda(z_0,z_1,\boldsymbol{\kappa},\omega\boldsymbol{\varepsilon}_1(\omega),\omega\boldsymbol{\varepsilon}_2(\omega),\omega\boldsymbol{\mu}_1(\omega),\omega\boldsymbol{\mu}_2(\omega)),\\ z_0,z_1\in[-d,d],\ z_0<z_1,\boldsymbol{\kappa}\in\mathbb{C}^2,\ \omega\in\mathbb{C}^+,\end{gathered} \tag{4.41}$$

which can be defined as the block operator matrix

$$\Lambda(z_0,z_1)\begin{bmatrix}\mathbf{E}\times\mathbf{n}|_{z=z_1}\\ \mathbf{E}\times\mathbf{n}|_{z=z_0}\end{bmatrix}=\begin{bmatrix}i\mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_1}\\ i\mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_0}\end{bmatrix}, \tag{4.42}$$

where $\mathbf{E},\mathbf{H}$ denote a solution of the time-harmonic Maxwell's equations (4.6) of the form (4.13), i.e., a function of the form (4.13) whose tangential components $\boldsymbol{\psi}$ with the form (4.14) satisfy the ODE (4.15) and whose normal components are given in terms of these tangential components $\boldsymbol{\psi}$ by (4.23).

A more explicit definition of this DtN map can be given as follows. First, on $\mathbb{C}^3$, with respect to the standard orthonormal basis vectors, we have the matrix representations

$$\mathbf{e}_3\times=\begin{bmatrix}0&-1&0\\1&0&0\\0&0&0\end{bmatrix},\ -\mathbf{e}_3\times\mathbf{e}_3\times=\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}, \tag{4.43}$$

and this allows us to write $\mathbf{E}\times\mathbf{n}=-\mathbf{n}\times\mathbf{E}$ and $\mathbf{n}\times\mathbf{H}\times\mathbf{n}=-\mathbf{n}\times\mathbf{n}\times\mathbf{H}$ as matrix multiplication so that we can write Λ as a 6×6 matrix which can be written in the 2×2 block matrix form as

$$\Lambda=\begin{bmatrix}\Lambda_{11}&\Lambda_{12}\\ \Lambda_{21}&\Lambda_{22}\end{bmatrix}. \tag{4.44}$$

We now want to get an explicit expression of this block form. Thus, we define the projections

$$\mathbf{P}_{\mathfrak{t}}=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix},\ \mathbf{Q}_{\mathfrak{t},1}=\begin{bmatrix}1&0&0&0\\0&1&0&0\end{bmatrix},\ \mathbf{Q}_{\mathfrak{t},2}=\begin{bmatrix}0&0&1&0\\0&0&0&1\end{bmatrix}. \tag{4.45}$$

It follows from this notation that

$$\mathbf{E}\times\mathbf{e}_3=-e^{i(k_1x+k_2y)}\mathbf{e}_3\times\mathbf{P}_{\mathfrak{t}}\left[\mathbf{Q}_{\mathfrak{t},1}\boldsymbol{\psi}(z)\right],\ \ \mathbf{n}\times\mathbf{H}\times\mathbf{n}=e^{i(k_1x+k_2y)}\mathbf{P}_{\mathfrak{t}}\left[\mathbf{Q}_{\mathfrak{t},2}\boldsymbol{\psi}(z)\right].$$

Hence, we have

$$\begin{aligned}
\begin{bmatrix} i\mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_1} \\ i\mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_0} \end{bmatrix} &= ie^{i(k_1x+k_2y)} \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},2}\psi(z_1) \\ \mathbf{Q}_{\mathfrak{t},2}\psi(z_0) \end{bmatrix} \\
&= ie^{i(k_1x+k_2y)} \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix} \mathbf{\Gamma}(z_0,z_1) \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},1}\psi(z_1) \\ \mathbf{Q}_{\mathfrak{t},1}\psi(z_0) \end{bmatrix} \\
&= i \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix} \mathbf{\Gamma}(z_0,z_1) \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix}^T \begin{bmatrix} \mathbf{n}\times\mathbf{E}\times\mathbf{n}|_{z=z_1} \\ \mathbf{n}\times\mathbf{E}\times\mathbf{n}|_{z=z_0} \end{bmatrix} \\
&= i \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix} \mathbf{\Gamma}(z_0,z_1) \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix}^T \begin{bmatrix} \mathbf{e}_3\times & 0 \\ 0 & -\mathbf{e}_3\times \end{bmatrix} \begin{bmatrix} \mathbf{E}\times\mathbf{n}|_{z=z_1} \\ \mathbf{E}\times\mathbf{n}|_{z=z_0} \end{bmatrix},
\end{aligned}$$

where we have used the fact that since

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},1}\boldsymbol{\psi}(z_1) \\ \mathbf{Q}_{\mathfrak{t},2}\boldsymbol{\psi}(z_1) \end{bmatrix}, \quad \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},1}\boldsymbol{\psi}(z_0) \\ \mathbf{Q}_{\mathfrak{t},2}\boldsymbol{\psi}(z_0) \end{bmatrix}, \quad \mathbf{T}(z_0,z_1) \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix},$$

then by Proposition 13 (given later in Section 4.2.3) we must have

$$\mathbf{\Gamma}(z_0,z_1) \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},1}\boldsymbol{\psi}(z_1) \\ \mathbf{Q}_{\mathfrak{t},1}\boldsymbol{\psi}(z_0) \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{\mathfrak{t},2}\boldsymbol{\psi}(z_1) \\ \mathbf{Q}_{\mathfrak{t},2}\boldsymbol{\psi}(z_0) \end{bmatrix},$$

where $\mathbf{\Gamma}(z_0,z_1)$ is defined in (4.53) [which is well-defined provided the matrix $\mathbf{T}_{12}(z_0,z_1)$ in the block decomposition of $\mathbf{T}(z_0,z_1)$ in (4.55) is invertible]. Therefore, the DtN map $\Lambda(z_0,z_1)$ can be defined explicitly as follows.

Definition 10 (Electromagnetic Dirichlet-to-Neumann map)**.** *The electromagnetic DtN map $\Lambda(z_0,z_1)$ is defined to be the 6×6 matrix (4.44) defined in terms of the 4×4 matrix $\mathbf{\Gamma}(z_0,z_1)$ in (4.53) and the 3×2 matrix $\mathbf{P}_{\mathfrak{t}}$ in (4.45) by*

$$\Lambda(z_0,z_1) = i \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix} \mathbf{\Gamma}(z_0,z_1) \begin{bmatrix} \mathbf{P}_{\mathfrak{t}} & 0 \\ 0 & \mathbf{P}_{\mathfrak{t}} \end{bmatrix}^T \begin{bmatrix} \mathbf{e}_3\times & 0 \\ 0 & -\mathbf{e}_3\times \end{bmatrix}, \tag{4.46}$$

and in the 2×2 block matrix form its entries are the 3×3 matrices

$$\Lambda_{11}(z_0,z_1) = i\mathbf{P}_{\mathfrak{t}}\mathbf{\Gamma}_{11}(z_0,z_1)\mathbf{P}_{\mathfrak{t}}^T\mathbf{e}_3\times, \tag{4.47}$$

$$\Lambda_{12}(z_0,z_1) = -i\mathbf{P}_{\mathfrak{t}}\mathbf{\Gamma}_{12}(z_0,z_1)\mathbf{P}_{\mathfrak{t}}^T\mathbf{e}_3\times, \tag{4.48}$$

$$\Lambda_{21}(z_0,z_1) = i\mathbf{P}_{\mathfrak{t}}\mathbf{\Gamma}_{21}(z_0,z_1)\mathbf{P}_{\mathfrak{t}}^T\mathbf{e}_3\times, \tag{4.49}$$

$$\Lambda_{22}(z_0,z_1) = -i\mathbf{P}_{\mathfrak{t}}\mathbf{\Gamma}_{22}(z_0,z_1)\mathbf{P}_{\mathfrak{t}}^T\mathbf{e}_3\times, \tag{4.50}$$

where $\mathbf{e}_3\times$ is the 3×3 matrix in (4.43).

Now for any $z_0, z_1 \in [-d,d]$, $z_0 < z_1$, we want to know whether the DtN map $\Lambda(z_0,z_1)$ is well-defined or not. The next theorem addresses this.

Theorem 11. *If $\operatorname{Im}\omega > 0$ and $\boldsymbol{\kappa} \in \mathbb{R}^2$ then for any 3×3 matrix-valued Herglotz functions $\omega\boldsymbol{\varepsilon}_j(\omega)$, $\omega\boldsymbol{\mu}_j(\omega)$, $j = 1,2$ with range in $M_3^+(\mathbb{C})$, the electromagnetic DtN map $\Lambda(z_0,z_1,\boldsymbol{\kappa},\omega\boldsymbol{\varepsilon}_1(\omega),\omega\boldsymbol{\varepsilon}_2(\omega),\omega\boldsymbol{\mu}_1(\omega),\omega\boldsymbol{\mu}_2(\omega))$ is well-defined.*

Proof. Let $\omega\boldsymbol{\varepsilon}_j(\omega)$, $\omega\boldsymbol{\mu}_j(\omega)$, $j = 1, 2$ be any 3×3 matrix-valued Herglotz functions with range in $M_3^+(\mathbb{C})$. Choose any values $\omega \in \mathbb{C}$ and $\boldsymbol{\kappa}$ with $\operatorname{Im}\omega > 0$ and $\boldsymbol{\kappa} \in \mathbb{R}^2$. Consider the time-harmonic Maxwell's equations (4.6) for the plane parallel layered media in **Figure 4.1** at the frequency ω for solutions of the form (4.13) with tangential wavevector $\boldsymbol{\kappa}$, where the dielectric permittivity $\boldsymbol{\varepsilon}$ and magnetic permeability $\boldsymbol{\mu}$ are defined in (4.2) and (4.3).

For $z_0, z_1 \in [-d, d]$ with $z_0 < z_1$, the transfer matrix (defined in Section 4.2.1) of the layered media with tensors $\boldsymbol{\varepsilon}(z, \omega)$, $\boldsymbol{\mu}(z, \omega)$ is $\mathbf{T}(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$. For simplicity we will suppress the dependency on the other parameters and denote this transfer matrix by $\mathbf{T}(z_0, z_1)$. It now follows from the passivity property (4.4) and Theorem 15, given below, that the matrix $\mathbf{J} - \mathbf{T}(z_0, z_1)^*\mathbf{J}\mathbf{T}(z_0, z_1)$ is positive definite. By Proposition 14, given below, it follows that the 2×2 matrices $\mathbf{T}_{ij}(z_0, z_1)$, $1 \leq i, j \leq 2$, that make up the blocks for the transfer matrix $\mathbf{T}(z_0, z_1)$ in the 2×2 block form in (4.52), are invertible. It follows from this that the matrix $\boldsymbol{\Gamma}(z_0, z_1)$ defined in (4.53) terms of these 2×2 matrices is well-defined. And therefore it follows from the fact that $\boldsymbol{\Gamma}(z_0, z_1)$ is well-defined that the electromagnetic DtN map $\Lambda(z_0, z_1) = \Lambda(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$, as given in Definition 10, is well-defined. This completes the proof. □

The main result of this section on the analytic properties of the DtN map is the following:

Theorem 12. *For any $\boldsymbol{\kappa} \in \mathbb{R}^2$ and any 3×3 matrix-valued Herglotz functions $\omega\boldsymbol{\varepsilon}_j(\omega)$, $\omega\boldsymbol{\mu}_j(\omega)$, $j = 1, 2$ with range in $M_3^+(\mathbb{C})$, the function*

$$\omega \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega)) \tag{4.51}$$

is analytic from $\mathbb{C}^+$ into $M_6^+(\mathbb{C})$ and, in particular, it is a matrix-valued Herglotz function. More generally, it is a Herglotz function in the variable $\mathbf{Z} = (\omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2) \in (M_3^+(\mathbb{C}))^4$ (see Definition 2).

Proof. Fix any 3×3 matrix-valued Herglotz functions $\omega\boldsymbol{\varepsilon}_j(\omega), \omega\boldsymbol{\mu}_j(\omega)$, $j = 1, 2$ with range in $M_3^+(\mathbb{C})$. Then for any electromagnetic field $\mathbf{E}, \mathbf{B}$ with tangential components $\boldsymbol{\psi}$ with $\operatorname{Im}\omega > 0$ and tangential wavevector $\boldsymbol{\kappa} \in \mathbb{R}^2$ we have, by Theorem 15 and its proof, that

$$\begin{aligned}
\left(\begin{bmatrix}\mathbf{E}\times\mathbf{n}|_{z=z_1}\\ \mathbf{E}\times\mathbf{n}|_{z=z_0}\end{bmatrix}, \operatorname{Im}\left[\Lambda\left(z_0, z_1\right)\right]\begin{bmatrix}\mathbf{E}\times\mathbf{n}|_{z=z_1}\\ \mathbf{E}\times\mathbf{n}|_{z=z_0}\end{bmatrix}\right) &= \operatorname{Re}\left(\begin{bmatrix}\mathbf{E}\times\mathbf{n}|_{z=z_1}\\ \mathbf{E}\times\mathbf{n}|_{z=z_0}\end{bmatrix}, \begin{bmatrix}\mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_1}\\ \mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_0}\end{bmatrix}\right)\\
&= \operatorname{Re}\left\{(\mathbf{E}\times\mathbf{n}|_{z=z_1}, \mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_1}) + (\mathbf{E}\times\mathbf{n}|_{z=z_0}, \mathbf{n}\times\mathbf{H}\times\mathbf{n}|_{z=z_0})\right\}\\
&= -\frac{1}{2}\left(\boldsymbol{\psi}\left(z_1\right), \mathbf{J}\boldsymbol{\psi}\left(z_1\right)\right) + \frac{1}{2}\left(\boldsymbol{\psi}\left(z_0\right), \mathbf{J}\boldsymbol{\psi}\left(z_0\right)\right)\\
&= \frac{1}{c}\int_{z_0}^{z_1}\left(\mathbf{H}, \operatorname{Im}\left[\omega\boldsymbol{\mu}\left(z, \omega\right)\right]\mathbf{H}\right) + \left(\mathbf{E}, \operatorname{Im}\left[\omega\boldsymbol{\varepsilon}\left(z, \omega\right)\right]\mathbf{E}\right) dz \geq 0,
\end{aligned}$$

with equality if and only if $\boldsymbol{\psi} \equiv 0$. It now follows from this and Theorem 15, which tells us that $\mathbf{J} - \mathbf{T}(z_0, z_1)^*\mathbf{J}\mathbf{T}(z_0, z_1)$ is positive definite, that we must have $\operatorname{Im}\Lambda(z_0, z_1) > 0$.

We will now prove that the function $\omega \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$ is analytic from $\mathbb{C}^+$ into $M_6^+(\mathbb{C})$. By Proposition 8 we know that the map

$$\omega \mapsto \mathbf{T}(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$$

is an analytic function into $M_4(\mathbb{C})$. This implies by (4.53), (4.54) and Theorem 6 that

$$\omega \mapsto \mathbf{\Gamma}(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$$

is an analytic function into $M_4(\mathbb{C})$ and so by (4.46) it follows that

$$\omega \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$$

is an analytic function into $M_6^+(\mathbb{C})$.

Now we introduce the variable $\mathbf{Z} = (\omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2) \in (M_3^+(\mathbb{C}))^4$. Here $M_3^+(\mathbb{C})$ is an open, connected, and convex subset of $M_3(\mathbb{C})$ as a Banach space in any normed topology (as all norms on a finite-dimensional vector space are equivalent) and hence so is $(M_3^+(\mathbb{C}))^4$ as a subset of $(M_3(\mathbb{C}))^4$. Our goal is to prove that the function $\mathbf{Z} \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \mathbf{Z})$ is analytic. Now as $(M_3(\mathbb{C}))^4$ equipped with any norm is a Banach space and is isomorphic to the Banach space $\mathbb{C}^{36}$ (by mapping the components of the 4-tuple and their matrix entries to a 36-tuple) equipped with standard inner product on $\mathbb{C}^{36}$. Thus, by Theorem 4 (Hartogs' Theorem) it suffices to prove that for each component Z_j of $\mathbf{Z}$ as an element of $\mathbb{C}^{36}$, the function $Z_j \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \mathbf{Z})$ is analytic for all other components of $\mathbf{Z} \in (M_3^+(\mathbb{C}))^4$ fixed. But this proof follows exactly as we did for proving $\omega \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega))$ is an analytic function into $M_6^+(\mathbb{C})$. Therefore, $\mathbf{Z} \mapsto \Lambda(z_0, z_1, \boldsymbol{\kappa}, \mathbf{Z})$ is analytic. This completes the proof. □

4.2.3 Auxiliary results

In this section we will derive some auxiliary results that are used in the preceding subsection. First, we write the transfer matrix $\mathbf{T}(z_0, z_1)$ in the 2×2 block matrix form

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \tag{4.52}$$

with respect to the decomposition $\mathbb{C}^4 = \mathbb{C}^2 \oplus \mathbb{C}^2$. We next define the 4×4 matrix $\mathbf{\Gamma}(z_0, z_1)$ in the 2×2 block matrix form by

$$\mathbf{\Gamma}(z_0, z_1) = \begin{bmatrix} \mathbf{\Gamma}_{11}(z_0, z_1) & \mathbf{\Gamma}_{12}(z_0, z_1) \\ \mathbf{\Gamma}_{21}(z_0, z_1) & \mathbf{\Gamma}_{22}(z_0, z_1) \end{bmatrix} \tag{4.53}$$

$$= \begin{bmatrix} \mathbf{T}_{22}(z_0, z_1)\mathbf{T}_{12}(z_0, z_1)^{-1} & \mathbf{T}_{21}(z_0, z_1) - \mathbf{T}_{22}(z_0, z_1)\mathbf{T}_{12}(z_0, z_1)^{-1}\mathbf{T}_{11}(z_0, z_1) \\ \mathbf{T}_{12}(z_0, z_1)^{-1} & -\mathbf{T}_{12}(z_0, z_1)^{-1}\mathbf{T}_{11}(z_0, z_1) \end{bmatrix}, \tag{4.54}$$

provided $\mathbf{T}_{12}(z_0, z_1)$ is invertible.

Let us now give an overview of the purpose of the results in this section. Using the next proposition, Proposition 13, we are able to give an explicit formula for the DtN map $\Lambda(z_0, z_1)$ in terms of the transfer matrix $\mathbf{T}(z_0, z_1)$ using the matrix $\mathbf{\Gamma}(z_0, z_1)$, the latter of which requires the invertibility of the matrix $\mathbf{T}_{12}(z_0, z_1)$. The proposition which follows after this one, i.e., Proposition 14, then tells us that the matrix $\mathbf{T}_{12}(z_0, z_1)$ is invertible, provided the matrix $\mathbf{J} - \mathbf{T}(z_0, z_1)^*\mathbf{J}\mathbf{T}(z_0, z_1)$ is positive definite. And, finally, Theorem 15 tells us that this matrix is positive definite (due to passivity).

Proposition 13. *If $\mathbf{T}_{12}(z_0, z_1)$ is invertible then for any $\mathbf{u}_0, \mathbf{u}_1 \in \mathbb{C}^2$ there exist unique $\mathbf{v}_0, \mathbf{v}_1 \in \mathbb{C}^2$ satisfying*

$$\mathbf{T}(z_0, z_1) \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix}. \tag{4.55}$$

These unique vectors $\mathbf{v}_0, \mathbf{v}_1$ *are given explicitly in terms of the vectors* $\mathbf{u}_0, \mathbf{u}_1$ *by the formula*

$$\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_0 \end{bmatrix} = \mathbf{\Gamma}(z_0, z_1) \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}. \tag{4.56}$$

Proof. Assume $\mathbf{T}_{12}(z_0, z_1)$ is invertible. Let $\mathbf{u}_0, \mathbf{u}_1 \in \mathbb{C}^2$. Then we have

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{v}_1 \end{bmatrix} = \mathbf{T}(z_0, z_1) \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11}(z_0, z_1)\mathbf{u}_0 + \mathbf{T}_{12}(z_0, z_1)\mathbf{v}_0 \\ \mathbf{T}_{21}(z_0, z_1)\mathbf{u}_0 + \mathbf{T}_{22}(z_0, z_1)\mathbf{v}_0 \end{bmatrix}$$

if and only if

$$\begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{T}_{22}(z_0, z_1) \\ 0 & \mathbf{T}_{12}(z_0, z_1) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{T}_{11}(z_0, z_1) \\ 0 & \mathbf{T}_{21}(z_0, z_1) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix},$$

and this holds if and only if

$$\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{T}_{22}(z_0, z_1)\mathbf{T}_{12}(z_0, z_1)^{-1} \\ 0 & \mathbf{T}_{12}(z_0, z_1)^{-1} \end{bmatrix} \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{T}_{11}(z_0, z_1) \\ 0 & \mathbf{T}_{21}(z_0, z_1) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{T}_{22}(z_0, z_1)\mathbf{T}_{12}(z_0, z_1)^{-1} & \mathbf{T}_{21}(z_0, z_1) - \mathbf{T}_{22}(z_0, z_1)\mathbf{T}_{12}(z_0, z_1)^{-1}\mathbf{T}_{11}(z_0, z_1) \\ \mathbf{T}_{12}(z_0, z_1)^{-1} & -\mathbf{T}_{12}(z_0, z_1)^{-1}\mathbf{T}_{11}(z_0, z_1) \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_0 \end{bmatrix}.$$

The proof of this proposition follows immediately from these equivalent statements. □

Proposition 14. *The matrix* $\mathbf{J} - \mathbf{T}^*\mathbf{J}\mathbf{T}$ *[dropping dependency on* (z_0, z_1) *for simplicity] has the block form*

$$\mathbf{J} - \mathbf{T}^*\mathbf{J}\mathbf{T} = \begin{bmatrix} 2\,\mathrm{Re}\,(\mathbf{T}_{11}^*\boldsymbol{\rho}\mathbf{T}_{21}) & \boldsymbol{\rho} - (\mathbf{T}_{21}^*\boldsymbol{\rho}^*\mathbf{T}_{12} + \mathbf{T}_{11}^*\boldsymbol{\rho}\mathbf{T}_{22}) \\ [\boldsymbol{\rho} - (\mathbf{T}_{21}^*\boldsymbol{\rho}^*\mathbf{T}_{12} + \mathbf{T}_{11}^*\boldsymbol{\rho}\mathbf{T}_{22})]^* & 2\,\mathrm{Re}\,(\mathbf{T}_{12}^*\boldsymbol{\rho}\mathbf{T}_{22}) \end{bmatrix}, \tag{4.57}$$

where $\mathrm{Re}(\mathbf{M}) = \frac{1}{2}(\mathbf{M} + \mathbf{M}^*)$ *denotes the real part of a square matrix* $\mathbf{M}$*. In particular, if* $\mathbf{J} - \mathbf{T}^*\mathbf{J}\mathbf{T} > 0$ *then* $\mathrm{Re}\,(\mathbf{T}_{11}^*\boldsymbol{\rho}\mathbf{T}_{21}) > 0$*,* $\mathrm{Re}\,(\mathbf{T}_{12}^*\boldsymbol{\rho}\mathbf{T}_{22}) > 0$*, and* $\mathbf{T}_{ij}$ *is invertible for* $1 \leq i, j \leq 2$*.*

Proof. The block representation (4.57) follows immediately from the block representations (4.16), (4.52) by block multiplication. Suppose $\mathbf{J} - \mathbf{T}^*\mathbf{J}\mathbf{T} > 0$. Then it follows immediately from the block representation (4.57) that $\mathrm{Re}\,(\mathbf{T}_{11}^*\boldsymbol{\rho}\mathbf{T}_{21}) > 0$, $\mathrm{Re}\,(\mathbf{T}_{12}^*\boldsymbol{\rho}\mathbf{T}_{22}) > 0$. Now it is a well-known fact from linear algebra that if $\mathrm{Re}\,\mathbf{M} > 0$ for a square matrix $\mathbf{M}$ then $\mathbf{M}$ is invertible. From this it immediately follows that $\mathbf{T}_{ij}$ is invertible for $1 \leq i, j \leq 2$. This completes the proof. □

Now we define the indefinite inner product $[\cdot, \cdot] : \mathbb{C}^4 \times \mathbb{C}^4 \to \mathbb{C}$ in terms of the standard inner product $(\cdot, \cdot) : \mathbb{C}^4 \times \mathbb{C}^4 \to \mathbb{C}$ by

$$[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2] = \frac{c}{16\pi}\,(\mathbf{J}\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\,, \quad \boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \in \mathbb{C}^4. \tag{4.58}$$

We also define the complex Poynting vector $\mathbf{S}$ for functions of the form (4.13) to be

$$\mathbf{S} = \frac{c}{8\pi}\mathbf{E} \times \overline{\mathbf{H}} = e^{-2(\mathrm{Im}(k_1)x + \mathrm{Im}(k_2)y)}\mathbf{S}\,(z)\,, \quad \mathbf{S}\,(z) = \frac{c}{8\pi}\mathbf{E}\,(z) \times \overline{\mathbf{H}\,(z)}$$

The energy conservation law for Maxwell's equations (4.6) for functions of the form (4.13) is now described by the next theorem.

Theorem 15. *Assume* $\operatorname{Im}\omega > 0$ *and* $\boldsymbol{\kappa} \in \mathbb{R}^2$. *Then for any* $z_0, z_1 \in [-d, d]$, $z_0 < z_1$ *and any solution* $\boldsymbol{\psi}$ *of the ODE (4.15) with* $\begin{bmatrix}\mathbf{E} & \mathbf{H}\end{bmatrix}^T$ *the corresponding solution of Maxwell's equations (4.6) of the form (4.13) whose tangential components (4.14) are* $\boldsymbol{\psi}$, *we have*

$$[\boldsymbol{\psi}(z_0), \boldsymbol{\psi}(z_0)] - [\boldsymbol{\psi}(z_1), \boldsymbol{\psi}(z_1)] = -\int_{z_0}^{z_1} \partial_z \left[\operatorname{Re}\mathbf{S}(z) \cdot e_3\right] dz = -\int_{z_0}^{z_1} \nabla \cdot \operatorname{Re}(\mathbf{S})\, dz \tag{4.59}$$

$$= \frac{1}{8\pi}\int_{z_0}^{z_1} \left(\mathbf{H}, \operatorname{Im}\left[\omega\boldsymbol{\mu}(z,\omega)\right]\mathbf{H}\right) + \left(\mathbf{E}, \operatorname{Im}\left[\omega\boldsymbol{\varepsilon}(z,\omega)\right]\mathbf{E}\right) dz \geq 0, \tag{4.60}$$

with equality if and only if $\boldsymbol{\psi} \equiv 0$. *In particular, this implies*

$$\mathbf{J} - \mathbf{T}(z_0, z_1, \boldsymbol{\kappa}, \omega)^* \mathbf{J} \mathbf{T}(z_0, z_1, \boldsymbol{\kappa}, \omega) > 0. \tag{4.61}$$

Proof. The equalities in (4.59) follow immediately from the equalities

$$\operatorname{Re}\mathbf{S}(z) \cdot e_3 = -\frac{1}{2}\left(\begin{bmatrix}\mathbf{E}(z)\\ \mathbf{H}(z)\end{bmatrix}, \begin{bmatrix}0 & \mathbf{e}_3\times\\ -\mathbf{e}_3\times & 0\end{bmatrix}\begin{bmatrix}\mathbf{E}(z)\\ \mathbf{H}(z)\end{bmatrix}\right) = \frac{1}{2}\left(\boldsymbol{\psi}(z), \mathbf{J}\boldsymbol{\psi}(z)\right).$$

The proof of the last term in (4.59) being equal to (4.60) is proved in almost the exact same way as the proof of Poynting's Theorem for time-harmonic fields [see Section 6.8 in Jackson (1999) and also Section V.A of Welters, Avniel, and Johnson (2014)] and so will be omitted. The inequality in (4.60) follows from passivity (4.4) and necessary and sufficient conditions for equality follow immediately from this. These facts imply immediately the inequality in (4.61). This completes the proof. □

4.3 Analyticity of the DtN map for bounded media

4.3.1 Formulation of the problem

For the sake of simplicity, we consider here an electromagnetic medium (see **Figure 4.2** for an example) composed of two isotropic homogeneous materials which fills an open connected bounded Lipschitz domain $\Omega \subset \mathbb{R}^3$ (we refer to the Section 5.1 of Kirsch and Hettlich 2015 for the definition of Lipschitz bounded domains which includes domains with nonsmooth boundary as polyhedra). However, our result could be easily extended to a medium composed of a finite number of anisotropic homogeneous materials, this is discussed in the last section. Thus, the dielectric permittivity ε and the magnetic permeability μ which characterized this medium are supposed to be piecewise constant functions which take respectively the complex values ε_1 and μ_1 in the first material, and ε_2 and μ_2 in the second one. Moreover, we assume that both materials are passive, thus these functions have to satisfy (see Milton 2002; Welters, Avniel, and Johnson 2014; Bernland, Luger, and Gustafsson 2011; Gustafsson and Sjöberg 2010):

$$\operatorname{Im}(\omega\varepsilon) > 0 \text{ and } \operatorname{Im}(\omega\mu) > 0 \text{ for } \operatorname{Im}\omega > 0, \tag{4.62}$$

where ω denotes the complex frequency.

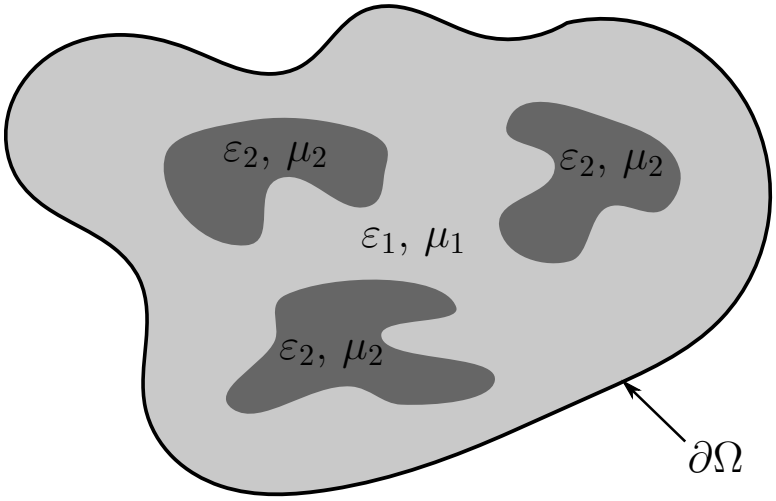


Figure 4.2: Example of the body Ω.

The time-harmonic Maxwell's equations (in Gaussian units) which link the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ in Ω are given by:

$$(\mathcal{P}) \begin{cases} \operatorname{curl} \mathbf{E} - i\omega\mu\, c^{-1}\mathbf{H} = 0 & \text{in } \Omega, \\ \operatorname{curl} \mathbf{H} + i\omega\varepsilon\, c^{-1}\mathbf{E} = 0 & \text{in } \Omega, \\ \mathbf{E} \times \mathbf{n} = \mathbf{f} & \text{on } \partial\Omega. \end{cases}$$

where $\mathbf{n}$ denotes here the outward normal vector on the boundary of Ω: $\partial\Omega$, c the speed of light in the vacuum and $\mathbf{f}$ the tangential electric field $\mathbf{E}$ on $\partial\Omega$.

Let us first introduce some classical functional spaces associated to the study of Maxwell's equations:

- $L^2(\Omega)$ which is a Hilbert space endowed with the inner product:

$$(\mathbf{f}, \mathbf{g})_{L^2(\Omega)} = \int_\Omega \mathbf{f}(x) \cdot \overline{\mathbf{g}(x)}\, \mathrm{d}\mathbf{x},$$

- $H(\operatorname{curl}, \Omega) = \left\{ \mathbf{u} \in L^2(\Omega) \mid \operatorname{curl} \mathbf{u} \in L^2(\Omega) \right\}$,
- $H_0(\operatorname{curl}, \Omega) = \{ \mathbf{u} \in H(\operatorname{curl}, \Omega) \mid \mathbf{u} \times \mathbf{n} = 0 \text{ on } \partial\Omega \}$,
- $H^{-\frac{1}{2}}(\operatorname{div}, \partial\Omega) = \{ (\mathbf{u} \times \mathbf{n})_{\partial\Omega} \mid \mathbf{u} \in H(\operatorname{curl}, \Omega) \}$,
- $H^{-\frac{1}{2}}(\operatorname{curl}, \partial\Omega) = \{ \mathbf{n} \times (\mathbf{u} \times \mathbf{n})_{\partial\Omega} \mid \mathbf{u} \in H(\operatorname{curl}, \Omega) \}$.

Here $H(\operatorname{curl}, \Omega)$ and $H_0(\operatorname{curl}, \Omega)$ are Hilbert spaces endowed with the norm $\|\cdot\|_{H(\operatorname{curl}, \Omega)}$ defined by

$$\|\mathbf{u}\|^2_{H(\operatorname{curl}, \Omega)} = \|\mathbf{u}\|^2_{L^2(\Omega)} + \|\operatorname{curl} \mathbf{u}\|^2_{L^2(\Omega)}.$$

Concerning the functional framework associated with the spaces of tangential traces and tangential trace components $H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)$ and $H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)$, we refer to the Section 5.1 of Kirsch and Hettlich (2015). These spaces are respectively Banach spaces for the norms $\|\cdot\|_{H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)}$ and $\|\cdot\|_{H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)}$ introduced in the Definition 5.23 of Kirsch and Hettlich (2015) and are linked by the duality relation: $(H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega))^* = H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)$. Moreover, their duality product $\langle\cdot,\cdot\rangle$ (see Theorem 5.26 of Kirsch and Hettlich (2015)) satisfies the following Green's identity:

$$\int_\Omega \mathbf{u}\cdot\mathrm{curl}\,\mathbf{v} - \mathbf{v}\cdot\mathrm{curl}\,\mathbf{u}\,\mathrm{d}\mathbf{x} = \langle \mathbf{n}\times(\mathbf{v}\times\mathbf{n}), \mathbf{u}\times\mathbf{n}\rangle\,,\ \forall \mathbf{u},\mathbf{v}\in H(\mathrm{curl}\,,\Omega). \tag{4.63}$$

Here we look for solutions $(\mathbf{E},\mathbf{H})\in H(\mathrm{curl}\,,\Omega)^2$ of the problem $(\mathcal{P})$ for data $\mathbf{f}\in H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)$.

4.3.2 The Dirichlet-to-Neumann map

We introduce the variable $\mathbf{Z}=(\omega\varepsilon_1,\omega\varepsilon_2,\omega\mu_1,\omega\mu_2)\in(\mathbb{C}^+)^4$. The electromagnetic Dirichlet-to-Neumann map $\Lambda_{\mathbf{Z}}: H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)\to H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)$ associated to the problem $(\mathcal{P})$ is defined as the linear operator:

$$\Lambda_{\mathbf{Z}}\,\mathbf{f} = i\,\mathbf{n}\times(\mathbf{H}\times\mathbf{n})_{\partial\Omega},\quad \forall \mathbf{f}\in H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega). \tag{4.64}$$

Remark 16. *This definition of the DtN map (4.64) is slightly different from the one introduced in Albanese and Monk (2006), Ola, Päivärinta, and Somersalo (2012) and Uhlmann and Zhou (2015). Here, the rotated tangential electric field $\mathbf{f}=\mathbf{E}\times\mathbf{n}$ is mapped (up to a constant) to the tangential component of the magnetic field $\mathbf{n}\times(\mathbf{H}\times\mathbf{n})=(\mathbf{I}-\mathbf{n}\mathbf{n}^T)\mathbf{H}$ and not to the rotated tangential magnetic field $\mathbf{H}\times\mathbf{n}$. This definition is closer to the one used in Chaulet (2014) to construct generalized impedance boundary conditions for electromagnetic scattering problems.*

We want to prove the following theorem:

Theorem 17. *The DtN map $\Lambda_{\mathbf{Z}}$ is well-defined, is a continuous linear operator with respect to the datum $\mathbf{f}$ and is an analytic function of $\mathbf{Z}$ in the open set $(\mathbb{C}^+)^4$. Moreover, the operator $\Lambda_{\mathbf{Z}}$ satisfies*

$$\mathrm{Im}\left\langle \Lambda_{\mathbf{Z}}\,\mathbf{f},\overline{\mathbf{f}}\right\rangle > 0,\ \forall \mathbf{f}\in H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)-\{0\}, \tag{4.65}$$

and as an immediate consequence, the function

$$h_{\mathbf{f}}(\mathbf{Z}) = \left\langle \Lambda_{\mathbf{Z}}\,\mathbf{f},\overline{\mathbf{f}}\right\rangle \text{ defined on } (\mathbb{C}^+)^4 \text{ for all } \mathbf{f}\in H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega) \tag{4.66}$$

is a Herglotz function of $\mathbf{Z}$ (see Definition 2).

Remark 18. *A similar theorem is obtained in the previous chapter of this book for a DtN map defined as the operator which maps the tangential electric field $\mathbf{n}\times(\mathbf{E}\times\mathbf{n})$ to $i\,\mathbf{n}\times\mathbf{H}$ on $\partial\Omega$. But for a regular boundary $\partial\Omega$ (for example $C^{1,1}$), this other definition of the DtN map can be rewritten as $-Q\Lambda_{\mathbf{Z}}\,Q$ with the isomorphism $Q: H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)\to H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)$ defined by $Q(\mathbf{g})=-\mathbf{n}\times\mathbf{g}$. Thus, one can show in the same way that the function*

$$h_{\mathbf{g}}(\mathbf{Z}) = \langle\overline{\mathbf{g}}, -Q\Lambda_{\mathbf{Z}}\,Q\,\mathbf{g}\rangle\,,\ \forall \mathbf{g}\in H^{-\frac{1}{2}}(\mathrm{curl}\,,\partial\Omega)$$

is a Herglotz function on $(\mathbb{C}^+)^4$. But, as it is mentioned in Remark 1, p. 30 and Corollary 2, p. 38 of Cessenat (1996), the isomorphism Q may not be well-defined if the function $\mathbf{n}$ is not regular enough. That is why, in order to make this connection, we assume that the boundary $\partial\Omega$ is slightly more regular than Lipschitz continuous.

4.3.3 Proof of the Theorem 17

We will first prove that the linear operator $T_{\mathbf{Z}} : H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega) \to H(\mathrm{curl}\,,\Omega)^2$ which associates the data $\mathbf{f}$ to the solution $(\mathbf{E},\mathbf{H}) \in H(\mathrm{curl}\,,\Omega)^2$ of $(\mathcal{P})$ is well-defined, continuous and analytic in $\mathbf{Z}$ in $(\mathbb{C}^+)^4$. In other words that the problem $(\mathcal{P})$ admits a unique solution $(\mathbf{E},\mathbf{H})$ which depends continuously on the data $\mathbf{f}$ and analytically on $\mathbf{Z}$. The approach we follow is standard, it uses a variational reformulation of the time-harmonic Maxwell's equations $(\mathcal{P})$ (see Cessenat 1996; Kirsch and Hettlich 2015; Monk 2003; Nedelec 2001).

The first step is to introduce a lifting of the boundary data $\mathbf{f}$. As $\mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega)$, there exists (see Theorem 5.24 of Kirsch and Hettlich 2015) a continuous lifting operator $R : H^{-\frac{1}{2}}(\mathrm{div}\,,\partial\Omega) \to H(\mathrm{curl}\,,\Omega)$ such that

$$R(\mathbf{f}) = \mathbf{E}_0, \tag{4.67}$$

that is a field $\mathbf{E}_0 \in H(\mathrm{curl}\,,\Omega)$ which depends continuously on $\mathbf{f}$ such that $\mathbf{E}_0 \times \mathbf{n} = \mathbf{f}$ on $\partial\Omega$. Thus, the field $\tilde{\mathbf{E}} = \mathbf{E} - \mathbf{E}_0$ satisfies the following problem with homogeneous boundary condition:

$$(\tilde{\mathcal{P}}) \begin{cases} \mathrm{curl}\,\tilde{\mathbf{E}} - i\omega\mu\, c^{-1}\mathbf{H} = -\mathrm{curl}\,\mathbf{E}_0 & \text{in } \Omega, \\ \mathrm{curl}\,\mathbf{H} + i\omega\varepsilon\, c^{-1}\tilde{\mathbf{E}} = -i\omega\varepsilon\, c^{-1}\mathbf{E}_0 & \text{in } \Omega, \\ \tilde{\mathbf{E}} \times \mathbf{n} = 0 & \text{on } \partial\Omega. \end{cases}$$

Now multiplying the second Maxwell's equation of $(\tilde{\mathcal{P}})$ by a test function $\boldsymbol{\psi} \in H_0(\mathrm{curl}\,,\Omega)$, integrating by parts and then eliminating the unknown $\mathbf{H}$ by using the first Maxwell's equation, we get the following variational formula for the electrical field $\tilde{\mathbf{E}}$:

$$\int_\Omega -c^2\,(\mu\omega)^{-1}\,\mathrm{curl}\,\tilde{\mathbf{E}} \cdot \overline{\mathrm{curl}\,\boldsymbol{\psi}} + \omega\varepsilon\,\tilde{\mathbf{E}} \cdot \overline{\boldsymbol{\psi}}\,\mathrm{d}\mathbf{x} = \int_\Omega c^2\,(\mu\omega)^{-1}\mathrm{curl}\,\mathbf{E}_0 \cdot \overline{\mathrm{curl}\,\boldsymbol{\psi}} - \omega\varepsilon\,\mathbf{E}_0 \cdot \overline{\boldsymbol{\psi}}\,\mathrm{d}\mathbf{x}, \tag{4.68}$$

satisfied by all $\boldsymbol{\psi} \in H_0(\mathrm{curl}\,,\Omega)$. The variational formula (4.68) and the problem $(\mathcal{P})$ are equivalent.

Proposition 19. *$\tilde{\mathbf{E}} \in H_0(\mathrm{curl}\,,\Omega)$ is a solution of the variational formulation (4.68) if and only if $(\mathbf{E} = \tilde{\mathbf{E}} + \mathbf{E}_0, \mathbf{H} = c\,(i\mu\omega)^{-1}\mathrm{curl}\,(\tilde{\mathbf{E}} + \mathbf{E}_0)) \in H(\mathrm{curl}\,,\Omega)^2$ satisfy the problem $(\mathcal{P})$.*

Proof. This proof is standard. For more details, we refer to the demonstration of the lemma 4.29 in Kirsch and Hettlich (2015). □

For all $\mathbf{Z} \in (\mathbb{C}^+)^4$, we introduce the sesquilinear form:

$$a_{\mathbf{Z}}(\boldsymbol{\phi},\boldsymbol{\psi}) = \int_\Omega -c^2(\mu\omega)^{-1}\,\mathrm{curl}\,\boldsymbol{\phi} \cdot \overline{\mathrm{curl}\,\boldsymbol{\psi}} + \omega\varepsilon\,\boldsymbol{\phi} \cdot \overline{\boldsymbol{\psi}}\,\mathrm{d}\mathbf{x},$$

defined on $H_0(\mathrm{curl}\,,\Omega)^2$. One easily proves by using the Cauchy–Schwarz inequality that:

$$|a_{\mathbf{Z}}(\boldsymbol{\phi},\boldsymbol{\psi})| \leq \max\left(c^2\left\|(\omega\mu)^{-1}\right\|_\infty, \|\omega\varepsilon\|_\infty\right)\,\|\boldsymbol{\phi}\|_{H(\mathrm{curl}\,,\Omega)}\,\|\boldsymbol{\psi}\|_{H(\mathrm{curl}\,,\Omega)}, \tag{4.69}$$

where $\|\cdot\|_\infty$ denotes the L^∞ norm. Thus, $a_{\mathbf{Z}}$ is continuous and as such it allows us to define a continuous linear operator $A_{\mathbf{Z}} \in L(H_0(\mathrm{curl}\,,\Omega), H_0(\mathrm{curl}\,,\Omega)^*)$ by

$$\langle A_{\mathbf{Z}}\boldsymbol{\phi},\boldsymbol{\psi}\rangle_{H_0(\mathrm{curl}\,,\Omega)} = a_{\mathbf{Z}}(\boldsymbol{\phi},\overline{\boldsymbol{\psi}}),\ \forall \boldsymbol{\phi},\boldsymbol{\psi} \in H_0(\mathrm{curl}\,,\Omega), \tag{4.70}$$

where $\langle \cdot, \cdot \rangle_{H_0(\mathrm{curl},\Omega)}$ stands for the duality product between $H_0(\mathrm{curl},\Omega)$ and its dual $H_0(\mathrm{curl},\Omega)^*$. We now introduce the antilinear form $l_{\mathbf{Z}}(\mathbf{E}_0)(\cdot)$:

$$l_{\mathbf{Z}}(\mathbf{E}_0)(\boldsymbol{\psi}) = \int_\Omega c^2\,(\mu\omega)^{-1}\mathrm{curl}\,\mathbf{E}_0 \cdot \overline{\mathrm{curl}\,\boldsymbol{\psi}} - \omega\varepsilon\,\mathbf{E}_0 \cdot \overline{\boldsymbol{\psi}}\,\mathrm{d}\mathbf{x},\ \forall \boldsymbol{\psi} \in H_0(\mathrm{curl},\Omega).$$

In the same way as (4.69), one can easily check:

$$|l_{\mathbf{Z}}(\mathbf{E}_0)(\boldsymbol{\psi})| \leq \max\left(c^2\ \left\|(\omega\mu)^{-1}\right\|_\infty, \|\omega\varepsilon\|_\infty\right)\ \|\mathbf{E}_0\|_{H(\mathrm{curl},\Omega)}\ \|\boldsymbol{\psi}\|_{H(\mathrm{curl},\Omega)}.$$

Hence, the linear operator $L_{\mathbf{Z}} : H(\mathrm{curl},\Omega) \to H_0(\mathrm{curl},\Omega)^*$ defined by

$$\langle L_{\mathbf{Z}}\mathbf{E}_0, \boldsymbol{\psi}\rangle_{H_0(\mathrm{curl},\Omega)} = l_{\mathbf{Z}}(\mathbf{E}_0)(\overline{\boldsymbol{\psi}}),\ \forall \mathbf{E}_0 \in H(\mathrm{curl},\Omega) \text{ and } \forall\, \boldsymbol{\psi} \in H_0(\mathrm{curl},\Omega), \tag{4.71}$$

is well-defined and continuous. Thus, we deduce from the relations (4.69) and (4.71) that the variational formula (4.68) is equivalent to solve the following infinite dimensional linear system

$$A_{\mathbf{Z}}\,\tilde{\mathbf{E}} = L_{\mathbf{Z}}\,\mathbf{E}_0. \tag{4.72}$$

Proposition 20. *If $\mathbf{Z} \in (\mathbb{C}^+)^4$, then the operator $A_{\mathbf{Z}}$ is an isomorphism from $H_0(\mathrm{curl},\Omega)$ to $H_0(\mathrm{curl},\Omega)^*$ and its inverse $A_{\mathbf{Z}}^{-1}$ depends analytically on $\mathbf{Z}$ in $(\mathbb{C}^+)^4$.*

Proof. Let $\mathbf{Z}$ be in $(\mathbb{C}^+)^4$. The invertibility of $A_{\mathbf{Z}}$ is an immediate consequence of the Lax–Milgram Theorem. Indeed, the coercivity of the sesquilinear form $a_{\mathbf{Z}}$ derives from the passivity hypothesis (4.62) of the material:

$$|a_{\mathbf{Z}}(\boldsymbol{\phi},\boldsymbol{\phi})| \geq \mathrm{Im}\,(a_{\mathbf{Z}}(\boldsymbol{\phi},\boldsymbol{\phi})) \geq \alpha\,\|\boldsymbol{\phi}\|^2_{H(\mathrm{curl},\Omega)},\ \forall \boldsymbol{\phi} \in H_0(\mathrm{curl},\Omega),$$

where $\alpha = \min(\,\mathrm{Im}(\omega\varepsilon_1), \mathrm{Im}(\omega\varepsilon_2), -c^2\,\mathrm{Im}(\,(\omega\mu_1)^{-1}), -c^2\,\mathrm{Im}(\omega\mu_2)^{-1})) > 0$.

Now the analyticity in $\mathbf{Z}$ of the operator $A_{\mathbf{Z}}^{-1}$ is proved as follows. First, one can verify easily that for all $\boldsymbol{\phi}, \psi \in H_0(\mathrm{curl},\Omega)$, the sesquilinear form $a_{\mathbf{Z}}(\boldsymbol{\phi},\overline{\psi})$ depends analytically on each component of $\mathbf{Z}$ when the others are fixed. It follows immediately from this and Theorem 6 that the operator $A_{\mathbf{Z}}$ [defined by the relation (4.70)] is analytic in the operator norm of $L(H_0(\mathrm{curl},\Omega), H_0(\mathrm{curl},\Omega)^*)$ and hence by Theorem 4 (Hartogs' Theorem) it is analytic in $\mathbf{Z}$ in the open set $(\mathbb{C}^+)^4$. Thus, since $A_{\mathbf{Z}}$ is an isomorphism which depends analytically on $\mathbf{Z}$ in the open set $(\mathbb{C}^+)^4$, then by Theorem 5 its inverse $A_{\mathbf{Z}}^{-1}$ depends analytically on $\mathbf{Z}$ in $(\mathbb{C}^+)^4$. □

Using Theorem 4 again, one can easily check in the same way as for the operator $A_{\mathbf{Z}}$ that the operator $L_{\mathbf{Z}}$ defined by (4.71) is also analytic in $\mathbf{Z}$ in $(\mathbb{C}^+)^4$. Hence, the variational formula (4.68) admits a unique solution:

$$\tilde{\mathbf{E}} = A_{\mathbf{Z}}^{-1}L_{\mathbf{Z}}\,\mathbf{E}_0 = A_{\mathbf{Z}}^{-1}L_{\mathbf{Z}}\,R(\mathbf{f}) \tag{4.73}$$

which depends continuously on the data $\mathbf{f}$ and analytically on $\mathbf{Z}$ in $(\mathbb{C}^+)^4$.

Corollary 21. *The linear operator $T_{\mathbf{Z}} : H^{-\frac{1}{2}}(\mathrm{div},\partial\Omega) \to H(\mathrm{curl},\Omega)^2$ which maps the data $\mathbf{f}$ to the solution $(\mathbf{E},\mathbf{H}) \in H(\mathrm{curl},\Omega)^2$ of $(\mathcal{P})$ is well-defined, continuous and depends analytically on $\mathbf{Z}$ in $(\mathbb{C}^+)^4$.*

Proof. This result is just a consequence of Propositions 19 and 20 which prove that the time-harmonic Maxwell's equations $(\mathcal{P})$ admits a unique solution $(\mathbf{E},\mathbf{H}) = T_{\mathbf{Z}}(\mathbf{f}) \in H(\mathrm{curl},\Omega)^2$ for data $\mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div},\partial\Omega)$ where the linear operator $T_{\mathbf{Z}}$ is defined by the following relation:

$$T_{\mathbf{Z}}(\mathbf{f}) = (R(\mathbf{f}) + \tilde{\mathbf{E}}, c\,(i\mu\omega)^{-1}\mathrm{curl}\,(\tilde{\mathbf{E}} + R(\mathbf{f})),\ \forall \mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div},\partial\Omega), \tag{4.74}$$

where $\tilde{\mathbf{E}} = A_{\mathbf{Z}}^{-1} L_{\mathbf{Z}}\, R(\mathbf{f})$ by the relation (4.73) and R stands for the lifting operator defined in (4.67). With the relation (4.74), the continuity of $T_{\mathbf{Z}}$ with respect to $\mathbf{f}$ and its analyticity with respect to $\mathbf{Z}$ follow immediately from the corresponding properties of the operator $A_{\mathbf{Z}}^{-1}$, $L_{\mathbf{Z}}$ and R. □

We now introduce the tangential component trace operator $\gamma_T : H(\mathrm{curl}\,, \Omega) \to H^{-\frac{1}{2}}(\mathrm{curl}\,, \partial\Omega)$ defined by:

$$\gamma_T(\mathbf{H}) = \mathbf{n} \times (\mathbf{H} \times \mathbf{n})_{\partial\Omega},\ \forall \mathbf{H} \in H(\mathrm{curl}\,, \Omega), \tag{4.75}$$

which is continuous (see Theorem 5.24 of Kirsch and Hettlich 2015) and the continuous linear operator $P : H(\mathrm{curl}\,, \Omega)^2 \to H(\mathrm{curl}\,, \Omega)$ defined by:

$$P(\mathbf{E}, \mathbf{H}) = \mathbf{H},\ \forall \mathbf{E}, \mathbf{H} \in H(\mathrm{curl}\,, \Omega). \tag{4.76}$$

This gives us the following operator representation of the electromagnetic DtN map defined in (4.64).

Proposition 22. *(Electromagnetic Dirichlet-to-Neumann map) The electromagnetic Dirichlet-to-Neumann map $\Lambda_{\mathbf{Z}} : H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega) \to H^{-\frac{1}{2}}(\mathrm{curl}\,, \partial\Omega)$ is the continuous linear operator defined by the composition of the continuous linear operators γ_T in (4.75), P in (4.76) and $T_{\mathbf{Z}}$ in (4.74) by*

$$\Lambda_{\mathbf{Z}}(\mathbf{f}) = i\, \gamma_T\, P\, T_{\mathbf{Z}}(\mathbf{f}),\ \forall \mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega). \tag{4.77}$$

Proof. Let $\mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega)$. Then $(\mathbf{E}, \mathbf{H}) = T_{\mathbf{Z}}(\mathbf{f})$ is the solution of problem $(\mathcal{P})$. Thus, by definition of γ_T and P we have $PT_{\mathbf{Z}}(\mathbf{f}) = \mathbf{H}$ and hence $i\gamma_T PT_{\mathbf{Z}}(\mathbf{f}) = i\mathbf{n} \times (\mathbf{H} \times \mathbf{n})_{\partial\Omega}$. Therefore, by the definition (4.64) of the DtN map we have $\Lambda_{\mathbf{Z}}(\mathbf{f}) = i\gamma_T PT_{\mathbf{Z}}(\mathbf{f})$. The fact that $\Lambda_{\mathbf{Z}}$ is a continuous linear operator follows immediately from this representation. This completes the proof. □

We can now derive the regularity properties of the DtN map $\Lambda_{\mathbf{Z}}$ by expressing this operator in terms of the operator $T_{\mathbf{Z}}$. The analyticity of the DtN map $\Lambda_{\mathbf{Z}}$ with respect to $\mathbf{Z}$ in $(\mathbb{C}^+)^4$ is now an immediate consequence of the fact that $\Lambda_{\mathbf{Z}}$ is the composition of two continuous linear operators $i\,\gamma_T$ and P independent of $\mathbf{Z}$ with the continuous linear operator $T_{\mathbf{Z}}$ which is analytic in $\mathbf{Z}$ (see Corollary 21).

Finally, to prove the positivity of $\mathrm{Im}\left\langle \Lambda_{\mathbf{Z}}\mathbf{f}, \overline{\mathbf{f}} \right\rangle$, we apply Green's identity (4.63) to the solution $(\mathbf{E}, \mathbf{H})$ of the problem $(\mathcal{P})$ for any nonzero data $\mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega)$. It yields

$$i \int_\Omega \overline{\mathbf{E}} \cdot \mathrm{curl}\, \mathbf{H} - \mathbf{H} \cdot \overline{\mathrm{curl}\, \mathbf{E}}\, \mathrm{d}\Omega = i \left\langle \mathbf{n} \times (\mathbf{H} \times \mathbf{n}), \overline{\mathbf{E}} \times \mathbf{n} \right\rangle = \left\langle \Lambda_{\mathbf{Z}}\, \mathbf{f}, \overline{\mathbf{f}} \right\rangle.$$

Since $(\mathbf{E}, \mathbf{H})$ is a solution of the time-harmonic Maxwell equations $(\mathcal{P})$, we can rewrite this last relation as:

$$\int_\Omega \omega\varepsilon\, c^{-1}\, |\mathbf{E}|^2 - \overline{\omega\mu}\, c^{-1}\, |\mathbf{H}|^2\, \mathrm{d}\mathbf{x} = \left\langle \Lambda_{\mathbf{Z}}\, \mathbf{f}, \overline{\mathbf{f}} \right\rangle. \tag{4.78}$$

By taking the imaginary part of (4.78) and using the passivity hypothesis (4.62) of the materials which compose the medium Ω, we get the positivity of $\mathrm{Im}\left\langle \Lambda_{\mathbf{Z}}\mathbf{f}, \overline{\mathbf{f}} \right\rangle$ (4.65) (since $(\mathbf{E}, \mathbf{H}) \neq (0, 0)$ for $\mathbf{f} \neq 0$) and it follows immediately that the function $h_{\mathbf{f}}$ defined by (4.66) is a Herglotz function of $\mathbf{Z}$.

4.3.4 Extensions of Theorem 17 to anisotropic and continuous media

Here we first extend Theorem 17 to the case of a medium Ω composed by N anisotropic homogeneous phases. Therefore, the dielectric permittivity $\boldsymbol{\varepsilon}(\mathbf{x})$ and magnetic permeability $\boldsymbol{\mu}(\mathbf{x})$ are now 3×3 tensor-valued functions of $\mathbf{x}$, which take for $j = 1, \cdots, N$ the constant values $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{\mu}_j$ in the jth material. Again, each material is supposed to be passive, in the sense that $\mathrm{Im}(\omega\boldsymbol{\varepsilon}_j)$ and $\mathrm{Im}(\omega\boldsymbol{\mu}_j)$ have to be positive tensors for all $j = 1, \cdots, N$ (see Milton 2002, Welters, Avniel, and Johnson 2014, Bernland, Luger, and Gustafsson 2011, Gustafsson and Sjöberg 2010).

First, we want to emphasize that besides the fact that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are now tensor-valued, the time-harmonic Maxwell's equations $(\mathcal{P})$ in Ω and its associated functional spaces remain the same. Moreover, as the vector space $M_3(\mathbb{C})$ is isomorphic to $\mathbb{C}^9$, we prove exactly in the same way that the DtN map $\Lambda_{\mathbf{Z}}$ defined by (4.64) is well-defined, is linear continuous with respect to $\mathbf{f}$, and is an analytic function with respect to $\mathbf{Z}$, where $\mathbf{Z}$ is here the vector of the $18N$ coefficients which are the elements (in some basis) of the tensors $\omega\boldsymbol{\varepsilon}_j$ and $\omega\boldsymbol{\mu}_j$ for $j = 1, \cdots, N$, in the open set $\mathcal{O}$ of $\mathbb{C}^{18N}$ characterized by the passivity relation (4.62). As $\mathcal{O}$ is isomorphic to the open set $(M_3^+(\mathbb{C}))^{2N}$, this is equivalent to say (as it is explained in the last paragraph of the subsection 4.2.2) that $\Lambda_{\mathbf{Z}}$ is an analytic function of the vector $\mathbf{Z}$, whose components are now those of the permittivity tensors $\omega\boldsymbol{\varepsilon}_j$ and permeability $\omega\boldsymbol{\mu}_j$ (for $j = 1, \cdots, N$) in each phase. Using the passivity assumption which is associated with the elements of $(M_3^+(\mathbb{C}))^{2N}$, one proves also identically the relation (4.65) on the DtN map for all $\mathbf{Z} \in (M_3^+(\mathbb{C}))^{2N}$.

The problem is now to define the notion of a Herglotz function. Indeed, when the tensors $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{\mu}_j$ of each composite are not all diagonal, it is not possible anymore to define the DtN map as a multivariate Herglotz function $h_{\mathbf{f}}$ on some copy of the upper-half plane: $(\mathbb{C}^+)^n$. The major obstruction to this construction is based on the simple observation that off-diagonal elements of a matrix in $M_3^+(\mathbb{C})$ will not necessarily have a positive imaginary part. Nevertheless, it is natural to define a Herglotz function which maps points $\mathbf{Z}$ represented by a $2N$-tuple of matrices $\mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N}$ with positive definite imaginary parts, i.e., $\mathbf{Z} = (\mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})$, to the upper half-plane.

One way to preserve the Herglotz property is to use a trajectory method (see Bergman 1978 and **Section 18.6** of Milton 2002), in other words, consider an analytic function $s \mapsto \mathbf{Z}(s)$ from $\mathbb{C}^+$ into $(M_3^+(\mathbb{C}))^{2N}$, i.e., a trajectory in one complex dimension (a surface in two real directions). Then, along this trajectory, we obtain immediately that the function

$$h_{\mathbf{f}}(s) = \left\langle \Lambda_{\mathbf{Z}(s)}\mathbf{f}, \overline{\mathbf{f}} \right\rangle, \ \forall \mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega), \tag{4.79}$$

is a Herglotz function (see Definition 2) of s in $\mathbb{C}^+$: analyticity follows from the fact that analyticity is preserved under composition of analytic functions, while, when s has positive imaginary part, nonnegativity of the imaginary part of $h_{\mathbf{f}}(s)$ follows from the fact that $\mathbf{Z}(s)$ lies in the domain where the imaginary part of the operator $\Lambda_{\mathbf{Z}(s)}$ is positive semi-definite.

A particularly interesting trajectory for electromagnetism, in an N-phase material, is the trajectory

$$s = \omega \to \mathbf{Z}(\omega) = (\omega\boldsymbol{\varepsilon}_1(\omega), \omega\boldsymbol{\varepsilon}_2(\omega), \ldots, \omega\boldsymbol{\varepsilon}_N(\omega), \omega\boldsymbol{\mu}_1(\omega), \omega\boldsymbol{\mu}_2(\omega), \ldots, \omega\boldsymbol{\mu}_N(\omega)),$$

where $\boldsymbol{\varepsilon}_j(\omega)$ and $\boldsymbol{\mu}_j(\omega)$ are the physical electric permittivity tensor and physical magnetic permeability tensor of the actual material constituting phase i as functions of the frequency ω. Due to the passive nature of these materials the trajectory maps ω in the upper half plane $\mathbb{C}^+$ into a trajectory in $(M_3^+(\mathbb{C}))^{2N}$. The physical interest about this trajectory is that one can in principle measure $\Lambda_{\mathbf{Z}(\omega)}$ along it, at least for real frequencies ω.

In the case of the trajectory method, one can easily generalize Theorem 17 to continuous anisotropic composites where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are matrix-valued functions of both variables $(\mathbf{x}, \omega) \in \Omega \times \mathbb{C}^+$. In this case, we suppose that

- (H1) For all $\omega \in \mathbb{C}^+$, $\boldsymbol{\varepsilon}(\cdot, \omega)$ and $\boldsymbol{\mu}(\cdot, \omega)$ are L^∞ matrix-valued functions on Ω which are locally bounded in the variable ω, in other words, we suppose that there exists $\delta > 0$ such that the open ball of center ω and radius δ: $B(\omega, \delta)$ is included in $\mathbb{C}^+$ and that

$$\sup_{z \in B(\omega,\delta)} \|\boldsymbol{\varepsilon}(\cdot, z)\|_\infty < \infty \text{ and } \sup_{z \in B(\omega,\delta)} \|\boldsymbol{\mu}(\cdot, z)\|_\infty < \infty \tag{4.80}$$

- (H2) The composite is passive which implies that for almost every $x \in \Omega$, the functions $\omega \mapsto \omega\boldsymbol{\varepsilon}(\mathbf{x}, \omega)$ and $\omega \mapsto \omega\boldsymbol{\mu}(\mathbf{x}, \omega)$ are analytic functions from $\mathbb{C}^+$ to $M_3^+(\mathbb{C})$ (see Section 11.1 of Milton 2002).

- (H3) For all $\omega \in \mathbb{C}^+$, there exists $C_\omega > 0$ such that

$$\operatorname*{ess\,inf}_{\mathbf{x} \in \Omega} \operatorname{Im}(\omega\boldsymbol{\varepsilon}(\cdot, \omega)) \geq C_\omega \operatorname{Id} \quad \text{and} \quad \operatorname*{ess\,inf}_{\mathbf{x} \in \Omega} -(\operatorname{Im}(\omega\boldsymbol{\mu}(\cdot, \omega))^{-1} \geq C_\omega \operatorname{Id}. \tag{4.81}$$

Remark 23. *The hypotheses (H1) and (H3) may seem complicated but they are satisfied for instance when $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are continuous functions of both variables $(\mathbf{x}, \omega) \in \operatorname{cl}\Omega \times \mathbb{C}^+$, where $\operatorname{cl}\Omega$ denotes the closure of Ω. In that case, one can see immediately that $(H1)$ is satisfied. Moreover, if we assume also that the passivity assumption (H2) holds on $\operatorname{cl}\Omega$ (instead of just Ω), the hypothesis (H3) is also satisfied since the functions $\operatorname{Im}(\omega\boldsymbol{\varepsilon}(\cdot, \omega))$ and $-(\operatorname{Im}(\omega\boldsymbol{\mu}(\cdot, \omega)))^{-1}$ are continuous functions on a compact set and thus reach their minimum value which is a positive matrix.*

Under these hypotheses, Theorem 17 remains valid: the function $h_{\mathbf{f}}(\omega)$ given by (4.66) is a Herglotz function of the frequency (by interpreting each formula of Section 4.3 with $\mathbf{Z} = \omega \in \mathbb{C}^+$ as a new analytic variable and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ as matrix valued functions of the variables $(\mathbf{x}, \omega)$). Moreover, the proof is basically the same as the one in Subsection 4.3.3. We just make precise here the justification of some technical points which appear when one reproduces this proof in this framework.

We first remark that the assumption (H1) on the tensors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ implies that the tensors $\omega\boldsymbol{\varepsilon}(\cdot, \omega)$ and $(\omega\boldsymbol{\mu}(\cdot, \omega))^{-1}$ are bounded functions of the space variable $\mathbf{x}$. Thus the bilinear form a_ω remains continuous and the operators A_ω and L_ω are still well-defined and continuous. With the coercivity hypothesis (H3), one can easily check that $\forall \boldsymbol{\phi} \in H_0(\operatorname{curl}, \Omega)$,

$$|a_\omega(\boldsymbol{\phi}, \boldsymbol{\phi})| \geq C_\omega \|\boldsymbol{\phi}\|^2_{H(\operatorname{curl},\Omega)},$$

and apply again the Lax–Milgram theorem to show the invertibility of A_ω. Then, the analyticity of the operators A_ω and L_ω with respect to ω in $\mathbb{C}^+$ is still obtained (thanks to the relations (4.70) and (4.71)) from their weak analyticity (see Theorem 6). This weak analyticity is proved by using Theorem 7 to show the analyticity of the integrals which appear in the expression of $\langle A_\omega\boldsymbol{\phi}, \boldsymbol{\psi}\rangle_{H_0(\operatorname{curl},\Omega)}$ and $\langle L_\omega\mathbf{E}_0, \boldsymbol{\psi}\rangle_{H_0(\operatorname{curl},\Omega)}$ for $\phi, \psi \in H_0(\operatorname{curl}, \Omega)$ and $\mathbf{E}_0 \in H(\operatorname{curl}, \Omega)$ (since the assumptions (H1) and (H2) imply the hypotheses of Theorem 7). Then the analyticity of A_ω^{-1} is proved by using again Theorem 5 and the rest of the proof follows by the same arguments as in the isotropic case.

4.4 Herglotz functions associated with anisotropic media

A theory of Herglotz functions directly defined on tensors and not only on scalar variables is particularly useful in the domain of bodies containing anisotropic materials such as, for instance, sea ice (see Carsey 1992, Stogryn 1987, Golden 1995, Golden 2009, Gully, Lin, Cherkaev, and Golden 2015) or in electromagnetism where it will even extend to complicated media such as gyrotropic materials for which the dielectric tensors and magnetic tensors are anisotropic but not symmetric (as there is no reciprocity principle in such media, see Landau, Lifshitz, and Pitaevskiĭ 1984). The idea for Herglotz representations of the effective moduli of anisotropic materials was first put forward in the appendix E of Milton (1981c), and was studied in depth in **Chapter 18** of Milton (2002), see also Barabash and Stroud (1999). In connection with sea ice, one is particularly interested in bounds where the moduli are complex: such bounds are an immediate corollary of appendix E of Milton (1981c) and series expansions of the effective conductivity (Willemse and Caspers 1979; Avellaneda and Bruno 1990) that are contingent on assumptions about the polycrystalline geometry, and more generally are obtained (even for viscoelasticity with anisotropic phases) in Milton (1987b) (to make the connection, see the discussion in Section 15 of the companion paper (Milton 1987a)), and also see the bounds (16.45) in Milton (1990). Explicit calculations were made by Gully, Lin, Cherkaev, and Golden (2015) and are in good agreement with sea ice measurements.

The trajectory method provides the desired representation, as shown in Section 18.8 of Milton (2002). We slightly modify that argument here. Given any tensors $\mathbf{L}_j = \omega\boldsymbol{\varepsilon}_j$ and $\mathbf{L}_{j+N} = \omega\boldsymbol{\mu}_j$, for $j = 1, 2, \dots N$, which we assume to have strictly positive definite imaginary parts, and given a reference tensor $\mathbf{L}_0$ which is real, but not necessarily positive definite, define the real matrices

$$\mathbf{A}_j = \mathrm{Re}[(\mathbf{L}_0 - \mathbf{L}_j)^{-1}], \quad \mathbf{B}_j = \mathrm{Im}[(\mathbf{L}_0 - \mathbf{L}_j)^{-1}], \quad j = 1, 2, \dots, 2N, \tag{4.82}$$

where, according to our assumption, $\mathbf{B}_j$ is positive definite for each j. Then consider the trajectory

$$\mathbf{Z}(s) = (\mathbf{L}'_1(s), \mathbf{L}'_2(s), \dots \mathbf{L}'_{2N}(s)), \quad \text{where } \mathbf{L}'_j(s) = \mathbf{L}_0 - (\mathbf{A}_j + s\mathbf{B}_j)^{-1}. \tag{4.83}$$

Each of the matrices $\mathbf{L}'_j(s)$ have positive definite imaginary parts when s is in $\mathbb{C}^+$, and so $\mathbf{Z}(s)$ maps $\mathbb{C}^+$ to $(M_3^+(\mathbb{C}))^{2N}$. Furthermore, by construction our trajectory passes through the desired point at $s = i$:

$$\mathbf{Z}(i) = (\omega\boldsymbol{\varepsilon}_1, \omega\boldsymbol{\varepsilon}_2, \dots, \omega\boldsymbol{\varepsilon}_N, \omega\boldsymbol{\mu}_1, \omega\boldsymbol{\mu}_2, \dots \omega\boldsymbol{\mu}_N). \tag{4.84}$$

Now $\Lambda_{\mathbf{Z}(s)}$ is an operator valued Herglotz function of s, and so has an integral representation involving a positive semi-definite operator-valued measure deriving from the values that $\mathbf{Z}(s)$ takes when s is just above the positive real axis. That measure in turn is linearly dependent on the measure derived from the values that $\Lambda_{\mathbf{Z}(\mathbf{L}'_1, \mathbf{L}'_2, \dots \mathbf{L}'_{2N})}$ takes as imaginary parts of $\mathbf{L}'_j$ become vanishingly small. Thus $\mathbf{Z}(i)$ can be expressed directly in terms of this latter measure, involving an integral kernel $\mathbf{K}_{\mathbf{L}_0}(\mathbf{L}_1, \mathbf{L}_2, \dots \mathbf{L}_{2N}, \mathbf{L}'_1, \mathbf{L}'_2, \dots \mathbf{L}'_{2N})$ that is singular with support that is concentrated on the trajectory traced by $(\mathbf{L}'_1(s), \mathbf{L}'_2(s), \dots \mathbf{L}'_{2N}(s))$ as s is varied along the real axis. The formula for $\Lambda_{\mathbf{Z}(i)}$ obtained from the above prescription can be rewritten (informally) as

$$\Lambda_{\mathbf{Z}(\mathbf{L}_1, \mathbf{L}_2, \dots \mathbf{L}_{2N})} = \int \mathbf{K}_{\mathbf{L}_0}(\mathbf{L}_1, \mathbf{L}_2, \dots \mathbf{L}_{2N}, \mathbf{L}'_1, \mathbf{L}'_2, \dots \mathbf{L}'_{2N})\, d\mathbf{m}(\mathbf{L}'_1, \mathbf{L}'_2, \dots \mathbf{L}'_{2N}). \tag{4.85}$$

With extra work, an explicit formula could be obtained for the kernel $\mathbf{K}_{\mathbf{L}_0}(\mathbf{L}_1, \mathbf{L}_2, \dots \mathbf{L}_{2N}, \mathbf{L}'_1, \mathbf{L}'_2, \dots \mathbf{L}'_{2N})$ which is non-zero except on the path traced out by $(\mathbf{L}'_1(s), \mathbf{L}'_2(s), \dots \mathbf{L}'_{2N}(s))$ as s varies over the reals, and

this path depends on $\mathbf{L}_0$ and the moduli of $\mathbf{L}_1, \mathbf{L}_2, \ldots \mathbf{L}_{2N}$. The measure $d\mathbf{m}(\mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})$ is derived from the values that $\Lambda_{\mathbf{Z}(\mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})}$ takes as imaginary parts of $\mathbf{L}'_j$ become vanishingly small.

The trajectory method has been unjustly criticised for failing to separate the dependence of the function (in this case $\Lambda_{\mathbf{Z}(\mathbf{L}_1, \mathbf{L}_2, \ldots \mathbf{L}_{2N})}$) on the moduli (in this case the tensors $\mathbf{L}_1, \mathbf{L}_2, \ldots \mathbf{L}_{2N}$) from the dependence on the geometry, which is contained in the measure (in this case derived from the values that $\Lambda_{\mathbf{Z}(\mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})}$ takes as imaginary parts of $\mathbf{L}'_j$ become vanishingly small.) But we see that (4.85) makes such a separation.

Now there are differences between this representation and standard representation formulas for multivariate Herglotz functions, but the main difference is that the kernel $\mathbf{K}_{\mathbf{L}_0}(\mathbf{L}_1, \mathbf{L}_2, \ldots \mathbf{L}_{2N}, \mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})$, unlike the Szegő kernel which enters the polydisk representation of Korányi and Pukánszky (1963), is singular, being concentrated on this trajectory. However, for each choice of $\mathbf{L}_0$ there is a representation, each involving a kernel with support on a different trajectory and so one can average the representations over the matrices $\mathbf{L}_0$ with any desired smooth nonnegative weighting, to obtain a family of representations with less singular kernels that are the average over $\mathbf{L}_0$ of $\mathbf{K}_{\mathbf{L}_0}(\mathbf{L}_1, \mathbf{L}_2, \ldots \mathbf{L}_{2N}, \mathbf{L}'_1, \mathbf{L}'_2, \ldots \mathbf{L}'_{2N})$. This nonuniqueness in the choice of kernel reflects the fact that the measure satisfies certain Fourier constraints on the polydisk (see Rudin (1969)).

Another approach to considering of the notion of Herglotz functions in anisotropic multicomponent media is as follows. It will consist of proving that $(M_3^+(\mathbb{C}))^{2N}$ is isometrically isomorphic to a tubular domain (defined below) of $\mathbb{C}^{18N}$ and use it to extend the definition of Herglotz functions via the theory of holomorphic functions on tubular domains with nonnegative imaginary part from Vladimirov 2002.

As we have already mentioned in the introduction to this chapter and at the end of Subsection 4.3.4, this extended definition is significant because these multivariate functions provide a deep connection to the theory of multivariate passive linear systems as described in Section 20 of Vladimirov 2002, for instance, and in the study of anisotropic composites (e.g., sea ice or gyrotropic media). In addition to this, such an extension may allow for a more general approach of the efforts of Golden and Papanicolaou (1985) and Milton and Golden (1990) to derive integral representations of multivariate Herglotz functions in $(\mathbb{C}^+)^N$, beyond that provided in Section 18.8 of Milton (2002), or for deriving bounds in the theory of composites using the analytic continuation method (see **Chapter 27** in Milton 2002).

Let us first introduce the definition of a tubular domain from **Chapter 2**, Section 9 of Vladimirov 2002.

Definition 24. *Let Γ be a closed convex acute cone in $\mathbb{R}^N$ with vertex at $\mathbf{0}$. We denote by $C = \mathrm{int}(\Gamma^*)$, where Γ^* stands for the dual of C (in the sense of cones' duality) and $\mathrm{int}(\Gamma^*)$ denotes the (topological) interior of Γ^*. Thus, C is an open, convex, nonempty cone. Then, a tubular domain in $\mathbb{C}^N$ with base C is defined as:*

$$\mathcal{T} = \mathbb{R}^N + iC = \left\{ \mathbf{z} = \mathbf{x} + i\mathbf{y} | \mathbf{x} \in \mathbb{R}^N \text{ and } \mathbf{y} \in C \right\}.$$

We will now show that $M_N^+(\mathbb{C})$ is isometrically isomorphic to a tubular domain $\mathcal{T}^C$ of $\mathbb{C}^{N^2}$ [see Proposition 25, Proposition 26, and Theorem 27 below (which we state without proof as they are easily verified)]. Toward this purpose, we first use the decomposition

$$\mathbf{M} = \frac{\mathbf{M} + \mathbf{M}^*}{2} + i\,\frac{\mathbf{M} - \mathbf{M}^*}{2i},\ \forall \mathbf{M} \in M_N(\mathbb{C}),$$

to parameterize the space $M_N^+(\mathbb{C})$ as

$$M_N^+(\mathbb{C}) = \left\{ \mathbf{M}_1 + i\,\mathbf{M}_2 | \mathbf{M}_1 \in H_N(\mathbb{C}) \text{ and } \mathbf{M}_2 \in H_N^+(\mathbb{C}) \right\},$$

where $H_N(\mathbb{C})$ and $H_N^+(\mathbb{C})$ denote the sets of Hermitian and positive definite Hermitian matrices, respectively.

Then, we recall with the two following propositions that $H_N(\mathbb{C})$ is a Euclidean space and $H_N^+(\mathbb{C})$ is a cone in $H_N(\mathbb{C})$ with some remarkable properties that we will use to construct the basis C of our tubular region. First, denote the standard orthonormal basis vectors of $\mathbb{R}^N$ by $\mathbf{e}_k$, $k = 1, \ldots, N$. With respect to this basis, let $\mathbf{E}_{kl}$, $1 \leq k, l \leq N$ denote the matrices in $M_N(\mathbb{C})$ such that as operators on $\mathbb{C}^N$ are equal to

$$\mathbf{E}_{kl} = \mathbf{e}_k \mathbf{e}_l^T \text{ for } k, l \in \{1, \ldots, N\}, \tag{4.86}$$

i.e., $\mathbf{E}_{kl}$ is the $N \times N$ matrix with 1 in the kth row, lth column and zeros everywhere else.

Proposition 25. *The Hermitian matrices $H_N(\mathbb{C})$ endowed with the inner product:*

$$(\mathbf{A}, \mathbf{B})_{H_N(\mathbb{C})} = \mathrm{Tr}(\mathbf{AB}), \quad \forall \mathbf{A}, \mathbf{B} \in H_N(\mathbb{C}) \tag{4.87}$$

is a Euclidean space of dimension N^2. Furthermore, an orthonormal basis of this space is given by the matrices:

$$\mathbf{E}_{kk} \text{ for } k \in \{1, 2, 3, ..., N\}, \tag{4.88}$$

$$\frac{1}{\sqrt{2}}(\mathbf{E}_{kl} + \mathbf{E}_{lk}),\ \frac{i}{\sqrt{2}}(\mathbf{E}_{kl} - \mathbf{E}_{lk}) \text{ for } l, k \in \{1, 2, 3, ..., N\} \text{ such that } l < k. \tag{4.89}$$

Moreover, if we denote by $I_N = \{1, \cdots, N\}$, then the linear map $\phi : H_N(\mathbb{C}) \mapsto \mathbb{R}^{N^2}$ given by

$$\phi(\mathbf{A}) = ((A_{kk})_{k \in I_N}, (\sqrt{2}\,\mathrm{Re}\, A_{kl})_{(k,l) \in I_N^2 | k<l}, (\sqrt{2}\,\mathrm{Im}\, A_{kl})_{(k,l) \in I_N^2 | k<l}) \in \mathbb{R}^{N^2}. \tag{4.90}$$

which represents the coordinates of $\mathbf{A}$ in the orthonormal basis (4.88) defines an isometry in the sense that

$$(\mathbf{A}, \mathbf{B})_{H_N(\mathbb{C})} = \phi(\mathbf{A}) \cdot \phi(\mathbf{B}),\ \forall \mathbf{A}, \mathbf{B} \in H_N(\mathbb{C}), \tag{4.91}$$

for $\cdot$ the standard dot product of $\mathbb{R}^{N^2}$.

Proposition 26. *In the Euclidean space $H_N(\mathbb{C})$, denote the closure of $H_N^+(\mathbb{C})$ by $\mathrm{cl}\, H_N^+(\mathbb{C})$ and its (topological) interior by $\mathrm{int}\left(\mathrm{cl}\, H_N^+(\mathbb{C})\right)$. Then*

$$\mathrm{cl}\, H_N^+(\mathbb{C}) = \{\mathbf{M} \in M_N(\mathbb{C}) | \mathrm{Im}\, \mathbf{M} \geq 0\}, \tag{4.92}$$

i.e., the set all positive semidefinite (Hermitian) matrices in $M_N(\mathbb{C})$. Furthermore, it is a closed, convex, acute cone with vertex at $\mathbf{0}$ and is self-dual (in sense of cones' duality). Moreover, $\mathrm{int}\left(\mathrm{cl}\, H_N^+(\mathbb{C})\right) = H_N^+(\mathbb{C})$ and it is an open, convex, nonempty cone (in the sense of the definition in Sec. 4.4 of Vladimirov 2002).

In particular, it follows immediately from the Propositions 25 and 26 that:

Theorem 27. *The set $M_N^+(\mathbb{C}) = H_N(\mathbb{C}) + iH_N^+(\mathbb{C})$ is isometrically isomorphic to a tubular domain $\mathcal{T}^C = \mathbb{R}^{N^2} + i\,C$ in $\mathbb{C}^{N^2}$ where $\mathbb{R}^{N^2}$ and C are respectively defined by the relations $\mathbb{R}^{N^2} = \phi(H_N(\mathbb{C}))$ and $C = \phi(H_N^+(\mathbb{C}))$ for ϕ the isometry defined in (4.90).*

As the Cartesian product of tubular domains is also a tubular domain, we obtain immediately that the space of tensors $(M_3^+(\mathbb{C}))^{2N}$ associated to a medium Ω composed of N anisotropic passive composites is isometrically isomorphic to the tubular region $(\mathcal{T}^C)^{2N} = \mathcal{T}^{C^{2N}}$ (where $\mathcal{T}^C$ is the tubular domain defined in the Theorem 27). Hence, identifying $(M_3^+(\mathbb{C}))^{2N}$ with $\mathcal{T}^{C^{2N}}$ via this isometry allows us to define in Theorem 17 the function

$$h_{\mathbf{f}}(\mathbf{Z}) = \left\langle \Lambda_{\mathbf{Z}}\, \mathbf{f}, \overline{\mathbf{f}} \right\rangle,\ \forall \mathbf{f} \in H^{-\frac{1}{2}}(\mathrm{div}\,, \partial\Omega),$$

on $(M_3^+(\mathbb{C}))^{2N}$ as an Herglotz function of $\mathbf{Z}$ in the sense that it is an holomorphic function on a tubular domain with a nonnegative imaginary part and it justifies Definition 2 given in the introduction.

Acknowledgments

Graeme Milton is grateful to the National Science Foundation for support through the Research Grant DMS-1211359. Aaron Welters is grateful for the support from the U.S. Air Force Office of Scientific Research (AFOSR) through the Air Force's Young Investigator Research Program (YIP) under the grant FA9550-15-1-0086.

5 The inverse problem: Obtaining information about what's inside a body

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

Using the fact that the Dirichlet-to-Neumann map derives from an effective tensor, which has the same abstract definition as effective tensors in composites, we employ bounds on these effective tensors to obtain bounds which correlate the Dirichlet-to-Neumann map at different frequencies. These bounds can be used in an inverse fashion to give information about the moduli inside the body. Also in two phase nonmagnetic bodies, with constant magnetic permeability, the Dirichlet-to-Neumann map is an analytic function of frequency, and there can be special complex frequencies where the electrical permittivity is the same for both phases. The Dirichlet-to-Neumann map can therefore be determined exactly (numerically) at these frequencies. Then measurements of the transient behavior of the response fields at the boundary of the body, for various time-dependent applied fields at the boundary can be correlated with perturbations to the Dirichlet-to-Neumann near these special frequencies, and information about the geometry can be extracted from bounds on these perturbations. In a similar vein, in quasistatic elastodynamics in a two-phase body, there may be special complex frequencies where the complex shear moduli of the phases coincide (though no necessarily their bulk moduli). With the right boundary conditions at these special frequencies one can extract information about the geometry from the Dirichlet-to-Neumann map, and that in turn can be correlated with the transient behavior of the traction fields at the boundary of the body, for various applied, time-dependent, displacement fields imposed at the boundary. For two-phase conducting, or dielectric, media in the quasistatic limit the analysis is even simpler and one does not have to look for these special frequencies. The transient response of the boundary fields can be directly correlated with perturbations of the Dirichlet-to-Neumann map for a nearly homogeneous material, and from bounds on these perturbations one can extract information about the geometry. Finally instead of measuring the transient response, one could measure the Dirichlet-to-Neumann map at a set of frequencies and results on the Nevanlinna–Pick interpolation problem for operator-valued functions give us bounds on the perturbations of the Dirichlet-to-Neumann map for a nearly homogeneous material, and thereby bounds on the geometry.

5.1 Introduction

Due to the enormous applications in medicine, geophysical prospecting, and homeland security, the field of inverse problems, where one tries to say something about what is inside a body from external measurements, has received enormous attention: see, for example, the books of Isakov (2006), Kirsch (2011), Mueller and Siltanen (2012), Kabanikhin (2012), Ammari, Garnier, Jing, Kang, Lim, Sølna, and Wang (2013), the conference proceedings Stefanov, Vasy, and Zworski (2014) in honor of Gunther Uhlmann, and of course the many articles in the journals "Inverse Problems", "Journal of Inverse and Ill-Posed Problems", "Inverse Problems and Imaging", and "Inverse Problems in Science and Engineering" devoted to the subject.

Here the waves we probe the body with could be acoustic (sound) waves, electromagnetic waves or elastic (seismic) waves. Or they could be static fields applied to the boundary of the body: electrical potentials, current fluxes, prescribed deformations (displacement fields) or prescribed tractions (force applied to the body). Or these applied fields could vary with time, perhaps sinusoidally (at one frequency) but more generally they could have a variety of dependencies with time. One of the simplest is when fields are suddenly applied to the body: perhaps one suddenly deforms the body and measures how the tractions at the boundary relax with time (stress relaxation), or perhaps one suddenly exerts a traction on the body and watches how it deforms with time (creep test). Or in acoustics one could apply a pressure field and watch how the fluid moves.

So the problem is: given measurements of the Dirichlet-to-Neumann map, or partial measurements perhaps for a variety of frequencies, or perhaps one measures the transient response for time-varying applied fields, how does one recover information about what's inside the body? Usually there is an overabundance of data. The question is, how do you interpret those data?

Unless the body is almost homogeneous, the difficulty is that the Dirichlet-to-Neumann map depends nonlinearly and in a complicated way (that is nonlocal) on the moduli of materials inside the body and on their positioning. Waves scatter, diffract, bend, interact with inclusions and the scattered waves they produce can scatter off other inclusions. In general it looks like an intractable mess.

What the results of this chapter do is find a path through this mess to allow one to systematically interpret data in new ways never envisaged before.

There are two main ideas which allow us to do this. The first is the recognition that mathematically speaking the Dirichlet-to-Neumann map is the exact analog of an effective tensor in a composite material and hence many results in the theory of composites can be directly mapped over to Dirichlet-to-Neumann operators. In particular, many bounds and the techniques for obtaining them can be carried over. These bounds correlate the Dirichlet-to-Neumann map at many different frequencies with the internal geometry and moduli (see Section 5.3). Also the analytic properties of the Dirichlet-to-Neumann map as a function of the component moduli, as established in the previous two chapters, allows one to correlate either the transient response of the body, or the response of the body at a set of frequencies with the internal geometry.

The second affiliated result is the recognition that the Dirichlet-to-Neumann map, when appropriately defined, is an analytic function of the frequency ω, having a positive semidefinite imaginary part when ω has positive imaginary part. Thus it is essentially an operator-valued "Herglotz function", and has an integral representation involving an operator-valued positive semi-definite measure. Also, in the cases considered here, one can also introduce $\tau = \omega^2$ and express the response of the body as a Stieltjes function of τ. This then enables one to bound the transient behavior of the body, using the procedure detailed in the following chapter. The bounds allow one to extrapolate the measured data at a variety of frequencies or measured transient responses, to special frequencies (possibly complex) where the medium is almost homogeneous — and where information about the internal geometry is much easier to recover (see Sections 5.7, 5.8, and 5.9).

Another approach, we which we do not explore here, but which has been very successful for two-phase

composites (Day and Thorpe 1999; Day, Grant, Sievers, and Thorpe 2000; Cherkaev 2001; Zhang and Cherkaev 2009; Cherkaev and Bonifasi-Lista 2011) is to take the experimental data of the response, perhaps at various frequencies or perhaps to other time varying applied fields, and find approximations fitting this data, as nearly as possible, for the (operator-, matrix-, or scalar-valued) measure that enters the integral representation of the response functions. Then one can use relations between this measure and the geometry (sum-rules) to extract information about the geometry. The inverse problem of finding the measure from experimental measurements is notorously ill-posed (Cherkaev 2001) so some regularization has to be made. This approach has the advantage of being suitable even when there are significant errors in the measurements, although (as is common in inverse problems) it is hard to assess how accurate the result is. Test cases can help give one confidence in the method to produce useful results.

An alternative hybrid approach, introduced by McPhedran, McKenzie, and Milton (1982), is to use the fact that the bounds imply that rational functions satisfying the required analytic constraints pass through the exact data, at a given finite set of frequencies, if the rational functions are of the appropriate degree. Then, if there are errors in the measurements (with known error bars), one can generate a family of such rational functions that are compatible with the data and in this way one obtains not only information about the geometry, but also an idea of the probable error associated with these predictions. It could happen that there is no rational function compatible with the data. This could be a sign that the errors have been underestimated, or that there is some physics involved which has not been captured by the underlying equations.

We caution that our results for bounds on the transient electrical response of bodies for the full Maxwell's equations, or for the quasistatic dielectric equations, might be of limited practical utility due to the fast electrical relaxation times of many materials, especially conductors. According to Table 7.7.1 in Haus and Melcher (1989), typical relaxation times are 1.5×10^{-19} seconds for copper; 3.6×10^{-6} seconds for distilled water; 0.55 seconds for corn-oil; and 5.1×10^{4} seconds for mica. Measurements would need to be made on those time scales to capture the transient response.

This chapter assumes the reader is familiar with the contents of **Chapters 1**, **2**, and **3** although not necessarily **Chapter 4**.

5.2 Transformations

Given any relation

$$\mathcal{G}(\mathbf{x}) = \mathbf{Z}(\mathbf{x})\mathcal{F}(\mathbf{x}), \tag{5.1}$$

we are free to rotate this in the complex plane and consider

$$\begin{aligned} \widetilde{\mathcal{G}}(\mathbf{x}) &= \widetilde{\mathbf{Z}}(\mathbf{x})\widetilde{\mathcal{F}}(\mathbf{x}), \qquad & \widetilde{\mathcal{G}}(\mathbf{x}) &= e^{i\theta}\mathcal{G}(\mathbf{x}), \\ \widetilde{\mathcal{F}}(\mathbf{x}) &= e^{i\alpha}\mathcal{F}(\mathbf{x}), \qquad & \widetilde{\mathbf{Z}}(\mathbf{x}) &= e^{i(\theta-\alpha)}\mathbf{Z}(\mathbf{x}). \end{aligned} \tag{5.2}$$

Also, as observed by Cherkaev and Gibiansky (1994), when the tensor $\mathbf{Z}(\mathbf{x})$ is symmetric, but complex, any equation of the form 5.1 can be expressed in terms of its real and imaginary parts,

$$\begin{pmatrix} \mathcal{G}' \\ \mathcal{G}'' \end{pmatrix} = \begin{pmatrix} -\mathbf{Z}'' & \mathbf{Z}' \\ \mathbf{Z}' & \mathbf{Z}'' \end{pmatrix} \begin{pmatrix} \mathcal{G}'' \\ \mathcal{F}' \end{pmatrix}, \tag{5.3}$$

where the primed quantities denote real parts, while the double primed quantities denote imaginary parts, and subsequently reformulated as

$$\underbrace{\begin{pmatrix}\mathcal{F}''\\ \mathcal{G}''\end{pmatrix}}_{\underline{\mathcal{G}}(\mathbf{x})} = \underbrace{\begin{pmatrix}[\mathbf{Z}'']^{-1} & [\mathbf{Z}'']^{-1}\mathbf{Z}'\\ \mathbf{Z}'[\mathbf{Z}'']^{-1} & \mathbf{Z}''+[\mathbf{Z}'][\mathbf{Z}'']^{-1}[\mathbf{Z}']\end{pmatrix}}_{\underline{\mathbf{Z}}(\mathbf{x})}\underbrace{\begin{pmatrix}-\mathcal{G}'\\ \mathcal{F}'\end{pmatrix}}_{\underline{\mathcal{F}}(\mathbf{x})}. \tag{5.4}$$

This manipulation is useful, because $\underline{\mathbf{Z}}(\mathbf{x})$ is real and symmetric and is positive definite if $\mathbf{Z}''$ is positive definite. Cherkaev and Gibiansky (1994) noticed that the quadratic form associated with the matrix entering the constitutive law in (5.3) is saddle shaped, and could be converted to a convex quadratic form by making a partial Legendre transform which is equivalent to rewriting the constitutive law in the form (5.4). Partial Legendre transforms are well known in thermodynamics for converting convex functions to saddle-shaped functions and vice versa (see, for example, Callen 1960a). Similar manipulations can be made for other non-self-adjoint problems (Milton 1990; Fannjiang and Papanicolaou 1994; Norris 1997; Carini and Mattei 2015; see also **Chapter 14** of this book). In **Chapter 5** of Cherkaev (2000) these ideas are extended to other optimization problems. Also the key identity holds and all of the analysis applies. Note that in (5.1) we are free to replace $\mathcal{G}=\mathcal{G}'+i\mathcal{G}''$ and $\mathcal{F}=\mathcal{F}'+i\mathcal{F}''$ with $i\mathcal{G}=i\mathcal{G}'-\mathcal{G}''$ and $i\mathcal{F}=i\mathcal{F}'-\mathcal{F}''$, i.e., make the replacements

$$\mathcal{G}'\to-\mathcal{G}'',\quad \mathcal{G}''\to\mathcal{G}',\quad \mathcal{F}'\to-\mathcal{F}'',\quad \mathcal{F}''\to\mathcal{F}'. \tag{5.5}$$

Naturally this invariance extends to the equation (5.4) and so we have

$$\begin{pmatrix}\mathcal{F}'\\ \mathcal{G}'\end{pmatrix} = \underline{\mathbf{Z}}(\mathbf{x})\begin{pmatrix}\mathcal{G}''\\ -\mathcal{F}''\end{pmatrix}. \tag{5.6}$$

I learned of this breakthrough result from Gibiansky and Cherkaev during a visit to Russia in 1987, and subsequently realized (Milton 1990) that the same procedure could be used to reformulate other problems with a non-self-adjoint $\mathbf{Z}(\mathbf{x})$, having a positive definite self-adjoint part, into an equivalent problem having a self-adjoint $\underline{\mathbf{Z}}(\mathbf{x})$ that is positive definite. Consider the equation (5.1) in conjunction with a solution to the adjoint problem, $\mathcal{G}'(\mathbf{x})=\mathbf{Z}^\dagger(\mathbf{x})\mathcal{F}'(\mathbf{x})$ where $\mathbf{Z}^\dagger$ denotes the adjoint of $\mathbf{Z}$ (and the primes no longer denote real parts), and let

$$\begin{aligned}\mathbf{Z}_s&=\mathbf{Z}+\mathbf{Z}^\dagger, & \mathbf{Z}_a&=\mathbf{Z}-\mathbf{Z}^\dagger,\\ \mathcal{F}_s&=\frac{1}{2}(\mathcal{F}+\mathcal{F}'), & \mathcal{G}_s&=\frac{1}{2}(\mathcal{G}+\mathcal{G}'),\\ \mathcal{F}_a&=\frac{1}{2}(\mathcal{F}-\mathcal{F}'), & \mathcal{G}_a&=\frac{1}{2}(\mathcal{G}-\mathcal{G}').\end{aligned} \tag{5.7}$$

Then the equation with its adjoint can be reformulated as

$$\underbrace{\begin{pmatrix}\mathcal{F}_s\\ \mathcal{G}_a\end{pmatrix}}_{\underline{\mathcal{G}}(\mathbf{x})} = \underbrace{\begin{pmatrix}\mathbf{Z}_s^{-1} & -\mathbf{Z}_s^{-1}\mathbf{Z}_a\\ \mathbf{Z}_a\mathbf{Z}_s^{-1} & \mathbf{Z}_s-\mathbf{Z}_a\mathbf{Z}_s^{-1}\mathbf{Z}_a\end{pmatrix}}_{\underline{\mathbf{Z}}(\mathbf{x})}\underbrace{\begin{pmatrix}\mathcal{G}_s\\ \mathcal{F}_a\end{pmatrix}}_{\underline{\mathcal{F}}(\mathbf{x})}. \tag{5.8}$$

Again the key identity holds and all of the analysis applies.

5.3 Bounds on effective tensors applied to Dirichlet-to-Neumann maps

Now $\mathbf{Z}(\mathbf{x})$ also depends on frequency ω, i.e., $\mathbf{Z}(\mathbf{x},\omega)$, and hence the associated $\mathbf{L}(\mathbf{x})$ also depends on frequency, i.e., $\mathbf{L}(\mathbf{x},\omega)$. Suppose we take measurements of the Dirichlet-to-Neumann map at frequencies $\omega_1,\omega_2,\ldots,\omega_s$. Lets make the transformations described in Section 5.2 to $\underline{\mathbf{Z}}(\mathbf{x},\omega_1),\ldots,\underline{\mathbf{Z}}(\mathbf{x},\omega_s)$, with associated Hermitian positive definite tensors $\underline{\mathbf{L}}(\mathbf{x},\omega_1),\ldots,\underline{\mathbf{L}}(\mathbf{x},\omega_s)$. The equations are now

$$\underline{\mathbf{J}}_0(\mathbf{x},\omega)+\underline{\mathbf{J}}(\mathbf{x},\omega)=\underline{\mathbf{L}}(\mathbf{x},\omega)[\underline{\mathbf{E}}_0(\mathbf{x},\omega)+\underline{\mathbf{E}}(\mathbf{x},\omega)], \tag{5.9}$$

with

$$\underline{\mathbf{J}}_0,\underline{\mathbf{E}}_0\in\underline{\mathcal{U}},\quad \underline{\mathbf{E}}\in\underline{\mathcal{E}},\quad \underline{\mathbf{J}}\in\underline{\mathcal{J}}, \tag{5.10}$$

where $\underline{\mathcal{U}}$, $\underline{\mathcal{E}}$, $\underline{\mathcal{J}}$ are defined, given some choice of reference medium $\underline{\mathbf{Z}}_0$, in a similar way the spaces $\mathcal{U}$, $\mathcal{E}$, and $\mathcal{J}$ are defined in Section 3.2 given the reference medium $\mathbf{Z}_0$. Now $\underline{\mathbf{E}}$ and $\underline{\mathbf{J}}$ each have some components that derive from potentials, and some components that are associated with fluxes.

Then, when the tensors $\underline{\mathbf{L}}(\mathbf{x},\omega_i)$ are Hermitian and positive definite, the simplest uncoupled bounds (see Section 2.7 in this book or, for example, Section 13.1 and **Chapter 22** of Milton 2002), are the classical bounds

$$\begin{aligned}\underline{\mathbf{L}}_*(\omega_i) &\geq 0,\\ \underline{\mathbf{L}}_*(\omega_i) &\leq \boldsymbol{\Gamma}_0\underline{\mathbf{L}}(\mathbf{x},\omega_i)\boldsymbol{\Gamma}_0,\\ [\underline{\mathbf{L}}_*(\omega_i)]^{-1} &\leq \boldsymbol{\Gamma}_0\underline{\mathbf{L}}^{-1}(\mathbf{x},\omega_i)\boldsymbol{\Gamma}_0,\end{aligned} \tag{5.11}$$

where $\boldsymbol{\Gamma}_0$ is the projection onto the space $\underline{\mathcal{U}}$. Note that since $\mathbf{Z}_0$ is given, one can in principle numerically calculate the fields in the space $\underline{\mathcal{U}}$: thus the operator $\boldsymbol{\Gamma}_0$ can be considered to be known.

The idea is to use the second and third bounds in an inverse fashion to say something about $\underline{\mathbf{L}}(\mathbf{x})$. The bounds tell us something about

$$\int_\Omega \underline{\mathbf{U}}_1(\mathbf{x})\underline{\mathbf{L}}(\mathbf{x},\omega_i)\underline{\mathbf{U}}_2(\mathbf{x}),\qquad \int_\Omega \underline{\mathbf{U}}_1(\mathbf{x})\,[\underline{\mathbf{L}}(\mathbf{x},\omega_i)]^{-1}\,\underline{\mathbf{U}}_2(\mathbf{x}), \tag{5.12}$$

as $\underline{\mathbf{U}}_1,\underline{\mathbf{U}}_2$ vary over all fields in $\underline{\mathcal{U}}$. It might be the case that we know something about how $\mathbf{L}(\mathbf{x},\omega)$ and hence $\underline{\mathbf{L}}(\mathbf{x},\omega)$ depend on frequency. For instance, for the quasistatic dielectric problem (or for electromagnetism), with isotropic constituent materials, it is common to assume the complex dielectric constant at low frequencies takes the form

$$\varepsilon(\mathbf{x},\omega)=\varepsilon_R(\mathbf{x})+\frac{i\sigma(\mathbf{x})}{\omega}. \tag{5.13}$$

So measurements at different frequencies will give us information on $\varepsilon_R(\mathbf{x})$ and $\sigma(\mathbf{x})$.

The second bound in (5.11) is also a direct consequence of a variational principle obtained by Milton, Seppecher, and Bouchitté (2009). that generalizes the variational principles Cherkaev and Gibiansky (1994) obtained for quasistatics. [The breakthrough observation which led to this generalization was the recognition that these equations could be written in the forms (3.7), (3.8), and (3.77) where the key identity holds and, after possibly multiplying it by complex number, the tensor entering the constitutive relation has a positive semi-definite imaginary part.] The variational principle (in the absence of any source term) involves the function

$$Y(\widetilde{\mathbf{u}}',\widetilde{\mathbf{G}}')=\int_\Omega \begin{pmatrix}\widetilde{\mathcal{F}}'\\ -\widetilde{\mathcal{G}}'\end{pmatrix}\cdot\underline{\mathbf{Z}}\begin{pmatrix}\widetilde{\mathcal{F}}'\\ -\widetilde{\mathcal{G}}'\end{pmatrix}, \tag{5.14}$$

where $\widetilde{\mathcal{F}}'$ and $\widetilde{\mathcal{G}}'$ are given by

$$\widetilde{\mathcal{F}}' = \begin{pmatrix} \nabla \widetilde{\mathbf{u}}' \\ \widetilde{\mathbf{u}}' \end{pmatrix}, \quad \widetilde{\mathcal{G}}' = \begin{pmatrix} \widetilde{\mathbf{G}}' \\ \nabla \cdot \widetilde{\mathbf{G}}' \end{pmatrix}, \tag{5.15}$$

and the real-valued trial fields $\widetilde{\mathbf{u}}'$ and $\widetilde{\mathbf{G}}'$ satisfy the boundary conditions

$$\widetilde{\mathbf{u}}' = \mathbf{u}'_0, \quad \mathbf{n} \cdot \widetilde{\mathbf{G}}' = \mathbf{t}'_0 \quad \text{for } \mathbf{x} \in \partial\Omega. \tag{5.16}$$

In terms of these functions and fields the variational principle is

$$Y(\mathbf{u}', \mathbf{G}') = \inf_{\widetilde{\mathbf{u}}'} \inf_{\widetilde{\mathbf{G}}'} Y(\widetilde{\mathbf{u}}', \widetilde{\mathbf{G}}'), \tag{5.17}$$

where $\mathbf{u}'$ and $\mathbf{G}'$ satisfy the boundary conditions $\mathbf{u}' = \mathbf{u}'_0$ and $\mathbf{n} \cdot \mathbf{G}' = \mathbf{t}'_0$ on $\partial\Omega$ and are the real parts of complex fields $\mathbf{u}$ and $\mathbf{G}$ solving

$$\begin{pmatrix} \mathbf{G} \\ \nabla \cdot \mathbf{G} \end{pmatrix} = \mathbf{Z} \begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix}, \tag{5.18}$$

with $\mathbf{Z}(\mathbf{x})$ being complex. We could take trial fields so that $\widetilde{\mathcal{F}}' = \mathcal{F}'_0$ and $\widetilde{\mathcal{G}}' = \mathcal{G}'_0$ where the fields $\mathcal{F}'_0$ and $\mathcal{G}'_0$ are the real parts of fields $\mathcal{F}_0 = \mathcal{F}'_0 + i\mathcal{F}''_0$ and $\mathcal{G}_0 = \mathcal{G}'_0 + i\mathcal{G}''_0$ such that

$$\mathcal{F}_0 = \begin{pmatrix} \nabla \mathbf{u}_0 \\ \mathbf{u}_0 \end{pmatrix}, \quad \mathcal{G}_0 = \begin{pmatrix} \mathbf{G}_0 \\ \nabla \mathbf{G}_0 \end{pmatrix}, \tag{5.19}$$

and

$$\underbrace{\begin{pmatrix} \mathcal{F}''_0 \\ \mathcal{G}''_0 \end{pmatrix}}_{\underline{\mathcal{G}}_0(\mathbf{x})} = \underline{\mathbf{Z}}_0 \underbrace{\begin{pmatrix} -\mathcal{G}'_0 \\ \mathcal{F}'_0 \end{pmatrix}}_{\underline{\mathcal{F}}_0(\mathbf{x})}. \tag{5.20}$$

We could set our boundary conditions (5.16) to match those of these given fields. Then we have

$$Y(\mathbf{u}'_0, \mathbf{G}'_0) = \int_\Omega \underline{\mathbf{U}}_0 \cdot \underline{\mathbf{L}}\, \underline{\mathbf{U}}_0, \tag{5.21}$$

where

$$\underline{\mathbf{U}}_0 = \underline{\mathbf{Z}}_0^{-1/2} \underline{\mathcal{G}}_0(\mathbf{x}) = \underline{\mathbf{Z}}_0^{1/2} \underline{\mathcal{F}}_0(\mathbf{x}) \in \underline{\mathcal{U}}, \quad \underline{\mathbf{L}}(\mathbf{x}) = \underline{\mathbf{Z}}_0^{-1/2} \underline{\mathbf{Z}}(\mathbf{x}) \underline{\mathbf{Z}}_0^{-1/2}. \tag{5.22}$$

With a bit more work, using an argument analogous to the way the identity (3.45) was obtained, it can be shown that

$$Y(\mathbf{u}', \mathbf{G}') = \int_\Omega \underline{\mathbf{U}}_0(\mathbf{x}) \cdot [\underline{\mathbf{L}}_* \, \underline{\mathbf{U}}_0](\mathbf{x}), \tag{5.23}$$

where to obtain $[\underline{\mathbf{L}}_* \, \underline{\mathbf{U}}_0](\mathbf{x})$ we need to apply the operator $\underline{\mathbf{L}}_*$ to the field $\underline{\mathbf{U}}_0$, and then evaluate the resulting field at the point $\mathbf{x}$. Hence the bound

$$Y(\mathbf{u}', \mathbf{G}') \le Y(\mathbf{u}'_0, \mathbf{G}'_0), \tag{5.24}$$

implied by the variational principle (5.17) implies the second bound in (5.11). Similarly the third inequality in (5.11) follows from the variational principle which is dual to (5.17), given in equation (2.51) in Milton, Seppecher, and Bouchitté (2009).

Of course the variational principles (5.17) allow a much wider choice of trial field, and consequently should yield even tighter bounds on the moduli when used in an inverse manner as discussed in Section 6 of Milton, Seppecher, and Bouchitté (2009) [except that the last bound in that paper, equation (6.8), is not useful at all, since the right side minus the left side is always positive no matter what the geometry]. This method for obtaining information about the geometry is similar to one proposed by Berryman and Kohn (1990) in the context of the conductivity equations. A variant of it has been implemented numerically by Borcea, Gray, and Zhang (2003) with some limited success. The inverse problem for the conductivity problem, known as electrical impedance tomography, is notoriously ill-posed and so we would hope for better resolution when one probes the medium with waves provided the medium is not too lossy (for electromagnetism) or viscous (for acoustics or elastodynamics). We remark that the variational principles of Cherkaev and Gibiansky (1994), in conjunction with the translation method (described in Section 5.5 below), have been used by Kang, Kim, Lee, Li, and Milton (2014) to obtain very tight bounds on the volume fraction of the phases in a two phase, two-dimensional body from a few measurements of complex potential and complex flux, associated with the quasistatic conductivity problem at a fixed frequency (when one or both of the two component conductivities are complex): see also Thaler and Milton (2015). For two-phase composite materials, bounds on the volume fraction have been obtained from known values of the effective complex dielectric constant, and the complex dielectric constants of the two-phases by McPhedran, McKenzie, and Milton (1982), McPhedran and Milton (1990), and Cherkaev and Golden (1998).

5.4 Choosing the reference tensor $\mathbf{Z}_0$

To gain some insight into how one may choose the reference tensor $\mathbf{Z}_0$ it is helpful to look for solutions to the equations

$$\begin{pmatrix} \mathbf{G} \\ \nabla \cdot \mathbf{G} \end{pmatrix} = \mathbf{Z}_0 \begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix}, \tag{5.25}$$

with a positive definite reference tensor $\mathbf{Z}_0$ of the form

$$\mathbf{Z}_0 = \begin{pmatrix} \lambda_1 \mathbf{I} & 0 \\ 0 & \lambda_2 \mathbf{I} \end{pmatrix}, \tag{5.26}$$

where the eigenvalues λ_1 and λ_2 are both positive. We then have

$$\mathbf{G} = \lambda_1 \nabla \mathbf{u}, \quad \nabla \cdot \mathbf{G} = \lambda_2 \mathbf{u}, \tag{5.27}$$

implying

$$\nabla^2 \mathbf{u} = (\lambda_2/\lambda_1)\mathbf{u}. \tag{5.28}$$

This, for example, has the solution

$$\mathbf{u} = \mathbf{v}_0 e^{\mathbf{k}_0 \cdot \mathbf{x}}, \quad \text{where } \mathbf{k}_0 = \mathbf{n}_0/\ell, \quad \ell = \sqrt{\lambda_1/\lambda_2}, \tag{5.29}$$

where $\mathbf{v}_0$ and $\mathbf{k}_0$ are constant vectors, and $\mathbf{n}_0$ is a unit vector. These solutions decay exponentially with a characteristic length $\ell = \sqrt{\lambda_1/\lambda_2}$. If ℓ is much smaller than the diameter of the body Ω then for most boundary conditions the fields $\mathbf{U}(\mathbf{x})$ associated with them are going to decay rapidly away from the boundary.

Similarly, if associated with the equation (5.4) we take a reference tensor

$$\underline{\mathbf{Z}}_0 = \begin{pmatrix} \mathbf{Z}_0^{-1} & 0 \\ 0 & \mathbf{Z}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{I}/\lambda_1 & 0 & 0 & 0 \\ 0 & \mathbf{I}/\lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_1\mathbf{I} & 0 \\ 0 & 0 & 0 & \lambda_2\mathbf{I} \end{pmatrix}, \tag{5.30}$$

then the characteristic length scale will still be ℓ. If ℓ is much smaller than the body Ω then quantities like those in (5.12) will be fairly insensitive to the values of $\underline{\mathbf{L}}(\mathbf{x}, \omega_i)$ that are well away from the edges: we can expect that generally the fields $\underline{\mathbf{U}}_1(\mathbf{x})$ and $\underline{\mathbf{U}}_2(\mathbf{x})$ will decay away from the boundary with a skin depth of the order of ℓ. Therefore we should choose a reference tensor $\underline{\mathbf{Z}}_0$ with λ larger than the size of the body.

However, since the deep interior of a body is most easily probed with waves it might be best to choose a reference tensor $\underline{\mathbf{Z}}_0$ which corresponds with a $\mathbf{Z}_0$ that is close to those associated with waves such as

$$\mathbf{Z}_0 = \begin{pmatrix} (-z_1 + i\delta_1)\mathbf{I} & 0 \\ 0 & (z_2 + i\delta_2)\mathbf{I} \end{pmatrix}, \tag{5.31}$$

where z_1 and z_2 are positive, while δ_1 and δ_2 are positive and fairly small. Then the fields (5.29) still solve the equations with a complex value for ℓ:

$$\ell = \ell' + i\ell'' = \sqrt{\lambda_1/\lambda_2} = \sqrt{(-z_1 + i\delta_1)/(z_2 + i\delta_2)}. \tag{5.32}$$

It might be useful to choose parameters such that the wavelength ℓ' is smaller than the size of the objects inside the body we wish to probe, while the attenuation length ℓ'' is larger than the size of the body. The associated reference tensor $\underline{\mathbf{Z}}_0$ will according to (5.4) be

$$\underline{\mathbf{Z}}_0 = \begin{pmatrix} \mathbf{I}/\delta_1 & 0 & -z_1\mathbf{I}/\delta_1 & 0 \\ 0 & \mathbf{I}/\delta_2 & 0 & z_2\mathbf{I}/\delta_2 \\ -z_1\mathbf{I}/\delta_1 & 0 & (\delta_1 + z_1^2/\delta_1)\mathbf{I} & 0 \\ 0 & z_2\mathbf{I}/\delta_2 & 0 & (\delta_2 + z_2^2/\delta_2)\mathbf{I} \end{pmatrix}. \tag{5.33}$$

5.5 Translation method bounds on Dirichlet-to-Neumann maps

The elementary bounds (5.11) can be improved by using the "Translation Method", which is one of the most powerful methods for bounding the effective tensors of composites. It is also known as the method of compensated compactness when it is used to say something about products of sequences of fields satisfying differential constraints, as arises in the theory of composites when one treats the composite as a sequence of materials, with possibly finer and finer microstructure. The method was introduced by Murat and Tartar (Tartar 1979b; Murat and Tartar 1985; Tartar 1985); see in particular theorem 8 of Tartar (1979b), and independently by Lurie and Cherkaev (1982, 1984). For nonlinear media the two approaches give different types of bounds, as discussed in Section 25.1 of Milton (2002): the compensated compactness method of Murat and Tartar gives bounds on the average fields, while the approach of Lurie and Cherkaev gives bounds on the energy. Since both approaches yield identical results for linear media, the term translation method [introduced in Milton (1990)] will be used to encompass both. The name arises because the bounds can be obtained by shifting, that is, translating, the tensor field by a constant tensor and applying the classical bounds. The method is discussed

in the books of Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009) and references therein.

The translation method was extended to the problem of bounding the Dirichlet-to-Neumann map of bodies, or more precisely in a two phase body to the problem of bounding the volume occupied by an inclusion phase in a series of papers by Kang, Kim, and Milton (2012), Kang, Milton, and Wang (2014), Milton and Nguyen (2012), Kang, Milton, and Wang (2014), Kang and Milton (2013), and Kang, Kim, Lee, Li, and Milton (2014): see also the related splitting method employed to bound the volume of an inclusions by Milton and Nguyen (2012) and Thaler and Milton (2015). At that time no one had made the connection between the Dirichlet-to-Neumann map, and the effective tensor of an associated abstract problem in the theory of composites as we did in **Chapter 3**.

In its simplest form, given a solution to say the equations (5.9) and (5.10) one looks for a constant tensor $\underline{\underline{\mathbf{T}}}$ such that

$$\underline{\underline{\mathbf{T}}}\,\underline{\mathcal{U}} \subset \underline{\mathcal{U}} \quad \text{and} \quad \underline{\underline{\mathbf{T}}}\,\underline{\mathcal{E}} \subset \underline{\mathcal{J}}, \tag{5.34}$$

by which we mean the fields

$$\underline{\mathbf{U}}'(\mathbf{x}) = \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{U}}(\mathbf{x}), \quad \underline{\mathbf{J}}'(\mathbf{x}) = \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{E}}(\mathbf{x}), \tag{5.35}$$

lie respectively in $\underline{\mathcal{U}}$ and $\underline{\mathcal{J}}$, for all fields $\underline{\mathbf{U}}$ and $\underline{\mathbf{E}}$ that lie respectively in $\underline{\mathcal{U}}$ and $\underline{\mathcal{E}}$.

Then using the fields which solve the equations (5.9) and (5.10) and taking $\underline{\underline{\mathbf{L}}}_t(\mathbf{x},\omega) = \underline{\underline{\mathbf{L}}}(\mathbf{x},\omega) - \underline{\underline{\mathbf{T}}}$ as our translated medium, we easily construct a solution to the equations in this medium:

$$\underline{\mathbf{J}}_{t0}(\mathbf{x},\omega) + \underline{\mathbf{J}}_t(\mathbf{x},\omega) = \underline{\underline{\mathbf{L}}}_t(\mathbf{x},\omega)[\underline{\mathbf{E}}_0(\mathbf{x},\omega) + \underline{\mathbf{E}}(\mathbf{x},\omega)], \tag{5.36}$$

with

$$\underline{\mathbf{J}}_{t0} = \underline{\mathbf{J}}_0 - \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{E}}_0 \in \underline{\mathcal{U}}, \quad \underline{\mathbf{J}}_t = \underline{\mathbf{J}} - \underline{\underline{\mathbf{T}}}\,\underline{\mathbf{E}} \in \underline{\mathcal{J}}. \tag{5.37}$$

This implies the effective tensor gets translated as

$$\underline{\underline{\mathbf{L}}}_{t*}(\omega) = \underline{\underline{\mathbf{L}}}_*(\omega) - \underline{\underline{\mathbf{T}}}_0, \tag{5.38}$$

where $\underline{\underline{\mathbf{T}}}_0$ is the operator mapping $\underline{\mathcal{U}}$ onto $\underline{\mathcal{U}}$ whose action is defined by the first equation in (5.35). If $\underline{\underline{\mathbf{T}}}$ is chosen to be Hermitian and so $\underline{\underline{\mathbf{L}}}_t(\mathbf{x}) = \underline{\underline{\mathbf{L}}}(\mathbf{x}) - \underline{\underline{\mathbf{T}}}$ is positive semidefinite then the classical bounds (5.11) applied to the translated medium imply

$$\underline{\underline{\mathbf{L}}}_*(\omega) - \underline{\underline{\mathbf{T}}} \geq 0, \quad \underline{\underline{\boldsymbol{\Gamma}}}_0[\underline{\underline{\mathbf{L}}}(\mathbf{x},\omega) - \underline{\underline{\mathbf{T}}}]^{-1}\underline{\underline{\boldsymbol{\Gamma}}}_0 \geq [\underline{\underline{\mathbf{L}}}_*(\mathbf{x},\omega) - \underline{\underline{\mathbf{T}}}_0]^{-1}, \tag{5.39}$$

which are known as the comparison bounds and translation bounds, respectively. Such translations are known as null-Lagrangians, as when one writes the minimizing variational principle for the equation, the quadratic form associated with $\underline{\underline{\mathbf{T}}}$ does not influence the Euler–Lagrange equation. It is also possible to use a wider class of translations whose quadratic forms are quasiconvex (Tartar 1979b; Murat and Tartar 1985; Tartar 1985) or Q_C^*-convex translations (see Section 1.5, **Chapter 13**, and Milton 2013b) but we do not explore this here: see, however, Kang and Milton (2013) for related investigations using quasiconvex functions.

Now with the choice (5.30) of reference tensor, with $\underline{\mathcal{F}}(\mathbf{x})$ defined by (5.4), consider a field in the space $\underline{\mathcal{E}}$ given by

$$\underbrace{\begin{pmatrix} -\mathbf{J}'(\mathbf{x}) \\ \mathbf{E}'(\mathbf{x}) \end{pmatrix}}_{\underline{\mathbf{E}}} \equiv \underline{\underline{\mathbf{Z}}}_0^{1/2}\underline{\mathcal{F}(\mathbf{x})} = \begin{pmatrix} -\mathbf{G}'/\sqrt{\lambda_1} \\ -\nabla\cdot\mathbf{G}'/\sqrt{\lambda_2} \\ \sqrt{\lambda_1}\nabla\mathbf{u}' \\ \sqrt{\lambda_2}\mathbf{u}' \end{pmatrix}, \tag{5.40}$$

with $\mathbf{u}' = 0$ and $\mathbf{n} \cdot \mathbf{G}' = 0$ on $\partial\Omega$. The matrix

$$\underline{\mathbf{T}} = \begin{pmatrix} 0 & t_1\mathbf{I} \\ t_2\mathbf{I} & 0 \end{pmatrix}, \tag{5.41}$$

where t_1 and t_2 are possibly complex numbers, has the property that for $\underline{\mathbf{E}} \in \underline{\mathcal{E}}$

$$\begin{pmatrix} 0 & t_1\mathbf{I} \\ t_2\mathbf{I} & 0 \end{pmatrix} \underline{\mathbf{E}} = \begin{pmatrix} t_1\mathbf{E}'(\mathbf{x}) \\ -t_2\mathbf{J}'(\mathbf{x}) \end{pmatrix} = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} \nabla(t_1\mathbf{u}') \\ t_1\mathbf{u}' \\ -t_2\mathbf{G}' \\ -\nabla\cdot(t_2\mathbf{G}') \end{pmatrix} \in \underline{\mathcal{J}}. \tag{5.42}$$

Also the space $\underline{\mathcal{U}}$ is constructed by looking for fields

$$\underline{\mathcal{G}}(\mathbf{x}) = \begin{pmatrix} \nabla\mathbf{u}'' \\ \mathbf{u}' \\ \mathbf{G}'' \\ \nabla\cdot\mathbf{G}'' \end{pmatrix}, \quad \underline{\mathcal{F}}(\mathbf{x}) = \begin{pmatrix} -\mathbf{G}' \\ -\nabla\cdot\mathbf{G}' \\ \nabla\mathbf{u}' \\ \mathbf{u}' \end{pmatrix}, \tag{5.43}$$

such that $\underline{\mathcal{G}}(\mathbf{x}) = \underline{\mathbf{Z}}_0\,\underline{\mathcal{F}}(\mathbf{x})$, which decouples into the equations,

$$\underbrace{\begin{pmatrix} -\mathbf{G}' \\ -\nabla\cdot\mathbf{G}' \end{pmatrix}}_{-\mathcal{G}'} = \mathbf{Z}_0 \underbrace{\begin{pmatrix} \nabla\mathbf{u}'' \\ \mathbf{u}' \end{pmatrix}}_{\mathcal{F}''}, \quad \underbrace{\begin{pmatrix} \mathbf{G}'' \\ \nabla\cdot\mathbf{G}'' \end{pmatrix}}_{\mathcal{G}''} = \mathbf{Z}_0 \underbrace{\begin{pmatrix} \nabla\mathbf{u}' \\ \mathbf{u}' \end{pmatrix}}_{\mathcal{F}'}, \tag{5.44}$$

where $\mathbf{Z}_0$ is given by (5.26). Consequently $\underline{\mathcal{U}}$ consists of fields

$$\underline{\mathbf{U}} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix}, \quad \text{with } \mathbf{U}_1, \mathbf{U}_2 \in \mathcal{U}, \tag{5.45}$$

where $\mathcal{U}$ is the space associated with $\mathbf{Z}_0$. Due to (5.42) and (5.45) the tensor $\underline{\mathbf{T}}$ given by (5.41) has the desired properties (5.34) of a translation, and the bounds (5.39) apply when $t_2 = \overline{t_1}$ and $\underline{\mathbf{L}}(\mathbf{x}) - \underline{\mathbf{T}} \geq 0$ for all $\mathbf{x}$.

More generally, with this choice (5.30) of reference tensor (without using the specific differential constraints on the fields), we can obtain translation method bounds which couple effective tensors by looking at

$$\mathbf{M}_* = \begin{pmatrix} \underline{\mathbf{L}}_*(\omega_1) & & & & & & \\ & \ddots & & & & & \\ & & \underline{\mathbf{L}}_*(\omega_k) & & & & \\ & & & [\underline{\mathbf{L}}_*(\omega_{k+1})]^{-1} & & & \\ & & & & \ddots & & \\ & & & & & [\underline{\mathbf{L}}_*(\omega_s)]^{-1} \end{pmatrix}, \tag{5.46}$$

where the blocks off the diagonal are zero, and the frequencies $\omega_1, \omega_2, \ldots, \omega_k, \omega_{k+1}, \ldots, \omega_s$ are not necessarily ordered, and not necessarily even distinct (they could even all be equal to a single frequency ω_0) if we are interested in bounds that couple the responses to different applied fields at the same frequency. For static

composite material problems (where $\omega_0 = 0$) such extended matrices (supertensors) $\mathbf{M}_*$, involving effective tensors and their inverses, were first introduced by Cherkaev and Gibiansky (1992, 1993) for bounding the possible moduli of two-phase composites as the geometry varies over all configurations. In fact, for static and quasistatic composite material problems, if one is given a set U of n component materials with tensors $\underline{\mathbf{L}}_1$, $\underline{\mathbf{L}}_2$, ..., $\underline{\mathbf{L}}_n$ and one obtains sharp microstructure independent bounds from below on the quadratic form associated with such supertensors then one can completely determine the set of all possible effective tensors $\underline{\mathbf{L}}_*$ as the microstructure varies, which is known as the G-closure of U, GU (Francfort and Milton 1994; Milton 1994; Milton and Cherkaev 1995; **Chapter 30** of Milton (2002)). In other words G-closures are characterized by the minimums of these quadratics forms in much the same way that convex sets can be characterized as the envelope of their tangent planes, i.e., through Legendre transforms.

Also we let

$$\mathbf{M}(\mathbf{x}) = \begin{pmatrix} \underline{\mathbf{L}}(\mathbf{x},\omega_1) & & & & & \\ & \ddots & & & & \\ & & \underline{\mathbf{L}}(\mathbf{x},\omega_k) & & & \\ & & & [\underline{\mathbf{L}}(\mathbf{x},\omega_{k+1})]^{-1} & & \\ & & & & \ddots & \\ & & & & & [\underline{\mathbf{L}}(\mathbf{x},\omega_s)]^{-1} \end{pmatrix}, \tag{5.47}$$

where again the blocks off the diagonal are zero. Then we take a "translation" $\mathbf{T}$ of the form (actually its a null-Lagrangian)

$$\mathbf{T} = \begin{pmatrix} \boldsymbol{\alpha}_{11} & \dots & \boldsymbol{\alpha}_{1k} & \boldsymbol{\beta}_{11} & \dots & \boldsymbol{\beta}_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_{k1} & \dots & \boldsymbol{\alpha}_{kk} & \boldsymbol{\beta}_{k1} & \dots & \boldsymbol{\beta}_{km} \\ \overline{\boldsymbol{\beta}}_{11} & \dots & \overline{\boldsymbol{\beta}}_{k1} & \boldsymbol{\gamma}_{11} & \dots & \boldsymbol{\gamma}_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{\boldsymbol{\beta}}_{1m} & \dots & \overline{\boldsymbol{\beta}}_{km} & \boldsymbol{\gamma}_{m1} & \dots & \boldsymbol{\gamma}_{mm} \end{pmatrix}, \tag{5.48}$$

where $m = s - k$ and the matrices $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_{ij}$, and $\boldsymbol{\gamma}_{ij}$ take the block form

$$\boldsymbol{\alpha}_{ij} = \begin{pmatrix} 0 & \alpha_{ij}^{(1)}\mathbf{I} \\ \alpha_{ij}^{(2)}\mathbf{I} & 0 \end{pmatrix}, \quad \boldsymbol{\beta}_{ij} = \begin{pmatrix} \beta_{ij}^{(1)}\mathbf{I} & 0 \\ 0 & \beta_{ij}^{(2)}\mathbf{I} \end{pmatrix}, \qquad \boldsymbol{\gamma}_{ij} = \begin{pmatrix} 0 & \gamma_{ij}^{(1)}\mathbf{I} \\ \gamma_{ij}^{(2)}\mathbf{I} & 0 \end{pmatrix}, \tag{5.49}$$

and satisfy

$$\boldsymbol{\alpha}_{ji} = \overline{\boldsymbol{\alpha}}_{ij}, \quad \boldsymbol{\gamma}_{ji} = \overline{\boldsymbol{\gamma}}_{ij}, \tag{5.50}$$

to ensure that $\mathbf{T}$ is Hermitian. These matrices $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_{ij}$, and $\boldsymbol{\gamma}_{ij}$ need to be chosen so that $\mathbf{M}(\mathbf{x}) - \mathbf{T} \geq 0$, which requires us to have some information on the maximum and minimum moduli inside the body. Let $\mathbf{T}_0$ be that operator acting on $\mathcal{U}^{2s}$ such that if $\mathbf{U}' = \mathbf{T}_0\mathbf{U}$ with $\mathbf{U}'$, $\mathbf{U} \in \mathcal{U}^{2s}$ then $\mathbf{U}'(\mathbf{x}) = \mathbf{T}\mathbf{U}(\mathbf{x})$ for all $\mathbf{x}$.

Then we have the "comparison bounds" $\mathbf{M}_* \geq \mathbf{T}_0$ for all $\mathbf{T}_0$ satisfying the constraint that $\mathbf{M}(\mathbf{x}) - \mathbf{T} \geq 0$ and we have the "translation bounds":

$$\boldsymbol{\Gamma}_0 \left[\mathbf{M}(\mathbf{x}) - \mathbf{T}\right]^{-1} \boldsymbol{\Gamma}_0 \geq \left[\mathbf{M}_* - \mathbf{T}_0\right]^{-1}. \tag{5.51}$$

Again the idea is that the right hand side can be approximately measured and this places bounds on the left hand side, and hence gives us direct information about the geometry inside the body.

The condition that $\mathbf{T}$ be chosen so that $\mathbf{M}(\mathbf{x}) - \mathbf{T} \geq 0$ for all $\mathbf{x}$ is easiest to determine if we have a n-phase body and we know the moduli of the phases at the frequencies $\omega_1, \omega_2, \ldots, \omega_s$. Then the condition will be satisfied if the matrices $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_{ij}$, and $\boldsymbol{\gamma}_{ij}$ are chosen so that

$$\mathbf{M}_k - \mathbf{T} \geq 0 \quad \text{for } k = 1, 2, \ldots, N, \tag{5.52}$$

where the $\mathbf{M}_k$ are the assumed known moduli of phase k.

With the choice (5.33) of reference tensor we have to take more care. In the context of the equations (5.9) and (5.10). Lets look for a translation of the form

$$\underline{\mathbf{T}} = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} 0 & t\mathbf{I} \\ s\mathbf{I} & 0 \end{pmatrix} \underline{\mathbf{Z}}_0^{-1/2}. \tag{5.53}$$

Fields in the space $\underline{\mathcal{U}}$ have the form

$$\underline{\mathbf{U}} = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} (\mathcal{F}_0)'' \\ (\mathcal{G}_0)'' \end{pmatrix} = \underline{\mathbf{Z}}_0^{+1/2} \begin{pmatrix} -(\mathcal{G}_0)' \\ (\mathcal{F}_0)' \end{pmatrix}, \tag{5.54}$$

where the subfields entering these equations satisfy the usual differential constraints. Due to the equivalence between (5.6) and (5.4), (5.54) implies

$$\underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} (\mathcal{F}_0)' \\ (\mathcal{G}_0)' \end{pmatrix} = \underline{\mathbf{Z}}_0^{+1/2} \begin{pmatrix} (\mathcal{G}_0)'' \\ -(\mathcal{F}_0)'' \end{pmatrix} \in \underline{\mathcal{U}}. \tag{5.55}$$

Now multiplying (5.54) on the left by $\underline{\mathbf{T}}$ gives

$$\underline{\mathbf{T}}\,\underline{\mathbf{U}} = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} t(\mathcal{F}_0)' \\ -s(\mathcal{G}_0)' \end{pmatrix}, \tag{5.56}$$

which from (5.55) will lie in $\underline{\mathcal{U}}$ provided we choose $s = -t$. ($\underline{\mathbf{T}}$ will then be Hermitian if t is proportional to i). With this choice of $\underline{\mathbf{T}}$, let us look at its action on a field $\underline{\mathbf{E}} \in \underline{\mathcal{E}}$ given by

$$\underline{\mathbf{E}} = \underline{\mathbf{Z}}_0^{1/2}\underline{\mathbf{F}}, \quad \text{with } \underline{\mathbf{F}} = \begin{pmatrix} -\mathbf{G}' \\ -\nabla \cdot \mathbf{G}' \\ \nabla \mathbf{u}' \\ \mathbf{u}' \end{pmatrix}, \tag{5.57}$$

with $\mathbf{n} \cdot \mathbf{G}' = 0$ and $\mathbf{u}' = 0$ on $\partial\Omega$. We have

$$\underline{\mathbf{T}}\,\underline{\mathbf{E}} = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} 0 & t\mathbf{I} \\ -t\mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} -\mathbf{G}' \\ -\nabla \cdot \mathbf{G}' \\ \nabla \mathbf{u}' \\ \mathbf{u}' \end{pmatrix} = t\underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} \nabla \mathbf{u}' \\ \mathbf{u}' \\ \mathbf{G}' \\ \nabla \cdot \mathbf{G}' \end{pmatrix} \in \underline{\mathcal{J}}. \tag{5.58}$$

In summary we have

$$\underline{\mathbf{T}}_1\underline{\mathcal{U}} \subset \underline{\mathcal{U}}, \quad \underline{\mathbf{T}}_1\underline{\mathcal{E}} \subset \underline{\mathcal{J}}, \quad \text{with } \underline{\mathbf{T}}_1 = \underline{\mathbf{Z}}_0^{-1/2} \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \underline{\mathbf{Z}}_0^{-1/2}. \tag{5.59}$$

Similarly it follows from the same line of reasoning (swapping the roles of $\underline{\mathcal{E}}$ and $\underline{\mathcal{J}}$, and the roles of $\underline{\mathbf{Z}}_0$ and $\underline{\mathbf{Z}}_0^{-1}$) that

$$\underline{\mathbf{T}}_2\underline{\mathcal{U}} \subset \underline{\mathcal{U}}, \quad \underline{\mathbf{T}}_2\underline{\mathcal{J}} \subset \underline{\mathcal{E}}, \quad \text{with } \underline{\mathbf{T}}_2 = \underline{\mathbf{Z}}_0^{1/2}\begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}\underline{\mathbf{Z}}_0^{1/2}. \tag{5.60}$$

Therefore the bounds (5.51) still hold with a choice of translation $\mathbf{T}$ of the form

$$\mathbf{T} = \begin{pmatrix} \alpha_{11}\underline{\mathbf{T}}_1 & \dots & \alpha_{1k}\underline{\mathbf{T}}_1 & \beta_{11}\mathbf{I} & \dots & \beta_{1m}\mathbf{I} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{k1}\underline{\mathbf{T}}_1 & \dots & \alpha_{kk}\underline{\mathbf{T}}_1 & \beta_{k1}\mathbf{I} & \dots & \beta_{km}\mathbf{I} \\ \overline{\beta}_{11}\mathbf{I} & \dots & \overline{\beta}_{k1}\mathbf{I} & \gamma_{11}\underline{\mathbf{T}}_2 & \dots & \gamma_{1m}\underline{\mathbf{T}}_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{\beta}_{1m}\mathbf{I} & \dots & \overline{\beta}_{km}\mathbf{I} & \gamma_{m1}\underline{\mathbf{T}}_2 & \dots & \gamma_{mm}\underline{\mathbf{T}}_2 \end{pmatrix}, \tag{5.61}$$

with parameters α_{ij}, β_{ij}, and γ_{ij} chosen with

$$\alpha_{ji} = -\overline{\alpha}_{ij}, \quad \beta_{ji} = \overline{\beta}_{ij}, \quad \gamma_{ji} = -\overline{\gamma}_{ij}, \tag{5.62}$$

to ensure $\mathbf{T}$ is Hermitian, and with $\mathbf{T}_0$ being that operator acting on $\underline{\mathcal{U}}^{k+m}$ such that if $\mathbf{U}' = \mathbf{T}_0\mathbf{U}$ with $\mathbf{U}'$, $\mathbf{U} \in \mathcal{U}^{k+m}$ then $\mathbf{U}'(\mathbf{x}) = \mathbf{T}\mathbf{U}(\mathbf{x})$ for all $\mathbf{x}$.

5.6 Analogous bounds for 2-phase composites

The question arises as to whether the bounds (5.51) can uniquely determine the microstructure for an appropriate choice of frequencies $\omega_1, \omega_2, ..., \omega_s$ in the limit as $s \to \infty$. To shed some light on this it is helpful to look at an analogous problem in the theory of periodic composites. Here we modify the analysis that Cherkaev and Gibiansky (1992) used to obtain sharp bounds coupling the effective electric permittivity and magnetic permeability tensors in two-dimensions. Since we are not incorporating differential constraints on the fields, beyond the orthogonality of $\mathcal{U}$, $\mathcal{E}$, and $\mathcal{J}$, we restrict the translations to those that only reflect this orthogonality.

Suppose, for simplicity, we have a two-phase composite of two isotropic phases with conductivities $\sigma_1^{(i)}\mathbf{I}$ and $\sigma_2^{(i)}\mathbf{I}$ where $i = 1, 2$ parameterizes the conductivities under 2 different experiments. Suppose also the composite is isotropic with effective conductivities $\sigma_*^{(i)}\mathbf{I}$. Let us define the tensors

$$\mathbf{M}_j = \begin{pmatrix} \sigma_j^{(1)}\mathbf{I} & 0 \\ 0 & \mathbf{I}/\sigma_j^{(2)} \end{pmatrix}, \tag{5.63}$$

for $j = 1, 2, *$. We take translations of the form

$$\mathbf{T} = \begin{pmatrix} 0 & t\mathbf{I} \\ t\mathbf{I} & 0 \end{pmatrix}, \tag{5.64}$$

where the constant t is real, and must be chosen such that $\mathbf{M}_1 - \mathbf{T} \geq 0$ and $\mathbf{M}_2 - \mathbf{T} \geq 0$. Defining

$$\mathbf{Y}_* = -f_2\mathbf{M}_1 - f_2\mathbf{M}_2 + f_1f_2(\mathbf{M}_1 - \mathbf{M}_2)(f_1\mathbf{M}_1 + f_2\mathbf{M}_2 - \mathbf{M}_*)^{-1}(\mathbf{M}_1 - \mathbf{M}_2), \tag{5.65}$$

in which f_1 and $f_2 = 1 - f_1$ are the volume fractions of the two phases, the bounds (5.51) take the form $\mathbf{Y}_* + \mathbf{T} \geq 0$, which for the special forms of $\mathbf{Y}_*$ and $\mathbf{T}$ given by (5.63)-(5.64) is equivalent to the bound $\mathbf{Y}_* - \mathbf{T} \geq 0$ (as can be seen by swapping the signs of the off diagonal blocks, which does not change the positive semidefiniteness of the matrix). In summary, the bound

$$\begin{pmatrix} y_*^{(1)} & -t \\ -t & 1/y_*^{(2)} \end{pmatrix} \geq 0, \tag{5.66}$$

in which

$$y_*^{(i)} = -f_2\sigma_1^{(i)} - f_2\sigma_2^{(i)} + f_1f_2(\sigma_1^{(i)} - \sigma_2^{(i)})(f_1\sigma_1^{(i)} + f_2\sigma_2^{(i)} - \sigma_*^{(i)})(\sigma_1^{(i)} - \sigma_2^{(i)}), \tag{5.67}$$

is satisfied for all real t such that

$$\begin{pmatrix} \sigma_j^{(1)} & -t \\ -t & 1/\sigma_j^{(2)} \end{pmatrix} \geq 0 \quad \text{for } j = 1, 2. \tag{5.68}$$

Let us suppose the phases have been labeled so that

$$\sigma_1^{(1)}/\sigma_1^{(2)} \geq \sigma_2^{(1)}/\sigma_2^{(2)}. \tag{5.69}$$

Then the inequalities (5.68) will be satisfied with

$$t^2 = \sigma_2^{(1)}/\sigma_2^{(2)}, \tag{5.70}$$

and (5.66) generates the bound

$$y_*^{(1)} \geq y_*^{(2)}\sigma_2^{(1)}/\sigma_2^{(2)}. \tag{5.71}$$

Now by switching experiment $i = 1$ with experiment $i = 2$ and taking

$$t^2 = \sigma_1^{(2)}/\sigma_1^{(1)}, \tag{5.72}$$

we get the additional bound

$$y_*^{(2)} \geq y_*^{(1)}\sigma_1^{(1)}/\sigma_1^{(2)}. \tag{5.73}$$

The bounds (5.71) and (5.73) are exactly those of (Prager 1969), who obtained them using variational principles. They also can be obtained using the analytic method [see, for example, (27.18) in Milton (2002), although Bergman (1978) could easily have derived them, as he derived even tighter bounds assuming isotropy: see equations (4.36) and (4.37) in his paper].

It is an open, and very interesting question as to whether with an appropriate choice of the matrices $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_{ij}$, and $\boldsymbol{\gamma}_{ij}$ one can recover the entire hierarchy of bounds derived using the analytic method that correlate the values of $\sigma_*^{(i)} = \sigma_*(\sigma_1^{(i)}, \sigma_2^{(i)})$ for $i = 1, 2,,..s$ with the volume fraction $f_1 = \partial\sigma_*(\sigma_1, 1)/\partial\sigma_1|_{\sigma_1=1}$. In applications to composites the hierarchy was derived by Milton (1981c) (and rederived by Bergman 1993). In a mathematical wider context it is essentially similar to the much studied problem of Nevanlinna–Pick interpolation, solved by Pick (1915) and Nevanlinna (1919, 1929): see for example Delsarte, Genin, and Kamp (1981). For the case when the ratios $\sigma_1^{(i)}/\sigma_2^{(i)}$ are all real and positive, the interpolation result in theorem 3.1 of **Chapter 5** of Kreĭn and Nudel'man (1974) is applicable. The Nevanlinna–Pick interpolation has been extended to matrix-valued functions (Fedčina 1972; Delsarte, Genin, and Kamp 1979); Chen and Koç 1994, 1995) and operator-valued functions (see Theorem 2.2 in **Chapter VIII** of Foiaş and Frazho 1990). This hierarchy of bounds converges to the exact analytic function $\sigma_*(\sigma_1, \sigma_2)$ and hence in this limit one can recover exactly the volume fraction. For bodies instead of composites the analog of the volume fraction would be the operator $\boldsymbol{\Gamma}_0\chi\boldsymbol{\Gamma}_0$, where $\chi(\mathbf{x})$ is the indicator function which is 1 in phase 1, and 0 in phase 2.

5.7 Transient response to electromagnetic fields in a two-phase body

Let us suppose we have a two-phase body with isotropic phases, having electrical permittivities $\varepsilon_1, \varepsilon_2$ and magnetic permeabilities μ_1, μ_2. Also suppose μ_1, μ_2 are real and frequency independent satisfying $\mu_1 = \mu_2$. We assume the electrical permittivities of both phases take the form

$$\varepsilon_1(\omega) = \varepsilon_{1R} + i\frac{\sigma_1}{\omega}, \quad \varepsilon_2(\omega) = \varepsilon_{2R} + i\frac{\sigma_2}{\omega}, \tag{5.74}$$

so that

$$\varepsilon(\omega, \mathbf{x}) = \chi(\mathbf{x})\varepsilon_1(\omega) + (1 - \chi(\mathbf{x}))\varepsilon_2(\omega), \tag{5.75}$$

where

$$\chi(\mathbf{x}) = \begin{cases} 1, & \text{in phase 1,} \\ 0, & \text{in phase 2.} \end{cases} \tag{5.76}$$

Now at the complex frequency $\omega = ip$, where p is real, we have

$$\varepsilon_1(\omega) = \varepsilon_{1R} + \frac{\sigma_1}{p}, \quad \varepsilon_2(\omega) = \varepsilon_{2R} + \frac{\sigma_2}{p}. \tag{5.77}$$

Consequently we have $\varepsilon_1(\omega) = \varepsilon_2(\omega)$ when $\omega = \omega_0$, where

$$\omega_0 = ip_0 = \underbrace{i\frac{\sigma_2 - \sigma_1}{\varepsilon_{1R} - \varepsilon_{2R}}}_{p_0}, \tag{5.78}$$

in which we assume $\sigma_2 - \sigma_1$ and $\varepsilon_{1R} - \varepsilon_{2R}$ have the same sign.

Now the equations of electromagnetism can be written [as follows from equations 4.16 and 4.20 in Milton, Seppecher, and Bouchitté (2009)] as

$$\underbrace{\begin{pmatrix} \mathbf{h} \\ i\omega\mathbf{d} \end{pmatrix}}_{\mathcal{G}(\mathbf{x})} = \underbrace{\begin{pmatrix} -(\omega^2\mu)^{-1} & 0 \\ 0 & \varepsilon \end{pmatrix}}_{\mathbf{Z}(\mathbf{x})} \underbrace{\begin{pmatrix} -\omega^2\mathbf{b} \\ i\omega\mathbf{e} \end{pmatrix}}_{\mathcal{F}(\mathbf{x})}. \tag{5.79}$$

At the frequency ω_0,

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_0 \equiv \varepsilon_{1R} + \frac{\sigma_1}{p_0} \tag{5.80}$$

is real, and we have $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_0$ with

$$\mathbf{Z}_0 = \begin{pmatrix} (p_0^2\mu_1)^{-1} & 0 \\ 0 & \varepsilon_0 \end{pmatrix}. \tag{5.81}$$

We take $\mathbf{Z}_0$ as our reference tensor. It should be quite easy to numerically calculate the fields $\mathcal{G}_0(\mathbf{x})$ and $\mathcal{F}_0(\mathbf{x})$ which solve the equations when $\mathbf{Z}(\mathbf{x}) = \mathbf{Z}_0$, as the boundary conditions vary, and they generate the space $\mathcal{U}$ of fields $\mathbf{Z}^{1/2}\mathcal{F}_0(\mathbf{x}) = \mathbf{Z}^{-1/2}\mathcal{G}_0(\mathbf{x}) = \mathbf{U}_0$. The spaces $\mathcal{E}$ and $\mathcal{J}$ are defined in Section 3.12, and our equations take the form

$$(\mathbf{J}_0 + \mathbf{J}) = \mathbf{L}(\mathbf{E}_0 + \mathbf{E}), \quad \text{with } \mathbf{E}_0, \mathbf{J}_0 \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \text{and } \mathbf{J} \in \mathcal{J}. \tag{5.82}$$

Now note the equations (5.79) in the time domain take the form

$$\begin{pmatrix} \mathbf{h} \\ -\frac{\partial \mathbf{d}}{\mathbf{d}t} \end{pmatrix} = \mathbf{K} * \begin{pmatrix} \frac{\partial^2 \mathbf{b}}{\partial t^2} \\ -\frac{\partial \mathbf{e}}{\mathbf{d}t} \end{pmatrix}, \tag{5.83}$$

where the asterisk denotes a convolution in time, and $\mathbf{Z}(\mathbf{x}, \omega)$ is the Fourier transform of the operator $\mathbf{K}$. Since the fields entering this relation are real, so too $\mathbf{K}$ must be real. It follows that $\mathbf{Z}(\mathbf{x}, \omega)$ satisfies the symmetry relations

$$\mathbf{Z}(\mathbf{x}, \omega) = \overline{\mathbf{Z}(\mathbf{x}, -\overline{\omega})}, \tag{5.84}$$

which in particular imply $\mathbf{Z}$ is real if $\omega = ip$ with p real and positive. Furthermore in the quadrant $\mathrm{Re}\,\omega > 0$, $\mathrm{Im}\,\omega > 0$ the imaginary part of $\mathbf{Z}$ must be positive definite in both phases. As $\mathbf{Z}(\mathbf{x}, \omega)$ is analytic in the upper half ω plane, we may introduce the variable $\tau = \omega^2$ it terms of which $\omega = \sqrt{\tau}$, with the square root having a branch cut just below the positive real axis in the $\tau = \omega^2$ plane. It follows that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are matrix-valued Herglotz functions of τ. [A matrix-valued Herglotz function has the property that the matrix elements are analytic functions of τ for τ in the upper half plane, $\mathrm{Im}(\tau) > 0$ and the imaginary part of the function takes positive semi-definite values in the upper half τ plane.] Then, from the analytic properties of $\mathbf{L}_*$ as a function of $\mathbf{L}_1$ and $\mathbf{L}_2$ it must inherit this property, and hence has the integral representation (see, for example, Theorem 5.4 in Gesztesy and Tsekanovskii 2000)

$$\mathbf{L}_*(\tau) = \mathbf{A}\tau + \mathbf{B} + \int_0^\infty d\boldsymbol{\mu}(\tau') \left(\frac{1}{\tau' - \tau} - \frac{\tau'}{1 + (\tau')^2} \right), \tag{5.85}$$

where, the variable τ' is real, $\mathbf{A}$ is a positive semidefinite operator, $\mathbf{B}$ is a real-valued operator, and $d\boldsymbol{\mu}$ is a positive operator-valued measure on the bounded Borel subsets of $\mathbb{R}$, such that

$$\int_0^\infty \boldsymbol{\mu}(\tau')(1 + (\tau')^2)^{-1} \tag{5.86}$$

is bounded. Think of it as approximately, i.e., as the limit of functions of the form,

$$\mathbf{L}_*(\tau) = \int_0^\infty \frac{\mathbf{R}(\tau')}{\tau' - \tau} d\tau', \tag{5.87}$$

where $\mathbf{R}(\tau')$ is a positive semidefinite operator-valued function.

Now perturb $\mathbf{L}(\mathbf{x})$ around $\mathbf{I}$ (corresponding to the solution with $\mathbf{Z}(\mathbf{x}) = \mathbf{Z}_0$):

$$\mathbf{L} = \mathbf{I} + \epsilon(\delta \mathbf{L}(\mathbf{x})), \tag{5.88}$$

where ϵ is a small parameter. We fix $\mathbf{E}_0$, which corresponds to taking boundary conditions with a fixed value of the tangential field $\mathbf{e}_T = (\mathbf{I} - \mathbf{n}\mathbf{n}^T)\mathbf{e}$. The remaining fields have the series expansions

$$\mathbf{J}_0 = \mathbf{E}_0 + \epsilon \mathbf{J}_0^{(1)} + \epsilon^2 \mathbf{J}_0^{(2)} + \ldots, \quad \mathbf{J} = \epsilon \mathbf{J}^{(1)} + \epsilon^2 \mathbf{J}^{(2)} + \ldots, \quad \mathbf{E} = \epsilon \mathbf{E}^{(1)} + \epsilon^2 \mathbf{E}^{(2)} + \ldots. \tag{5.89}$$

Substituting these in the constitutive law and collecting those terms of order ϵ gives

$$\mathbf{J}_0^{(1)} + \mathbf{J}^{(1)} = \delta \mathbf{L}(\mathbf{x}) \mathbf{E}_0 + \mathbf{E}^{(1)}. \tag{5.90}$$

Applying the projection operator $\mathbf{\Gamma}_0$ onto the space $\mathcal{U}$ to both sides of this equation, gives the expression

$$\mathbf{L}_* = \mathbf{I} + \epsilon \mathbf{\Gamma}_0 \delta \mathbf{L} \mathbf{\Gamma}_0 + \ldots, \tag{5.91}$$

for the effective tensor to first order in ϵ. It then follows that

$$\left.\frac{\partial \mathbf{L}_*}{\partial \tau}\right|_{\tau=-p_0^2} = \mathbf{\Gamma}_0 \left.\frac{\partial \mathbf{L}(\mathbf{x})}{\partial \tau}\right|_{\tau=-p_0^2} \mathbf{\Gamma}_0 \equiv \mathbf{F}_0, \tag{5.92}$$

where, from (5.74),

$$\begin{aligned} \left.\frac{\partial \mathbf{L}(\mathbf{x})}{\partial \tau}\right|_{\tau=-p_0^2} &= \mathbf{Z}_0^{-1/2} \left.\frac{\partial \mathbf{Z}(\mathbf{x})}{\partial \tau}\right|_{\tau=-p_0^2} \mathbf{Z}_0^{-1/2} \\ &= \mathbf{Z}_0^{-1/2} \begin{pmatrix} 1/(p_0^2\mu) & 0 \\ 0 & [\sigma_1 \chi(\mathbf{x}) + \sigma_2(1-\chi(\mathbf{x}))]/(2p_0^3) \end{pmatrix} \mathbf{Z}_0^{-1/2}. \end{aligned} \tag{5.93}$$

We have the constraints

$$\begin{aligned} \mathbf{L}_*(-p_0^2) &= \mathbf{I}, \\ \left.\frac{\partial \mathbf{L}_*(\tau)}{\partial \tau}\right|_{\tau=-p_0^2} &= \mathbf{F}_0. \end{aligned} \tag{5.94}$$

These provide constraints that are linear in the measure (sum rules). If $\mathbf{P}$ denotes the projection onto a finite dimensional space $\mathcal{M}$, we have

$$\mathbf{P}\mathbf{L}_*(\tau)\mathbf{P}^T = \int_0^\infty \frac{\mathbf{P}\mathbf{R}(\tau')\mathbf{P}^T}{\tau' - \tau} d\tau', \tag{5.95}$$

and the sum rules become

$$\begin{aligned} \mathbf{P}\mathbf{L}_*(-p_0^2)\mathbf{P}^T &= \mathbf{P}\mathbf{P}^T, \\ \mathbf{P}\left.\frac{\partial \mathbf{L}_*(\tau)}{\partial \tau}\right|_{\tau=-p_0^2} \mathbf{P}^T &= \mathbf{P}\mathbf{F}_0\mathbf{P}^T. \end{aligned} \tag{5.96}$$

Of course measuring or computing $\mathbf{P}\mathbf{Y}_*(\tau)\mathbf{P}^T$ at a set of values of τ is much easier since we only need to consider "applied fields" in $\mathcal{M}$.

Now given an "applied field" $\widetilde{\mathbf{E}}_0(t)$ in $\mathcal{U}$ that depends on time, there is some response $\widetilde{\mathbf{J}}_0(t)$ in $\mathcal{U}$ given by some kernel $\mathbf{K}_*$, i.e.,

$$\widetilde{\mathbf{J}}_0(t) = \int_{-\infty}^t \mathbf{K}_*(t-t')\widetilde{\mathbf{E}}_0(t')\, dt'. \tag{5.97}$$

Basically $\mathbf{K}_*$ is the Fourier (or Laplace) transform of $\mathbf{L}_*(\omega)$ and depends *linearly* on the measure $\mathbf{P}\mathbf{R}(\tau)\mathbf{P}^T$. Therefore we may use the linear programming theory methods in the next chapter to bound the response which can then be used in an inverse fashion to bound $\mathbf{P}\mathbf{F}_0\mathbf{P}^T$.

Note that there may be other frequencies $\omega_0^1, \omega_0^2, \ldots, \omega_0^J$ for general materials, with some other dependence of $\varepsilon_1(\omega), \varepsilon_2(\omega)$ where $\varepsilon_1(\omega_0^j) = \varepsilon_2(\omega_0^j)$, and these need not necessarily lie on the imaginary $\omega-$axis. Of

course there could be no such frequencies. In the case there are other such frequencies this, with $\tau_j = (\omega_0^j)^2$, will give other sum rules of the form

$$\begin{aligned} \mathbf{P}\mathbf{L}_*(\tau_j)\mathbf{P}^T &= \mathbf{P}\mathbf{L}_0^{(j)}\mathbf{P}^T, \\ \mathbf{P}\frac{\partial \mathbf{L}_*(\tau)}{\partial \tau}|_{\tau=\tau_j}\mathbf{P}^T &= \mathbf{P}\mathbf{F}_0^{(j)}\mathbf{P}^T. \end{aligned} \tag{5.98}$$

with consequent improvement in the bounds.

5.8 Quasistatic dielectric response in a body with two isotropic phases and related problems

In quasistatics the wavelength is very large compared to the body and we can just use the usual dielectric equations, but with complex fields. Also we can take a homogeneous reference tensor, which is also isotropic $\mathbf{Z}_0 = \varepsilon_0\mathbf{I}$ and by choosing the scales of our dimensions we can assume $\varepsilon_0 = 1$ so that we do not have to distinguish fields that are multiplied by $\mathbf{Z}_0^{1/2}$ or $\mathbf{Z}_0^{-1/2}$

Now $\mathbf{L}_*$ is an analytic function of $\varepsilon_1, \varepsilon_2$, which is homogeneous

$$\lambda\mathbf{L}_*(\varepsilon_1, \varepsilon_2) = \mathbf{L}_*(\lambda\varepsilon_1, \lambda\varepsilon_2), \tag{5.99}$$

for all real and complex λ. We also have $\mathrm{Im}(\mathbf{L}_*(\varepsilon_1, \varepsilon_2)) \geq 0$ when $\mathrm{Im}(\varepsilon_i) \geq 0$, $i = 1, 2$.

This analyticity and the subsequent results in this section of course also apply in a body with two isotropic phases for the isomorphic problems of antiplane elasticity (in the viscoelastic case) and magnetic response (in quasistatics where the magnetic permeability tensor is complex). It also applies to quasistatic elasticity with two isotropic incompressible phases (for which the Dirichlet-to-Neumann map is an analytic function of the shear moduli μ_1 and μ_2 of the phases).

Following Bergman (1978) we introduce

$$s = \frac{\varepsilon_2}{\varepsilon_2 - \varepsilon_1}. \tag{5.100}$$

Then $\mathbf{L}_*(\varepsilon_1, \varepsilon_2)/\varepsilon_2$ is an analytic function of $\varepsilon_1/\varepsilon_2$, and also of s which is real when $\varepsilon_1, \varepsilon_2$ are real and positive. Also following Bergman (1978) (but replacing his symbol t with u as here t represents time) we introduce

$$u = \frac{1/\varepsilon_2}{1/\varepsilon_2 - 1/\varepsilon_1} = \frac{\varepsilon_1}{\varepsilon_1 - \varepsilon_2} = 1 - s, \tag{5.101}$$

and let us set

$$\widehat{\mathbf{L}}_*\left(\frac{1}{\varepsilon_1}, \frac{1}{\varepsilon_2}\right) = \left[\mathbf{L}_*(\varepsilon_1, \varepsilon_2)\right]^{-1}. \tag{5.102}$$

Define

$$\mathbf{F}(s) = \mathbf{I} - \mathbf{L}_*\left(1 - \frac{1}{s}, 1\right), \qquad \mathbf{G}(u) = \mathbf{I} - \widehat{\mathbf{L}}_*\left(1 - \frac{1}{u}, 1\right). \tag{5.103}$$

The functions $\mathbf{F}(s)$ and $\mathbf{G}(u)$ (in Bergman's notation $\mathbf{G}(u)$ is $\mathbf{\Phi}(t)$) have the approximate spectral representations:

$$\mathbf{F}(s) = \sum_{i=0}^{n} \frac{\mathbf{B}_i}{s - s_i}, \qquad \mathbf{G}(u) = \sum_{i=0}^{m} \frac{\mathbf{P}_i}{u - u_i}, \tag{5.104}$$

where the operator-valued residues $\mathbf{B}_i$ and $\mathbf{P}_i$ are positive semidefinite and self-adjoint and the poles s_i, u_i are real with

$$\begin{aligned} 0 \le s_0 \le s_1 \le \ldots \le s_n \le 1, \\ 0 \le u_0 \le u_1 \le \ldots \le u_m \le 1. \end{aligned} \tag{5.105}$$

These operator-valued residues and poles must be such that

$$\sum_{i=0}^{n} \frac{\mathbf{B}_i}{1-s_i} \le \mathbf{I}, \quad \sum_{i=0}^{m} \frac{\mathbf{P}_i}{1-u_i} \le \mathbf{I}, \tag{5.106}$$

to ensure $\mathbf{L}_*$ and $\widehat{\mathbf{L}}_*$ are positive semidefinite when $\epsilon_1 = 0$ and $\epsilon_2 = 0$ respectively.

Now given an "applied electric field" $\widetilde{\mathbf{e}}_0(t)$ in $\mathcal{U}$ that depends on time, there is some response displacement field $\widetilde{\mathbf{d}}_0(t)$ taking values in $\mathcal{U}$. We have

$$\widetilde{\mathbf{d}}_0(t) = \varepsilon_2(t) * \widetilde{\mathbf{e}}_0(t) - \sum_{i=0}^{n} \mathbf{B}_i \mathcal{L}^{-1}\left[\frac{\varepsilon_2(\lambda)}{s(\lambda)-s_i}\right](t) * \widetilde{\mathbf{e}}_0(t), \tag{5.107}$$

where $*$ denotes a convolution, $\mathcal{L}^{-1}$ denotes the inverse Laplace transform and λ is the Laplace transform parameter. The $\varepsilon_2(\lambda)$ and $s(\lambda)$ are known from the response of the phases:

$$\mathbf{d}(\mathbf{x},t) = \sum_{i=1,2} \chi_i(\mathbf{x}) \mathcal{L}^{-1}\left[\varepsilon_i(\lambda)\right](t) * \mathbf{e}(\mathbf{x},t), \tag{5.108}$$

where as usual $\chi_i(\mathbf{x})$ is the indicator function of phase i, and $\mathbf{d}$ and $\mathbf{e}$ denote the electric displacement field and electric field respectively. We assume here that $\mathcal{L}^{-1}(\varepsilon_i(\lambda))$ is known. Since $\widetilde{\mathbf{d}}_0(t)$, at any fixed time $t = t_0$, is a *linear function* of the residues $\mathbf{B}_i$, we may apply linear programming theory and all of the analysis in the next chapter applies. That is we can obtain bounds on $\widetilde{\mathbf{d}}_0(t_0)$ and even bounds correlating $\widetilde{\mathbf{d}}_0(t_0), \widetilde{\mathbf{d}}_0(t_1), \ldots, \widetilde{\mathbf{d}}_0(t_k)$. Moreover it may be easier to measure the current field $\mathbf{j} = \partial\mathbf{d}/\partial t$ and we can use the same method to obtain bounds correlating $\widetilde{\mathbf{j}}_0(t_0), \widetilde{\mathbf{j}}_0(t_1), \ldots, \widetilde{\mathbf{j}}_0(t_k)$. Note that equivalently these bound the currents $\mathbf{n} \cdot \widetilde{\mathbf{j}}_0(t_0), \mathbf{n} \cdot \widetilde{\mathbf{j}}_0(t_1), \ldots, \mathbf{n} \cdot \widetilde{\mathbf{j}}_0(t_k)$, which can be measured at the surface of the body

Dually, given an "applied field" $\widetilde{\mathbf{d}}_0(t)$ in $\mathcal{U}$ that depends on time, there is a response $\widetilde{\mathbf{e}}_0(t)$ taking values in $\mathcal{U}$. We have

$$\widetilde{\mathbf{e}}_0(t) = [\varepsilon_2(\lambda)]^{-1} * \widetilde{\mathbf{d}}_0(t) - \sum_{i=0}^{m} \mathbf{P}_i \mathcal{L}^{-1}\left[\frac{(\varepsilon_2(\lambda))^{-1}}{u(\lambda)-u_i}\right](t) * \widetilde{\mathbf{d}}_0(t), \tag{5.109}$$

and since this is a *linear function* of the residues $\mathbf{P}_i$, we may again apply all of the analysis of the next chapter. Thus we can obtain bounds on $\widetilde{\mathbf{e}}_0(t_0)$ and even bounds correlating $\widetilde{\mathbf{e}}_0(t_0), \widetilde{\mathbf{e}}_0(t_1), \ldots, \widetilde{\mathbf{e}}_0(t_k)$.

These bounds may be used in an inverse way to bound

$$\left.\frac{\partial \mathbf{L}_*(\varepsilon_1, 1)}{\partial \varepsilon_1}\right|_{\varepsilon_1 = 1}, \tag{5.110}$$

and thus extract information about the geometry.

By direct analogy with (5.91) we have

$$\left.\frac{\partial \mathbf{L}_*(\varepsilon_1,1)}{\partial \varepsilon_1}\right|_{\varepsilon_1=1} = \mathbf{\Gamma}_0\chi\mathbf{\Gamma}_0 \equiv \mathbf{F}_0. \tag{5.111}$$

This derivative is essentially the Dirichlet-to-Neumann map for an almost homogeneous material. Calderón (1980) showed that small perturbations in the scalar-valued dielectric constant (and in particular the indicator function $\chi(\mathbf{x})$ in our problem) could be uniquely determined, from this Dirichlet-to-Neumann map and furthermore gave a reconstruction method which was implemented in three dimensions by Boverman, Isaacson, Kao, Saulnier, and Newell (2008) and Delbary, Hansen, and Knudsen (2011) [see also Section 16.2 of Mueller and Siltanen (2012)]. We remark that the scalar-valued dielectric constant can be recovered even if the perturbation is not small [Sylvester and Uhlmann 1987; Astala and Päivärinta 2006]. On the other hand, if we have a small anisotropic perturbation of the conductivity tensor of a homogeneous isotropic medium, then this anisotropic perturbation cannot be recovered from the first order perturbation of the Dirichlet-to-Neumann map (Sylvester 1993).

Also one can bound

$$\left.\frac{\partial \mathbf{P}_0^T\mathbf{L}_*(\varepsilon_1,1)\mathbf{P}_0}{\partial \varepsilon_1}\right|_{\varepsilon_1=1} = \mathbf{P}_0^T\chi\mathbf{P}_0, \tag{5.112}$$

where $\mathbf{P}_0$ projects onto a finite dimensional space $\widetilde{\mathcal{U}}_0 \subset \mathcal{U}$, by taking "applied fields" $\widetilde{\mathbf{e}}_0(t)$ that only take values in $\widetilde{\mathcal{U}}_0$.

5.9 Quasistatic viscoelastic response in a two-phase body

Here, rather than analyze the response of the body in terms of the effective operator $\mathbf{L}_*$, we use the alternate description of using the Dirichlet-to-Neumann map $\mathbf{\Lambda}$, mapping displacements $\mathbf{u}_0$ at the boundary $\partial\Omega$ to tractions $\mathbf{t} = \mathbf{n}\cdot\boldsymbol{\sigma}$ at $\partial\Omega$. The Dirichlet-to-Neumann map $\mathbf{\Lambda}$ inherits the analytic properties of the effective operator $\mathbf{L}_*$ as functions of the component tensors $\mathbf{L}_1$ and $\mathbf{L}_2$, or equivalently as functions of the bulk moduli κ_1, κ_2 and shear moduli μ_1 and μ_2 of the phases. These moduli usually depend on the frequency ω. Since they are Fourier transforms of real integral kernels , they satisfy the symmetry relations

$$\kappa_i(\mathbf{x},\omega) = \overline{\kappa_i(\mathbf{x},-\overline{\omega})}, \quad \mu_i(\mathbf{x},\omega) = \overline{\mu_i(\mathbf{x},-\overline{\omega})}, \quad i=1,2, \tag{5.113}$$

and are analytic functions of ω when $\operatorname{Im}\omega > 0$, taking real values when ω is purely imaginary. Furthermore they have nonnegative imaginary parts when ω is in the quadrant $\operatorname{Im}\omega > 0, \operatorname{Re}\omega > 0$. As a function of $\tau = \omega^2$ they have nonnegative (or nonpositive) imaginary part according to whether τ has a positive (or, respectively, negative) imaginary part, and the Dirichlet-to-Neumann map $\mathbf{\Lambda}$ has accordingly a positive semidefinite (or negative semidefinite) imaginary part and is an analytic function of τ except possibly on the positive τ-axis

These facts imply that the Dirichlet-to-Neumann map $\mathbf{\Lambda}(\tau)$ has the representation of an operator-valued Herglotz function with no measure on the negative real τ-axis, i.e,

$$\mathbf{\Lambda}(\tau) = \mathbf{A}\tau + \mathbf{B} + \int_0^\infty d\boldsymbol{\mu}(\tau')\left(\frac{1}{\tau'-\tau} - \frac{\tau'}{1+(\tau')^2}\right), \tag{5.114}$$

where τ' is real, $\mathbf{A}$ is a positive semidefinite operator, $\mathbf{B}$ is a real-valued operator, and $d\boldsymbol{\mu}$ is a positive operator-valued measure on the bounded Borel subsets of $\mathbb{R}$, such that

$$\int_0^\infty \boldsymbol{\mu}(\tau')(1+(\tau')^2)^{-1} \tag{5.115}$$

is bounded.

In the case of quasistatic elasticity, where the wavelength is long compared to the body, we look for (generally complex) frequencies such that $\mu_1(\omega_0^k) = \mu_2(\omega_0^k)$, where μ is the shear modulus. It is *not necessary* that the bulk moduli are equal. At these frequencies ω_0^j, the operator $\mathbf{\Lambda}([\omega_0^k]^2) = \mathbf{\Lambda}^{(k)}$ still depends on the geometry but for *appropriate boundary conditions* it follows from the work of Hill (1963) and the generalization of Thaler and Milton (2014), that there exist boundary conditions for $\mathbf{u}(\mathbf{x})$ on the boundary of Ω such that the displacement field inside the body is the gradient of a scalar potential ϕ, i.e., $\mathbf{u} = \nabla\phi$ inside Ω. This is not true for all boundary conditions, just these special ones. To find these special boundary conditions, one imagines embedding the body in an infinite homogeneous medium with appropriate "applied fields" at infinity, i.e., $\mathbf{u} \approx \nabla g$ as $|\mathbf{x}| \to \infty$. Then from the known exterior Dirichlet-to-Neumann map, which can be calculated numerically, and from $\mathbf{\Lambda}^{(k)}$, one can find these special boundary conditions (see Thaler and Milton 2014) on $\mathbf{u}(\mathbf{x})$ at the boundary $\partial\Omega$ such that $\mathbf{u} = \nabla\phi$ inside Ω. Note when the body Ω is absent, the exterior Dirichlet-to-Neumann map takes the displacement on $\partial\Omega$ to the traction on $\partial\Omega$ when fields are applied at infinity. We remark that for two-phase composites having the same real bulk modulus in both phases, Cherkaev and Bonifasi-Lista (2011) have used an analytic representation for the response as a function of the shear modulus of one (possibly viscoelastic) phase, assuming the other phase had a real shear modulus, to estimate the volume fraction of bone from numerically simulated measurements of the Young's modulus at various frequencies (they assumed the bone had a cylindrical structure, with properties constant along the axial direction, which may be unrealistic.)

The following analysis provides a test which must be satisfied if we have the right boundary conditions. Suppose we are in three dimensions, $d = 3$. Take any scalar potential ψ inside Ω and consider the matrix-valued field $\mathbf{G}$ with elements

$$G_{il} = e_{ijl}\frac{\partial}{\partial x_j}\psi(\mathbf{x}). \tag{5.116}$$

Here e_{ijl} is the completely antisymmetric Levi-Civita tensor, taking a value of 1 if ijl is an even permutation of $1, 2, 3$; -1 if it's an odd permutation; 0 otherwise. Clearly G_{il} is an antisymmetric tensor, while $\nabla\mathbf{u} = \nabla\nabla\phi$ is a symmetric tensor. So we see that

$$\begin{aligned} 0 = \int_\Omega G_{il}\frac{\partial}{\partial x_i}u_l &= \int_\Omega -\left[\frac{\partial}{\partial x_i}G_{il}\right]u_l + \int_{\partial\Omega} n_i G_{il}u_l \\ &= \int_\Omega -\cancelto{0}{\left[\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}e_{ijl}\psi\right]}u_l + \int_{\partial\Omega} n_i e_{ijl}u_l\frac{\partial}{\partial x_j}\psi, \end{aligned} \tag{5.117}$$

i.e.,

$$\int_{\partial\Omega} n_i e_{ijl}u_l\frac{\partial}{\partial x_j}\psi = 0, \tag{5.118}$$

for all functions ψ. If this condition is not exactly satisfied (due to some numerical error) it makes sense to project $\mathbf{u}(\mathbf{x})$ at $\partial\Omega$ onto the space of fields satisfying this condition. It may be the case that if this condition is satisfied for all ψ then necessarily $\mathbf{u} = \nabla\phi$ inside Ω for some ϕ but I am unsure about that. Having found boundary conditions on $\mathbf{u}(\mathbf{x})$ at $\partial\Omega$ such that $\mathbf{u}(\mathbf{x}) = \nabla\phi$ inside Ω, then the strain is

$$\boldsymbol{\epsilon}(\mathbf{x}) = \nabla\nabla\phi, \tag{5.119}$$

and the stress is

$$\begin{aligned} \boldsymbol{\sigma}(\mathbf{x}) &= \lambda(\mathbf{x})[\mathrm{Tr}\,\boldsymbol{\epsilon}(\mathbf{x})]\mathbf{I} + 2\mu\boldsymbol{\epsilon}(\mathbf{x}) \\ &= \lambda(\mathbf{x})[\Delta\phi]\mathbf{I} + 2\mu\nabla\nabla\phi, \end{aligned} \tag{5.120}$$

where $\lambda = \kappa - 2\mu/d$ is the Lamé modulus, κ is the bulk modulus and d is the dimension. Since $\nabla \cdot \boldsymbol{\sigma}(\mathbf{x}) = 0$ inside Ω we get

$$\nabla(\lambda(\mathbf{x})\Delta\phi) + 2\mu\nabla(\Delta\phi) = 0. \tag{5.121}$$

Integrating this gives

$$\lambda(\mathbf{x})\Delta\phi + 2\mu\Delta\phi = \alpha, \tag{5.122}$$

where α is some constant. Then we have

$$\Delta\phi = \frac{\alpha}{\lambda(\mathbf{x}) + 2\mu}. \tag{5.123}$$

Note that

$$\int_\Omega \Delta\phi = \int_{\partial\Omega} (\nabla\phi) \cdot \mathbf{n} = \int_{\partial\Omega} \mathbf{u} \cdot \mathbf{n}, \tag{5.124}$$

implies

$$\alpha = \int_{\partial\Omega} \mathbf{u} \cdot \mathbf{n} \left[\int_\Omega 1/[\lambda(\mathbf{x}) + 2\mu] \right]^{-1}. \tag{5.125}$$

So the source term is piecewise constant:

$$\Delta\phi = \begin{cases} k_1, & \text{in phase 1,} \\ k_2, & \text{in phase 2,} \end{cases} \tag{5.126}$$

where

$$k_i = \frac{1}{\lambda_i + 2\mu} \int_{\partial\Omega} \mathbf{u} \cdot \mathbf{n} \left[\frac{f_1}{\lambda_1 + 2\mu} + \frac{f_2}{\lambda_2 + 2\mu} \right]^{-1}. \tag{5.127}$$

Also in the exterior medium we have imagined, suppose $\lambda(\mathbf{x}) = \lambda_2$. From the "applied field" at infinity $\Delta g = C_g$ is known and since $\nabla\phi$ must be constant in this exterior medium too, $\nabla\phi = C_g$ there and $C_g = k_2$. So as observed by Thaler and Milton (2014) we can determine the volume fraction f_1 and $f_2 = 1 - f_1$. So now from $\mathbf{u}$ at the boundary of Ω we can determine $\nabla\phi$, and hence ϕ and the flux $\mathbf{n} \cdot \nabla\phi$ where $\mathbf{n}$ is the outward normal. Also from $\boldsymbol{\sigma} \cdot \mathbf{n}$ we know $\mathbf{n} \cdot \nabla\nabla\phi$ (note $\Delta\phi$ is known).

The problem of recovering the geometry is known as the gravimetric problem, because of its origins of finding the mass inside a body given the gravitational field outside. Without additional restrictions on the inclusion shape there is no unique solution to this problem: a spherical shell of material of finite thickness produces the same gravitational field outside as a solid sphere with the same center and appropriate radius. In Section 3.1 of Isakov (1990) [see also Section 4.1 of Isakov (2006)] it is proven that the three-dimensional inclusion shape is unique if either it is convex, or star shaped with respect to its center of gravity. In two-dimensions an algorithm for recovering the shape of a simply connected star shaped inclusion was developed and implemented by Ring (1995). If the "inclusion phase" consists of a number of very small islands, then their location and size in two and three dimensions can be obtained using the algorithms of Kang and Lee (2004) and Kolokolnikov and Lindsay (2015).

As in the next chapter, one considers the response

$$\widetilde{\mathbf{J}}_1(t) = \int_{-\infty}^{t} \mathbf{K}_\Lambda(t - t')\widetilde{\mathbf{E}}_1(t')\, dt', \tag{5.128}$$

for some applied field $\widetilde{\mathbf{E}}_1(t)$, or set of applied fields $\widetilde{\mathbf{E}}_1^{(j)}(t)$, and since the response (or set of responses) is linear in the measure one can bound $\widetilde{\mathbf{J}}_1^{(j)}(t)$.

Now the constraints on the measure become

$$\mathbf{\Lambda}([\omega_0^k]^2) = \mathbf{\Lambda}^{(k)}. \tag{5.129}$$

As these are linear constraints, one can apply linear programming theory to bound $\widetilde{\mathbf{J}}_1^{(1)}(t)$, $\widetilde{\mathbf{J}}_1^{(2)}(t)$,...,$\widetilde{\mathbf{J}}_1^{(n)}(t)$ in terms of all the $\mathbf{Y}_0^{(k)}$. The idea is to use these in an inverse fashion to bound the $\mathbf{Y}_0^{(k)}$ from all the experimental data, then to extract information about the geometry from the $\mathbf{Y}_0^{(k)}$. In practice one will let $\widetilde{\mathcal{E}}_0$ be the space spanned by the $\widetilde{\mathbf{E}}_1^{(j)}(t)$ as j and t vary. Let $\mathbf{P}_0$ be the projection onto this space, then we just measure the $\mathbf{P}_0\widetilde{\mathbf{J}}_1^{(j)}(t)$:

$$\mathbf{P}_0\widetilde{\mathbf{J}}_1^{(j)}(t) = \int \mathbf{P}_0\mathbf{K}_\Lambda(t-t')\mathbf{P}_0^T\mathbf{P}_0\widetilde{\mathbf{E}}_1^{(j)}(t)'dt'. \tag{5.130}$$

The problem then involves just an $m \times m$ matrix-valued measure where $m \geq n$ is the dimension of $\widetilde{\mathcal{E}}_0$.

5.10 Probing the body using measurements at a discrete set of frequencies

Let us return to the problem of extracting information about the geometry in a two-phase system probed by quasistatic electromagnetic fields with the two isotropic phases having complex permittivities $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$. Suppose we have measured the $m \times m$ response matrix $\mathbf{P}_0^T\mathbf{L}_*\mathbf{P}_0$ of the body at a set of frequencies ω_1, ω_2....ω_s. Rather than using the variable $\varepsilon_2/(\varepsilon_2-\varepsilon_1)$ it is preferable to begin by using the variable $h = \varepsilon_1/\varepsilon_2$. Then we have the approximate representation

$$\mathbf{P}_0^T[\mathbf{L}_*(h,1)]\mathbf{P}_0 = \mathbf{A}_0 h + \mathbf{A}_1 - \sum_{i=2}^{N} \frac{\mathbf{A}_i}{h+h_i}, \tag{5.131}$$

where $h_i > 0$ for all i, for all i the matrices $\mathbf{A}_i$ are positive semidefinite and

$$\mathbf{D} \equiv \mathbf{A}_1 - \sum_{i=2}^{N} \frac{\mathbf{A}_i}{h_i} \geq 0, \tag{5.132}$$

to ensure $\mathbf{P}_0^T\mathbf{L}_*(0,1)\mathbf{P}_0 \geq 0$. The constraint that $\mathbf{L}_*(1,1) = \mathbf{I}$ implies

$$\mathbf{P}_0^T\mathbf{P}_0 = \mathbf{A}_0 + \mathbf{A}_1 - \sum_{i=2}^{N} \frac{\mathbf{A}_i}{1+h_i}. \tag{5.133}$$

Subtracting this from (5.131) gives

$$\begin{aligned} \mathbf{P}_0^T[\mathbf{L}_*(h,1) - \mathbf{I}]\mathbf{P}_0 &= \mathbf{A}_0(h-1) + \sum_{i=2}^{N} \frac{\mathbf{A}_i}{1+h_i} - \sum_{i=2}^{N} \frac{\mathbf{A}_i}{h+h_i} \\ &= \mathbf{A}_0(h-1) + \sum_{i=2}^{N} \frac{(h-1)\mathbf{A}_i}{(1+h_i)(h+h_i)}, \end{aligned} \tag{5.134}$$

giving

$$\frac{\mathbf{P}_0^T[\mathbf{L}_*(h,1)-\mathbf{I}]\mathbf{P}_0}{h-1} = \mathbf{A}_0 + \sum_{i=2}^{N} \frac{\mathbf{A}_i'}{h+h_i}, \quad \text{where } \mathbf{A}_i' = \mathbf{A}_i/(1+h_i). \tag{5.135}$$

We now introduce the variables $\eta = \sqrt{-h}$ and $\eta_i = \sqrt{h_i}$, where the square root function $\sqrt{t}$ is taken with a branch cut just below the positive real $t = -h$ axis, so its range is the upper half η-plane. Then the function

$$\begin{aligned} \mathbf{G}(\eta) \equiv \eta\frac{\mathbf{P}_0^T[\mathbf{L}_*(-\eta^2,1)-\mathbf{I}]\mathbf{P}_0}{-\eta^2-1} &= \eta\mathbf{A}_0 + \sum_{i=2}^{N} \frac{\eta\mathbf{A}_i'}{\eta_i^2-\eta^2} \\ &= \eta\mathbf{A}_0 + \sum_{i=2}^{N} \left[\frac{\mathbf{A}_i'/2}{\eta_i-\eta} - \frac{\mathbf{A}_i'/2}{\eta_i+\eta}\right], \end{aligned} \tag{5.136}$$

from this representation evidently has a positive semidefinite imaginary part when η has positive imaginary part. So $\mathbf{G}(\eta)$ is a matrix-valued Herglotz function of η. Furthermore it has the symmetry properties

$$\overline{\mathbf{G}(-\overline{\eta})} = -\eta\frac{\mathbf{P}_0^T[\overline{\mathbf{L}_*(-\overline{\eta}^2,1)}-\mathbf{I}]\mathbf{P}_0}{-\eta^2-1} = -\mathbf{G}(\eta), \tag{5.137}$$

where we have used the fact that $\mathbf{L}_*(h,1)$ is real symmetric, i.e., $\overline{\mathbf{L}_*(\overline{h},1)} = \mathbf{L}_*(h,1)$. Now introducing the Cayley transform

$$z = \frac{\eta-i}{\eta+i}, \tag{5.138}$$

which maps the upper half plane to the unit disk, we see that the function

$$\mathbf{W}(z) = -i\mathbf{G}[i(1+z)/(1-z)] \tag{5.139}$$

is in the Carathéodory class of matrix-valued rational functions, i.e., having a positive-semidefinite real part when z is in the unit disk. Given knowledge of the functions $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$, then from measurements of $\mathbf{P}_0^T\mathbf{L}_*\mathbf{P}_0$ at a set of frequencies ω_1, ω_2....ω_s, we can determine the values of the matrices

$$\mathbf{W}_i = \mathbf{W}(z_i), \quad \text{with } z_i = \frac{\sqrt{-h_i^\circ}-i}{\sqrt{-h_i^\circ}+i}, \quad \text{where } h_i^\circ = \varepsilon_1(\omega_i)/\varepsilon_2(\omega_i), \tag{5.140}$$

for $i = 1, 2$,...s. Furthermore (5.137) implies $\mathbf{W}(z)$ is real symmetric, i.e., $\overline{\mathbf{W}(\overline{z})} = \mathbf{W}(z)$. So if z_i is not real (i.e., η is not purely imaginary) then we also know $\mathbf{W}(z)$ at the complex conjugate points $\overline{z}_i$:

$$\overline{\mathbf{W}}_i = \mathbf{W}(\overline{z}_i), \tag{5.141}$$

and these can be incorporated among our known function values.

The Nevanlinna–Pick interpolation problem for matrix-valued functions consists of finding a matrix-valued function $\mathbf{W}(z)$ in the Carathéodory class that interpolates the known matrix values, i.e., is such that $\mathbf{W}_i = \mathbf{W}(z_i)$. The criterion for the existence of such a function (Delsarte, Genin, and Kamp 1979) is that the block-Pick matrix defined by

$$\mathbf{P} = \begin{pmatrix} \frac{\overline{\mathbf{W}}_1+\mathbf{W}_1}{1-\overline{z}_1 z_1} & \cdots & \frac{\overline{\mathbf{W}}_1+\mathbf{W}_s}{1-\overline{z}_1 z_s} \\ \vdots & \ddots & \vdots \\ \frac{\overline{\mathbf{W}}_s+\mathbf{W}_1}{1-\overline{z}_s z_1} & \cdots & \frac{\overline{\mathbf{W}}_s+\mathbf{W}_s}{1-\overline{z}_s z_s} \end{pmatrix}, \tag{5.142}$$

(i.e., with block entries $\mathbf{P}_{ij} = (\overline{\mathbf{W}}_i + \mathbf{W}_j)/(1 - \overline{z}_i z_j)$ which are $m \times m$ matrices) is positive semidefinite. Of course the experiments have errors and it may be the case that the experimentally measured $\mathbf{W}_i^\circ$ are not compatible with the Pick criterion. The Pick criterion (5.142) defines an admissible region in $\mathbf{W}_1, \mathbf{W}_2 ... \mathbf{W}_i ...$ space. We should pick a point in this admissible region which is the closest (in some appropriate norm) to the experimentally measured point ($\mathbf{W}_1^\circ$, $\mathbf{W}_2^\circ$,...,$\mathbf{W}_i^\circ$,...). Given that this criterion is satisfied, algorithms are available (Chen and Koç 1994, 1995) for finding an interpolating function $\mathbf{W}(z)$. Note that if an interpolating function $\mathbf{W}(z)$ does not satisfy the symmetry $\overline{\mathbf{W}(\overline{z})} = \mathbf{W}(z)$ then we can replace it by $[\overline{\mathbf{W}(\overline{z})} + \mathbf{W}(z)]/2$ to obtain one that does.

Also it may happen that the interpolating function does not satisfy the inequality

$$\begin{aligned} \lim_{\delta \to 0} \frac{\mathbf{W}(-1+\delta)}{\delta} &= \lim_{\delta \to 0} \frac{\mathbf{G}(i\delta/2)}{2i\delta/2} \\ &= \lim_{\delta \to 0} \frac{\mathbf{P}_0^T[\mathbf{L}_*(\delta^2/4, 1) - \mathbf{I}]\mathbf{P}_0}{2(\delta^2/4 - 1)} \\ &= \mathbf{P}_0^T[\mathbf{I} - \mathbf{L}_*(0,1)]\mathbf{P}_0/2 \\ &\leq \mathbf{P}_0^T\mathbf{P}_0/2, \end{aligned} \tag{5.143}$$

implied by the positive semidefiniteness of $\mathbf{L}_*(0,1)$. Then, following the same argument as used later in 6.6, we can without loss of generality assume that $\mathbf{P}_0^T[\mathbf{L}_*(0,1)]\mathbf{P}_0 = 0$. To see this we pick an extremely small constant ϵ and add to the right hand side of (5.131) an extra pole at $h = -h_{N+1} = -\epsilon$ with a tiny residue $\mathbf{A}_{N+1} = \epsilon\mathbf{D}$, where $\mathbf{D}$ is given by (5.132). Then (5.131) becomes

$$\mathbf{P}_0^T[\mathbf{L}_*(h,1)]\mathbf{P}_0 = \mathbf{A}_0 h + \mathbf{A}_1 - \sum_{i=2}^{N} \frac{\mathbf{A}_i}{h + h_i} - \frac{\epsilon\mathbf{D}}{h + \epsilon}, \tag{5.144}$$

where the additional term we have added ensures $\mathbf{P}_0^T[\mathbf{L}_*(0,1)]\mathbf{P}_0 = 0$, but at the same time has little effect on the function when h is not extremely small. In practice we could pick a small positive real value for δ and take as an additional known value

$$\mathbf{W}(-1+\delta) = (1-r)\delta\mathbf{P}_0^T\mathbf{P}_0/2, \tag{5.145}$$

where r is a small parameter which ensures one has strict inequality but approximate equality in (5.143) when δ is small but not infinitesimal.

Having found an interpolating function we can calculate

$$\mathbf{W}(0) = -i\mathbf{G}(i) = \lim_{h \to 1} \frac{\mathbf{P}_0^T[\mathbf{L}_*(h,1) - \mathbf{I}]\mathbf{P}_0}{h-1} = \left.\frac{\partial \mathbf{P}_0^T\mathbf{L}_*(h,1)\mathbf{P}_0}{\partial h}\right|_{h=1} = \mathbf{P}_0^T\chi\mathbf{P}_0, \tag{5.146}$$

and from knowledge of this hopefully extract the microstructure.

6 Bounds for the response of viscoelastic composites under antiplane loadings in the time domain

Ornella Mattei* **and Graeme W. Milton****

*DICATAM, University of Brescia, Via Branze, 43-25123, Brescia, Italy
**Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA
emails: ornella.mattei@ing.unibs.it, milton@math.utah.edu

Abstract

In order to derive bounds on the strain and stress response of a two-component composite material with viscoelastic phases, we revisit the so-called analytic method (Bergman 1978), which allows one to approximate the complex effective tensor, function of the ratio of the component shear moduli, as the sum of poles weighted by positive semidefinite residue matrices. The novelty of the present investigation lies in the application of such a method, previously applied (Milton 1980; Bergman 1980) to problems involving cyclic loadings in the frequency domain, to derive bounds in the *time domain* for the antiplane viscoelasticity case.

The position of the poles and the residues matrices are the variational parameters of the problem: the aim is to determine such parameters in order to have the minimum (or maximum) response at any given moment in time. All the information about the composite, such as the knowledge of the volume fractions or the transverse isotropy of the composite, is translated for each fixed pole configuration into (linear) constraints on the residues, the so-called sum rules. Further constraints can be obtained from the knowledge of the response of the composite at specific times (in this paper, for instance, we show how one can include information about the instantaneous and the long-term response of the composite).

The linearity of the constraints, along with the observation that the response at a fixed time is linear in the residues, enables one to use the theory of linear programming to reduce the problem to one involving relatively few nonzero residues. Finally, bounds on the response are obtained by numerically optimizing over the pole positions. In the examples studied, the results turn out to be very accurate estimates: if sufficient information about the composite is available, the bounds can be quite tight over the entire range of time, allowing one to predict the transient behavior of the composite. Furthermore, the bounds incorporating the volume fractions (and possibly transverse isotropy) can be extremely tight at certain specific times: thus measuring the response at such times, and using the bounds in an inverse fashion, gives

very tight bounds on the volume fraction of the phases in the composite. The work in this chapter forms part of the PhD thesis of Ornella Mattei (Mattei 2016).

6.1 Introduction

The problem of calculating the mechanical response of a composite material has been extensively investigated in the literature, with particular attention being paid to the derivation of approximation formulas and bounds on the effective properties of the composite. This has taken precedence over the determination of the exact response of the material, which represents a difficult task even in the rare situations where the microstructure is known.

Historically, in the elasticity case, the determination of bounds on the overall properties of the composite followed from the formulation of suitable extremum variational principles, as illustrated, for instance, in the pioneering work of Hill (1952), Hashin and Shtrikman (1962,1963), and Hashin (1965a), which paved the way to the calculation of rigorous geometry-independent bounds. Such bounds have proven to be useful benchmarks for testing experimental results and for setting limits on the range of possible responses, which is relevant when one is optimizing the microstructure to maximize performance. Variational principles are useful even when some of the moduli are negative: in Kochmann and Milton 2014, the authors have ruled out the possibility of achieving very stiff statically stable composites by combining materials with positive and negative moduli, as suggested by Lakes and Drugan (2002). The variational method is especially powerful when coupled with the translation method of Tartar and Murat, and Lurie and Cherkaev [see **Chapter 24** of Milton 2002 for relevant references], who used it to derive optimal bounds on the possible effective conductivity tensors of two and three dimensional two-phase conducting composites. The translation method can also be used to bound the response of inhomogeneous bodies or, inversely, to bound the volume fraction of the phases from measurements of the fields at the surface of the body (Kang and Milton 2013). For surveys of bounds on the effective properties of composites (and the various methods used to derive them) see the review articles of Willis (1981a) and Hashin (1983) and the books of Nemat-Nasser and Hori (1999), Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009) and references therein.

For the viscoelasticity case, instead, the lack in the time domain of variational formulations analogous to the ones for the elasticity problem first prompted several authors to apply the correspondence principle (Hashin 1965b; Christensen 1971) to the well-established results of the elastic problem in order to study the linear viscoelastic response of composites subject to a cyclic loading with a certain frequency. In fact, for low frequency harmonic vibrations, where the inertia effects can be neglected and the viscoelastic loss is small compared to the elastic moduli (thus ruling out phases which are viscous fluids or gel-like), the bounds that have been obtained which couple the effective properties with the derivatives of the effective properties with respect to the moduli (such as those obtained by Prager (1969)) when the moduli are real, imply the correspondence principle bounds on the complex effective properties (Schulgasser and Hashin 1976). The correspondence principle itself requires justification, and this justification is provided by the analyticity of the effective moduli as functions of the component moduli [see Section 11.4 in Milton 2002]. This analyticity was first recognized by Bergman (1978) in the context of the dielectric problem for composites of two isotropic components. Some of the assumptions underlying his initial analysis were incorrect (Milton 1979): in particular, he assumed that for periodic media the effective dielectric constant is a rational function of the component moduli. This is not true in two-phase checkerboard geometries, where the function has a branch cut, and if branch cuts can appear one may ask: why cannot they occur when the dielectric constants have positive imaginary parts, and not just when the ratio of the dielectric constants is real and negative?

A plausible justification for Bergman's approach was first provided by Milton (1981a), based on the assumption that the composite could be approximated by a resistor network containing two types of impedances, where the length scale of the network grid in the circuit is much smaller than that of the composite microstructure (see, for example, **Figure 8.5**(a)). Later, Golden and Papanicolaou (1983, 1985) gave a rigorous proof of the analytic properties and, moreover, they established the analyticity for multicomponent media and obtained a integral representation formula for the effective conductivity tensor as a function of the component conductivities which separates the dependence on the component conductivities (contained in an appropriate integral kernel) from the dependence on the microstructure (the relevant information about which is contained in a positive measure). These analytic properties enabled Milton (1980, 1981a, 1981c) and Bergman (1980, 1982), in independent works, to show that the complex effective dielectric constant (no matter how lossy the materials are, provided only that the quasistatic approximation is valid) is confined to a nested set of lens–shaped regions in the complex plane, where the relevant lens–shaped region is determined by what information is known about the composite (such as the volume fractions of the phases, whether it is isotropic or transversely isotropic, the values of real or complex dielectric constant at a set of other frequencies).

One can consider a composite containing inclusions (possibly at random orientations) in a matrix and take the dilute limit of these bounds as the volume fraction goes to zero. This gives bounds on the quasistatic average complex polarizability (Milton, McPhedran, and McKenzie 1981, Figures 2 and 3) that imply the bounds on the absorption of radiation by dispersions of small particles (governed by the imaginary part of the average complex polarizability) recently derived by Miller, Hsu, Reid, Qiu, DeLacy, Joannopoulos, Soljačić, and Johnson (2014). In a major advance, using a strikingly simple argument, such bounds have also been obtained by Miller, Polimeridis, Reid, Hsu, DeLacy, Joannopoulos, Soljačić, and Johnson (2016) for the full time-harmonic Maxwell's equations, without making the quasistatic approximation, and thus are applicable when the particle size is comparable to that of the wavelength.

Many of these bounds on the complex effective dielectric constant are implied by bounds on Stieltjes functions: see Milton (1986a), the discussion in the Introduction of Milton 1987b and references therein, and **Chapter V** in Kreĭn and Nudel'man (1974). To emphasize the remark made in Section 2.10 of **Chapter 2**, we reiterate that there is also a close connection to the Nevanlinna–Pick interpolation problem, solved by Nevanlinna (1919, 1929) and Pick (1915), of obtaining sharp bounds which correlate the values a Herglotz-function takes at a set of points in the upper half of the complex plane (a Herglotz function is a function which is analytic in the upper half of the complex plane, and has positive imaginary part there). Generalizations of the Nevanlinna–Pick interpolation problem, and different ways to obtain these generalizations, have been the subject of much research [see Ball and Trent (1998), the book of Agler and McCarthy (2002), the appendix of Charina, Putinar, Scheiderer, and Stöckler (2015), and references therein]. Of particular relevance to the theory of composites (as for anisotropic materials the effective tensor is represented by a matrix) is that the Nevanlinna–Pick interpolation problem has been solved for matrix-valued Herglotz functions (Delsarte, Genin, and Kamp 1979) and that algorithms are available for computing interpolations (Chen and Koç 1994, 1995). Nevanlinna–Pick interpolation and its generalizations to multivariate functions are also important in circuit and system theory, network synthesis, and control theory (Delsarte, Genin, and Kamp 1979; Kummert 1989; Ball and ter Horst 2010).

With a small modification the bounds in Milton (1981c) also apply to the related problem of bounding the viscoelastic moduli of homogeneous materials at one frequency, given the viscoelastic moduli at several other frequencies (Eyre, Milton, and Lakes 2002). An alternate approach to this problem is to directly recover the relaxation spectrum from measurements of viscoelastic moduli at various frequencies (Zhang, Lamoureux, Margrave, and Cherkaev 2011). A separate and interesting question is whether the relaxation models assumed in both approaches, although commonly used with much success, have any sound theoretical basis. The

models are justified if one assumes the relaxation kernel has a completely monotonic behavior (Beris and Edwards 1993), but there seems to be no physical reason for assuming this monotonicity of all derivatives of the relaxation kernel. Certainly the complete monotonicity fails if there are substructures in the material that can resonate. There is also a parallel between linear elastic moduli and dielectric moduli, and no one would suggest that dielectric relaxation kernels are necessarily completely monotonic: indeed substances such as gold, silver, and silicon carbide can have a negative real part of the dielectric constant at infrared and visible frequencies, which is prohibited by these models. The approach we use in this chapter is valid even if the relaxation kernel is not completely monotonic.

The bounds in the two-dimensional case immediately imply bounds for the mathematically equivalent problem of antiplane elasticity (in this connection it is to be noted that the claim of Bergman (1980) that the two-dimensional bounds of Milton (1980) were not attained by assemblages of doubly coated cylinders was, in fact, wrong: curiously an earlier version of his paper, which did not reference the doubly coated cylinder geometry in Milton (1980), but which did reference the paper, had claimed that the three-dimensional bounds were attained by doubly coated spheres, which is incorrect). Interestingly in the two dimensional case (i.e, the antiplane elastic or antiplane viscoelastic case) for two component media (and polycrystals of a single crystal), the characterization of the analytic properties is complete, and moreover the functional dependence of the matrix-valued effective dielectric tensor on the component moduli (or on the crystal tensor) can be mimicked to an arbitrarily high degree of approximation by a hierarchical laminate structure (Milton 1986b; Clark and Milton 1994) [see also Section 18.5 in Milton 2002] or when the two-component composite is isotropic by an assemblage of multicoated cylinders (Milton 1981a) [see the paragraph preceding Section VI]. Consequently the entire hierarchy of (antiplane viscoelasticity) bounds for two-dimensional transversely isotropic composites derived by Milton (1981c) are sharp, as are those obtained by Clark and Milton (1995).

For two-component media, Kantor and Bergman (1984) obtained a integral representation formula for the analytic properties of the effective elasticity tensor along a one-parameter trajectory in the moduli space (later generalized to two-parameter trajectories by Ou (2012)). A general framework for representation formulas, which yields representation formulas for the effective tensor for dielectrics, elasticity, piezoelectricity, thermoelasticity, thermoelectricity, and other coupled field problems in multicomponent (possibly polycrystalline) nonlossy or lossy media (with possibly nonsymmetric local tensors or having real and imaginary parts which do not necessarily commute) was developed by Milton [see Sections 18.6, 18.7, and 18.8 of Milton 2002]. When more than two (real or complex) moduli are involved, another powerful approach, the field equation recursion method which is based on subspace collections, generates a whole hierarchy of bounds on effective tensors (not just on their associated quadratic forms), including the effective dielectric tensors of multicomponent (possibly polycrystalline) dielectric media with real or complex moduli, and the effective elastic or viscoelastic tensors of multicomponent (possibly polycrystalline) phases (Milton 1987a, 1987b) [see also **Chapter 29** of Milton (2002)]. These bounds are applicable provided the real and imaginary parts of the local dielectric tensor, or viscoelasticity tensor, commute (i.e., can be simultaneously diagonalized in an appropriate basis).

Another breakthrough came when Cherkaev and Gibiansky (1994) derived variational principles for electromagnetism with lossy materials and for viscoelasticity, assuming quasistatic equations and (fixed frequency) time-harmonic fields. This provided a powerful tool for obtaining bounds on the complex dielectric constant of multicomponent (possibly anisotropic) media (Milton 1990) and for obtaining bounds on the complex bulk and shear moduli of two- and three-dimensional two-phase composites (Gibiansky and Milton 1993, Gibiansky and Lakes 1993, 1997, Milton and Berryman 1997, and Gibiansky, Milton, and Berryman 1999), using both Hashin–Shtrikman method and the translation method. These variational principles of Cherkaev and Gibiansky have been extended to media with nonsymmetric tensors by Milton (1990) [such as occur in

conduction when a magnetic field is present: see Briane and Milton 2011, where bounds are developed using these variational principles] and to beyond the quasistatic regime, to the full time-harmonic equations of electromagnetism, acoustics, and elastodynamics in lossy inhomogeneous bodies (Milton, Seppecher, and Bouchitté 2009; Milton and Willis 2010).

By contrast, very few results have been obtained regarding bounds on the creep and relaxation functions in the time domain: Schapery (1974) provided some interesting results via pseudo-elastic approximations; Huet (1995) using the concept of a pseudo-convolutive bilinear form derived useful unilateral and bilateral bounds for the relaxation function tensor; Vinogradov and Milton (2005) obtained bounds which correlate the very short time response with the long time asymptotic behavior; and Carini and Mattei (2015) have derived some elementary bounds from their novel variational principles in the time domain, which exploit the positive definiteness of a part of the constitutive law operator, combined with the transformation technique of Cherkaev and Gibiansky (1994) and Milton (1990) (note that Milton's work was based on that of Cherkaev and Gibiansky). The bounds of Carini and Mattei correlate the response at different times, while we are primarily interested in bounds on the response at a fixed given time.

Here we use the analytic integral representation formula developed by Bergman (1978) (justified by Milton (1981a) and proved by Golden and Papanicolaou (1983)) to obtain bounds on the macroscopic response of a two component composite (with microstructure independent of x_1) in the time domain for antiplane viscoelasticity. Our objective is to bound the transient response of the composite material. The key point which leads to the bounds is the observation that the response at a fixed time is linear in the residues (or eigenvalues of the residues when they are matrix-valued) which enter the representation formula. This enables one to use linear programming theory to reduce the problem to one involving relatively few nonzero residues and then the optimization over the pole positions (and orientation of the residue matrices if they are anisotropic) can be done numerically.

There are two main conclusions that follow from our work. The first is that if sufficient information about the composite is incorporated in the bounds, such as the volume fractions of the phases and the fact that the geometry is transversely isotropic, the bounds can be quite tight over the entire range of time. This should be very useful for predicting the transient behavior of composites. The second very significant point is that the bounds incorporating the volume fractions (and possibly transverse isotropy) can be extremely tight at certain specific times: thus measuring the response at such times, and using the bounds in an inverse fashion, could give very tight (and presumably useful) bounds on the volume fraction of the phases in the composite. As suggested to us by Yuhang Hu, of the University of Illinois at Urbana-Champaign, these bounds can be used in other ways too: if the volume fraction was known, but the elastic moduli of the purely elastic phase were not, then from the response at these specific times one could get tight bounds on the shear modulus of this elastic phase. The bounds we derive could be tightened further, for example, by incorporating information about the complex effective tensor measured at one or more frequencies (with cyclical loading).

An important application of our work is to the torsion of two-phase beams (or other cylindrical structures not necessarily with circular cross-sections) with microstructure, such as fiberous materials with very long parallel fibers surrounded by matrix material. It is assumed that the moduli are independent of the x_1-coordinate, where the x_1 axis is chosen to be parallel to the beam. Then provided there is a wide separation of length scales the beam behaves as a homogeneous elastic material with monoclinic symmetry (as it has reflection symmetry about any plane $x_1 = c$, where c is a constant) and thus (see, for example, Love 1906) locally undergoes antiplane shear, the only non-zero components of the macroscopic shear $\overline{\boldsymbol{\epsilon}}(\mathbf{x}, t)$ being the $\overline{\epsilon}_{12}(\mathbf{x}, t)$ and $\overline{\epsilon}_{13}(\mathbf{x}, t)$ components. The applicability of homogenization theory is established in more detail in the two-scale analyses of Tokarzewski, Telega, and Galka (2001) and Bonifasi-Lista and Cherkaev (2008). [There are also results on the torsion of composite beams where homogenization theory does not apply: see

Benveniste and Chen (2003) and Barretta, Luciano, and Willis (2015). In this case our analysis is not directly applicable.] Note that even a homogeneous circular cylinder under torsion locally undergoes antiplane shearing, even though it does not look like an antiplane deformation since $u_1(\mathbf{x}) = 0$ for all $\mathbf{x}$. Let us suppose the beam undergoes some time varying torsion, and let us ignore Saint-Venant end effects and ignore the travel time for disturbances to propagate along the beam. Then assuming the microstructure was independent of x_1, transversely isotropic, and the same throughout a cross-section of the beam, each representative volume element would within a proportionality factor and moduli rotations undergo the same response as a function of time. In this way our analysis for bounding the response of each representative volume element will directly translate into bounds on the response of the whole beam, such as bounds on the torque generated in response to time varying rotations applied at the ends of the cylindrical structure. Therefore it would be interesting if the predictions of our theory could be tested via such torsion experiments.

We remark that the method we use here is immediately applicable to bounding the transient response of three-dimensional two-component composites of lossy dielectric materials (or mixtures of a lossy material with a nonlossy one)(the case of two-dimensional two-component composites is of course mathematically isomorphic to the antiplane viscoelastic case studied here). This will be presented in a separate paper, directed towards physicists and electrical engineers. The observation made at the end of the Introduction of the previous chapter is also relevant here: such an analysis might be of limited practical utility due to the fast electrical relaxation times of many materials, especially conductors, and measurements would need to be made on those time scales to capture the transient response. We also believe the method can be extended to obtain bounds on the transient response of fully three-dimensional viscoelastic composites, not just in the antiplane case. In this setting, it is likely that the representation formulas for the effective elasticity tensor derived by Kantor and Bergman (1984) and Ou (2012) and in Sections 18.6, 18.7, and 18.8 of Milton (2002), or their generalizations, will prove useful.

This chapter is mostly self-contained and can be read separately from the rest of the book.

6.2 Summary of the results

The results here presented concern bounds on the response, in terms of stresses and strains, of a two-component viscoelastic composite material in the time domain. We suppose that the external loadings are applied in such a way as to generate an *antiplane shear* state within the material. We recall that such a state is achieved when the components $u_2(\mathbf{x}, t)$ and $u_3(\mathbf{x}, t)$ of the displacement field $\mathbf{u}(\mathbf{x}, t)$ are zero ($\mathbf{x}$ is the coordinate with respect to a Cartesian orthogonal reference system), for every $\mathbf{x} \in \Omega$ and every $t \in [0, +\infty)$, and the corresponding strain and stress states are of pure shear in the 12- and 13-planes, that is, by means of Voigt notation, they are represented by the two-component vectors $\boldsymbol{\epsilon}(\mathbf{x}, t) = [2\epsilon_{12}(\mathbf{x}, t) \;\; 2\epsilon_{13}(\mathbf{x}, t)]^{\mathrm{T}}$ and $\boldsymbol{\sigma}(\mathbf{x}, t) = [\sigma_{12}(\mathbf{x}, t) \;\; \sigma_{13}(\mathbf{x}, t)]^{\mathrm{T}}$. To ensure that a state of antiplane shear exists we assume that the microgeometry and, hence, the moduli depend only on x_2 and x_3.

We assume that both phases have an isotropic behavior, so that the direct and inverse constitutive laws, ruled by the 2×2 matrices $\mathbf{C}(\mathbf{x}, t)$ and $\mathbf{M}(\mathbf{x}, t)$, read as follows

$$\boldsymbol{\sigma}(\mathbf{x}, t) = \mathbf{C}(\mathbf{x}, t) * \boldsymbol{\epsilon}(\mathbf{x}, t) \quad \text{with} \quad \mathbf{C}(\mathbf{x}, t) = \sum_{i=1,2} \chi_i(\mathbf{x})\, \mu_i(t)\mathbf{I}, \tag{6.1}$$

$$\boldsymbol{\epsilon}(\mathbf{x}, t) = \mathbf{M}(\mathbf{x}, t) * \boldsymbol{\sigma}(\mathbf{x}, t) \quad \text{with} \quad \mathbf{M}(\mathbf{x}, t) = \sum_{i=1,2} \chi_i(\mathbf{x})\, \zeta_i(t)\mathbf{I}, \tag{6.2}$$

where $*$ indicates a time convolution, $\mathbf{I}$ is the identity matrix, $\chi_i(\mathbf{x})$ is the indicator function of phase i, and $\mu_i(t)$ and $\zeta_i(t)$ are, respectively, the shear stiffness and the shear compliance of phase i, both functions of time. A word about the notation may be helpful: on the left hand side of (6.1) [and (6.2)] $\boldsymbol{\sigma}(\mathbf{x},t)$ [respectively $\boldsymbol{\epsilon}(\mathbf{x},t)$] refers to the stress [strain] at a specific time t, while on the right hand side $\mathbf{C}(\mathbf{x},t)$ and $\boldsymbol{\epsilon}(\mathbf{x},t)$ [$\mathbf{M}(\mathbf{x},t)$ and $\boldsymbol{\sigma}(\mathbf{x},t)$] refer to the relaxation kernel and strain [creep kernel and stress] as functions of time from time 0 (before which there is no stress or strain) up to time t, which are convolved together to produce the stress [strain] at the specific time t.

In this investigation, we are interested in determining the effective behavior of the composite (we consider the most general case for which the composite does not have any specific symmetry), described by the effective direct and inverse constitutive law operators $\mathbf{C}_*(t)$ and $\mathbf{M}_*(t)$ as follows

$$\overline{\boldsymbol{\sigma}}(t) = \mathbf{C}_*(t) * \overline{\boldsymbol{\epsilon}}(t), \qquad \overline{\boldsymbol{\epsilon}}(t) = \mathbf{M}_*(t) * \overline{\boldsymbol{\sigma}}(t), \tag{6.3}$$

where here and henceforth the bar denotes the volume average operation. In particular, we seek estimates for the shear stress and strain components $\overline{\sigma}_{12}(t)$ and $\overline{\epsilon}_{12}(t)$ for each time $t \in [0,\infty)$.

By applying the so-called *analytic method*, based on the analyticity properties of the Laplace transforms $\mathbf{C}_*(\lambda)$ and $\mathbf{M}_*(\lambda)$ (λ is the Laplace transform parameter) of the operators $\mathbf{C}_*(t)$ and $\mathbf{M}_*(t)$ as functions of the Laplace transforms $\mu_i(\lambda)$ and $\zeta_i(\lambda)$ of $\mu_i(t)$ and $\zeta_i(t)$, $i = 1, 2$ (see Section 6.3), the effective constitutive laws (6.3) turn into:

$$\overline{\boldsymbol{\sigma}}(t) = \mu_2(t) * \overline{\boldsymbol{\epsilon}}(t) - \sum_{i=0}^{m} \mathbf{B}_i \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s(\lambda) - s_i}\right](t) * \overline{\boldsymbol{\epsilon}}(t), \tag{6.4}$$

$$\overline{\boldsymbol{\epsilon}}(t) = \zeta_2(t) * \overline{\boldsymbol{\sigma}}(t) - \sum_{i=0}^{m} \mathbf{P}_i \, \mathcal{L}^{-1}\left[\frac{\zeta_2(\lambda)}{u(\lambda) - u_i}\right](t) * \overline{\boldsymbol{\sigma}}(t), \tag{6.5}$$

where $\mathcal{L}^{-1}$ represents the inverse of the Laplace transform, s_i and u_i are, respectively, the poles of the functions

$$\mathbf{F}(s) = \mathbf{I} - \frac{\mathbf{C}_*(\lambda)}{\mu_2(\lambda)}, \qquad \mathbf{G}(u) = \mathbf{I} - \frac{\mathbf{M}_*(\lambda)}{\zeta_2(\lambda)}, \tag{6.6}$$

with residues $\mathbf{B}_i$ and $\mathbf{P}_i$, respectively, where the parameters $s(\lambda)$ and $u(\lambda)$ are defined as follows

$$s(\lambda) = \frac{\mu_2(\lambda)}{\mu_2(\lambda) - \mu_1(\lambda)}, \qquad u(\lambda) = \frac{\zeta_2(\lambda)}{\zeta_2(\lambda) - \zeta_1(\lambda)}. \tag{6.7}$$

The poles s_i and u_i lie in the semi-closed interval $[0, 1)$ and the residues $\mathbf{B}_i$ and $\mathbf{P}_i$ are positive semi-definite matrices. It must be noted that equations (6.4) and (6.5) hold only in the case when $\mathbf{C}_*(\lambda)$ and $\mathbf{M}_*(\lambda)$ are rational functions of the eigenvalues $\mu_i(\lambda)$ and $\zeta_i(\lambda)$, $i = 1, 2$, respectively. There is no lack of generality in considering only rational functions, since, from a physical viewpoint, the antiplane elasticity problem under consideration is mathematically equivalent to one for electrical conductivity and as pointed out by Milton (1981a) one can approximate the composite by a discrete resistor network, for which these functions are rational (to get the idea, see **Figure 2.6**). Mathematically, the representation of these functions involve Stieltjes functions (i.e., functions of the form (1.184)) and irrational Stieltjes functions can be approximated to an arbitrarily high degree of approximation by rational ones, except in the near vicinity of the negative real axis and far from the origin. (Then the kernel entering the representation formula is a smoothing operation and consequently the measure associated with an irrational Stieltjes function can be approximated by a discrete measure).

All the information about the composite, such as the knowledge of the volume fractions or the eventual isotropy of the material, is then transformed into constraints on the residues $\mathbf{B}_i$ and $\mathbf{P}_i$, the so-called *sum rules*, introduced by Bergman (1978) and discussed in Section 6.3. Such constraints are then contextualized in Section 6.4 so that bounds on the components $\overline{\sigma}_{12}(t)$ and $\overline{\epsilon}_{12}(t)$ are derived by means of the theory of linear programming. In particular, for each information available about the composite, that is, for each sum rule that is taken into account, we provide analytic expressions for the maximum and minimum values of the field components $\overline{\sigma}_{12}(t)$ and $\overline{\epsilon}_{12}(t)$ at each instant in time, when the applied fields are respectively $\overline{\boldsymbol{\epsilon}}(t) = [\overline{\epsilon}_{12}(t)\ \ 0]^T$ and $\overline{\boldsymbol{\sigma}}(t) = [\overline{\sigma}_{12}(t)\ \ 0]^T$ (see Section 6.5).

In this section we present some numerical results by specifying the models used for the behavior of the two phases, so that the inverse of the Laplace transform in (6.4) and (6.5) can be calculated explicitly.

6.2.1 Bounds on the stress response

For the sake of simplicity, we suppose that phase 2 is characterized by a linear elastic behavior, with shear modulus $\mu_2(t) = G_2\delta(t)$, $\delta(t)$ being the Dirac delta function, and that phase 1 is described by the Maxwell model. We recall that such a model is represented by a purely Newtonian viscous damper (viscosity coefficient η_M) and a purely Hookean elastic spring (elastic modulus G_M) connected in series so that the shear modulus of phase 1 is $\mu_1(t) = G_M\delta(t) - G_M^2/\eta_M \exp[-G_M t/\eta_M]$. We caution that the Maxwell model has only one relaxation time η_M/G_M, while real materials may have an extraordinarily broad range of relaxation times extending over many decades (Lakes and Quackenbush 1996). Our methods extend to them too: the Maxwell model is just used for simplicity, and because some of the calculations can be done analytically.

To capture the most interesting case, we suppose that the material is not "well-ordered", that is, the product of the difference of the instantaneous moduli (very close to $t = 0$) and the long time moduli (as t tends to infinity) is negative, i.e., $G_2 < G_M$. Nevertheless, for completeness, in the following we will show also some results concerning the "well-ordered" case, that is, when the product of these differences is positive ($G_2 > G_M$).

We consider the classical relaxation test in which the applied average stain is held constant after being initially applied, i.e., $\overline{\boldsymbol{\epsilon}}(t) = \boldsymbol{\epsilon}_0 = [\epsilon_0\ \ 0]^{\mathrm{T}}$. From (6.4) we derive the following expression for $\overline{\sigma}_{12}(t)$:

$$\overline{\sigma}_{12}(t) = G_2\epsilon_0 - G_2\epsilon_0 \sum_{i=0}^{m} \left\{ 1 - \frac{\exp\left[-\frac{G_2(1-s_i)t}{\eta_M\left(\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M}-1\right)\right)} \right]}{\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M}-1\right)} \right\} \frac{B_{11}^{(i)}}{1-s_i}, \tag{6.8}$$

where $B_{11}^{(i)}$ are the 11-components of the 2×2 matrices $\mathbf{B}_i$.

Now suppose that no information about the geometry of the composite is available. As shown in Subsection 6.5.1, in order to optimize $\overline{\sigma}_{12}(t)$ for any given time $t \in [0, \infty)$, it suffices (by linear programming theory) to take only one element $B_{11}^{(0)}$ to be nonzero. In particular, it turns out that $B_{11}^{(0)} = 1 - s_0$ and the expression of $\overline{\sigma}_{12}(t)$ is then given by

$$\overline{\sigma}_{12}(t) = G_2\epsilon_0 \frac{\exp\left[-\frac{G_2(1-s_0)t}{\eta_M[s_0(1-G_2/G_M)+G_2/G_M]} \right]}{s_0\,(1 - G_2/G_M) + G_2/G_M}. \tag{6.9}$$

The maximum (or minimum) value of $\overline{\sigma}_{12}(t)$ is obtained by varying the pole s_0 over its domain of validity, i.e., $[0, 1)$. Since the response (6.9) corresponds to that of a laminate oriented with the x_2 axis normal to the layer planes, varying s_0 corresponds to varying the volume fraction of the phases in the laminate (since no

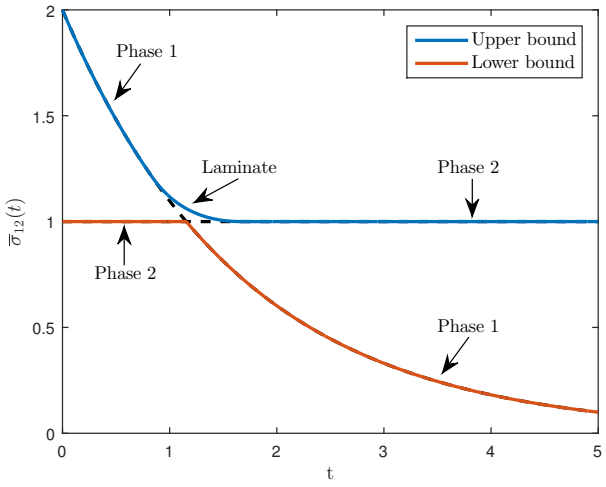


Figure 6.1: Lower and upper bounds on $\overline{\sigma}_{12}(t)$ when no information about the composite is given. The stress $\overline{\sigma}_{12}(t)$ is normalized with respect to the elastic stress in phase 2, equal to $\epsilon_0 G_2$. The material purely made of phase 1 provides the upper bound for $t \leq t_1 = 0.83$ and the lower bound for $t \geq t_2 = 1.15$, whereas the material purely made of phase 2 attains the lower bound for $t \leq t_2 = 1.15$ and the upper for $t \geq t_1 = 1.67$. For $t_1 \leq t \leq t_3$ the upper bound is realized by a laminate of the two components.

information about the composite is available, the volume fraction f_1 of phase 1 can be varied from 0 to 1). In particular, the case $s_0 = 0$ corresponds to a "composite" which contains only phase 1, while the case $s_0 \to 1$ corresponds to a "composite" which contains only phase 2.

As shown in **Figure 6.1**, where $\overline{\sigma}_{12}(t)$ is normalized with respect to the stress state in the elastic phase, $G_2\epsilon_0$, the material purely made of phase 1 ($s_0 = 0$) attains the upper bound for $t \leq t_1 = \eta_M/G_M(1 - G_2/G_M)$ (equal to 0.83 in **Figure 6.1**) and the lower bound for $t \geq t_2 = \eta_M/G_M \log(G_M/G_2)$ (equal to 1.15 in **Figure 6.1**), whereas the material purely made of phase 2 ($s_0 \to 1$) attains the lower bound for $t \leq t_2$ and the upper one for $t \geq t_3 = \eta_M/G_2(1 - G_2/G_M)$ (equal to 1.67 in **Figure 6.1**): the same microstructure can provide both the maximum and the minimum response depending on the interval of time considered. Furthermore, for $t_1 \leq t \leq t_3$ the upper bound is realized by a laminate of the two components corresponding to the pole s_0 positioned at

$$s_0^{opt} = \frac{\frac{tG_2}{\eta_K} - \frac{G_2}{G_K}\left(1 - \frac{G_2}{G_K}\right)}{\left(1 - \frac{G_2}{G_K}\right)^2}. \tag{6.10}$$

Due to the dependence of s_0^{opt} on time t, it follows that the volume of the phases in the laminate attaining the bounds needs to be adjusted according to the time at which one is optimizing the response.

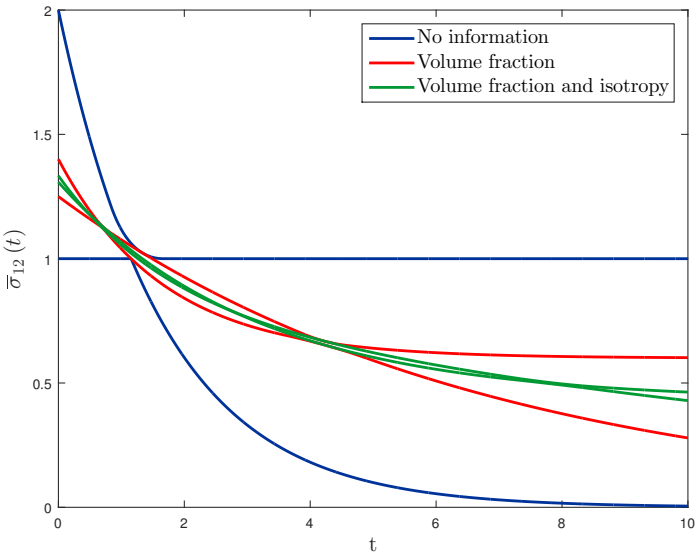


Figure 6.2: Comparison between the lower and upper bounds on $\overline{\sigma}_{12}(t)$ (normalized with respect to the elastic stress in phase 2, equal to $\epsilon_0 G_2$) in the following three cases: no information about the composite is given; the volume fraction of the components is known ($f_1 = 0.4$); and the composite is isotropic with given volume fractions. The bounds become tighter and tighter as more information on the composite structure is included, so that if color is missing from the figure the outermost pair of bounds are those with no information, the middle pair include just the volume fraction, and the innermost pair include both volume fraction and isotropy.

Specifically, the upper bound is given by

$$\overline{\sigma}_{12}^{\max}(t) = \begin{cases} \epsilon_0 G_M \exp\left[-\frac{G_M}{\eta_M} t\right] & s_0 = 0 \quad , t \le t_1, \\ \epsilon_0 \frac{\eta_M}{t}\left(1 - \frac{G_2}{G_M}\right) \exp\left[1 - \frac{t}{\eta_M}\frac{G_2 G_M}{G_M - G_2}\right] & s_0 = s_0^{opt} \quad , t_1 \le t \le t_3, \\ \epsilon_0 G_2 & s_0 \to 1 \quad , t \ge t_3, \end{cases} \tag{6.11}$$

whereas the lower bound corresponds to

$$\overline{\sigma}_{12}^{\min}(t) = \begin{cases} \epsilon_0 G_2 & s_0 \to 1 \quad , t \le t_2, \\ \epsilon_0 G_M \exp\left[-\frac{G_M}{\eta_M} t\right] & s_0 = 0 \quad , t \ge t_2. \end{cases} \tag{6.12}$$

In the case when the volume fractions of the components are known, tighter bounds can be obtained. In particular, in **Figure 6.2** we compare the results obtained by considering the following situations: no information about the composite is available (the case analyzed in detail above), the volume fraction of the constituents is known (two poles), and the composite is transversely isotropic with given volume fractions (three poles). It is worth noting that the bounds corresponding to the latter case are very tight and therefore the response of the composite in terms of $\overline{\sigma}_{12}(t)$ is almost completely determined.

Significantly, the bounds in **Figure 6.2** which include the volume fraction (and possibly, transverse isotropy) are extremely tight at particular times t, and so, if the volume fraction is unknown, we can measure the value of $\overline{\sigma}_{12}(t)$ at these times, and then use the bounds in an inverse fashion to determine (almost exactly) the volume fraction. To understand why the bounds are extremely tight at these times we rewrite the relation (6.8) in the form

$$\overline{\sigma}_{12}(t) = G_2\epsilon_0 - G_2\epsilon_0 \sum_{i=0}^{m} K(s_i, t) B_{11}^{(i)}, \tag{6.13}$$

with coefficients

$$K(s_i, t) = \left\{ 1 - \frac{\exp\left[-\frac{G_2(1-s_i)t}{\eta_M\left(\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M} - 1\right)\right)} \right]}{\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M} - 1\right)} \right\} \frac{1}{1 - s_i}. \tag{6.14}$$

If at a time $t = \tau_0$ the coefficients $K(s_i, t)$ were almost independent of s_i, i.e., $K(s_i, \tau_0) \approx c_0$ for all i, then by substituting this in (6.13) and using the sum rule given later in (6.39), we see that

$$\overline{\sigma}_{12}(\tau_0) \approx G_2\epsilon_0 - G_2\epsilon_0 c_0 f_1. \tag{6.15}$$

Alternatively, if at another time $t = \tau_1$ the coefficients $K(s_i, t)$ depend almost linearly on s_i, i.e., $K(s_i, \tau_1) \approx c_0 + c_1 s_i$ for all i, and the geometry is transversely isotropic, then by substituting this in (6.13) and using the sum rules given later in (6.39) and (6.40) we see that

$$\overline{\sigma}_{12}(\tau_1) \approx G_2\epsilon_0 - G_2\epsilon_0(c_0 f_1 + c_1 f_1 f_2/2). \tag{6.16}$$

Video 1[1] shows $K(s_i, t)$ as a function of time for our example, and we see indeed that the coefficients $K(s_i, t)$ are almost independent of s_i at the times when the bounds which incorporate only the volume fraction are very tight (for example, at $\tau_0 = 0.78$ and at $\tau_0 = 4.3$ — see also **Figure 6.2**), and they depend almost linearly on s_i at the times when the bounds which incorporate the volume fractions and the transverse isotropy are very tight (for instance, at $\tau_1 = 2.8$ and $\tau_1 = 8.21$ — see also **Figure 6.2**).

Other information about the composite can be considered, such as the knowledge of the value of $\overline{\sigma}_{12}(t)$ at a specific time. Figs. 6.3 and 6.4 show the results obtained in the case when the value of $\overline{\sigma}_{12}(t)$ at $t = 0$ and $t \to \infty$, respectively, is given.

With reference to **Figure 6.3**, notice that the combination of the knowledge of the volume fraction and of the value of $\overline{\sigma}_{12}(t)$ at $t = 0$, $\overline{\sigma}_{12}(0)$, provides very tight bounds on $\overline{\sigma}_{12}(t)$.

Concerning **Figure 6.4**, a few remarks should be made. First of all, notice that for the case when only the value of $\overline{\sigma}_{12}(t)$ for $t \to \infty$, $\overline{\sigma}_{12}(\infty)$, is prescribed, the upper bound seems to not reach such a value: it provides a constant stress state equal to the one in the material purely composed of phase 2. This is due to the fact that the only nonzero residue $B_{11}^{(0)} = (1 - s_0)\left(1 - \overline{\sigma}_{12}(\infty)/(G_2\epsilon_0)\right)$ in (6.8) takes a value very close to zero when the corresponding pole s_0 tends to 1 and, consequently, the predicted response is that of phase 2 (see equation (6.8)). However, in the near vicinity of $t \to \infty$, the upper bound rapidly converges to the prescribed value $\overline{\sigma}_{12}(\infty)$. Regarding the upper bound obtained by considering both the values of $\overline{\sigma}_{12}(t)$ at $t \to \infty$ and the volume fractions to be known, notice that it presents a small slope which allows it to slowly reach the value $\overline{\sigma}_{12}(\infty)$ for $t \to \infty$.

[1] All videos in this chapter are available at `http://www.math.utah.edu/books/milton`, along with high-resolution copies of the figures in this chapter.

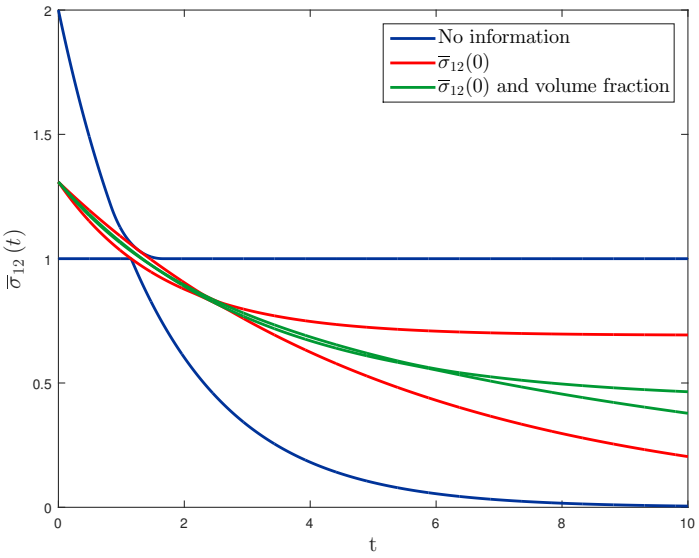


Figure 6.3: Comparison between the lower and upper bounds on $\overline{\sigma}_{12}(t)$ (normalized with respect to the elastic stress in phase 2, equal to $\epsilon_0 G_2$) in the following three cases: no information about the composite is given; the value of $\overline{\sigma}_{12}(t)$ at $t = 0$ is prescribed; and the value of $\overline{\sigma}_{12}(t)$ at $t = 0$ and the volume fractions are known ($f_1 = 0.4$). The bounds become tighter and tighter as more information about the composite structure is included, and this allows the bounds to be identified when their colors have not been reproduced.

The "well ordered" case, corresponding to the choice $G_2 > G_M$, is less interesting due to the fact that the curves representing the behavior of phase 1 and phase 2 do not intersect. However, for completeness, in **Figure 6.5**, we provide bounds on $\overline{\sigma}_{12}(t)$ for $G_2 > G_M$ in the following cases: no information about the composite is available; the volume fraction is known; and the composite is transversely isotropic with given volume fraction. Again, the bounds become tighter the more information about the composite is considered. Nevertheless, the bounds are wide compared to the case $G_2 < G_M$, and are tightest at $t = 0$.

Besides optimizing the component $\overline{\sigma}_{12}(t)$ of the averaged stress field $\overline{\boldsymbol{\sigma}}(t)$, one would like to determine also what are the possible values the vector $\overline{\boldsymbol{\sigma}}(t) = [\overline{\sigma}_{12}(t)\ \ \overline{\sigma}_{13}(t)]^T$ can take as time evolves. One way to get some information about this is to look for the maximum or minimum value attained by a linear combination of the components $\overline{\sigma}_{12}(t)$ and $\overline{\sigma}_{13}(t)$ of $\overline{\boldsymbol{\sigma}}(t)$. Let us consider, then, the following scalar objective function, for each fixed angle α:

$$\mathcal{F}(t) = \sin\alpha\,\overline{\sigma}_{12}(t) + \cos\alpha\,\overline{\sigma}_{13}(t), \tag{6.17}$$

where, in general, $\overline{\sigma}_{12}(t)$ and $\overline{\sigma}_{13}(t)$ are given by (6.4).

Let us assume that the same hypotheses valid for the bounds on $\overline{\sigma}_{12}(t)$ still hold, i.e., phase 1 is described

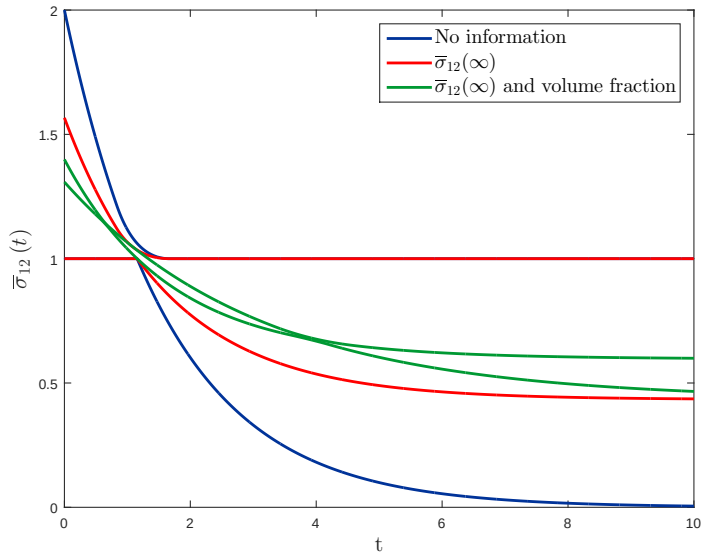


Figure 6.4: Comparison between the lower and upper bounds on $\overline{\sigma}_{12}(t)$ (normalized with respect to the elastic stress in phase 2, equal to $\epsilon_0 G_2$) in the following three cases: no information about the composite is given; the value of $\overline{\sigma}_{12}(t)$ at $t \to \infty$ is prescribed; and the value of $\overline{\sigma}_{12}(t)$ at $t \to \infty$ and the volume fractions are known ($f_1 = 0.4$). In the last two cases, the upper bound attains the assigned value of $\overline{\sigma}_{12}(t)$ at $t \to \infty$ only in the near vicinity of $t \to \infty$, whereas the lower bound converges very fast. In the first two cases the two upper bounds are almost indistinguishable (and essentially independent of t) for a large range of t greater than about 1.5, and the two lower bounds are almost indistinguishable (and essentially independent of t) for t less than about 1: this, and the fact that the bounds do not get worse as more information is included, allows one to identify them when their color is missing.

by the Maxwell model, phase 2 is elastic, and $\overline{\boldsymbol{\epsilon}}(t) = \boldsymbol{\epsilon}_0 = [\epsilon_0 \;\; 0]^{\mathrm{T}}$. Then, we have

$$\overline{\boldsymbol{\sigma}}(t) = G_2\boldsymbol{\epsilon}_0 - G_2 \sum_{i=0}^{m} \left\{ 1 - \frac{\exp\left[-\frac{G_2(1-s_i)t}{\eta_M\left(\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M}-1\right)\right)} \right]}{\frac{G_2}{G_M} - s_i\left(\frac{G_2}{G_M}-1\right)} \right\} \frac{\mathbf{B}_i}{1-s_i}\boldsymbol{\epsilon}_0. \tag{6.18}$$

Furthermore, we suppose that the microstructure has *reflective symmetry*, that is, it is symmetric with respect to reflection about a certain plane. Such an assumption implies that all residues $\mathbf{B}_i$ in (6.18) are diagonal matrices with respect to the same basis (i.e., they commute). In general, optimizing the quantity $\mathcal{F}(t)$ for a fixed α and, then, varying α will only allow us to find the convex hull of the set of possible vectors $\overline{\boldsymbol{\sigma}}(t)$ at each time t. However, in the case of reflective symmetry, we can first fix the orientation of the residues (i.e., the orientation of the composite) [2] and, then, we can find the minimum value of $\mathcal{F}(t)$, and a (possibly

[2]Fixing the orientation of the composite means fixing the value of the angle $\theta = \theta_i$, for each i, in equation (6.42). In particular, when the orientation is fixed, two possible configurations of the microstructure are admissible: one corresponds to the angle θ and the other, reflected with respect to the first one, corresponds to the angle $\theta + \pi/2$. Strictly speaking the microstructure does not necessarily have this additional reflective symmetry, but the associated effective tensor does have it.

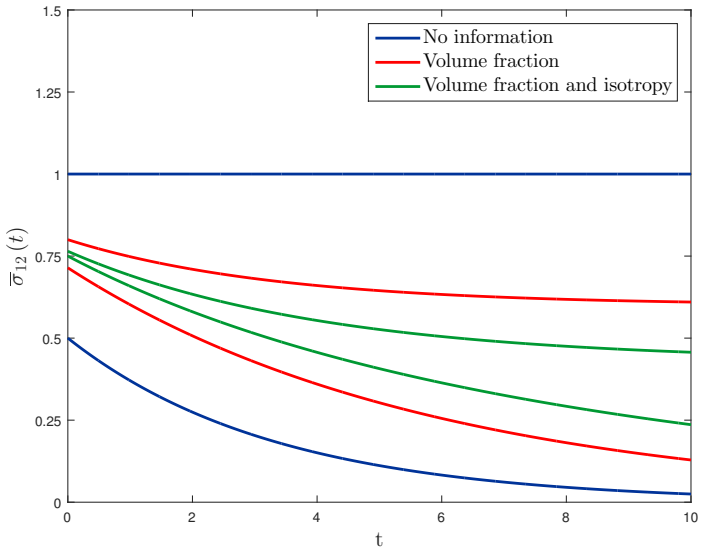


Figure 6.5: Comparison between the lower and upper bounds on $\overline{\sigma}_{12}(t)$ (normalized with respect to the elastic stress in phase 2, equal to $\epsilon_0 G_2$) in the "well-ordered case" $G_2 > G_M$. The following three subcases are considered: no information about the composite is given; the volume fraction of the components is known ($f_1 = 0.4$); and the composite is isotropic with given volume fractions. The bounds become tighter as more information on the composite structure is included, and this allows them to be identified when their color is missing. However the bounds remain quite wide except near $t = 0$.

nonunique) function $\overline{\boldsymbol{\sigma}}(t, \alpha)$ which realizes it (observe that finding the maximum value of $\mathcal{F}(t)$ is the same as finding the minimum when α is replaced by $\alpha + \pi$). Next, for each t construct the set which is the union of the points $\overline{\boldsymbol{\sigma}}(t, \alpha)$ as α varies between 0 and 2π, and take its convex hull: the boundary of this convex hull is the trajectory of $\overline{\boldsymbol{\sigma}}(t, \alpha)$ as α increases, except if there is a jump in the value of $\overline{\boldsymbol{\sigma}}(t, \alpha)$, for which the successive values of $\overline{\boldsymbol{\sigma}}(t, \alpha)$ are joined by a straight line. Finally, we take the union of these convex hulls as the orientation is varied. In this way we obtain bounds which at any instant of time t confine the pair $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$ to a region which is not necessarily convex.

In the case when no information about the geometry of the composite is available, apart from the reflective symmetry, the optimum value of $\mathcal{F}(t)$ is attained when a maximum of two residues are nonzero.

Video 2, plots $\overline{\sigma}_{12}(t)$ against $\overline{\sigma}_{13}(t)$ (both normalized by $G_2\epsilon_0$, the stress state in phase 2) for each moment of time, in the case when the orientation of the composite is fixed (blue curve). To enrich the results, the video also plots the domain $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$ corresponding to the stress state in a laminate with the prescribed orientation (red curve). We recall that for a laminate the stress state is unequivocally determined, since the eigenvalues of the two nonnull residues are related to the harmonic and arithmetic means of the moduli of the two phases. Note that, since no information about the composite is available, the volume fraction f_1 of phase 1 can vary from 0 to 1. In the initial frame of the video, at $t = 0$, the point $(0, 1)$ (corresponding to $s_0 \to 1$, $s_1 \to 1$) represents the instantaneous stress state within phase 2, whereas the point $(0, 2)$ (corresponding to $s_0 = s_1 = 0$) represents the stress state within phase 1. Obviously, both points belong also to the red curve representing the laminate behavior. As time goes by, the domain becomes smaller and smaller with the upper

vertex still representing the behavior of phase 1 while the lower vertex, the point $(0,1)$, remaining fixed as it represents the elastic behavior of phase 2. For times t between $t_1 = 0.83$ and $t_3 = 1.67$ a change takes place: the upper vertex does not represent the response of a phase 1, nor even that of a laminate. Then for times $t > t_3 = 1.67$, the upper vertex coincides with the point $(0,1)$, representing the behavior of phase 2. The lower vertex describes the behavior of phase 2 until $t = t_2 = 1.15$, after which it represents the behavior of phase 1.

When the orientation of the composite is not known, one has to perform the previous analysis for each possible orientation and, then, take the union of the resulting domains, as shown in Video 3.

For each fixed value of the angle α, sharp bounds on the function $\mathcal{F}(t)$ (6.17) give the straight lines forming an angle equal to α, with respect to the $\overline{\sigma}_{13}(t)$-axis, which are tangent to the domain of possible $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$. For each time, the values of $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$ which attain the bounds on $\mathcal{F}(t)$ correspond to those points where the tangent line intersects this domain.

We do not provide numerical results bounding the function $\mathcal{F}(t)$ for the case in which the volume fractions of the components are known, due to the large number of variables involved.

6.2.2 Bounds on the strain response

In this case, we suppose that phase 2 is still elastic, with $\zeta_2(t) = 1/G_2\delta(t)$, while we represent the behavior of phase 1 by means of the Kelvin–Voigt model, composed by a purely viscous damper (η_K) and purely elastic spring (G_K) connected in parallel, so that $\zeta_1(t) = \exp(-G_K t/\eta_K)/\eta_K$. The most interesting results correspond to the non "well-ordered" case, corresponding to $G_2 < G_K$.

Moreover, if we consider the classical creep test, for which the applied averaged stress field is constant in time after it has been initially imposed, i.e., $\overline{\boldsymbol{\sigma}}(t) = \boldsymbol{\sigma}_0$, and we set $\boldsymbol{\sigma}_0 = [\sigma_0 \;\; 0]^{\mathrm{T}}$, then equation (6.5) yields:

$$\overline{\epsilon}_{12}(t) = \frac{\sigma_0}{2G_2} - \frac{\sigma_0}{2G_2}\sum_{i=0}^{m}\left\{ G_K - G_2 + G_2\frac{\exp\left[-\frac{(G_K(1-u_i)+u_iG_2)t}{\eta_K(1-u_i)}\right]}{1-u_i}\right\}\frac{P_{11}^{(i)}}{G_K - u_i(G_K-G_2)}, \quad (6.19)$$

where the $P_{11}^{(i)}$ are the 11-components of the residues $\mathbf{P}_i$.

In the case when no information about the geometry of the composite is available, bounds on $\overline{\epsilon}_{12}(t)$ are obtained by taking only one residue to be nonzero (see Subsection 6.5.2). In particular, it turns out that $P_{11}^{(0)} = 1 - u_0$ and $\overline{\epsilon}_{12}(t)$, from (6.19), takes the following form:

$$\overline{\epsilon}_{12}(t) = \frac{\sigma_0}{2G_2}\left\{1 - \frac{(1-u_0)(G_K-G_2) + G_2\exp\left[-\frac{G_K - u_0(G_K-G_2)}{\eta_K(1-u_0)}t\right]}{G_K - u_0(G_K-G_2)}\right\}. \quad (6.20)$$

As shown in **Figure 6.6**, the material purely made of phase 1 ($u_0 = 0$) attains the lower bound for $t \leq t_I = \frac{\eta_K}{G_K}\log\left(\frac{G_2}{G_2-G_K}\right) = 2.78$ and the upper bound for $t \geq t_{II} = 5.14$, whereas the material purely made of phase 2 ($u_0 \to 1$) attains the lower bound for $t \geq t_I = 2.78$. For $t_I \leq t \leq t_{II}$ the upper bound is achieved by a laminate. Figs. 6.7, 6.8, and 6.9 depict the bounds on $\overline{\epsilon}_{12}(t)$ for different combinations of information about the composite. In particular, **Figure 6.7** shows the results when the volume fraction is known and the composite is transversely isotropic, **Figure 6.8** when f_1 and $\overline{\epsilon}_{12}(0)$ are assigned, and **Figure 6.9** when f_1 and $\overline{\epsilon}_{12}(\infty)$ are prescribed. For each case, very tight bounds on $\overline{\epsilon}_{12}(t)$ are obtained.

With reference to **Figure 6.8**, it is worth noting that the upper bound attains the value $\overline{\epsilon}_{12}(0)$ by converging to such a value only in the near vicinity of $t = 0$. This is due to the fact that the only nonzero residue

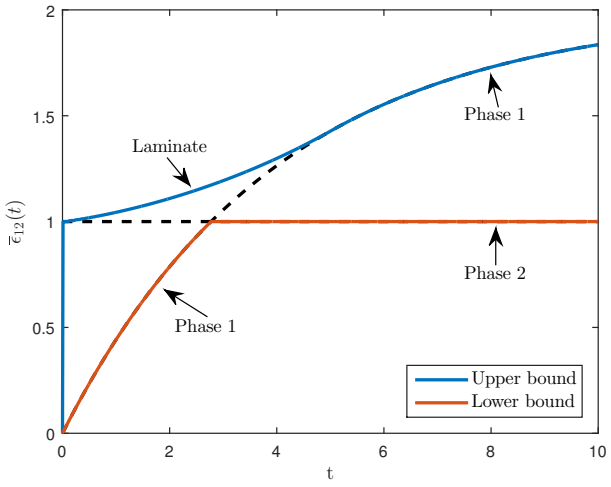


Figure 6.6: Lower and upper bounds on $\bar{\epsilon}_{12}(t)$ (normalized with respect to the elastic strain in phase 2, equal to $\sigma_0/(2G_2)$) in the case when no information about the composite is given. The material purely made of phase 1 provides the lower bound for $t \leq t_I = 2.84$ and the upper bound for $t \geq t_{II} = 5.16$, whereas the material purely made of phase 2 attains the lower bound for $t \geq t_I = 2.84$. For $t \leq t_{II} = 5.16$ the upper bound is realized by a laminate of the two components.

$P_{11}^{(0)} = (1-u_0)(1 - G_2\bar{\epsilon}_{12}(0)/\sigma_0)$ tends to zero as $u_0 \to 1$.

One would like also to seek bounds on the values of the vector $\bar{\boldsymbol{\epsilon}}(t) = [\bar{\epsilon}_{12}(t) \;\; \bar{\epsilon}_{13}(t)]^T$ as time evolves. We do this by seeking bounds on a linear combination of the components $\bar{\epsilon}_{12}(t)$ and $\bar{\epsilon}_{13}(t)$ of $\bar{\boldsymbol{\epsilon}}(t)$. Let us consider, then, the following objective function, at a fixed angle α:

$$\mathcal{G}(t) = \sin\alpha\,\bar{\epsilon}_{12}(t) + \cos\alpha\,\bar{\epsilon}_{13}(t). \tag{6.21}$$

We suppose that the following hypotheses still hold: phase 1 is described by the Kelvin–Voigt model, phase 2 has an elastic behavior, and the applied stress history is constant in time for $t > 0$. Then, equation (6.5) turns into

$$\bar{\boldsymbol{\epsilon}}(t) = \frac{\boldsymbol{\sigma_0}}{2G_2} - \frac{1}{2G_2}\sum_{i=0}^{m}\left\{G_K - G_2 + G_2\frac{\exp\left[-\frac{(G_K(1-u_i)+u_iG_2)t}{\eta_K(1-u_i)}\right]}{1-u_i}\right\}\frac{\mathbf{P}_i}{G_K - u_i(G_K-G_2)}\boldsymbol{\sigma}_0. \tag{6.22}$$

We assume the composite has reflection symmetry and following the same argument adopted for deriving bounds on $\mathcal{F}(t)$ (6.17), we first fix the orientation of the composite (i.e., residues), then for each time t we minimize the function $\mathcal{G}(t)$ (6.21), where the components $\bar{\epsilon}_{13}(t)$ of $\bar{\boldsymbol{\epsilon}}(t)$ are given by (6.22), and we look for a function $\bar{\boldsymbol{\epsilon}}(t,\alpha)$ which achieves the minimum. Next, for each t we construct the set which is the union of the

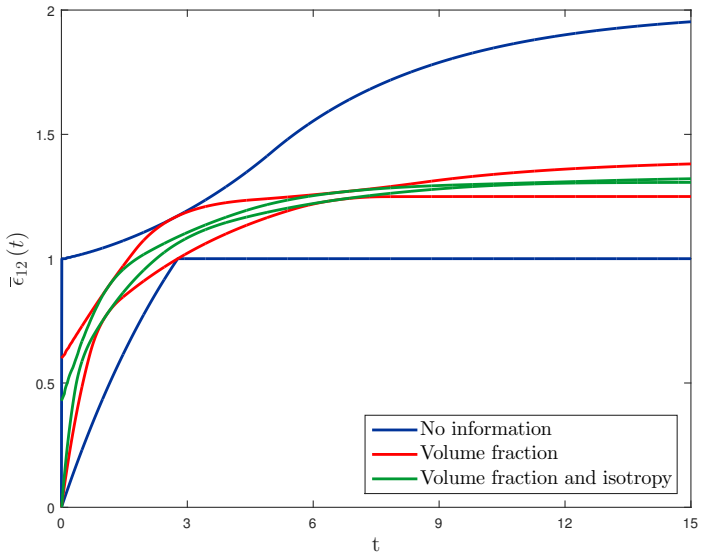


Figure 6.7: Comparison between the lower and upper bounds on $\overline{\epsilon}_{12}(t)$ (normalized with respect to the elastic strain in phase 2, equal to $\sigma_0/(2G_2)$) in the following three cases: no information about the composite is given; the volume fraction of the components is known ($f_1 = 0.4$); and the composite is isotropic with given volume fractions. Again they get tighter as more information is included which allows them to be identified in the absence of color.

points $\overline{\boldsymbol{\sigma}}(t, \alpha)$ as α varies between 0 and 2π, and take its convex hull. Finally, we take the union of the results as the orientation of the composite is varied.

In Videos 4 and 5, we plot the domain $\overline{\epsilon}_{13}(t)$–$\overline{\epsilon}_{12}(t)$ (where both strains have been normalized by the strain field in the elastic phase $\sigma_0/(2G_2)$) for each time $t \in [0, \infty)$, for the case when no information about the composite is available. In particular, in Video 4 we suppose one knows the orientation of the composite, while in Video 5 we suppose that such information is not available and, therefore, we consider the union of the domains calculated for each fixed orientation. Once again, the results are enriched by considering also the exact solution provided by a laminate.

The optimum value of $\mathcal{G}(t)$ is attained when a maximum of two residues are nonzero. At $t = 0$, the strain field turns out to be $\overline{\boldsymbol{\epsilon}}(0) = (0, 0)$ and, therefore, it does not depend on the position of the poles u_0 and u_1. For times $t > 0$, instead, we maximize (or minimize) $\mathcal{G}(t)$ by varying the position of the two poles. The point $(0, 1)$, corresponding to $u_0 \to 1$, $u_1 \to 1$, keeps fixed since it represents the elastic response of phase 2. As the time goes by, the domain becomes smaller and smaller converging towards this point. At $t = t_I = 2.78$, the domain coincides with the one representing the laminate response and, then, for $t > t_I = 2.78$ it becomes bigger and bigger above the point $(0, 1)$.

Assigned the angle α (see equation (6.21)), bounds on $\mathcal{G}(t)$ are derived by considering the points of intersection between the domain $(\overline{\epsilon}_{13}(t), \overline{\epsilon}_{12}(t))$ and the tangents having slope equal to $\tan\alpha$, for each time t.

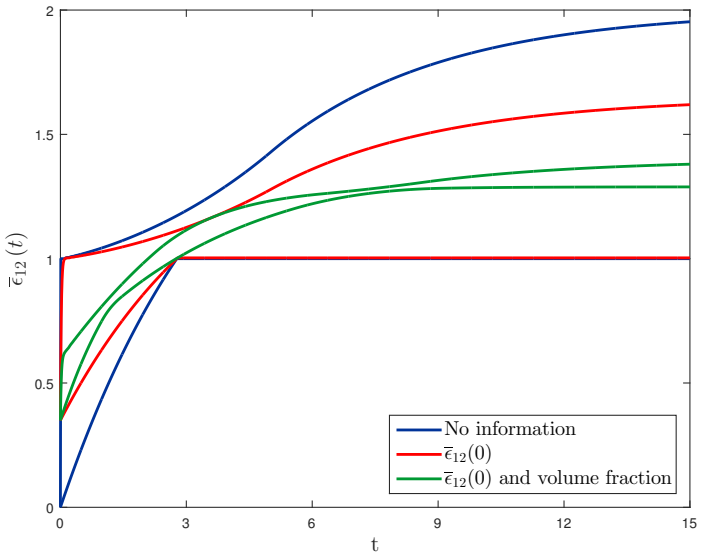


Figure 6.8: Comparison between the lower and upper bounds on $\overline{\epsilon}_{12}(t)$ (normalized with respect to the elastic strain in phase 2, equal to $\sigma_0/(2G_2)$) in the following three cases: no information about the composite is given; the value of $\overline{\epsilon}_{12}(t)$ at $t = 0$ is prescribed; and the value of $\overline{\epsilon}_{12}(t)$ at $t = 0$ and the volume fractions are known ($f_1 = 0.4$). In the last two cases, the upper bound attains the assigned value of $\overline{\epsilon}_{12}(t)$ at $t = 0$ only in the near vicinity of $t = 0$. In the first two cases the lower bounds are indistinguishable (and essentially independent of t) for t greater than about 2. This and the ordering of the bounds allows them to be identified when their color is missing.

6.3 Formulation of the problem

We consider a 3D body Ω made of a statistically homogeneous two-phase composite material with a length scale of inhomogeneities much smaller than the length scale of the body (that is, Ω can be interpreted as the Representative Volume Element (RVE) of the composite), and subject on the boundary Γ either to prescribed displacements or to assigned tractions, applied in such a way as to generate a *shear antiplane* state within the solid.

In the case when the volume average of the strain field $\overline{\boldsymbol{\epsilon}}(t)$ is assigned, we choose kinematic boundary conditions of the affine type all over the surface Γ:

$$u_1(\mathbf{x},t) = 2H(t)\ \left(\overline{\epsilon}_{12}(t)\, x_2 + \overline{\epsilon}_{13}(t)\, x_3\right), \quad u_2(\mathbf{x},t) = u_3(\mathbf{x},t) = 0, \tag{6.23}$$

with $H(t)$ the Heaviside unit-step function of time, whereas in the case when the volume average of the stress field $\overline{\boldsymbol{\sigma}}(t)$ is prescribed, we apply homogeneous tractions $\mathbf{p}(\mathbf{x},t)$ on Γ:

$$p_1(\mathbf{x},t) = H(t)\ \left(\overline{\sigma}_{12}(t)\, n_2(\mathbf{x}) + \overline{\sigma}_{13}(t)\, n_3(\mathbf{x})\right), \quad p_2(\mathbf{x},t) = p_3(\mathbf{x},t) = 0, \tag{6.24}$$

where $\mathbf{n}(\mathbf{x})$ is the unit outward normal to $\partial\Omega$.

The local constitutive equations are given by (6.1) and (6.2), while the effective constitutive laws are expressed by equation (6.3).

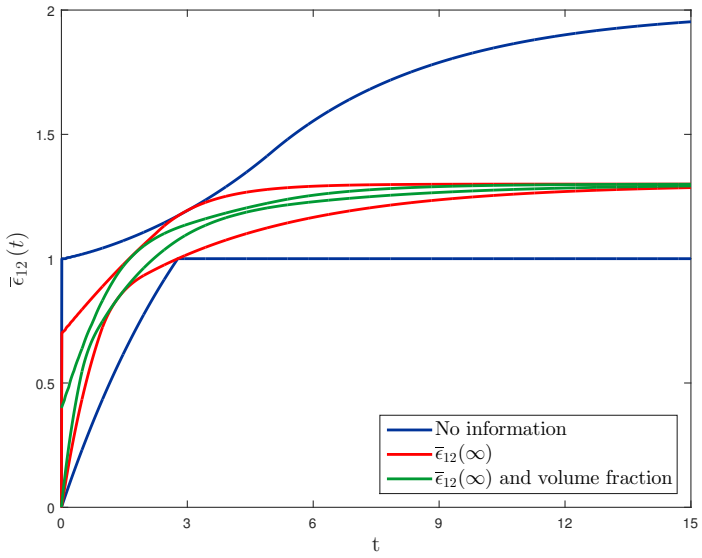


Figure 6.9: Comparison between the lower and upper bounds on $\bar{\epsilon}_{12}(t)$ (normalized with respect to the elastic strain in phase 2, equal to $\sigma_0/(2G_2)$) in the following three cases: no information about the composite is given; the value of $\bar{\epsilon}_{12}(t)$ at $t \to \infty$ is prescribed; and the value of $\bar{\epsilon}_{12}(t)$ at $t \to \infty$ and the volume fractions are known ($f_1 = 0.4$). The ordering of the bounds allows them to be identified when their color is missing.

By applying the Laplace transform to (6.3), we obtain

$$\overline{\boldsymbol{\sigma}}(\lambda) = \mathbf{C}_*(\lambda)\overline{\boldsymbol{\epsilon}}(\lambda), \qquad \overline{\boldsymbol{\epsilon}}(\lambda) = \mathbf{M}_*(\lambda)\overline{\boldsymbol{\sigma}}(\lambda), \tag{6.25}$$

where the matrices $\mathbf{C}_*(\lambda)$ and $\mathbf{M}_*(\lambda)$ are analytic functions of the eigenvalues $\mu_i(\lambda)$ and $\zeta_i(\lambda)$, $i = 1, 2$ (Bergman 1978, Milton 1981a, Golden and Papanicolaou 1983). Consequently, by exploiting such analytic properties, an integral representation formula for the operators $\mathbf{C}_*(\lambda)$ and $\mathbf{M}_*(\lambda)$ can be derived (for a rigorous mathematical proof, refer to the papers by Golden and Papanicolaou (1983, 1985)).

In particular, let us focus on the operator $\mathbf{C}_*(\lambda)$. By introducing the parameter $s(\lambda)$, defined by (6.7), and the function $\mathbf{F}(s)$, given by (6.6), Golden and Papanicolaou (1983) enunciated and proved the so-called *Representation theorem*, which asserts that there exists a finite Borel measure $\boldsymbol{\eta}(y)$, defined over the interval $[0, 1]$ such that the measure is positive semi-definite matrix-valued satisfying

$$\mathbf{F}(s) = \int_0^1 \frac{\mathrm{d}\boldsymbol{\eta}(y)}{s - y}, \tag{6.26}$$

for all $s \notin [0, 1]$.

In the case when $\mathbf{C}_*(\lambda)$, and hence $\mathbf{F}(s)$, are rational functions the measure is concentrated at the poles $s_0, s_1, ..., s_m$ of the rational function $\mathbf{F}(s)$ and equation (6.26) turns into

$$\mathbf{F}(s) = \sum_{i=0}^{m} \frac{\mathbf{B}_i}{s - s_i}, \tag{6.27}$$

in which the poles s_i lie on the semi-closed interval $[0,1)$ and the residues $\mathbf{B}_i$ are positive semi-definite matrices, that is

$$0 \leq s_0 \leq s_1 \leq ... \leq s_m < 1 \quad \text{and} \quad \mathbf{B}_i \geq 0 \ \text{ for all } i. \tag{6.28}$$

Notice that, since $\mathbf{C}_*(\lambda)$ is real and positive definite when the ratio $\mu_1(\lambda)/\mu_2(\lambda)$ is real and positive, and in particular as such a ratio tends to zero, from the definition (6.6) of $\mathbf{F}(s)$ it follows that, as $s \to 1$

$$\mathbf{F}(1) = \int_0^1 \frac{\mathrm{d}\boldsymbol{\eta}(y)}{1-y} \leq \mathbf{I}, \tag{6.29}$$

and, in the case of rational functions, the latter reduces to the following constraint on the poles and residues of $\mathbf{F}(s)$:

$$\mathbf{F}(1) = \sum_{i=0}^{m} \frac{\mathbf{B}_i}{1-s_i} \leq \mathbf{I}. \tag{6.30}$$

In order to further reduce the number of free parameters s_i and $\mathbf{B}_i$, all the available information about the composite microstructure has to be translated into constraints, the so-called *sum rules*, on such parameters. In particular, the sum rules are obtained by expanding the representation (6.26) of $\mathbf{F}(s)$ in powers of $1/s$ as $s \to \infty$, which corresponds to consider the case $\mu_1(\lambda) = \mu_2(\lambda) = 1$, that is, when the microscopic structure is nearly homogeneous. When $s \to \infty$, the denominator in (6.26) can be expanded as a series expansion in powers of $1/s$ to give

$$\mathbf{F}(s) = \sum_{j=0}^{\infty} \frac{\mathbf{A}_j}{s^{j+1}} \qquad \text{with } \mathbf{A}_j = \int_0^1 y^j \,\mathrm{d}\boldsymbol{\eta}(y). \tag{6.31}$$

It is clear that constraints on the moments of the measure are provided by the knowledge of the leading terms in the series, such as $\mathbf{A}_0$ and $\mathbf{A}_1$, which were derived through perturbation analysis by Brown, Jr. (1955) and Bergman (1978): see also equation (28) in Milton (1981a). In particular, if the volume fractions f_1 and $f_2 = 1 - f_1$ of the constituents are known, the first and second moments of the measure are given by

$$\mathbf{A}_0 = \int_0^1 \mathrm{d}\boldsymbol{\eta}(y) = f_1 \mathbf{I}, \tag{6.32}$$

$$\mathrm{Tr}\mathbf{A}_1 = \int_0^1 y \,\mathrm{d}\boldsymbol{\eta}(y) = f_1 \, f_2, \tag{6.33}$$

and the consequent constraints on the residues $\mathbf{B}_i$ and poles s_i read

$$\sum_{i=0}^{m} \mathbf{B}_i = f_1 \mathbf{I}, \tag{6.34}$$

$$\mathrm{Tr}\left(\sum_{i=0}^{m} \mathbf{B}_i s_i \right) = f_1 \, f_2. \tag{6.35}$$

Concerning the inverse constitutive law operator $\mathbf{M}_*(\lambda)$, an analogous procedure leads to the following spectral representation:

$$\mathbf{G}(u) = \sum_{i=0}^{m} \frac{\mathbf{P}_i}{u - u_i}, \tag{6.36}$$

where the parameter $u(\lambda)$ is defined by (6.7) and the function $\mathbf{G}(u)$ is given by (6.6).

The residues $\mathbf{P}_i$ and poles u_i satisfy the same constraints fulfilled by $\mathbf{B}_i$ and s_i. In particular, they satisfy inequalities (6.28) and (6.30), and equations (6.34) and (6.35), provided one replaces $\mathbf{B}_i$ and s_i with $\mathbf{P}_i$ and u_i.

6.4 Sum rules

The sum rules we develop here are implicit in the work of Bergman (1978), but we reproduce them here for completeness. Let us consider the $\overline{\sigma}_{12}(t)$ component of the averaged stress field $\overline{\boldsymbol{\sigma}}(t)$ that from (6.4) is given in the most general case by :

$$\overline{\sigma}_{12}(t) = \mu_2(t) * \overline{\epsilon}_{12}(t) - \sum_{i=0}^{m} B_{11}^{(i)} \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s - s_i}\right](t) * \overline{\epsilon}_{12}(t), \tag{6.37}$$

where for simplicity we set $\overline{\epsilon}_{13}(t) = 0$. In order to optimize the value of $\overline{\sigma}_{12}(t)$ for each $t \in [0, \infty)$ as a function of the 11-components $B_{11}^{(i)}$ of the residues $\mathbf{B}_i$, the constraints illustrated in Section 6.3 must be translated into constraints on $B_{11}^{(i)}$, nonnegative quantities by virtue of (6.28). In particular, inequality (6.30) rephrased as $\sum_{i=0}^{m} \mathbf{e}^{\mathrm{T}} \mathbf{B}_i \mathbf{e}/(1 - s_i) \leq 1$ delivers with $\mathbf{e} = [1 \;\; 0]^{\mathrm{T}}$,

$$1 - \sum_{i=0}^{m} \frac{B_{11}^{(i)}}{1 - s_i} \geq 0. \tag{6.38}$$

We remark that given any set of poles $0 \leq s_0 \leq s_1 \leq s_2 \leq \ldots \leq s_m < 1$ and any set of nonnegative residues $B_{11}^{(0)}, B_{11}^{(1)}, B_{11}^{(2)}, \ldots, B_{11}^{(m)}$ one can find a composite (which is a laminate of laminates) which realizes the response (6.37) for all times [see Appendix B of Milton (1981c) and Section 18.5 of Milton (2002)]. This implies that all our bounds based on the representation (6.37) will be optimal (and attained within this class of laminates of laminates), except those bounds that assume transverse isotropy. The bounds assuming transverse isotropy will likely not be optimal as they fail to take into account the phase interchange relation of Keller (1964), which places a nonlinear constraint on the residues.

By rephrasing the constraint (6.34) as $\sum_{i=0}^{m} \mathbf{e}^{\mathrm{T}} \mathbf{B}_i \mathbf{e} = f_1$, we have

$$\sum_{i=0}^{m} B_{11}^{(i)} = f_1. \tag{6.39}$$

Finally, by introducing the hypothesis of a transversely isotropic material (for which the residues $\mathbf{B}_i$ are diagonal matrices with $B_{11}^{(i)} = B_{22}^{(i)}$), the constraint (6.35) turns into

$$\sum_{i=0}^{m} B_{11}^{(i)} s_i = \frac{f_1 f_2}{2}. \tag{6.40}$$

Due to the linearity with respect to $B_{11}^{(i)}$ of $\overline{\sigma}_{12}(t)$ and of the above constraints, we can apply the theory of linear programming (Dantzig 1998) to optimize $\overline{\sigma}_{12}(t)$, as shown in Section 6.5.

In the case when the function to optimize is the scalar quantity $\mathcal{F}(t)$ defined by (6.17), the sum rules must be written in terms of the four components of the 2×2 matrices $\mathbf{B}_i$.

The constraint (6.28) on the positive semi-definiteness of the residues $\mathbf{B}_i$ yields a condition on the determinant of $\mathbf{B}_i$, which is quadratic with respect to the components of $\mathbf{B}_i$. In order to have only linear constraints, we express the residues in the following form:

$$\mathbf{B}_i = \mathbf{R}_i^{\mathrm{T}} \mathbf{b}_i \mathbf{R}_i, \qquad i = 0, 1, ..., m, \tag{6.41}$$

with

$$\mathbf{R}_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}, \qquad \mathbf{b}_i = \begin{bmatrix} b_{Ai} & 0 \\ 0 & b_{Bi} \end{bmatrix}. \tag{6.42}$$

Consequently, the condition on the positive semi-definiteness of the residues is translated into the following linear constraint on the elements b_{Ai} and b_{Bi}, for $i = 0, 1, ..., m$:

$$b_{Ai} \geq 0 \quad \text{and} \quad b_{Bi} \geq 0. \tag{6.43}$$

Regarding the constraint (6.30), in order to avoid the condition of nonnegativity of the determinant of the matrix $\mathbf{I} - \sum_{i=0}^{m} \mathbf{B}_i/(1-s_i)$, which is quadratic with respect to b_{Ai} and b_{Bi}, we initially restrict our attention to the case of composites endued with *reflective symmetry*. In such composites the angles of rotation θ_i (6.42) take the same value for each residue, that is, the residues are diagonal matrices with respect to the same basis, so that $\theta_i = \theta$ for every $i = 0, 1, ..., m$, and the constraint (6.30) turns into the following linear conditions on b_{Ai} and b_{B_i}:

$$1 - \sum_{i=0}^{m} \frac{b_{Ai}}{1-s_i} \geq 0, \qquad 1 - \sum_{i=0}^{m} \frac{b_{Bi}}{1-s_i} \geq 0. \tag{6.44}$$

Furthermore, under the reflective symmetry property, relations (6.34), (6.35) lead to

$$\sum_{i=0}^{m} b_{Ai} = f_1, \qquad \sum_{i=0}^{m} b_{Bi} = f_1 f_2, \tag{6.45}$$

$$\sum_{i=0}^{m} (b_{Ai} + b_{Bi}) s_i = f_1 f_2. \tag{6.46}$$

It is understood that in the case when one would like to optimize the strain response, such as the $\overline{\epsilon}_{12}(t)$ component of the average stress field (6.5):

$$\overline{\epsilon}_{12}(t) = \zeta_2(t) * \overline{\sigma}_{12}(t) - \sum_{i=0}^{m} P_{11}^{(i)} \, \mathcal{L}^{-1}\left[\frac{\zeta_2(\lambda)}{u - u_i}\right](t) * \overline{\sigma}_{12}(t), \tag{6.47}$$

where we set $\overline{\sigma}_{13}(t) = 0$, or the function $\mathcal{G}(t)$ (6.21), the constraints above still hold, provided we rephrase them in terms of the residues $\mathbf{P}_i$ and poles u_i of the function $\mathbf{G}(u)$ (6.6). Again, it is true that given any set of poles $0 \leq u_0 \leq u_1 \leq u_2 \leq \ldots \leq u_m < 1$ and any set of nonnegative residues $P_{11}^{(0)}, P_{11}^{(1)}, P_{11}^{(2)}, \ldots, P_{11}^{(m)}$ one can find a composite (which is a laminate of laminates) which realizes the response (6.47) for all times [see the last paragraph in Section 18.5 of Milton (2002)]. This implies that all our bounds based on the representation (6.47) will be optimal (and attained within this class of laminates of laminates), except those bounds that assume transverse isotropy.

6.5 Derivation of bounds in the time domain

The spectral representations (6.27) and (6.36) of the matrix-valued functions $\mathbf{F}(s)$ and $\mathbf{G}(u)$ respectively provide bounds on the response of the material expressed in terms of bounds on the stress component $\overline{\sigma}_{12}(t)$ (6.37) and on $\mathcal{F}(t)$ (6.17) or on the strain component $\overline{\epsilon}_{12}(t)$ (6.47) and on $\mathcal{G}(t)$ (6.21). These bounds are found by suitably varying the associated residues and poles in order to satisfy the sum rules shown in Section 6.4. Since the parameters $\mu_i(\lambda)$ and $\zeta_i(\lambda)$, $i = 1, 2$, are real it follows that $s(\lambda)$ and $u(\lambda)$ (6.7) are also real.

6.5.1 Bounds on the stress response

By virtue of equations (6.6) and (6.27) the direct complex effective constitutive law (6.25) can be rephrased as follows

$$\overline{\boldsymbol{\sigma}}(\lambda) = \mu_2(\lambda) \left[\overline{\boldsymbol{\epsilon}}(\lambda) - \sum_{i=0}^{m} \frac{\mathbf{B}_i}{s - s_i} \overline{\boldsymbol{\epsilon}}(\lambda) \right], \tag{6.48}$$

and by applying the inverse of the Laplace transform, the averaged stress field in the time domain is given by (6.4). Notice that in (6.4) the inverse of the Laplace transform of $\mu_2(\lambda)/(s(\lambda) - s_i)$ can be calculated explicitly, provided we know the functions $\mu_i(\lambda)$, $i = 1, 2$.

Now the problem is to bound $\overline{\sigma}_{12}(t)$ (6.37) for each fixed value of t. The idea is to take a fixed but large value of m and find the maximum (or minimum) value of $\overline{\sigma}_{12}(t)$ as the poles s_i and the nonnegative components $B_{11}^{(i)}$ of the residues $\mathbf{B}_i$ are varied subject to the constraints (6.38), (6.39) and (6.40). Since the resulting maximum (or minimum) could depend on m, we should ideally take the limit as m tends to infinity. However, it turns out that the extremum does not depend on m, provided m is large enough, and therefore there is no need to take limits.

It is worth noting that varying the poles s_i and the residues $\mathbf{B}_i$ corresponds, roughly speaking, to varying the microgeometry of the composite. Therefore, the procedure described above may be compared to finding the maximum (or minimum) value of $\mathcal{F}(t)$ as the geometry of the composite is varied over all configurations. Strictly speaking this is not quite correct as not all combinations of poles s_i and the residues $\mathbf{B}_i$ correspond to composites, as composites satisfy the phase interchange relation of Keller (1964), which we have ignored as it places a nonlinear constraint on the residues. This implies that the bounds we obtain assuming transverse isotropy, or the bounds we obtain by minimizing $\mathcal{F}(t)$ (6.17) or $\mathcal{G}(t)$ (6.21), are probably not optimal (though we emphasize that our bounds on $\overline{\sigma}_{12}(t)$ and $\overline{\epsilon}_{12}(t)$ which do not assume transverse isotropy are optimal).

No available information about the composite In this case the maximum (or minimum) value of $\overline{\sigma}_{12}(t)$ is achieved when either one residue is nonzero or all residues are zero. In particular, the extremum occurs either when the constraint (6.38) is satisfied as an equality by $B_{11}^{(0)}$, which takes the value $B_{11}^{(0)} = 1 - s_0$, while $B_{11}^{(i)} = 0$, for $i = 1, ..., m$, or when $B_{11}^{(i)} = 0$ for every $i = 0, 1, ..., m$. Consequently, either

$$\overline{\sigma}_{12}(t) = \mu_2(t) * \overline{\epsilon}_{12}(t) - (1 - s_0)\, \mathcal{L}^{-1} \left[\frac{\mu_2(\lambda)}{s(\lambda) - s_0} \right] (t) * \overline{\epsilon}_{12}(t), \tag{6.49}$$

with $s_0 \in [0, 1)$, or

$$\overline{\sigma}_{12}(t) = \mu_2(t) * \overline{\epsilon}_{12}(t). \tag{6.50}$$

It is clear that the latter case is a subcase of (6.49) when $s_0 \to 1$, and corresponds to an isotropic material purely composed of phase 2, whereas when $s_0 = 0$ in (6.49), by means of the definition (6.7) of $s(\lambda)$, (6.49)

provides the stress state in an isotropic material purely composed of phase 1, i.e., $\overline{\sigma}_{12}(t) = \mu_1(t) * \overline{\epsilon}_{12}(t)$. All that remains (and in general this is best done numerically) is to find, for each time t, the position of the pole s_0 which maximizes or minimizes (6.49).

The upper and lower limits of the function (6.49) are given by equations (6.11) and (6.12) and they are shown in **Figure 6.1** for the specific case when the response of one phase is given by the Maxwell model and the other having purely elastic behavior, with constant applied strain history.

The volume fraction of the constituents is known If f_1 is prescribed, then $\overline{\sigma}_{12}(t)$ is optimized by considering either only one nonzero residue satisfying constraint (6.39) or only two nonzero residues fulfilling the constraint (6.39) and relation (6.38) as an equality. In the first case $B_{11}^{(0)} = f_1$ and

$$\overline{\sigma}_{12}(t) = \mu_2(t) * \overline{\epsilon}_{12}(t) - f_1 \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s(\lambda) - s_0}\right](t) * \overline{\epsilon}_{12}(t), \tag{6.51}$$

with $s_0 \in [0, f_2]$, whereas in the second case $B_{11}^{(0)} = \frac{(1-s_0)(s_1-f_2)}{s_1-s_0}$, $B_{11}^{(1)} = \frac{(1-s_1)(f_2-s_0)}{s_1-s_0}$ and

$$\begin{aligned}\overline{\sigma}_{12}(t) = \mu_2(t) * \overline{\epsilon}_{12}(t) &- \frac{(1-s_0)(s_1-f_2)}{s_1-s_0} \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s(\lambda) - s_0}\right](t) * \overline{\epsilon}_{12}(t) \\ &- \frac{(1-s_1)(f_2-s_0)}{s_1-s_0} \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s(\lambda) - s_1}\right](t) * \overline{\epsilon}_{12}(t),\end{aligned} \tag{6.52}$$

with $s_0 \in [0, f_2]$ and $s_1 \in [f_2, 1)$.

We point out that equation (6.51) is a specific case of (6.52), when the pole s_1 approaches 1. The remaining optimization over the position of the poles in general needs to be done numerically.

Figure 6.2 shows the bounds obtained from equation (6.52), in the case when phase 1 is modeled by the Maxwell model and phase 2 has an elastic behavior, with the further assumption that the strain history is constant.

The composite is isotropic with known volume fractions Bounds on $\overline{\sigma}_{12}(t)$ can then be derived by either considering two nonzero residues satisfying equations (6.39) and (6.40), so that $B_{11}^{(0)} = f_1 \, \frac{s_1 - f_2/2}{s_1 - s_0}$, $B_{11}^{(1)} = f_1 \, \frac{f_2/2 - s_0}{s_1 - s_0}$ (subject to the constraint that the inequality (6.38) is satisfied) or by taking only three residues to be nonzero, with (6.38) holding as an equality, so that

$$\begin{aligned}B_{11}^{(0)} &= \frac{(1-s_0)(1-s_1)(1-s_2)}{(s_1-s_0)(s_2-s_0)}\left[1 - \frac{f_1}{1-s_2} - f_1 \, \frac{s_2 - f_2/2}{(1-s_1)(1-s_2)}\right], \\ B_{11}^{(1)} &= \frac{(1-s_0)(1-s_1)(1-s_2)}{(s_1-s_0)(s_2-s_1)}\left[\frac{f_1}{1-s_0} + f_1 \, \frac{f_2/2 - s_0}{(1-s_0)(1-s_2)} - 1\right], \\ B_{11}^{(2)} &= \frac{(1-s_0)(1-s_1)(1-s_2)}{(s_2-s_0)(s_2-s_1)}\left[1 - \frac{f_1}{1-s_0} - f_1 \, \frac{f_2/2 - s_0}{(1-s_0)(1-s_1)}\right].\end{aligned} \tag{6.53}$$

Again the remaining optimization over the position of the poles in general needs to be done numerically. This case is shown in **Figure 6.2** for the Maxwell model-Elastic model case with constant strain history.

Apart from the knowledge of the volume fractions and of the possible isotropy of the composite, other information may be given. For instance, the value of $\overline{\sigma}_{12}(t)$ at $t = 0$ or at $t \to \infty$ may be known. In such a case, we can derive bounds on $\overline{\sigma}_{12}(t)$ as follows:

Given the value of $\overline{\sigma}_{12}(t)$ at $t = 0$ or at $t \to \infty$ The maximum (or minimum) value of the 12-component of the averaged stress field can be obtained either by considering only one nonzero residue satisfying equation (6.37) evaluated at $t = 0$ or at $t \to \infty$, respectively, or only two nonzero residues fulfilling constraint (6.37) (evaluated at $t = 0$) and relation (6.38) as an equality.

It is worth noting that tighter bounds can be derived by considering combinations of information, such as the value of $\overline{\sigma}_{12}(t)$ at zero or infinity and the volume fraction of the material (see Figs.6.3 and 6.4). For the sake of brevity we do not report here the explicit results for that case but it is understood that they are derived following the same procedure applied above.

Now let us look at the problem of bounding the function $\mathcal{F}(t)$ (6.17) for a composite with reflective symmetry, with the angles α and θ fixed.

Bounds in the case when no information about the composite is available In the case when the only available information about the composite is the shear modulus $\mu_i(\lambda)$ of each constituent, then bounds on $\mathcal{F}(t)$ (6.17) have to be sought by considering the constraints (6.43) and (6.44). The optimum value of $\mathcal{F}(t)$ is attained when maximum two residues are nonzero. In particular, the representative case can be considered as the one for which both the constraints given by (6.44) are fulfilled as equalities. Then, only one of the b_{Ai} elements and only one of the b_{Bj} elements, with $i \neq j$, are nonzero, that is, either $b_{A0} = 1 - s_0$ and $b_{B1} = 1 - s_1$ or $b_{A1} = 1 - s_1$ and $b_{B0} = 1 - s_0$, where s_0 has to be varied over $[0, 1)$ and s_1 over $[s_0, 1)$ to give the optimum value of $\mathcal{F}(t)$. Note that the second case can be recovered from the first one, by switching the angle θ to $\theta + \pi/2$ (see equation (6.42)). Let us consider, then, the first option. The corresponding expression for the averaged stress field $\overline{\boldsymbol{\sigma}}(t)$ (6.4) reads:

$$\begin{aligned}\overline{\boldsymbol{\sigma}}(t) = \mu_2(t) * \overline{\boldsymbol{\epsilon}}(t) - (1 - s_0) \begin{bmatrix} \cos^2\theta & -\sin\theta\,\cos\theta \\ -\sin\theta\,\cos\theta & \sin^2\theta \end{bmatrix} \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s - s_0}\right](t) * \overline{\boldsymbol{\epsilon}}(t) \\ - (1 - s_1) \begin{bmatrix} \sin^2\theta & \sin\theta\,\cos\theta \\ \sin\theta\,\cos\theta & \cos^2\theta \end{bmatrix} \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s - s_1}\right](t) * \overline{\boldsymbol{\epsilon}}(t),\end{aligned} \tag{6.54}$$

and the maximum (or minimum) value of $\mathcal{F}(t)$ has to be determined by varying the poles s_0 and s_1 over the respective validity intervals. Finally the union of the resulting possible values of $\overline{\boldsymbol{\sigma}}(t)$ is taken as θ is varied (see Video 3). This case can be considered as the representative combination because, when either the poles approach 1 (with the associated residue tending to zero) or take the same value, all the other possible combinations can be derived consequently.

Bounds in the case when the volume fractions are known In the case when the volume fractions f_1 and f_2 of the constituents are known, bounds on $\mathcal{F}(t)$ (6.17) can be derived by considering also the constraints provided by equations (6.45) and (6.46). Specifically, the maximum (or minimum) value of the function $\mathcal{F}(t)$ is attained by one of the combinations which range from the two poles case to the five poles case. In the former situation, the bound is realized by considering either two nonzero b_{Ai} and one nonzero b_{Bj}, where j is equal to one of the two i, or vice versa. In the five poles case, instead, the bound on $\mathcal{F}(t)$ is attained by considering those b_{Ai} and b_{Bj} which satisfy (6.45)-(6.46) and constraints (6.44) as equalities, that is, by considering either three nonzero b_{Ai} and two nonzero b_{Bj}, with $i \neq j$, or vice versa. We stress the fact that the five poles case is the representative one (and the only one which needs to be considered) in the sense that

all the other combinations can be consequently recovered by letting some poles collapse to the same value or approach 1.

6.5.2 Bounds on the strain response

Let us consider the complex effective inverse constitutive law (6.25). Thanks to the relation between $\mathbf{M}_*(\lambda)$ and $\mathbf{G}(u)$, given by (6.6), and the spectral representation (6.36) of the function $\mathbf{G}(u)$, the averaged strain field in the complex domain is then described by the following equation:

$$\overline{\boldsymbol{\epsilon}}(\lambda) = \zeta_2(\lambda) \left[\overline{\boldsymbol{\sigma}}(\lambda) - \sum_{i=0}^{m} \frac{\mathbf{P}_i}{u - u_i} \overline{\boldsymbol{\sigma}}(\lambda) \right], \tag{6.55}$$

while in the time domain, by applying the inverse of the Laplace transform, $\overline{\boldsymbol{\epsilon}}(t)$ is given by (6.5).

In this case, the problem consists in bounding the $\overline{\epsilon}_{12}(t)$ component (6.47) of the averaged strain field. Alternatively, the aim could be the optimization of the function $\mathcal{G}(t)$ (6.21). In both cases, following the same arguments adopted in Subsection 6.5.1, bounds analogous to those obtained for $\mathcal{F}(t)$ and $\overline{\sigma}_{12}(t)$ can be deduced also for $\mathcal{G}(t)$ and $\overline{\epsilon}_{12}(t)$, respectively.

6.6 Composites without reflective symmetry

Bounds on the functions $\mathcal{F}(t)$ (6.17) and $\mathcal{G}(t)$ (6.21) have been derived under the hypothesis of reflective symmetry. In particular, such an assumption allows one to derive linear constraints on the diagonal elements b_{Ai} and b_{Bi} of the matrices $\mathbf{b}_i$ (6.42). Nevertheless, in the case when the composite is not symmetric with respect to a certain plane, that is, the reflective symmetry assumption does not hold, we can still derive linear constraints on the elements b_{Ai} and b_{Bi}.

To see this, let us introduce an additional pole $s_{m+1} = 1 - \delta$, where δ is a sufficiently small parameter, with residue

$$\mathbf{B}_{m+1} = \delta \mathbf{D}, \qquad \mathbf{D} = \mathbf{I} - \sum_{i=0}^{m} \frac{\mathbf{B}_i}{1 - s_i}.$$

Then, the introduction of a fictitious pole with very small residue does not affect the bounds on the analytic function, except in the near vicinity of $s = 1$. Consequently, inequality (6.30) can be replaced by the following equality:

$$\sum_{i=0}^{m+1} \frac{\mathbf{B}_i}{1 - s_i} = \mathbf{I}, \tag{6.56}$$

which provides three linear constraints with respect to the b_{Ai} and b_{Bi}:

$$\sum_{i=0}^{m+1} \frac{b_{Ai} \cos^2 \theta_i + b_{Bi} \sin^2 \theta_i}{1 - s_i} = 1, \qquad \sum_{i=0}^{m+1} \frac{b_{Ai} \sin^2 \theta_i + b_{Bi} \cos^2 \theta_i}{1 - s_i} = 1,$$

$$\sum_{i=0}^{m+1} \frac{(b_{Ai} - b_{Bi}) \cos \theta_i \sin \theta_i}{1 - s_i} = 0. \tag{6.57}$$

Finally, relations (6.34) and (6.35) written in terms of the b_{Ai} and b_{Bi} lead, respectively, to

$$\sum_{i=0}^{m} b_{Ai}\cos^2\theta_i + b_{Bi}\sin^2\theta_i = f_1, \qquad \sum_{i=0}^{m} b_{Ai}\sin^2\theta_i + b_{Bi}\cos^2\theta_i = f_1, \tag{6.58}$$

$$\sum_{i=0}^{m} (b_{Ai} - b_{Bi})\cos\theta_i \sin\theta_i = 0, \tag{6.59}$$

and

$$\sum_{i=0}^{m} (b_{Ai} + b_{Bi})\, s_i = f_1 f_2. \tag{6.60}$$

In contrast to the case with reflective symmetry, the bounds on $\mathcal{F}(t)$ (as α is varied), for fixed t, necessarily restrict $\overline{\boldsymbol{\sigma}}(t)$ to a convex region in the $(\overline{\sigma}_{12}(t), \overline{\sigma}_{13}(t))$ plane. However the range of values of $\overline{\boldsymbol{\sigma}}(t)$, as the poles and residue matrices are varied (subject to the constraints (6.6), and, if the volume fractions are known, (6.58) and (6.60)) is in fact a convex set in the $(\overline{\sigma}_{12}(t), \overline{\sigma}_{13}(t))$ plane. To see this, suppose m is enormously large. Then there is no loss of generality if we take the poles to be evenly spaced: $s_i = i/(m+2)$, and take the angles θ_i to increase by small amounts going in total many times "around the clock": $\theta_i = 2\pi(m \bmod k)/k$, where k is chosen with $m \gg k \gg 1$, and only vary the b_{Ai} and b_{Bi}. Then, if a set of parameters b_{Ai} and b_{Bi}, $i = 0, 1, \ldots, m$ satisfy the constraints, and another set b'_{Ai} and b'_{Bi} also satisfy it, so will the linear combination $wb_{Ai} + (1-w)b'_{Ai}$ and $wb_{Ai} + (1-w)b'_{Ai}$, for any weight $w \in (0,1)$ and the resulting response vector $\overline{\boldsymbol{\sigma}}_w(t)$ will be a linear combination of the two response vectors, $\overline{\boldsymbol{\sigma}}(t)$ and $\overline{\boldsymbol{\sigma}}'(t)$ associated with the original two sets of parameters.

In the following, we show the procedure to be adopted in order to derive bounds on the function $\mathcal{F}(t)$ (6.17). In contrast to the case with reflective symmetry, the bounds on $\mathcal{F}(t)$ (as α is varied) for fixed t necessarily restrict $\overline{\boldsymbol{\sigma}}(t)$ to a convex region in the $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$ plane. Another method needs to be devised to obtain bounds that confine $\overline{\boldsymbol{\sigma}}(t)$ to regions that are not-necessarily convex in the $(\overline{\sigma}_{13}(t), \overline{\sigma}_{12}(t))$ plane.

Bounds in the case when no information about the composite is available For the sake of brevity, we do not report the explicit expression taken by the stress field (6.4) for each combination of poles related to this case but we consider only the representative case. In particular, the optimal value of $\mathcal{F}(t)$ is attained when either only one or only three residues are nonzero. In particular, the representative combination of residues corresponds to the case for which the three constraints given by (6.56) are fulfilled. Such a condition holds when either only three elements among the b_{Ai} are nonzero, while $b_{Bi} = 0$ for every $i = 0, 1, ..., m$, and vice versa, or when only two elements among the b_{Ai} and one element among the b_{Bj}, with $i \neq j$, are nonzero, and vice versa. It is worth noting that, by suitably choosing the angles θ_i (6.42), the latter case is equivalent to the former one.

Bounds in the case when the volume fractions are known The combinations of residues which provide the maximum (or minimum) value of $\mathcal{F}(t)$ are those which satisfy the seven equations given by the constraints (6.56), (6.58) and (6.60). In particular, the combination with the minimum number of poles is given by three nonzero b_{Ai} and the corresponding three nonzero b_{Bi} (three poles in total), while the combination with the maximum number of poles consists of seven poles and can be achieved either considering six nonzero b_{Ai} and one nonzero b_{Bj}, $i \neq j$, and vice versa, or five nonzero b_{Ai} and two nonzero b_{Bj}, $i \neq j$, and vice versa, or four nonzero b_{Ai} and three nonzero b_{Bj}, $i \neq j$, and vice versa. We remark that all combinations

corresponding to the same number of poles are equivalent, since we are free to replace each rotation angle θ_i (6.42) by $\theta_i + \pi/2$. We emphasize that the seven pole case is the representative one (and the only one which needs to be considered) in the sense that all the other combinations can be consequently recovered by letting some poles collapse to the same value or approach 1 (implying that the associated residue tends to zero).

6.7 Bounding the homogenized relaxation and creep kernels

Note that the relation (6.4) when $\overline{\boldsymbol{\epsilon}}(t)$ is chosen to be a constant $\boldsymbol{\epsilon}_0$ for all $t > 0$ can be written in the form

$$\overline{\boldsymbol{\sigma}}(t) = \boldsymbol{C}^h(t)\boldsymbol{\epsilon}_0, \tag{6.61}$$

where $\boldsymbol{C}^h(t)$, the homogenized relaxation kernel, is given by

$$\boldsymbol{C}^h(t) = \mu_2(t) - \sum_{i=0}^{m} \mathbf{B}_i \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s(\lambda) - s_i}\right](t). \tag{6.62}$$

The same arguments that were used in the previous section to show that the range of values of $\overline{\boldsymbol{\sigma}}(t)$, as the poles and residue matrices are varied is in fact a convex set, can also be applied here: the range of values of the matrix-valued relaxation kernel $\boldsymbol{C}^h(t)$ as the poles and residue matrices are varied (subject to any linear sum rules on the residues, implied by the known information about the composite) is also a convex set.

To find this convex set we consider for each fixed time t the objective function

$$\mathcal{F}(\mathbf{V}) = \mathrm{Tr}(\mathbf{V}\boldsymbol{C}^h(t)), \tag{6.63}$$

where $\mathbf{V}$ is any 2×2 real-valued symmetric matrix. By substituting (6.62) in this expression we see that the objective function depends linearly on the residue matrices $\mathbf{B}_i$, and thus we can use the same techniques as before to find the minimum values of $\mathcal{F}$ for a given matrix $\mathbf{V}$ (incorporating, if desired, known information about the composite which impose sum rules on the residues): let us call this minimum $\mathcal{F}^{\min}(\mathbf{V})$. The constraint that

$$\mathrm{Tr}(\mathbf{V}\boldsymbol{C}^h(t)) \geq \mathcal{F}^{\min}(\mathbf{V}) \tag{6.64}$$

confines $\boldsymbol{C}^h(t)$ to lie on one side of a "hyperplane" in a 3-dimensional space with the elements of $\boldsymbol{C}^h(t)$ as coordinates (as it is a symmetric 2×2 matrix there are only 3 independent elements). Finally, by varying $\mathbf{V}$ we constrain $\boldsymbol{C}^h(t)$ to the desired convex set in this 3-dimensional space.

In a similar way the relation (6.5) when $\overline{\boldsymbol{\sigma}}(t)$ is chosen to be a constant, $\boldsymbol{\sigma}_0$, for all $t > 0$ can be written in the form

$$\overline{\boldsymbol{\epsilon}}(t) = \boldsymbol{M}^h(t)\boldsymbol{\sigma}_0, \tag{6.65}$$

where $\boldsymbol{M}^h(t)$, the homogenized creep kernel, is given by

$$\boldsymbol{M}^h(t) = \zeta_2(t) - \sum_{i=0}^{m} \mathbf{P}_i \, \mathcal{L}^{-1}\left[\frac{\zeta_2(\lambda)}{u - u_i}\right](t). \tag{6.66}$$

As $\boldsymbol{M}^h(t)$ depends linearly on the residues $\mathbf{P}_i$ we can also use the same approach to bound it (subject to any linear sum rules on the residues, implied by the known information about the composite).

6.8 Correlating the transient response to different applied fields at different times

We have been focusing on deriving bounds on the transient response of the composite at a single time t, and for a single applied field. However, if desired, the method allows one to obtain coupled bounds which correlate the responses at a set of different times $t = t_1, t_2, \ldots, t_n$, and for different applied fields (which may or may not be all the same). To see this, suppose for example that we are interested in coupling the stresses $\overline{\boldsymbol{\sigma}}^{(j)}(t^{(j)})$, for $j = 1, 2, \ldots, n$ that arise respectively in response to the applied strains $\overline{\boldsymbol{\epsilon}}^{(j)}(t)$, for $j = 1, 2, \ldots, n$. From (6.4) it directly follows that

$$\overline{\boldsymbol{\sigma}^{(j)}}(t^{(j)}) = \mu_2(t^{(j)}) * \overline{\boldsymbol{\epsilon}}(t^{(j)}) - \sum_{i=0}^{m} \mathbf{B}_i \, \mathcal{L}^{-1}\left[\frac{\mu_2(\lambda)}{s - s_i}\right](t^{(j)}) * \overline{\boldsymbol{\epsilon}}^{(j)}(t^{(j)}). \tag{6.67}$$

The same arguments that were used in Section 6.6 to show that the range of values of $\overline{\boldsymbol{\sigma}}(t)$, as the poles and residue matrices are varied is in fact a convex set, can also be applied here: the range of values of the n-tuple $(\overline{\boldsymbol{\sigma}}^{(1)}(t^{(1)}), \overline{\boldsymbol{\sigma}}^{(2)}(t^{(2)}), \ldots, \overline{\boldsymbol{\sigma}}^{(n)}(t^{(n)}))$ as the poles and residue matrices are varied (subject to any linear sum rules on the residues, implied by the known information about the composite) is also a convex set.

To find this convex set, consider the objective function

$$\mathcal{F}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}) = \sum_{j=1}^{n} \mathbf{v}^{(j)} \cdot \boldsymbol{\sigma}^{(j)}(t^{(j)}). \tag{6.68}$$

By substituting (6.67) in this expression we see that the objective function depends linearly on the residue matrices $\mathbf{B}_i$, and thus we can use the same techniques as before to find the minimum values of $\mathcal{F}$ for a given set of vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}$ (incorporating, if desired, known information about the composite which impose sum rules on the residues): let us call this minimum $\mathcal{F}^{\min}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)})$. The constraint that

$$\sum_{j=1}^{n} \mathbf{v}^{(j)} \cdot \boldsymbol{\sigma}^{(j)}(t^{(j)}) \geq \mathcal{F}^{\min}(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}) \tag{6.69}$$

confines the n-tuple $(\overline{\boldsymbol{\sigma}}^{(1)}(t^{(1)}), \overline{\boldsymbol{\sigma}}^{(2)}(t^{(2)}), \ldots, \overline{\boldsymbol{\sigma}}^{(n)}(t^{(n)}))$ to lie on one side of a "hyperplane" in a $2n$-dimensional space with the elements of the $\overline{\boldsymbol{\sigma}}^{(j)}(t^{(j)})$ as coordinates. Finally by varying the vectors $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$, ..., $\mathbf{v}^{(n)}$ we constrain the n-tuple to the desired convex set in this multidimensional space.

Note that the applied strains $\overline{\boldsymbol{\epsilon}}^{(j)}(t)$ could all be identical, and in this case the bounds will correlate the values of the resulting stress field $\overline{\boldsymbol{\sigma}}(t)$ at times $t = t_1, t_2, \ldots, t_n$. These bounds, correlating the transient response to different applied fields at a set of different times, might be very useful for predicting the response to a new applied field, given measurements (at specific times) for the response to a set of test applied fields. Or they could be very useful if used in an inverse fashion to determine information about the composite, such as the volume fractions of the phases.

It is clear that the method can easily be extended in the obvious way to obtain bounds which correlate the matrix values of the relaxation kernel $\boldsymbol{C}^h(t)$ (6.62) at different times or the creep kernel $\boldsymbol{M}^h(t)$ (6.66) at different times.

6.9 Concluding remarks

In this chapter we proposed a new approach to derive bounds on the response of a two-component viscoelastic composite under antiplane loadings, in the time domain. The starting point is represented by the so-called analytic method, first proposed by Bergman (1978) to bound effective conductivities when the component conductivities are real, and later extended to bound the complex effective tensor of a two-component dielectric composite in the frequency domain (see, for instance, Milton (1980, 1981a, 1981c), and Bergman (1980)) but, to the best of our knowledge, the method until now has been applied only in the frequency domain, for cyclic external actions at a certain frequency. This work may be the first to extend the field of applicability of the analytic method to problems defined in the time domain with noncyclic external actions.

The core of the analytic method is based on the fact that, by virtue of the analyticity property of the complex effective tensor of the viscoelastic composite with respect to the complex moduli of the components, one can write the complex effective tensor as the sum of poles weighted by positive semi-definite matrix-valued residues. Consequently, the response of the material, in terms of stresses or strains, turns out to depend only on the position of the poles and on the value of the associated residues, which are the variational parameters of the problem. The aim is to find the combinations of such parameters which provide the maximum (or minimum) response of the composite for each moment of time.

The optimization of the response of the material is performed in two steps. First, all the available information about the composite, such as the knowledge of the volume fraction of the constituents or of the value of the response of the material at a certain moment of time, is translated into (linear) constraints on the poles and residues. Then, the response of the material being linear in the residues, allows one to apply the theory of linear programming to limit the number of nonzero residues, so that the problem is reduced to a new one with a relatively small number of nonzero residues. Finally, the optimization over the positions of the poles is performed numerically for two specific cases: when the stress response has to be bounded, we consider a composite made of an elastic phase and a phase with a behavior describable by the Maxwell model, whereas when we bound the strain response, we consider a composite made of an elastic phase and a phase modeled by the Kelvin–Voigt model.

The estimates given by the numerical results prove to be increasingly accurate the more information about the composite is incorporated. In particular, when information such as the volume fraction of the components or the value of the response at a specific time is considered, the bounds are quite tight over the entire range of time, thus allowing one to predict the transient behavior of the composite. Most noticeably, when combinations of information are considered, such as the knowledge of the volume fractions and the eventual transverse isotropy of the composite, the bounds are extremely tight at certain specific times, suggesting the possibility of measuring the response of such times and, by using the bounds in an inverse fashion, almost exactly determining the volume fraction of the components of the composite.

Acknowledgments

Ornella Mattei is grateful for support from the Italian Ministry of Education, University, and Research (MIUR), from the University of Brescia, and from the University of Utah. Graeme Milton is grateful to the American National Science Foundation (Research Grant DMS-1211359) and the University of Utah for support.

7 Superfunctions and the algebra of subspace collections and their association with rational functions of several complex variables

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

A natural connection between rational functions of several real or complex variables, and subspace collections is explored. A new class of function, superfunctions, are introduced which are the counterpart to functions at the level of subspace collections. Operations on subspace collections are found to correspond to various operations on rational functions, such as addition, multiplication and substitution. It is established that every function that is matrix valued, with matrix elements that are rational and homogeneous of degree 1, can be generated from an appropriate, but not necessarily unique, subspace collection: the mapping from subspace collections to rational functions is onto, but not one to one. For some applications superfunctions may be more important than functions, as they incorporate more information about the physical problem, yet can be manipulated in much the same way as functions. Previously subspace collections had been introduced when there was an inner product on the vector (or Hilbert) space, and appropriate subspaces were mutually orthogonal. In that setting certain normalization and reduction operations on subspace collections led to a continued fraction expansion of the associated function, which allowed one to bound the function in terms of a set of weight matrices and normalization matrices that are derived from series expansions. Here we also initiate the theory of normalization and reduction operations, appropriate when there is no inner product on the space.

7.1 Introduction

Subspace collections have a rich algebraic structure, and a close connection with rational functions of several real or complex variables. Here we are interested in three types of subspace collections. The first type of

subspace collection is finite-dimensional vector spaces $\mathcal{H}$ (over the real or complex numbers) that have the decomposition

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.1}$$

which we call a $Z(n)$ subspace collection. Here $\mathcal{U}$, $\mathcal{E}$, and $\mathcal{J}$ could be any vector subspaces, the only condition being that there is no nonzero vector in common to any pair of them and that their span is $\mathcal{H}$. Similarly $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_n$ could be any n subspaces, the only condition being that there is no nonzero vector in common to any pair of them and that their span is $\mathcal{H}$.

The second type of subspace collection is finite-dimensional vector spaces $\mathcal{K}$ (over the real or complex numbers) that have the decomposition

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.2}$$

which we call a $Y(n)$ subspace collection, where the $\mathcal{E}$ and $\mathcal{J}$ entering (7.2) are not to be confused with the subspaces $\mathcal{E}$ and $\mathcal{J}$ entering (7.1). Here $\mathcal{E}$, and $\mathcal{J}$ could be any pair of vector subspaces, the only condition being that there is no nonzero vector in their intersection and that their span is $\mathcal{K}$. For technical reasons it is usually convenient to assume $\mathcal{V}$ has no nonzero vector in common with either $\mathcal{E}$ or $\mathcal{J}$. Apart from that restriction, $\mathcal{V}$, $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_n$ could be any $n+1$ subspaces, the only condition being that there is no nonzero vector in common to any pair of them and that their span is $\mathcal{K}$.

The third type of subspace collection is finite-dimensional vector spaces $\mathcal{K}$ (over the real or complex numbers) that have the decomposition

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V}^I \oplus \mathcal{V}^O \oplus \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.3}$$

which we call a superfunction $F^s(n)$. In a superfunction the space $\mathcal{V}^I$ and the space $\mathcal{V}^O$ are called the input and output subspaces respectively, and they have the same dimension. Again, $\mathcal{E}$, and $\mathcal{J}$ could be any pair of vector subspaces, the only condition being that there is no nonzero vector in their intersection and that their span is $\mathcal{K}$. For superfunctions we require the technical condition that for any choice of vectors $\mathbf{E}^I, \mathbf{J}^I \in \mathcal{V}^I$ and $\mathbf{E}^O, \mathbf{J}^O \in \mathcal{V}^O$ there exist vectors $\mathbf{E}^P$ and $\mathbf{J}^P$ in $\mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n$ such that

$$\mathbf{E} = \mathbf{E}^I + \mathbf{E}^O + \mathbf{E}^P \in \mathcal{E}, \quad \mathbf{J} = \mathbf{J}^I + \mathbf{J}^O + \mathbf{J}^P \in \mathcal{J}. \tag{7.4}$$

Apart from that restriction, $\mathcal{V}^I$, $\mathcal{V}^O$, $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_n$ could be any $n+2$ subspaces, the only condition being that there is no nonzero vector in common to any pair of them and that their span is $\mathcal{K}$.

As we will see there is a close connection between a superfunction $F^s(n)$ and a $Y(n)$ subspace collection, and also many connections between them and $Z(n)$ subspace collections. All are intertwined and that is the beauty of the theory. $Z(3)$ and $Y(2)$ subspace collections and superfunctions $F^s(1)$ can be visualized in 3-dimensional space: examples of these are given **Figure 7.1**.

One reason $Y(n)$ subspace collections, $Z(n)$ subspaces collections, and superfunctions $F^s(n)$ are important is because they arise in many physical problems. For examples in network theory and in the theory of the effective moduli of composite materials, see the review in **Chapter 2** and Milton (2002). There are also many other physical problems where subspace collections arise as is apparent in **Chapters 1**, **3**, **8**, **9**, **12**, **13**, and **14** of this book. In physics applications the subspaces are usually orthogonal with respect to some inner product on the space $\mathcal{H}$ or $\mathcal{K}$ but as this chapter shows the theory of them can be developed without reference to an inner product. This generalization is important to make contact between general rational functions of complex variables, thus extending the notion of a function: hence the name superfunction. The generalization is also important for applications, such as speeding up numerical methods for calculating the fields that solve the problem: we will see an example of this in the next chapter.

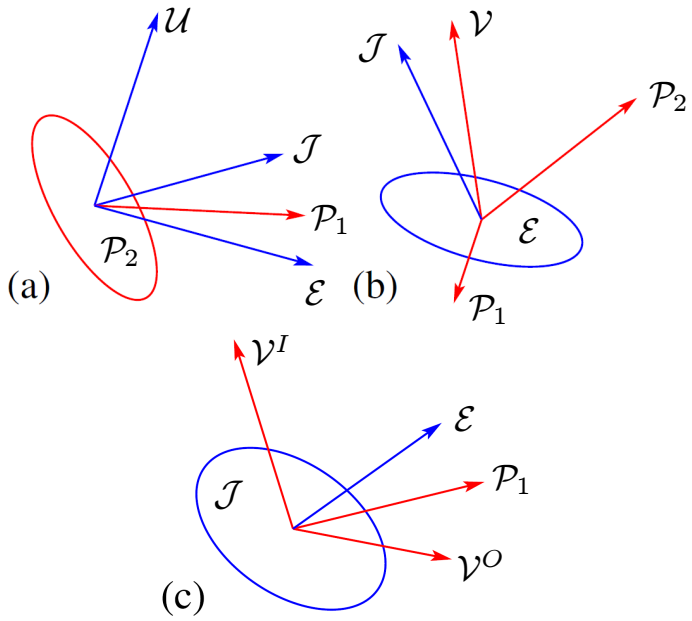


Figure 7.1: Shown in (a) is an example of a $Z(3)$ subspace collection, in (b) a $Y(2)$ subspace collection, and in (c) a superfunction $F^s(1)$. The rays denote one-dimensional subspaces: they should really be drawn as lines, but for clarity they are drawn as rays and should be extended in the opposite direction as the ray. The circles, which look like ellipses as they are tilted, represent two-dimensional subspaces.

It may very well be the case that superfunctions become more important than functions in some applications, as suggested by the flow chart of **Figure 7.2**. The reason is that when one extracts the function from a superfunction, which we will see how to do shortly, one generally loses information that is contained in a superfunction. For example, in the context of physical problems where there is an inner product on the space this information may came in the form of a series expansion for the fields up to a given order, and from this series expansion one can extract the "weight matrices" and "normalization matrices", introduced by Milton and Golden (1985) and Milton (1987a, 1987b). These matrices basically encode the information about the "angles" between the various subspaces (when there is an inner product). One can then develop a continued fraction expansion for the function associated with the superfunction, with the normalization factors and weight matrices that enter it at each level having the property that they are positive semidefinite, with the weight matrices summing to one. Truncating the continued fraction gives approximations to the function, that are similar in some respects the diagonal Padé approximants, and in fact give bounds on the function if the truncation is done appropriately. The information contained in the weight matrices and normalization matrices, cannot in general be recovered (at least when $n \geq 4$) from the series expansion of the associated function. (Although one can potentially determine these matrices from the series expansion of the functions associated with coupled field problems, as shown in **Chapter 9**). This theory was established by Milton and Golden (1985) and Milton (1987a, 1987b, 1991). (see also **Chapters 19**, **20**, and **29** in Milton (2002)) for the case of $Z(n)$ subspace collections, for any integer $n \geq 1$. In this paper we develop the basic theory of

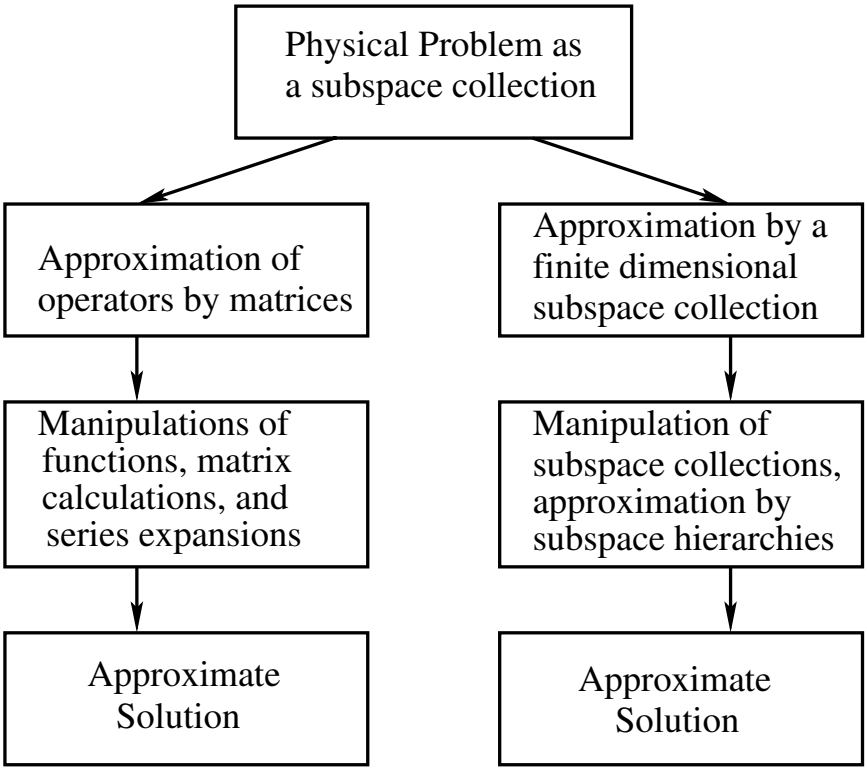


Figure 7.2: Two routes to solving a physical problem formulated in terms of subspace collections. It is suggested that the route on the right may result in a better approximation as more information is kept.

subspace collections in the case where there is no inner product on the vector space $\mathcal{H}$ or $\mathcal{K}$. We also make the first steps towards generating continued fraction expansions in the case where there is no inner product on the vector space $\mathcal{H}$ or $\mathcal{K}$.

Let us first suppose $\mathcal{V}$ and $\mathcal{U}$ are one-dimensional. We will see that there are generally homogeneous (of degree 1) rational functions $Y(z_1, z_2, \ldots, z_n)$ and $Z(z_1, z_2, \ldots, z_n)$ (over the real or complex numbers) of degree 1 that are associated respectively with these $Y(n)$ and $Z(n)$ subspace collections, where $Z(z_1, z_2, \ldots, z_n)$ satisfies the additional constraint that $Z(1, 1, \ldots, 1) = 1$. Conversely, we will see that given any rational functions $Y(z_1, z_2, \ldots, z_n)$ and $Z(z_1, z_2, \ldots, z_n)$ with these properties, then there exists at least one subspace collection realizing these functions as its associated function. There are also operations on these subspace collections that correspond to operations on the associated function, such as substitution.

For superfunctions the simplest case is when the input and output spaces $\mathcal{V}^I$ and $\mathcal{V}^O$ are one-dimensional. Then with a specific basis for $\mathcal{V}^I$ and $\mathcal{V}^O$ the corresponding function $\mathbf{F}(z_1, z_2, \ldots, z_n)$ is 2 by 2 matrix-valued with the elements $F_{11}(z_1, z_2, \ldots, z_n)$ and $F_{22}(z_1, z_2, \ldots, z_n)$ being homogeneous of degree zero, the element $F_{12}(z_1, z_2, \ldots, z_n)$ being homogeneous of degree minus 1, and $F_{21}(z_1, z_2, \ldots, z_n)$ being homogeneous of degree 1. There are operations on superfunctions that correspond to addition, multiplication and forming an inverse (and hence division) of the associated functions. So superfunctions form an algebra. Also one can

do substitutions at the level of subspace collections. Actually the operation of addition of superfunctions is naturally done with the associated Y-problem, although one could equally do addition of superfunctions with the associated inverse Y-problem (where the spaces $\mathcal{E}$ and $\mathcal{J}$ are interchanged). Thus there is an inherent ambiguity of how one wants to define addition of superfunctions. The definitions of addition, multiplication. and substitution of subspace collections may seem a little complicated and abstract, yet they are the exact counterpart of similar operations one may do on multiterminal electrical networks, and they do produce the corresponding action on the associated functions. (In fact it was thinking about electrical circuits which guided the construction of these operations in a more general setting).

When $\mathcal{V}$ and $\mathcal{U}$ have dimension greater than 1, then $Y(z_1, z_2, \ldots, z_n)$ and $Z(z_1, z_2, \ldots, z_n)$ get replaced by linear operator-valued functions $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ and $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ which map $\mathcal{V}$ to $\mathcal{V}$ and $\mathcal{U}$ to $\mathcal{U}$ respectively. Similarly, the function $\mathbf{F}(z_1, z_2, \ldots, z_n)$ should really be thought of as a linear operator mapping $\mathcal{V}^I$ to $\mathcal{V}^O$

The original motivation for studying subspace collections, and their associated functions, arose from the study of the effective conductivity tensor $\mathbf{Z}$ of periodic composite materials. For a composite with n isotropic phases, with scalar conductivities $z_1, z_2, \ldots, z_n$, the effective conductivity tensor was found to be a homogeneous (of degree 1) analytic function $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ of the component conductivities with positive definite imaginary part when the component conductivities have positive imaginary part [Bergman 1978; Milton 1979, 1981a, Golden and Papanicolaou 1983] (see also **Chapter 18** of Milton (2002)). It was also recognized (Milton 1987a, 1990) that the problem of determining the effective conductivity function could be formulated in terms of three mutually orthogonal spaces in the Hilbert space $\mathcal{H}$ of square integrable functions: namely the space $\mathcal{U}$ of constant fields, the space $\mathcal{E}$ of periodic square integrable electric fields (having zero curl), and the space $\mathcal{J}$ of square integrable current fields (having zero divergence), and if the composite had n isotropic phases, with conductivities $z_1, z_2, \ldots, z_n$, then it was also natural to decompose $\mathcal{H}$ into the direct sum of n mutually orthogonal subspaces $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$ where $\mathcal{P}_i$ consists of those square integrable fields which are nonzero only within component i. This formulation, in terms of a $Z(n)$ subspace collection, evolved out of earlier Hilbert space formulations of the problem (Fokin 1982; Kohler and Papanicolaou 1982; Papanicolaou and Varadhan 1982; Golden and Papanicolaou 1983; Kantor and Bergman 1984; Dell'Antonio, Figari, and Orlandi 1986) and can easily be extended to the elastic, thermoelastic, piezoelectric, and poroelastic equations of multiphase and polycrystalline materials (see, for example, **Chapter 12** in Milton (2002)). The formulation has proved to be particularly important in the theory of exact relations of composite materials (Grabovsky 1998; Grabovsky and Sage 1998: Grabovsky and Milton 1998; Grabovsky, Milton, and Sage 2000; Grabovsky 2004) (see also **Chapter 17** in Milton (2002)) where one seeks microstructure independent relations satisfied by effective tensors. For two-dimensional polycrystals a complete correspondence was established between subspace collections and a representative class of multiple rank laminate polycrystal geometries (Clark and Milton 1994), thus showing that the subspace collection of any two-dimensional polycrystal, with any configuration of crystal grains, could be approximated arbitrarily closely by the subspace collection of one of these multiple rank laminate polycrystal geometries.

Curiously the connection between $Z(n)$ subspace collections and the effective conductivity allowed the effective conductivity function $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ to be expanded as a new type of continued fraction, involving matrices of increasing dimension as one proceeds down the continued fraction when $n > 2$ (Milton 1987a, 1987b, 1991; see also **Chapters 19**, **20**, and **29** in Milton 2002). The coefficients in the weight and normalization matrices entering the continued fraction can be expressed in terms of inner products between fields that enter the series expansion of the solution field in a nearly homogeneous medium (with all the conductivities $z_1, z_2, \ldots, z_n$ being close to one another). One application of the continued fraction expansion has been to obtain bounds on the diagonal elements of the complex effective conductivity tensor of a three phase con-

ducting composite, with complex conductivities z_1, z_2 and z_3, that were tighter than bounds obtained by any other method (see **Figure 4** in Milton 1987b). This procedure essentially extended to multivariate functions the procedure (using successive fractional linear transformations) that was used to obtain bounds (Baker, Jr. 1969) on the values in the complex plane that Stieltjes functions can take when a finite number of Taylor series coefficients are known (see also Golden and Papanicolaou 1983; Bergman 1986) where essentially the same transformation is used to derive bounds on the complex dielectric constant of two component media using series expansion coefficients, as noted in the appendix in Milton 1986a, and see Milton 1981c, where the same set of bounds is derived using a different procedure, namely the method of variation of poles and zeros.)

In the case $n = 2$ the continued fraction reduces to a usual continued fraction expansion, like those continued fractions associated with Padé approximants (see **Chapter 4** of **Part I** of Baker, Jr. and Graves-Morris 1981). $Y(n)$ subspace collections enter, for example, if one eliminates from the Hilbert space the constant fields and then reformulates the conductivity equations in terms of the remaining fields: the driving fields are then fields which are constant in each phase but have zero average value (see **Chapter 19** in Milton 2002 and references therein). The interrelationship between $Z(n)$ subspace collections and $Y(n)$ subspace collections is what gives rise to these novel continued fractions.

Finite-dimensional $Z(n)$ and $Y(n)$ subspace collections also arise naturally in the study of the effective resistance of electrical circuits constructed from n types of resistors having conductances z_1, z_2, $\ldots z_n$ (see **Chapter 20** in Milton 2002). This is not surprising as periodic resistor networks can be seen as discrete approximations to conducting composite materials (see, for example, Milton 1981a and **Figure 8.5**(a) in this book). **Figure 7.3** illustrates a discrete network of impedances, and gives an indication of the physical meaning of the $Z(n)$ and $Y(n)$ subspace collections in this context.

In this figure, the vector space $\mathcal{H}$ is 6-dimensional, and is the direct sum of the two-dimensional space $\mathcal{P}_1$ consisting of fields that are nonzero only along the resistors $c_1 z_1$ and $c_3 z_1$; the two-dimensional space $\mathcal{P}_2$ consisting of fields that are nonzero only along the resistors $c_2 z_2$ and $c_5 z_2$; and the one-dimensional space $\mathcal{P}_3$ consisting of fields that are nonzero only along the resistor $c_4 z_3$. The response of the network, when one terminal is grounded (with zero voltage) is a 3×3 matrix. When it acts on the vector, having as elements the voltages at the three remaining terminals, it gives the three currents flowing through these terminals. The 3×3 matrix-valued function $\mathbf{Z}(z_1, z_2, z_3)$ gives the matrix-valued response relative to the response when $z_1 = z_2 = z_3 = 1$. Now, let us imagine all the resistors, or impedances, in (a) are on one side of the circuit board, with the terminals being conducting posts that penetrate the board. On the other side of the board these posts are connected to a tree-like graph of batteries (or alternating current sources if the fields vary sinusoidally in time) shown in (b). The three fields in these batteries constitute the space $\mathcal{V}$. The $Y(3)$ subspace collection contains fields on both sides of the board in $\mathcal{K} = \mathcal{H} \oplus \mathcal{V}$. The associated 3×3 matrix-valued Y-function $\mathbf{Y}(z_1, z_2, z_3)$ gives the current going through the three batteries, in response to the voltages across them. Note that $\mathbf{Y}(z_1, z_2, z_3)$ is not diagonal: a voltage across one battery, sends current through the other two batteries, even when they have zero voltage across them.

Superfunctions are a natural generalization of multiport electrical circuits with input ports and output ports, as illustrated in **Figure 7.4**. The function $\mathbf{F}$ gives the currents and potential drops across the output batteries/resistors that are generated in response to currents and potential drops across the input batteries. Note that the networks associated with superfunctions automatically satisfy the "port condition" that the net flow of current from the input terminals to the output terminals is zero.

In this chapter we show that the connection between finite-dimensional $Z(n)$ and $Y(n)$ subspace collections and homogeneous (degree 1) operator-valued rational functions $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ and $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ persists even when the subspaces in each decomposition are not necessarily mutually orthogonal, and indeed even in the absence of an inner product (on the space $\mathcal{H}$ or $\mathcal{K}$). The results developed in (Milton, 1987a,

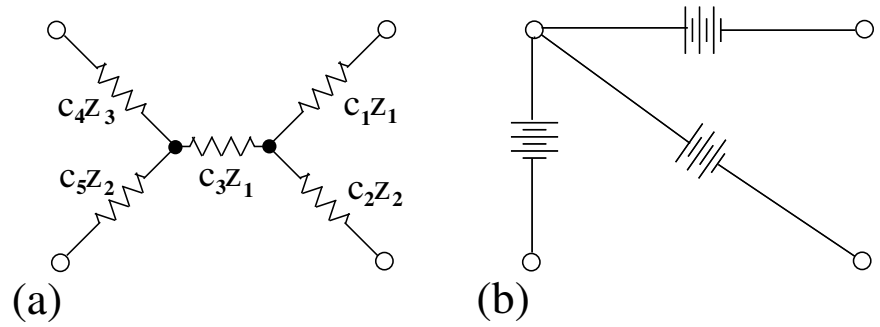


Figure 7.3: Shown in (a) is a 4 terminal electrical network, which is representative of a $Z(3)$ subspace collection. Here the c_i are real positive scaling constants: the conductance of each element is $c_j z_k$ where z_k is real or complex (when z is complex we should refer to $c_j z_k$ as an admittance rather than as a conductance). Complex values of z are appropriate when the applied potentials vary sinusoidally with time, and some of the impedence elements are capacitors or inductors. Figure (b) shows the batteries on the back side of the circuit board, representing the space $\mathcal{V}$, which combined with the resistors on the front side is representative of a $Y(3)$ subspace collection. The Y-function $\mathbf{Y}(z_1, z_2, z_3)$ gives the current going through the three batteries, in response to the voltages across them.

1987b, 1991 and in **Chapters 19**, **20**, and **29** of Milton, 2002) are extended to the case where there is no inner product. Accordingly, some steps in the analysis and some assumptions need to be revised. In this more general setting we can generate, from an appropriate $Z(n)$ subspace collection, any desired scalar-valued rational function $Z(z_1, z_2, \ldots, z_n)$ satisfying the homogeneity property $Z(1, 1, \ldots, 1) = 1$.

It is to be emphasized that subspace collections, with the associated rules for addition, multiplication, division, subtraction and substitution, are algebraic objects in their own right: there is no need to think of the associated analytic functions (that are in general operator-valued), except that the correspondence makes it easier to think about subspace collections. The resistor network examples of $Y(n)$ subspace collections made it possible for me to see how the operations of addition, multiplication and substitution of subspace collections should be defined in the general case. Although these operations seem a bit foreign and convoluted, they are quite natural when seen through the lens of the resistor network example.

My belief is that the geometrical structure of subspace collections (and in particular superfunctions) will be reflected in the algebraic geometrical structure of their associated rational functions. If this is the case, understanding the topological features of subspace collections might shed light on the geometrical features of algebraic varieties. While this paper does not directly address this issue, it sheds the first light on the relation between finite-dimensional subspace collections and rational functions of several complex variables, in the case where the subspaces are not mutually orthogonal, and it introduces superfunctions. The functions derived from superfunctions are well studied and have widespread applications in signal processing, control theory, network synthesis and design, and in optics, acoustics and elastodynamics (usually in layered media), where they are called a variety of names including transfer matrices, transmission matrices, transfer functions, system functions, and network functions. In these contexts it is the function that is studied, but people do not think of the superfunction. I thank Aaron Welters and Mihai Putinar for drawing my attention to the connection between transfer functions and response functions (such the effective conductivity tensor of composites).

We remark that for $Z(3)$ orthogonal subspace collections, with $\mathcal{U}$ being one-dimensional, it is still an open

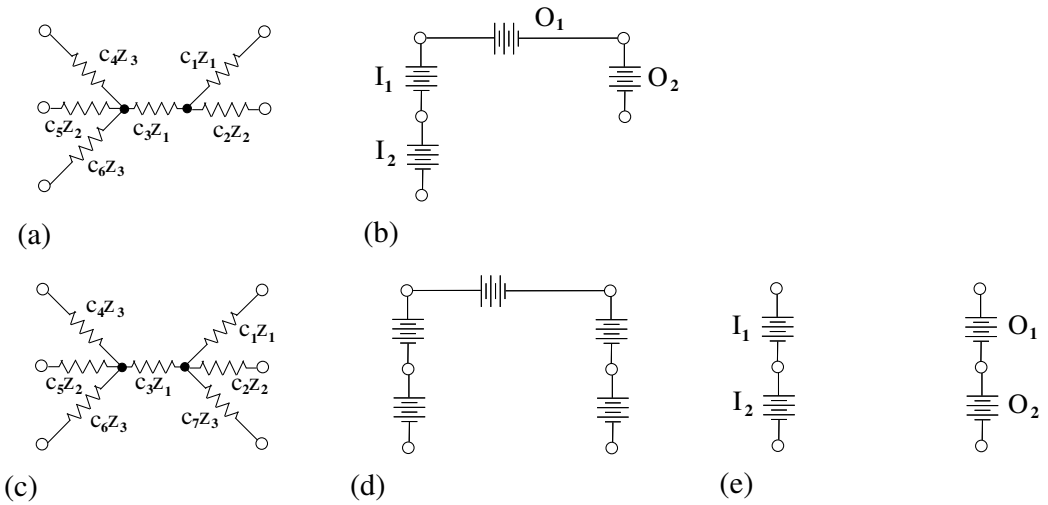


Figure 7.4: Shown in (a) is a 5 terminal electrical network, which is representative of a $Z(3)$ subspace collection. Here the c_i are real positive scaling constants: the admittance of each element is $c_j z_k$ where z_k is real or complex. Figure (b) shows the batteries on the back side of the circuit board, representing the space $\mathcal{V}$, which is divided into the input space $\mathcal{V}^I$, consisting of those vectors in $\mathcal{K}$ that are nonzero only in the batteries I_1 and I_2 and the output space $\mathcal{V}^O$, consisting of those vectors in $\mathcal{K}$ that are nonzero only in the batteries/resistors 0_1 and 0_2. Figure (c) shows a 6 terminal electrical network, and the naturally associated subspace $\mathcal{V}$ represented by the batteries in Figure (d). To convert this to a problem where the dimension of $\mathcal{V}$ is even we remove the battery at the top, and accordingly reduce the dimension of both $\mathcal{V}$ and $\mathcal{J}$ by one. Figure (e) shows the input space $\mathcal{V}^I$, consisting of those vectors that are nonzero only in the batteries I_1 and I_2 and the output space $\mathcal{V}^O$, consisting of those vectors in $\mathcal{K}$ that are nonzero only in the batteries/resistors 0_1 and 0_2.

and intriguing question as to whether there could be a one-to-one correspondence between them (assuming they are pruned as described in Section 7.15 and modulo trivial equivalences between subspace collections) and scalar functions $Z(z_1, z_2, z_3)$ satisfying the homogeneity, Herglotz and normalization properties. The Z-problem described the next section provides a nonlinear map from the $Z(3)$ orthogonal subspace collection to an associated scalar function $Z(z_1, z_2, z_3)$ satisfying the homogeneity, Herglotz and normalization properties, but the question is whether one can uniquely recover the pruned subspace collection, modulo trivial equivalences, given only the function $Z(z_1, z_2, z_3)$? The intriguing counting argument given in Section 29.2 of Milton (2002) suggests the possibility of a one-to-one correspondence. There is a similar counting argument for nonorthogonal subspace collections given in Section 7.18, but in this case we will see in an explicit example that a one-to-one correspondence does not hold.

This chapter is mostly self-contained, but it is suggested that the reader browse **Chapters 1**, **2** and **3** to understand the application of (nonorthogonal) subspace collections to physical problems.

7.2 Subspace collections and their associated functions

Let $\mathcal{K}$ be a vector space which has a decomposition into two different direct sums of subspaces

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{H}, \tag{7.5}$$

where $\mathcal{H}$ itself is a direct sum of n subspaces

$$\mathcal{H} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n. \tag{7.6}$$

Any vector $\mathbf{K} \in \mathcal{K}$ has a unique decomposition into component vectors,

$$\mathbf{K} = \mathbf{E} + \mathbf{J} = \mathbf{v} + \mathbf{H}, \quad \mathbf{H} = \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_n, \tag{7.7}$$

each in the associated subspaces:

$$\mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{v} \in \mathcal{V}, \quad \mathbf{H} \in \mathcal{H}, \quad \mathbf{P}_i \in \mathcal{P}_i \text{ for } i = 1, 2, \ldots, n. \tag{7.8}$$

This decomposition serves to define projection operators $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ onto $\mathcal{E}$ and $\mathcal{J}$, projection operators $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ onto $\mathcal{V}$ and $\mathcal{H}$, and projection operators $\mathbf{\Lambda}_i$ onto the subspaces $\mathcal{P}_i$. By definition we have

$$\mathbf{E} = \mathbf{\Gamma}_1\mathbf{K}, \quad \mathbf{J} = \mathbf{\Gamma}_2\mathbf{K}, \quad \mathbf{v} = \mathbf{\Pi}_1\mathbf{K}, \quad \mathbf{H} = \mathbf{\Pi}_2\mathbf{K}, \quad \mathbf{P}_i = \mathbf{\Lambda}_i\mathbf{K}. \tag{7.9}$$

Associated with this subspace collection is a linear operator-valued function $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ acting on the space $\mathcal{V}$, which is a homogeneous function of degree 1 of the n complex variables $z_1, z_2, \ldots, z_n$. To obtain the function we take each field $\mathbf{E}_1 \in \mathcal{V}$ and look for vectors $\mathbf{J}$ and $\mathbf{E}$ that solve the equations

$$\mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2, \quad \text{where } \mathbf{J}_2 = \mathbf{\Pi}_2\mathbf{J}, \quad \mathbf{E}_2 = \mathbf{\Pi}_2\mathbf{E}, \tag{7.10}$$

with $\mathbf{E}_1 = \mathbf{\Pi}_1\mathbf{E}$, where

$$\mathbf{L} = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i. \tag{7.11}$$

We call this problem the Y-problem. The associated operator $\mathbf{Y}$, by definition, governs the linear relation

$$\mathbf{J}_1 = -\mathbf{Y}\mathbf{E}_1, \quad \text{where } \mathbf{J}_1 = \mathbf{\Pi}_1\mathbf{J}. \tag{7.12}$$

A necessary condition for $\mathbf{J}_1$ to be uniquely defined given $\mathbf{E}_1$ is that

$$\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}, \tag{7.13}$$

since if $\mathbf{J}$ and $\mathbf{E}$ solve (7.10) so too will $\mathbf{J} + \mathbf{v}$ and $\mathbf{E}$, for any $\mathbf{v} \in \mathcal{V} \cap \mathcal{J}$. The inverse Y-problem is to solve (7.10) for each field $\mathbf{J}_1 = \mathbf{\Pi}_1\mathbf{J} \in \mathcal{V}$. A necessary condition for $\mathbf{E}_1$ to be uniquely defined given $\mathbf{J}_1$ is that

$$\mathcal{V} \cap \mathcal{E} = \{\mathbf{0}\}. \tag{7.14}$$

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is a basis of $\mathcal{V}$, then the operator $\mathbf{Y}$ can be represented by a matrix, the Y-matrix, also denoted by $\mathbf{Y}$ with elements Y_{ik} such that

$$\mathbf{Y}\mathbf{v}_k = \sum_{i=1}^{m} Y_{ik}\mathbf{v}_i. \tag{7.15}$$

If m is even and $\mathcal{V}$ has the decomposition

$$\mathcal{V} = \mathcal{V}^I \oplus \mathcal{V}^O, \tag{7.16}$$

where $\mathcal{V}^I$ and $\mathcal{V}^O$ have the same dimension $(m/2)$ then we have a superfunction F^s. The superfunction is the collection of subspaces and there is a function $\mathbf{F}$ associated with it. The fields $\mathbf{E}_1$ and $\mathbf{J}_1$ have the unique decomposition

$$\mathbf{E}_1 = \mathbf{E}^I + \mathbf{E}^O, \quad \mathbf{J}_1 = \mathbf{J}^I + \mathbf{J}^O, \tag{7.17}$$

with

$$\mathbf{E}^I, \mathbf{J}^I \in \mathcal{V}^I, \quad \mathbf{E}^O, \mathbf{J}^O \in \mathcal{V}^O, \tag{7.18}$$

where the superscripts I and O refer to input and output respectively. We write

$$\mathbf{E}^I = \mathbf{\Pi}^I \mathbf{E}_1, \quad \mathbf{E}^O = \mathbf{\Pi}^O \mathbf{E}_1, \quad \mathbf{J}^I = \mathbf{\Pi}^I \mathbf{J}_1, \quad \mathbf{J}^O = \mathbf{\Pi}^O \mathbf{J}_1, \tag{7.19}$$

which defines the projections $\mathbf{\Pi}^I$ and $\mathbf{\Pi}^O$ onto the input and output spaces. Now the relation (7.12) can be written as

$$\begin{pmatrix} \mathbf{J}^I \\ \mathbf{J}^O \end{pmatrix} = \begin{pmatrix} \mathbf{Y}^{II} & \mathbf{Y}^{IO} \\ \mathbf{Y}^{OI} & \mathbf{Y}^{OO} \end{pmatrix} \begin{pmatrix} \mathbf{E}^I \\ \mathbf{E}^O \end{pmatrix}, \tag{7.20}$$

and manipulated into the form

$$\begin{pmatrix} \mathbf{E}^O \\ \mathbf{J}^O \end{pmatrix} = \mathbf{F} \begin{pmatrix} \mathbf{E}^I \\ \mathbf{J}^I \end{pmatrix}, \tag{7.21}$$

which defines the linear operator-valued function

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^{EE} & \mathbf{F}^{EJ} \\ \mathbf{F}^{JE} & \mathbf{F}^{JJ} \end{pmatrix} = \begin{pmatrix} -(\mathbf{Y}^{IO})^{-1}\mathbf{Y}^{II} & -(\mathbf{Y}^{IO})^{-1} \\ [\mathbf{Y}^{OO}(\mathbf{Y}^{IO})^{-1}\mathbf{Y}^{II} - \mathbf{Y}^{OI}] & \mathbf{Y}^{OO}(\mathbf{Y}^{IO})^{-1} \end{pmatrix}, \tag{7.22}$$

which, provided the operator $\mathbf{Y}^{IO}$ is nonsingular, is the function associated with the superfunction. This relation can be inverted to yield $\mathbf{Y}$ in terms of $\mathbf{F}$,

$$\mathbf{Y} = \begin{pmatrix} (\mathbf{F}^{EJ})^{-1}\mathbf{F}^{EE} & -(\mathbf{F}^{EJ})^{-1} \\ [\mathbf{F}^{JJ}(\mathbf{F}^{EJ})^{-1}\mathbf{F}^{EE} - \mathbf{F}^{JE}] & -\mathbf{F}^{JJ}(\mathbf{F}^{EJ})^{-1} \end{pmatrix}, \tag{7.23}$$

provided the operator $\mathbf{F}^{EJ}$ can be inverted. The superfunction problem is for given input fields $\mathbf{E}^I$ and $\mathbf{J}^I$ to find fields $\mathbf{E}$ and $\mathbf{J}$ that solve the Y-problem (7.10) and (7.11), with $\mathbf{\Pi}^I\mathbf{E} = \mathbf{E}^I$ and $\mathbf{\Pi}^I\mathbf{J} = \mathbf{J}^I$. It may happen that the superfunction problem has a solution when the Y-problem does not (this happens if and only if $\mathbf{F}^{EJ}$ is singular), and conversely the Y-problem may have a solution when the superfunction problem does not (this happens if and only if $\mathbf{Y}^{IO}$ is singular).

Another association between subspace collections and functions comes if a vector space $\mathcal{H}$ has the decomposition

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.24}$$

where $\mathcal{E}$ and $\mathcal{J}$ are not to be confused with the spaces in (7.5). Any vector $\mathbf{H} \in \mathcal{H}$ has a unique decomposition into component vectors,

$$\mathbf{H} = \mathbf{u} + \mathbf{E} + \mathbf{J} = \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_n, \tag{7.25}$$

each in the associated subspaces:

$$\mathbf{u} \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{P}_i \in \mathcal{P}_i \text{ for } i = 1, 2, \ldots, n. \tag{7.26}$$

This decomposition serves to define projection operators $\mathbf{\Gamma}_0$, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ onto $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$, and projection operators $\mathbf{\Lambda}_i$ onto the subspaces $\mathcal{P}_i$. Associated with this subspace collection is a linear operator-valued function $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ acting on the space $\mathcal{U}$, which is a homogeneous function of degree 1 of the n complex variables $z_1, z_2, \ldots, z_n$. To obtain the function we take each vector $\mathbf{e} \in \mathcal{U}$ and look for vectors $\mathbf{j}$, $\mathbf{J}$ and $\mathbf{E}$ that solve the equations

$$\mathbf{j} \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{j} + \mathbf{J} = \mathbf{L}(\mathbf{e} + \mathbf{E}), \quad \text{where } \mathbf{L} = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i. \tag{7.27}$$

We call this problem the Z-problem. The associated operator $\mathbf{Z}$, by definition, governs the linear relation

$$\mathbf{j} = \mathbf{Z}\mathbf{e}. \tag{7.28}$$

If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ is a basis of $\mathcal{U}$, then the operator $\mathbf{Z}$ can be represented by a matrix, also denoted by $\mathbf{Z}$ with elements Z_{ik} such that

$$\mathbf{Z}\mathbf{u}_k = \sum_{i=1}^{m} Z_{ik} \mathbf{u}_i. \tag{7.29}$$

When $z_1 = z_2 = \cdots = z_n = 1$ (7.27) has the trivial solution

$$\mathbf{j} = \mathbf{e}, \quad \mathbf{J} = \mathbf{E} = 0, \tag{7.30}$$

and so we deduce that

$$\mathbf{Z}(1, 1, \ldots, 1) = \mathbf{I}. \tag{7.31}$$

The inverse Z-problem is to solve the equations (7.27) for each given vector $\mathbf{j} \in \mathcal{U}$.

7.3 Some simple examples

Consider a $Y(n)$ subspace collection

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.32}$$

where $\mathcal{E}, \mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_n$ are all one-dimensional, and $\mathcal{J}$ is n-dimensional. Choose as our basis for $\mathcal{K}$ the $n + 1$ vectors $\mathbf{p}_0 \in \mathcal{V}$, and $\mathbf{p}_i \in \mathcal{P}_i$, $i = 1, 2, \ldots n$. Vectors $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ can be expanded in this basis:

$$\mathbf{E} = \sum_{i=0}^{n} E_i \mathbf{p}_i, \quad \mathbf{J} = \sum_{i=0}^{n} J_i \mathbf{p}_i. \tag{7.33}$$

The relation $\mathbf{\Pi}_2 \mathbf{J} = \mathbf{L}\mathbf{\Pi}_2 \mathbf{E}$ implies

$$J_i = z_i E_1. \tag{7.34}$$

Let us suppose that $E_0 = 1$. Then E_1 and E_2 are determined by the orientation of the one-dimensional subspace $\mathcal{E}$ with respect to the subspaces $\mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_n$. Also since $\mathcal{J}$ has codimension 1, there exist

constants $W_0, W_1, \ldots W_n$, determined by the orientation of the n-dimensional subspace $\mathcal{J}$ with respect to the subspaces $\mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots \mathcal{P}_n$ such that

$$\sum_{i=0}^{n} W_i J_i = 0. \tag{7.35}$$

Let us suppose that $W_0 = 1$. Then we have

$$J_0 = -\sum_{i=1}^{n} W_i J_i = -\sum_{i=1}^{n} W_i E_i z_i, \tag{7.36}$$

which since $E_0 = 1$ implies $J_0 = -Y E_0$, with

$$Y = \sum_{i=1}^{n} \alpha_i z_i, \quad \text{where } \alpha_i = W_i E_i. \tag{7.37}$$

As the E_i and W_i are arbitrary constants, we see that Y can be any linear combination of the z_i. In particular, with $W_1 E_1 = 1$ and $W_i E_i = 0$ when $i \neq 1$ we obtain

$$Y = z_1. \tag{7.38}$$

As a second example consider a $Y(1)$ subspace collection

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}_1, \tag{7.39}$$

where all the spaces $\mathcal{E}$, $\mathcal{J}$, $\mathcal{V}$, and $\mathcal{P}_1$ are all two-dimensional. Choose as our basis for $\mathcal{K}$ two vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathcal{V}$ and two vectors $\mathbf{p}_3$ and $\mathbf{p}_4$ in $\mathcal{P}_1$. Then since $\mathcal{E}$ is two-dimensional, there generically exist constants e_{13}, e_{14}, e_{23} and e_{24} such that

$$\mathbf{p}_1 + e_{13}\mathbf{p}_3 + e_{14}\mathbf{p}_4 \in \mathcal{E}, \quad \mathbf{p}_2 + e_{23}\mathbf{p}_3 + e_{24}\mathbf{p}_4 \in \mathcal{E}. \tag{7.40}$$

Also since $\mathcal{J}$ is two-dimensional, there generically exist constants j_{31}, j_{32}, j_{41} and j_{42} such that

$$\mathbf{p}_3 + j_{31}\mathbf{p}_1 + j_{32}\mathbf{p}_2 \in \mathcal{J}, \quad \mathbf{p}_4 + j_{41}\mathbf{p}_1 + e_{42}\mathbf{p}_2 \in \mathcal{J}. \tag{7.41}$$

So the Y-problem is solved with vectors

$$\begin{aligned}
\mathbf{E} &= \mathbf{p}_1 + e_{13}\mathbf{p}_3 + e_{14}\mathbf{p}_4, \\
\mathbf{E}_1 &= \mathbf{p}_1, \quad \mathbf{E}_2 = e_{13}\mathbf{p}_3 + e_{14}\mathbf{p}_4, \\
\mathbf{J}_2 &= z_1(e_{13}\mathbf{p}_3 + e_{14}\mathbf{p}_4), \\
\mathbf{J} &= z_1[e_{13}(\mathbf{p}_3 + j_{31}\mathbf{p}_1 + j_{32}\mathbf{p}_2) + e_{14}(\mathbf{p}_4 + j_{41}\mathbf{p}_1 + e_{42}\mathbf{p}_2)], \\
\mathbf{J}_1 &= z_1[(e_{13}j_{31} + e_{14}j_{41})\mathbf{p}_1 + (e_{13}j_{32} + e_{42}j_{42})\mathbf{p}_2,
\end{aligned} \tag{7.42}$$

and is also solved with vectors

$$\begin{aligned}
\mathbf{E} &= \mathbf{p}_2 + e_{23}\mathbf{p}_3 + e_{24}\mathbf{p}_4, \\
\mathbf{E}_1 &= \mathbf{p}_2, \quad \mathbf{E}_2 = e_{23}\mathbf{p}_3 + e_{24}\mathbf{p}_4, \\
\mathbf{J}_2 &= z_1(e_{23}\mathbf{p}_3 + e_{24}\mathbf{p}_4), \\
\mathbf{J} &= z_1[e_{23}(\mathbf{p}_3 + j_{31}\mathbf{p}_1 + j_{32}\mathbf{p}_2) + e_{24}(\mathbf{p}_4 + j_{41}\mathbf{p}_1 + e_{42}\mathbf{p}_2)], \\
\mathbf{J}_1 &= z_1[(e_{23}j_{31} + e_{24}j_{41})\mathbf{p}_1 + (e_{23}j_{32} + e_{24}j_{42})\mathbf{p}_2.
\end{aligned} \tag{7.43}$$

From these equations in follows that $\mathbf{Y}(z_1)$ in this basis is the 2 by 2 matrix

$$\mathbf{Y}(z_1) = z_1\mathbf{A}, \quad \text{with } \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \tag{7.44}$$

where

$$\begin{aligned} a_{11} = e_{13}j_{31} + e_{14}j_{41}, \quad a_{12} = e_{13}j_{32} + e_{42}j_{42}, \\ a_{21} = e_{23}j_{31} + e_{24}j_{41}, \quad a_{22} = e_{23}j_{32} + e_{24}j_{42}. \end{aligned} \tag{7.45}$$

As the coefficients $e_{13}, e_{14}, e_{23}, e_{24}, j_{31}, j_{32}, j_{41}$ and j_{42} can be any complex numbers we desire it follows that we can realize any desired complex matrix $\mathbf{A}$. By taking $\mathcal{V}^I$ to be the one-dimensional space spanned by $\mathbf{p}_1$ and taking $\mathcal{V}^O$ to be the one-dimensional space spanned by $\mathbf{p}_2$ we obtain a superfunction Y^S where the associated function takes the form

$$\mathbf{F}(z_1) = \begin{pmatrix} b_{11} & b_{12}/z_1 \\ b_{21}z_1 & b_{22} \end{pmatrix}, \tag{7.46}$$

in which the parameters b_{11}, b_{12}, b_{21} and b_{22} can be any complex numbers we choose.

As a third example consider a $Z(2)$ subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2, \tag{7.47}$$

where the subspaces $\mathcal{U}, \mathcal{E}, \mathcal{J}$ and $\mathcal{P}_2$ are all one-dimensional, while $\mathcal{P}_1$ is two-dimensional. Choose, as our basis for $\mathcal{H}$, 3 vectors $\mathbf{U}_0 \in \mathcal{U}$, $\mathbf{E}_0 \in \mathcal{E}$ and $\mathbf{J}_0 \in \mathcal{J}$, and take a vector $\mathbf{P}$ as a basis for $\mathcal{P}_2$. The coefficients P_U, P_E and P_J in the expansion

$$\mathbf{P} = P_U\mathbf{U}_0 + P_E\mathbf{E}_0 + P_J\mathbf{J}_0 \tag{7.48}$$

determine the orientation of $\mathcal{P}_2$ with respect to the subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$. In the basis $\mathbf{U}_0$, $\mathbf{E}_0$, and $\mathbf{J}_0$ the equations

$$\mathbf{e} + \mathbf{E} = \mathbf{Q} + \alpha\mathbf{P}, \quad \mathbf{j} + \mathbf{J} = z_1\mathbf{Q} + z_2\alpha\mathbf{P}, \tag{7.49}$$

with

$$\mathbf{e}, \mathbf{j} \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{Q} \in \mathcal{P}_1, \tag{7.50}$$

take the form

$$\begin{aligned} \begin{pmatrix} e \\ E \\ 0 \end{pmatrix} &= \begin{pmatrix} Q_U \\ Q_E \\ Q_J \end{pmatrix} + \alpha \begin{pmatrix} P_U \\ P_E \\ P_J \end{pmatrix}, \\ \begin{pmatrix} j \\ 0 \\ J \end{pmatrix} &= z_1 \begin{pmatrix} Q_U \\ Q_E \\ Q_J \end{pmatrix} + z_2\alpha \begin{pmatrix} P_U \\ P_E \\ P_J \end{pmatrix}, \end{aligned} \tag{7.51}$$

and since $\mathbf{Q} \in \mathcal{P}_1$ there exist constants W_U, W_E and W_J, which determine the orientation of $\mathcal{P}_1$ with respect to $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$, such that

$$W_UQ_U + W_EQ_E + W_JQ_J = 0. \tag{7.52}$$

Hence we obtain the equations

$$\begin{aligned} W_U e + W_E E &= \alpha(W_U P_U + W_E P_E + W_J P_J) \equiv \alpha \mathbf{W}\cdot\mathbf{P}, \\ 0 &= z_1(E - \alpha P_E) + z_2 \alpha P_E, \\ j &= z_1(e - \alpha P_U) + z_2 \alpha P_U. \end{aligned} \tag{7.53}$$

Eliminating E and α from these equations gives $j = Ze$, with

$$Z = z_1 + \frac{(z_2 - z_1) W_U P_U}{\mathbf{W}\cdot\mathbf{P} + W_E P_E (z_2 - z_1)/z_1}. \tag{7.54}$$

In particular if the subspaces are oriented so that

$$\mathbf{W}\cdot\mathbf{P} = W_E P_E = -W_U P_U, \tag{7.55}$$

then (7.54) gives

$$Z = z_1^2/z_2, \tag{7.56}$$

which with $z_2 = 1$ produces the function z_1^2 and with $z_1 = 1$ produces the function $1/z_2$. Also, with $W_E P_E = 0$ we obtain

$$Z = z_1 + \frac{(z_2 - z_1) W_U P_U}{\mathbf{W}\cdot\mathbf{P}}, \tag{7.57}$$

which is a "weighted average" of z_1 and z_2, $Z = w_1 z_1 + w_2 z_2$ with "weights" w_1 and w_2 that sum to 1 but which are not necessarily positive, nor even real.

7.4 Formulas for the associated functions

Following Section 12.8 of Milton (2002) a formula for the effective tensor $\mathbf{Z}$ results by applying the operator $\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2$ (which projects on the space $\mathcal{U} \oplus \mathcal{J}$) to both sides of the constitutive law $\mathbf{e} + \mathbf{E} = \mathbf{L}^{-1}(\mathbf{j} + \mathbf{J})$. Solving the resulting equation

$$\mathbf{e} = (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)(\mathbf{j} + \mathbf{J}), \tag{7.58}$$

for $\mathbf{j} + \mathbf{J}$ gives

$$\mathbf{j} + \mathbf{J} = [(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)]^{-1}\mathbf{e}, \tag{7.59}$$

where the last inverse is to be taken on the subspace $\mathcal{U} \oplus \mathcal{J}$. By applying $\mathbf{\Gamma}_0$ to both sides of this equation we see that

$$\mathbf{Z} = \mathbf{\Gamma}_0[(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)]^{-1}\mathbf{\Gamma}_0, \tag{7.60}$$

which is the result given in (12.59) of Milton (2002).

Another formula for $\mathbf{Z}$ follows from noting that for any arbitrary constant $z_0 \neq 0$,

$$[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})](\mathbf{e} + \mathbf{E}) = z_0\mathbf{e} + z_0\mathbf{E} - \mathbf{\Gamma}_1\mathbf{J} - z_0\mathbf{\Gamma}_1\mathbf{E} = z_0\mathbf{e}. \tag{7.61}$$

Solving this for $\mathbf{e} + \mathbf{E}$ gives

$$\mathbf{e} + \mathbf{E} = z_0[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})]^{-1}\mathbf{e}, \tag{7.62}$$

and applying $\mathbf{\Gamma}_0\mathbf{L}$ to both sides yields

$$\mathbf{j} = z_0\mathbf{\Gamma}_0\mathbf{L}[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})]^{-1}\mathbf{e}. \tag{7.63}$$

Thus we have a formula for the $\mathbf{Z}$ operator,

$$\mathbf{Z} = z_0\mathbf{\Gamma}_0\mathbf{L}[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})]^{-1}\mathbf{\Gamma}_0 = z_0\mathbf{\Gamma}_0 + z_0\mathbf{\Gamma}_0(\mathbf{L} - z_0\mathbf{I})[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})]^{-1}\mathbf{\Gamma}_0, \tag{7.64}$$

where we have used the identity

$$\mathbf{\Gamma}_0 = z_0\mathbf{\Gamma}_0[z_0\mathbf{I} - \mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})]^{-1}\mathbf{\Gamma}_0, \tag{7.65}$$

obtained by applying $\mathbf{\Gamma}_0$ to both sides of (7.62). This formula (7.64) is a special case of the formula (12.60) given in Milton (2002), and is well known in different contexts (Kröner 1977).

To obtain a formula for $\mathbf{Y}$ notice that (7.10) and (7.12) imply that

$$0 = \mathbf{\Gamma}_2\mathbf{E}' = \mathbf{\Gamma}_2\mathbf{E}_1 + \mathbf{\Gamma}_2\mathbf{E}_2 = \mathbf{\Gamma}_2\mathbf{E}_1 + \mathbf{\Gamma}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2\mathbf{J}', \tag{7.66}$$

where the inverse of $\mathbf{L}$ is to be taken on the subspace $\mathcal{H}$. Solving for $\mathbf{J}'$ gives

$$\mathbf{J}' = -(\mathbf{\Gamma}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}\mathbf{\Gamma}_2\mathbf{E}_1, \tag{7.67}$$

where the inverse is to be taken on the subspace $\mathcal{J}$. Then by applying $\mathbf{\Pi}_1$ to both sides of this equation and equating $\mathbf{\Pi}_1\mathbf{J}' = \mathbf{J}_1$ with $-\mathbf{Y}\mathbf{E}_1$ we obtain the desired formula

$$\mathbf{Y} = \mathbf{\Pi}_1\mathbf{\Gamma}_2(\mathbf{\Gamma}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}\mathbf{\Gamma}_2\mathbf{\Pi}_1 \tag{7.68}$$

for $\mathbf{Y}$, as given in formula (19.29) of Milton (2002).

Another formula for $\mathbf{Y}$ is obtained by taking an arbitrary constant $z_0 \neq 0$, and defining

$$\mathbf{P}' = \mathbf{J}' - z_0\mathbf{E}'. \tag{7.69}$$

Applying $\mathbf{\Gamma}_1$ to both sides of (7.69) gives

$$\mathbf{\Gamma}_1\mathbf{P}' = -z_0\mathbf{E}' = -z_0(\mathbf{E}_1 + \mathbf{E}_2), \tag{7.70}$$

and applying $\mathbf{\Pi}_2$ to both sides of (7.70) gives

$$\mathbf{\Pi}_2\mathbf{P}' = \mathbf{J}_2 - z_0\mathbf{E}_2 = (\mathbf{L} - z_0\mathbf{I})\mathbf{E}_2. \tag{7.71}$$

Combining these results we see that $\mathbf{P}'$ satisfies

$$[\mathbf{\Gamma}_1 + z_0(\mathbf{L} - z_0\mathbf{I})^{-1}\mathbf{\Pi}_1]\mathbf{P}' = -z_0\mathbf{E}_1. \tag{7.72}$$

Assuming that the operator $[\mathbf{\Gamma}_1 + z_0(\mathbf{L} - z_0\mathbf{I})^{-1}\mathbf{\Pi}_1]$ is nonsingular this gives

$$\mathbf{P}' = -z_0[\mathbf{\Gamma}_1 + z_0(\mathbf{L} - z_0\mathbf{I})^{-1}\mathbf{\Pi}_1]^{-1}\mathbf{E}_1. \tag{7.73}$$

Applying $\mathbf{\Pi}_1 = \mathbf{I} - \mathbf{\Pi}_2$ to both sides yields

$$\mathbf{J}_1 - z_0\mathbf{E}_1 = -(\mathbf{Y} + z_0\mathbf{I})\mathbf{E}_1 = -z_0\mathbf{\Gamma}_1[\mathbf{\Gamma}_1 + z_0(\mathbf{L} - z_0\mathbf{I})^{-1}\mathbf{\Pi}_1]^{-1}\mathbf{E}_1 \tag{7.74}$$

As this holds for all $\mathbf{E}_1 \in \mathcal{V}$ we obtain the formula

$$\mathbf{Y} = -z_0\mathbf{\Pi}_1 + z_0\mathbf{\Gamma}_1[\mathbf{\Gamma}_1 + z_0(\mathbf{L} - z_0\mathbf{I})^{-1}\mathbf{\Pi}_1]^{-1}\mathbf{\Pi}_1 \tag{7.75}$$

which is a special case of the formula (19.37) obtained in Section 19.5 of Milton (2002).

7.5 Multiplying superfunctions

Multiplying superfunctions is similar the way electrical circuits, each with $2m$ terminal can be combined. An example is shown in **Figure 7.5**.

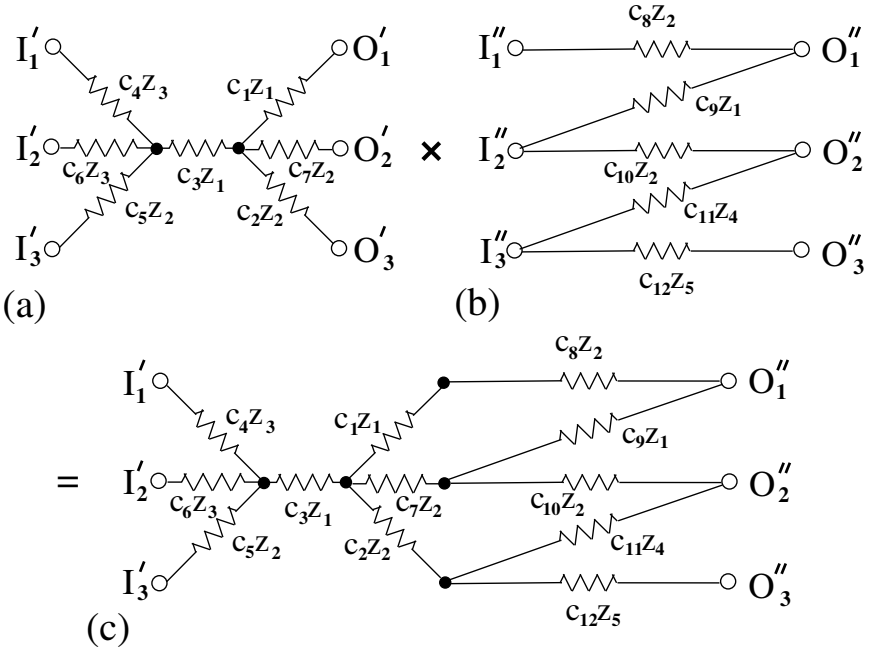


Figure 7.5: Multiplying superfunctions is like hooking networks, with an equal number of input and output terminals, together in series. Shown in (a) and (b) are 6 terminal electrical networks, each (along with their respective tree-like battery configurations on the opposite side of the circuit board that are not shown here) represent a superfunction as the terminals have been divided into input terminals (I'_1, I'_2, and I'_3 for the circuit (a), and I''_1, I''_2, and I''_3 for the circuit (b)) and output terminals (O'_1, O'_2, and O'_3 for the circuit (a), and O''_1, O''_2, and O''_3 for the circuit (b)). The product superfunction is the 6 terminal electrical network (along with its tree-like battery configurations on the opposite side of the circuit board) shown in (c). Note there is some flexibility in how one multiplies superfunctions: instead of connecting the terminals O'_i with I''_i for $i = 1, 2, 3$, one could for example, connect O'_1,O'_2, and O'_3 with any permutation of I''_1, I''_2 and I''_3. This is why, when taking a product, one needs to specify the maps ($\mathbf{M}^E$ and $\mathbf{M}^J$) one is using between the output space of one superfunction, and the input space of the second superfunction by which one is multiplying it.

Suppose we have two superfunctions, $(F^s)'$ and $(F^s)''$:

$$\begin{aligned} \mathcal{K}' &= \mathcal{E}' \oplus \mathcal{J}' = (\mathcal{V}^I)' \oplus (\mathcal{V}^O)' \oplus \mathcal{H}' \quad \text{with } \mathcal{H}' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_j, \\ \mathcal{K}'' &= \mathcal{E}'' \oplus \mathcal{J}'' = (\mathcal{V}^I)'' \oplus (\mathcal{V}^O)'' \oplus \mathcal{H}'' \quad \text{with } \mathcal{H}'' = \mathcal{P}''_1 \oplus \mathcal{P}''_2 \oplus \cdots \oplus \mathcal{P}''_k, \end{aligned} \tag{7.76}$$

where the spaces $(\mathcal{V}^I)', (\mathcal{V}^O)', (\mathcal{V}^I)'', (\mathcal{V}^O)''$ all have the same dimension m. To take their product one needs to first find two nonsingular linear operators $\mathbf{M}^E$ and $\mathbf{M}^J$ which map $(\mathcal{V}^O)'$ to $(\mathcal{V}^I)''$. The resulting product

superfunction

$$F^s = (F^s)' \times_{\mathbf{M}} (F^s)'', \tag{7.77}$$

is the subspace collection

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = (\mathcal{V}^I)' \oplus (\mathcal{V}^O)'' \oplus \mathcal{H}, \tag{7.78}$$

where

$$\mathcal{H} = \mathcal{P}_1' \oplus \mathcal{P}_2' \oplus \cdots \oplus \mathcal{P}_j' \oplus \mathcal{P}_1'' \oplus \mathcal{P}_2'' \oplus \cdots \oplus \mathcal{P}_k''', \tag{7.79}$$

and the operator $\mathbf{L}$ acting on $\mathcal{H}$ is

$$\mathbf{L} = \sum_{i=1}^{j} z_i' \mathbf{\Lambda}_i' + \sum_{\ell=1}^{k} z_\ell'' \mathbf{\Lambda}_\ell'', \tag{7.80}$$

in which $\mathbf{\Lambda}_i'$ and $\mathbf{\Lambda}_\ell''$ are the projections onto $\mathcal{P}_i'$ and $\mathcal{P}_\ell''$. A vector $\mathbf{E}$ is in $\mathcal{E}$ if and only if we can find vectors

$$\begin{aligned} \mathbf{E}' &= (\mathbf{E}^I)' + (\mathbf{E}^O)' + \mathbf{E}_2' \in \mathcal{E}', \\ \mathbf{E}'' &= (\mathbf{E}^I)'' + (\mathbf{E}^O)'' + \mathbf{E}_2'' \in \mathcal{E}'', \end{aligned} \tag{7.81}$$

such that

$$(\mathbf{E}^I)'' = \mathbf{M}^E (\mathbf{E}^O)', \quad \mathbf{E} = (\mathbf{E}^I)' + (\mathbf{E}^O)'' + \mathbf{E}_2' + \mathbf{E}_2'', \tag{7.82}$$

with

$$(\mathbf{E}^I)' \in (\mathcal{V}^I)', \quad (\mathbf{E}^O)' \in (\mathcal{V}^O)', \quad \mathbf{E}_2' \in \mathcal{H}', \quad (\mathbf{E}^I)'' \in (\mathcal{V}^I)'', \quad (\mathbf{E}^O)'' \in (\mathcal{V}^O)'', \quad \mathbf{E}_2'' \in \mathcal{H}''. \tag{7.83}$$

A vector $\mathbf{J}$ is in $\mathcal{J}$ if and only if we can find vectors

$$\begin{aligned} \mathbf{J}' &= (\mathbf{J}^I)' + (\mathbf{J}^O)' + \mathbf{J}_2' \in \mathcal{J}', \\ \mathbf{J}'' &= (\mathbf{J}^I)'' + (\mathbf{J}^O)'' + \mathbf{J}_2'' \in \mathcal{J}'', \end{aligned} \tag{7.84}$$

such that

$$(\mathbf{J}^I)'' = \mathbf{M}^J (\mathbf{J}^O)', \quad \mathbf{J} = (\mathbf{J}^I)' + (\mathbf{J}^O)'' + \mathbf{J}_2' + \mathbf{J}_2'', \tag{7.85}$$

with

$$(\mathbf{J}^I)' \in (\mathcal{V}^I)', \quad (\mathbf{J}^O)' \in (\mathcal{V}^O)', \quad \mathbf{J}_2' \in \mathcal{H}', \quad (\mathbf{J}^I)'' \in (\mathcal{V}^I)'', \quad (\mathbf{J}^O)'' \in (\mathcal{V}^O)'', \quad \mathbf{J}_2'' \in \mathcal{H}''. \tag{7.86}$$

To ensure that the two spaces $\mathcal{E}$ and $\mathcal{J}$ are independent we need to make the technical assumption that $\mathbf{M}^E$ and $\mathbf{M}^J$ are chosen so that the operator $\mathbf{A}$ mapping $(\mathcal{V}^O)'$ to $(\mathcal{V}^I)''$, defined by

$$\mathbf{A} = \mathbf{M}^E (\mathbf{\Pi}^O)' \mathbf{\Gamma}_1' - (\mathbf{\Pi}^I)'' \mathbf{\Gamma}_1'' [\mathbf{M}^E (\mathbf{\Pi}^O)' \mathbf{\Gamma}_1' + \mathbf{M}^J (\mathbf{\Pi}^O)' \mathbf{\Gamma}_2'], \tag{7.87}$$

is nonsingular (i.e., the null–space of the operator contains only the zero vector). Our aim is to show that if $\mathbf{A}$ is nonsingular and

$$\mathbf{E} = (\mathbf{E}^I)' + (\mathbf{E}^O)'' + \mathbf{E}_2' + \mathbf{E}_2'' = \mathbf{J} = (\mathbf{J}^I)' + (\mathbf{J}^O)'' + \mathbf{J}_2' + \mathbf{J}_2'', \quad \text{with } \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J} \tag{7.88}$$

then $\mathbf{E} = \mathbf{J} = 0$. First note that by resolving (7.88) into components in the spaces $(\mathcal{V}^I)'$, $(\mathcal{V}^I)''$, $\mathcal{H}'$, and $\mathcal{H}''$ we obtain

$$(\mathbf{E}^I)' = (\mathbf{J}^I)', \quad (\mathbf{E}^O)'' = (\mathbf{J}^O)'', \quad \mathbf{E}_2' = \mathbf{J}_2', \quad \mathbf{E}_2'' = \mathbf{J}_2''. \tag{7.89}$$

Also since $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ there exist vectors $(\mathbf{E}^O)', (\mathbf{J}^O)' \in (\mathcal{V}^O)'$ and $(\mathbf{E}^I)'', (\mathbf{J}^I)'' \in (\mathcal{V}^I)''$ such that (7.81) and (7.84) hold. Since $\mathcal{E}' \cap \mathcal{J}' = \{\mathbf{0}\}$ and $\mathcal{E}'' \cap \mathcal{J}'' = \{\mathbf{0}\}$ it follows that

$$\mathbf{P} \equiv (\mathbf{E}^O)' - (\mathbf{J}^O)' = \mathbf{E}' - \mathbf{J}' \neq 0 \quad \text{or} \quad \mathbf{E}' = \mathbf{J}' = 0, \tag{7.90}$$

and

$$\mathbf{Q} \equiv (\mathbf{E}^I)'' - (\mathbf{J}^I)'' = \mathbf{E}'' - \mathbf{J}'' \neq 0 \quad \text{or} \quad \mathbf{E}'' = \mathbf{J}'' = 0. \tag{7.91}$$

Now we have

$$\begin{aligned} (\mathbf{\Pi}^O)'\mathbf{\Gamma}_1'\mathbf{P} &= (\mathbf{\Pi}^O)'\mathbf{E}' = (\mathbf{E}^O)', \quad (\mathbf{\Pi}^O)'\mathbf{\Gamma}_2'\mathbf{P} = -(\mathbf{\Pi}^O)'\mathbf{J}' = -(\mathbf{J}^O)' \\ (\mathbf{\Pi}^I)''\mathbf{\Gamma}_1''\mathbf{Q} &= (\mathbf{\Pi}^I)''\mathbf{E}'' = (\mathbf{E}^I)'', \quad (\mathbf{\Pi}^I)''\mathbf{\Gamma}_2''\mathbf{Q} = -(\mathbf{\Pi}^I)''\mathbf{J}'' = -(\mathbf{J}^I)''. \end{aligned} \tag{7.92}$$

Since $(\mathbf{E}^I)'' = \mathbf{M}^E(\mathbf{E}^O)'$ and $(\mathbf{J}^I)'' = \mathbf{M}^J(\mathbf{J}^O)'$ we get from the first pair of equations in (7.92) the result that

$$(\mathbf{E}^I)'' = \mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1'\mathbf{P}, \quad (\mathbf{J}^I)'' = -\mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2'\mathbf{P}, \tag{7.93}$$

which implies

$$\mathbf{Q} = [\mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1' + \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2']\mathbf{P}. \tag{7.94}$$

Substituting this back in the second pair of equations in (7.92), and using (7.93), gives

$$\begin{aligned} (\mathbf{\Pi}^I)''\mathbf{\Gamma}_1''[\mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1' + \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2']\mathbf{P} &= \mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1'\mathbf{P} \\ (\mathbf{\Pi}^I)''\mathbf{\Gamma}_2''[\mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1' + \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2']\mathbf{P} &= \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2'\mathbf{P}. \end{aligned} \tag{7.95}$$

These two equations are not independent since by adding them we obtain

$$[\mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1' + \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2']\mathbf{P} = \mathbf{M}^E(\mathbf{\Pi}^O)'\mathbf{\Gamma}_1'\mathbf{P} + \mathbf{M}^J(\mathbf{\Pi}^O)'\mathbf{\Gamma}_2'\mathbf{P} \tag{7.96}$$

which is obviously satisfied. Also the first equation in (7.95) says $\mathbf{P}$ is in the null space of $\mathbf{A}$, which by our assumption implies $\mathbf{P} = 0$. Then (7.94) implies $\mathbf{Q} = 0$ and this rules out the first possibilities in (7.90) and (7.91), implying $\mathbf{E}' = \mathbf{J}' = 0$ and $\mathbf{E}'' = \mathbf{J}'' = 0$. We conclude that $\mathbf{E} = \mathbf{J} = 0$.

To check that the space $\mathcal{E} \oplus \mathcal{J}$ spans $(\mathcal{V}^I)' \oplus (\mathcal{V}^O)'' \oplus \mathcal{H}$, we just need to count dimensions. The dimension of the space on the right is $2m$+dim($\mathcal{H}$). The dimension of $\mathcal{E}$ according to (7.81) is dim($\mathcal{E}'$)+dim($\mathcal{E}''$) less m because of the m constraints $(\mathbf{E}^I)'' = \mathbf{M}^E(\mathbf{E}^O)'$. Similarly the dimension of $\mathcal{J}$ is dim($\mathcal{J}'$)+dim($\mathcal{J}''$)-m. Adding these up, we get the dimension of $\mathcal{E} \oplus \mathcal{J}$ is dim$\mathcal{K}'$+dim$\mathcal{K}''$-$2m$=$2m$+dim($\mathcal{H}'$)+dim($\mathcal{H}'$)=$2m$+dim($\mathcal{H}$).

Let $\mathbf{F}'$ and $\mathbf{F}'$ be the functions associated with the superfunctions $(F^s)'$ and $(F^s)''$. Given operators

$$\mathbf{L}' = \sum_{i=1}^{j} z_i'\mathbf{\Lambda}_i', \quad \mathbf{L}'' = \sum_{i=1}^{k} z_i''\mathbf{\Lambda}_i'', \tag{7.97}$$

where $\mathbf{\Lambda}_i'$ projects onto $\mathcal{P}_i'$ and $\mathbf{\Lambda}_i''$ projects onto $\mathcal{P}_i''$, and given input fields $(\mathbf{E}^I)'$ and $(\mathbf{J}^I)'$ we can calculate

$$\begin{aligned} \begin{pmatrix} (\mathbf{E}^O)' \\ (\mathbf{J}^O)' \end{pmatrix} &= \mathbf{F}' \begin{pmatrix} (\mathbf{E}^I)' \\ (\mathbf{J}^I)' \end{pmatrix}, \\ (\mathbf{E}^I)' &= \mathbf{M}^E(\mathbf{E}^O)'', \quad (\mathbf{J}^I)' = \mathbf{M}^J(\mathbf{J}^O)'', \\ \begin{pmatrix} (\mathbf{E}^O)'' \\ (\mathbf{J}^O)'' \end{pmatrix} &= \mathbf{F}'' \begin{pmatrix} (\mathbf{E}^I)'' \\ (\mathbf{J}^I)'' \end{pmatrix}. \end{aligned} \tag{7.98}$$

From the knowledge of $(\mathbf{E}^O)'$ and $(\mathbf{E}^I)'$, and of $(\mathbf{E}^O)''$ and $(\mathbf{E}^I)''$, we can calculate the fields $\mathbf{E}'$, $\mathbf{E}''$, $\mathbf{J}'$, and $\mathbf{J}''$ of the form (7.81) and (7.84) solving the Y' problem and the Y'' problem:

$$\begin{aligned} \mathbf{E}' \in \mathcal{E}', \quad \mathbf{J}' \in \mathcal{J}', \quad \mathbf{J}'_1 = \mathbf{L}'\mathbf{E}'_1, \\ \mathbf{E}'' \in \mathcal{E}'', \quad \mathbf{J}'' \in \mathcal{J}'', \quad \mathbf{J}''_1 = \mathbf{L}''\mathbf{E}''_1. \end{aligned} \tag{7.99}$$

Then the fields $\mathbf{E}$ and $\mathbf{J}$ given by (7.82) and (7.85) solve the Y problem in the space $\mathcal{K}$, and the function associated to the superfunction F^s is given by the product rule

$$\mathbf{F} = \mathbf{F}' \begin{pmatrix} \mathbf{M}^E & 0 \\ 0 & \mathbf{M}^J \end{pmatrix} \mathbf{F}''. \tag{7.100}$$

Let us choose a basis $(\mathbf{v}_1^I)'', (\mathbf{v}_2^I)'', \ldots, (\mathbf{v}_m^I)''$ for $(\mathcal{V}^I)''$, choose a basis $(\mathbf{v}_1^O)'', (\mathbf{v}_2^O)'', \ldots, (\mathbf{v}_m^O)''$ for $(\mathcal{V}^O)''$, take $\mathbf{M}^E(\mathbf{v}_1^O)'', \mathbf{M}^E(\mathbf{v}_2^O)'', \ldots, \mathbf{M}^E(\mathbf{v}_m^O)''$ as our basis for $(\mathcal{V}^I)'$, and choose a basis $(\mathbf{v}_1^O)', (\mathbf{v}_2^O)', \ldots, (\mathbf{v}_m^O)'$ for $(\mathcal{V}^O)'$. Then the operator $\mathbf{M}^E$ is represented as the identity matrix in the basis. Let us also choose the operator $\mathbf{M}^J$ so it is represented by *minus* the identity matrix in this basis. Then in this basis the relation (7.100) takes the form

$$\mathbf{F} = \mathbf{F}' \begin{pmatrix} \mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{pmatrix} \mathbf{F}''. \tag{7.101}$$

Note that we could have avoided this slightly awkward multiplication rule if we had replaced the definition (7.21) of the associated function by

$$\begin{pmatrix} \mathbf{E}^O \\ -\mathbf{J}^O \end{pmatrix} = \mathbf{F} \begin{pmatrix} \mathbf{E}^I \\ \mathbf{J}^I \end{pmatrix}. \tag{7.102}$$

Then the multiplication rule (with this choice of $\mathbf{M}^E$ and $\mathbf{M}^J$) would have become simply $\mathbf{F} = \mathbf{F}'\mathbf{F}''$. We chose not to do this in the interest of preserving the symmetric roles of the spaces $\mathcal{E}$ and $\mathcal{J}$ in the definition of the function associated with the superfunction.

In passing, let us suppose there is an inner product on the vector spaces $\mathcal{K}'$ and $\mathcal{K}''$, and that the sets of spaces $\{\mathcal{E}', \mathcal{J}'\}$, $\{(\mathcal{V}^I)', (\mathcal{V}^O)', \mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_j\}$, $\{\mathcal{E}'', \mathcal{J}''\}$, $\{(\mathcal{V}^I)'', (\mathcal{V}^O)'', \mathcal{P}''_1, \mathcal{P}''_2, \ldots, \mathcal{P}''_k\}$ all contain mutually orthogonal spaces. For any two fields

$$\mathbf{P} = \mathbf{P}^I + \mathbf{P}^O + \mathbf{P}' + \mathbf{P}'', \quad \mathbf{Q} = \mathbf{Q}^I + \mathbf{Q}^O + \mathbf{Q}' + \mathbf{Q}'', \tag{7.103}$$

in the vector space $\mathcal{K}$, with

$$\mathbf{P}^I, \mathbf{Q}^I \in (\mathcal{V}^I)', \quad \mathbf{P}^O, \mathbf{Q}^O \in (\mathcal{V}^O)'', \quad \mathbf{P}', \mathbf{Q}' \in \mathcal{H}', \quad \mathbf{P}'', \mathbf{Q}'' \in \mathcal{H}'', \tag{7.104}$$

let us define the inner product of them to be

$$(\mathbf{P}, \mathbf{Q}) = (\mathbf{P}^I, \mathbf{Q}^I)' + (\mathbf{P}^O, \mathbf{Q}^O)'' + (\mathbf{P}', \mathbf{Q}')' + (\mathbf{P}'', \mathbf{Q}'')'', \tag{7.105}$$

in which $(\ ,\ )'$ and $(\ ,\ )''$ denote the inner product on the spaces $\mathcal{K}'$ and $\mathcal{K}''$ respectively. It is immediately clear from this definition that the subspaces $(\mathcal{V}^I)'$, $(\mathcal{V}^O)''$, $\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_j, \mathcal{P}''_1, \mathcal{P}''_2, \ldots, \mathcal{P}''_k$ are mutually orthogonal in the new superfunction. Now take a field $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$. By the definition of these subspaces there

must exist fields $\mathbf{E}' \in \mathcal{E}'$ and $\mathbf{E}'' \in \mathcal{E}''$ such that (7.81) to (7.83) hold, and fields $\mathbf{J}' \in \mathcal{J}'$, $\mathbf{J}'' \in \mathcal{J}''$ such that (7.84) to (7.86) hold. Consequently we have

$$\begin{aligned}
(\mathbf{J},\mathbf{E}) &= (\mathbf{J}'+\mathbf{J}''-(\mathbf{J}^O)'-(\mathbf{J}^I)'',\mathbf{E}'+\mathbf{E}''-(\mathbf{E}^O)'-(\mathbf{E}^I)'')\\
&= ((\mathbf{J}^O)',(\mathbf{E}^O)')'+((\mathbf{J}^I)'',(\mathbf{E}^I)'')''-\\
&\quad (\mathbf{J}',(\mathbf{E}^O)')'-(\mathbf{J}'',(\mathbf{E}^I)'')''-((\mathbf{J}^O)',\mathbf{E}')'-((\mathbf{J}^I)'',\mathbf{E}'')''\\
&= -((\mathbf{J}^O)',(\mathbf{E}^O)')'-((\mathbf{J}^I)'',(\mathbf{E}^I)'')''\\
&= -((\mathbf{J}^O)',(\mathbf{E}^O)')'-(\mathbf{M}^J(\mathbf{J}^O)',\mathbf{M}^E(\mathbf{E}^O)')''\\
&= -((\mathbf{J}^O)',(\mathbf{E}^O)')'-((\mathbf{M}^E)^\dagger\mathbf{M}^J(\mathbf{J}^O)',(\mathbf{E}^O)')', \qquad (7.106)
\end{aligned}$$

in which $(\mathbf{M}^E)^\dagger$ is the adjoint of $\mathbf{M}^E$. So we see that the spaces $\mathcal{J}$ and $\mathcal{E}$ will be orthogonal if we choose

$$(\mathbf{M}^E)^\dagger\mathbf{M}^J = -\mathbf{I}. \qquad (7.107)$$

Note that the orthogonality of the spaces $\mathcal{J}$ and $\mathcal{E}$ immediately implies that they have no nonzero vector in their intersection.

In the case of nonorthogonal subspace collections, we are free to choose the maps $\mathbf{M}^E$ and $\mathbf{M}^J$ that map $(\mathcal{V}^O)'$ to $(\mathcal{V}^I)''$, so long as they and the map $\mathbf{A}$ are nonsingular. However, in view of (7.107), it would be quite natural to restrict our definition of multiplication by requiring that $\mathbf{M}^J = -\mathbf{M}^E$, i.e., one can pick a nonsingular map $\mathbf{M}$ mapping $(\mathcal{V}^O)'$ to $(\mathcal{V}^I)''$ and set

$$\mathbf{M}^E = \mathbf{M}, \quad \mathbf{M}^J = -\mathbf{M}. \qquad (7.108)$$

With this choice, subtracting the equations in (7.95) gives

$$(\mathbf{\Pi}^I)''(\mathbf{\Gamma}_1''-\mathbf{\Gamma}_2'')\mathbf{M}(\mathbf{\Pi}^O)'(\mathbf{\Gamma}_1'-\mathbf{\Gamma}_2')\mathbf{P} = \mathbf{M}\mathbf{P}. \qquad (7.109)$$

Returning to the case where the subspaces are orthogonal, (7.107) is satisfied if $\mathbf{M}\mathbf{M}^\dagger = -\mathbf{I}$. An alternative way to see that $\mathcal{J}$ and $\mathcal{E}$ have no nonzero vector in their intersection is as follows. Choose an orthonormal basis $(\mathbf{v}_1^O)', (\mathbf{v}_2^O)', \ldots (\mathbf{v}_m^O)'$ for $(\mathcal{V}^O)'$ and take $\mathbf{M}^E = -\mathbf{M}^J$ as a map such that $\mathbf{M}^E(\mathbf{v}_1^O)', \mathbf{M}^E(\mathbf{v}_2^O)', \ldots \mathbf{M}^E(\mathbf{v}_m^O)'$ form an orthonormal basis for $(\mathcal{V}^I)''$. Then the operator $\mathbf{M}^E$ is represented as the identity matrix in the basis, and $\mathbf{M}^J$ is represented by $-\mathbf{I}$. Now, recalling the definition (2.37) of the norm $|\mathbf{Q}| = (\mathbf{Q},\mathbf{Q})^{1/2}$ of a vector $\mathbf{Q}$ recall that the action of the operators $(\mathbf{\Pi}^O)'$, $(\mathbf{\Pi}^I)''$ cannot increase the norm of a vector, while $\mathbf{\Gamma}_1' - \mathbf{\Gamma}_2'$ and $\mathbf{\Gamma}_1'' - \mathbf{\Gamma}_2''$ preserve the norm (as can be seen if we take a basis where these are diagonal). Hence (7.109) can be satisfied only when there is a $\mathbf{P} \in (\mathcal{V}^O)'$ such that

$$(\mathbf{\Gamma}_1' - \mathbf{\Gamma}_2')\mathbf{P} \in (\mathcal{V}^O)', \quad (\mathbf{\Gamma}_1'' - \mathbf{\Gamma}_2'')\mathbf{M}\mathbf{P} \in (\mathcal{V}^I)''. \qquad (7.110)$$

Then as $\mathbf{\Gamma}_1' + \mathbf{\Gamma}_2' = \mathbf{I}$ and $\mathbf{\Gamma}_1'' + \mathbf{\Gamma}_2'' = \mathbf{I}$ we obtain

$$(\mathbf{\Gamma}_1' + \mathbf{\Gamma}_2')\mathbf{P} \in (\mathcal{V}^O)', \quad (\mathbf{\Gamma}_1'' + \mathbf{\Gamma}_2'')\mathbf{M}\mathbf{P} \in (\mathcal{V}^I)''. \qquad (7.111)$$

Adding and subtracting (7.110) and (7.111) then implies

$$\mathbf{\Gamma}_1'\mathbf{P} \in (\mathcal{V}^O)', \quad \mathbf{\Gamma}_2'\mathbf{P} \in (\mathcal{V}^O)', \quad \mathbf{\Gamma}_1''\mathbf{P} \in (\mathcal{V}^I)'', \quad \mathbf{\Gamma}_2''\mathbf{P} \in (\mathcal{V}^I)'', \qquad (7.112)$$

which is excluded by our assumption that $\mathcal{V}'$ has no vector in common with $\mathcal{E}'$ or $\mathcal{J}'$ and that $\mathcal{V}''$ has no vector in common with $\mathcal{E}''$ or $\mathcal{J}''$.

7.6 Multiplicative identity superfunctions

Suppose we are given nonsingular maps $\mathbf{M}^E$ and $\mathbf{M}^J$ which map the m-dimensional space $(\mathcal{V}^O)'$ to the m-dimensional space $(\mathcal{V}^I)''$. Let $\mathcal{K}''$ denote the $2m$-dimensional space

$$\mathcal{K}'' = (\mathcal{V}^O)' \oplus (\mathcal{V}^I)''. \tag{7.113}$$

Within this space define $\mathcal{E}''$ as the subspace consisting of all vectors of the form $\mathbf{E} = \mathbf{v} + (\mathbf{M}^E)^{-1}\mathbf{v}$ with $\mathbf{v} \in (\mathcal{V}^I)''$ and define $\mathcal{J}''$ as the subspace consisting of all vectors of the form $\mathbf{J} = \mathbf{w} + (\mathbf{M}^J)^{-1}\mathbf{w}$ with $\mathbf{w} \in (\mathcal{V}^I)''$. If these subspaces have a vector in common then

$$\mathbf{v} + (\mathbf{M}^E)^{-1}\mathbf{v} = \mathbf{w} + (\mathbf{M}^J)^{-1}\mathbf{w}, \quad \text{i.e., } \mathbf{v} - \mathbf{w} = (\mathbf{M}^J)^{-1}\mathbf{w} - (\mathbf{M}^E)^{-1}\mathbf{v}. \tag{7.114}$$

In this last equation the fields on the left and on the right lie respectively in $(\mathcal{V}^I)''$ and $(\mathcal{V}^O)'$. As the intersection of these subspaces consists of only the zero vector, we conclude that both sides must be zero, i.e., $\mathbf{w} = \mathbf{v}$ and

$$\mathbf{u} \equiv= (\mathbf{M}^E)^{-1}\mathbf{v} = (\mathbf{M}^J)^{-1}\mathbf{v}. \tag{7.115}$$

Thus, $\mathbf{M}^E\mathbf{u} = \mathbf{v} = \mathbf{M}^J\mathbf{u}$ and if we assume that $\mathbf{M}^J - \mathbf{M}^E$ is nonsingular, then $0 = \mathbf{u} = \mathbf{v} = \mathbf{w}$. So under this assumption the subspaces have only the zero vector in their intersection. Then, since they each have dimension m we conclude that

$$\mathcal{K}'' = (\mathcal{V}^O)' \oplus (\mathcal{V}^I)'' = \mathcal{E}'' \oplus \mathcal{J}'', \tag{7.116}$$

which defines a superfunction $(F^s)''$ in which $\mathcal{H}$ is empty.

We now look at the associated superfunction problem. As the space $\mathcal{H}$ is empty, if we are given vectors $\mathbf{E}^I$ and $\mathbf{J}^I$ in the input space $(\mathcal{V}^I)''$, the superfunction problem then consists of finding vectors $\mathbf{E}^O$ and $\mathbf{J}^O$ in the output space $(\mathcal{V}^O)'$ such that

$$\mathbf{E}^I + \mathbf{E}^O \in \mathcal{E}'', \quad \mathbf{J}^I + \mathbf{J}^O \in \mathcal{J}''. \tag{7.117}$$

From our definition of the subspaces $\mathcal{E}''$ and $\mathcal{J}''$ we immediately see that the superfunction problem is solved with fields

$$\mathbf{E}^0 = (\mathbf{M}^E)^{-1}\mathbf{E}^I, \quad \mathbf{J}^0 = (\mathbf{M}^J)^{-1}\mathbf{J}^I, \tag{7.118}$$

implying, through (7.21), that the associated function is

$$\mathbf{F}'' = \begin{pmatrix} (\mathbf{M}^E)^{-1} & 0 \\ 0 & (\mathbf{M}^J)^{-1} \end{pmatrix}. \tag{7.119}$$

So if we take another superfunction $(F^s)'$ and multiply it by this superfunction $(F^s)''$, the product rule (7.100) implies that the resulting superfunction F^s has the associated function

$$\mathbf{F} = \mathbf{F}'. \tag{7.120}$$

We conclude that this superfunction $(F^s)''$ is the multiplicative identity, when multiplication is defined with the maps $\mathbf{M}^E$ and $\mathbf{M}^J$.

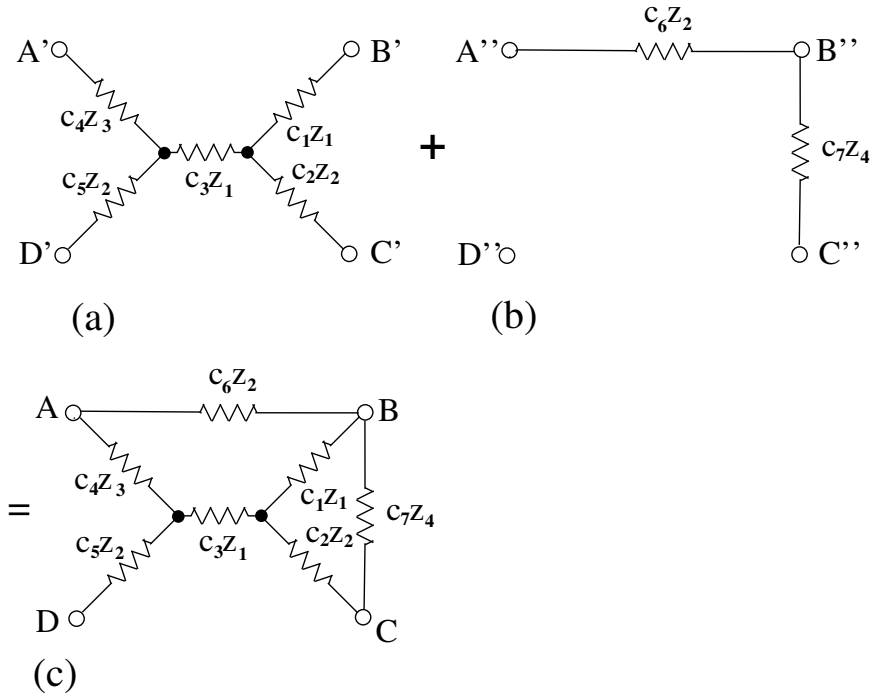


Figure 7.6: Adding Y-subspace collections is like hooking networks together in parallel. The 4 terminal networks in (a) and (b), each representing (along with their respective tree-like battery configurations on the opposite side of the circuit board that are not shown here) $Y(3)$ and $Y(2)$ subspace collections, are added together to form the 4 terminal network in (c) which is a $Y(4)$ subspace collection. Note that the circuit in (b) is really only a 3 terminal network, so it has been embedded in a 4 terminal network (with no electrical connections to the 4th terminal). Also note there is some flexibility in how one adds together these subspace collections: we connected the terminals A', B', C', and D', to respectively the terminals A'', B'', C'', and D'', but we could have connected them to any permutation of these terminals. This flexibility is reflected in the need to introduce nonsingular operators $\mathbf{S}'$ and $\mathbf{S}''$ which respectively map $\mathcal{V}'$ and $\mathcal{V}''$ to $\mathcal{V}$, before addition can defined.

7.7 Addition of Y-subspace collections and embeddings

Adding superfunctions is similar the way electrical circuits, each with n terminals can be combined. An example is shown in **Figure 7.6**.

Suppose we have $Y(j)$ and $Y(k)$ subspace collections:

$$\begin{aligned} \mathcal{K}' &= \mathcal{E}' \oplus \mathcal{J}' = \mathcal{V}' \oplus \mathcal{H}' \quad \text{with } \mathcal{H}' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_j, \\ \mathcal{K}'' &= \mathcal{E}'' \oplus \mathcal{J}'' = \mathcal{V}'' \oplus \mathcal{H}'' \quad \text{with } \mathcal{H}'' = \mathcal{P}''_1 \oplus \mathcal{P}''_2 \oplus \cdots \oplus \mathcal{P}''_k, \end{aligned} \tag{7.121}$$

where the spaces $\mathcal{V}'$ and $\mathcal{V}''$ have the same dimension n. To define the sum of the subspace collections we need to introduce another n-dimensional space $\mathcal{V}$ and nonsingular operators $\mathbf{S}'$ and $\mathbf{S}''$ which respectively

map $\mathcal{V}'$ and $\mathcal{V}''$ to $\mathcal{V}$. Then the sum of the subspace collections

$$\mathcal{K} = \mathcal{K}' +_{\{\mathbf{S}',\mathbf{S}''\}} \mathcal{K}'' \tag{7.122}$$

is the subspace collection

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{H}, \tag{7.123}$$

where

$$\mathcal{H} = \mathcal{H}' \oplus \mathcal{H}'' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_j \oplus \mathcal{P}''_1 \oplus \mathcal{P}''_2 \oplus \cdots \oplus \mathcal{P}''_k. \tag{7.124}$$

Here a field $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$, with $\mathbf{E}_1 \in \mathcal{V}$ and $\mathbf{E}_2 \in \mathcal{H}$, is in $\mathcal{E}$ if and only if there exist fields

$$\mathbf{E}' = \mathbf{E}'_1 + \mathbf{E}'_2 \in \mathcal{E}', \quad \mathbf{E}'' = \mathbf{E}''_1 + \mathbf{E}''_2 \in \mathcal{E}'', \tag{7.125}$$

with

$$\mathbf{E}'_1 \in \mathcal{V}', \ \ \mathbf{E}'_2 \in \mathcal{H}', \ \ \mathbf{E}''_1 \in \mathcal{V}'', \ \ \mathbf{E}''_2 \in \mathcal{H}'', \tag{7.126}$$

such that

$$\mathbf{S}'\mathbf{E}'_1 = \mathbf{S}''\mathbf{E}''_1 = \mathbf{E}_1. \tag{7.127}$$

Also a field $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$, with $\mathbf{J}_1 \in \mathcal{V}$ and $\mathbf{J}_2 \in \mathcal{H}$, is in $\mathcal{J}$ if and only if there exist fields

$$\mathbf{J}' = \mathbf{J}'_1 + \mathbf{J}'_2 \in \mathcal{E}', \quad \mathbf{J}'' = \mathbf{J}''_1 + \mathbf{J}''_2 \in \mathcal{E}'', \tag{7.128}$$

with

$$\mathbf{J}'_1 \in \mathcal{V}', \ \ \mathbf{J}'_2 \in \mathcal{H}', \ \ \mathbf{J}''_1 \in \mathcal{V}'', \ \ \mathbf{J}''_2 \in \mathcal{H}'', \tag{7.129}$$

such that

$$\mathbf{S}'\mathbf{J}'_1 + \mathbf{S}''\mathbf{J}''_1 = \mathbf{J}_1. \tag{7.130}$$

So given $\mathbf{E}_1 \in \mathcal{V}$, we let $\mathbf{E}'_1 = (\mathbf{S}')^{-1}\mathbf{E}$ and $\mathbf{E}''_1 = (\mathbf{S}'')^{-1}\mathbf{E}_1$, and we solve the Y-problem in each of the two subspace collections $Y(j)$ and $Y(k)$, finding fields satisfying (7.125), (7.126), (7.128), and (7.129) with

$$\mathbf{J}'_2 = \mathbf{L}'\mathbf{E}_2, \quad \mathbf{J}''_2 = \mathbf{L}''\mathbf{E}''_2, \tag{7.131}$$

where

$$\mathbf{L}' = \sum_{i=1}^{j} z'_i \mathbf{\Lambda}'_i, \quad \mathbf{L}'' = \sum_{i=1}^{k} z''_i \mathbf{\Lambda}''_i, \tag{7.132}$$

and $\mathbf{\Lambda}'_i$ projects onto $\mathcal{P}'_i$ while $\mathbf{\Lambda}''_i$ projects onto $\mathcal{P}''_i$. Hence we have

$$\mathbf{J}_2 = \mathbf{J}'_2 + \mathbf{J}''_2 = \mathbf{L}(\mathbf{E}'_2 + \mathbf{E}''_2), \quad \text{with } \mathbf{L} = \mathbf{L}' + \mathbf{L}''. \tag{7.133}$$

Then (7.130) implies

$$\mathbf{J}_1 = \mathbf{S}'\mathbf{J}'_1 + \mathbf{S}''\mathbf{J}''_1 = \mathbf{S}'\mathbf{Y}'\mathbf{E}'_1 + \mathbf{S}''\mathbf{Y}''\mathbf{E}''_1 = \mathbf{Y}\mathbf{E}_1, \tag{7.134}$$

where

$$\mathbf{Y} = \mathbf{S}'\mathbf{Y}'(\mathbf{S}')^{-1} + \mathbf{S}''\mathbf{Y}''(\mathbf{S}'')^{-1}. \tag{7.135}$$

If we have a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ for $\mathcal{V}$, then it is natural to take $(\mathbf{S}')^{-1}\mathbf{v}_1$, $(\mathbf{S}')^{-1}\mathbf{v}_2$, …, $(\mathbf{S}')^{-1}\mathbf{v}_n$ as a basis for $\mathcal{V}'$, and to take $(\mathbf{S}'')^{-1}\mathbf{v}_1$, $(\mathbf{S}'')^{-1}\mathbf{v}_2$, …, $(\mathbf{S}'')^{-1}\mathbf{v}_n$ as a basis for $\mathcal{V}''$. Then the operators $\mathbf{S}'$ and $\mathbf{S}''$ are represented by identity matrices, and in these bases (7.135) becomes $\mathbf{Y} = \mathbf{Y}' + \mathbf{Y}''$.

In the case where either or both of the subspaces $\mathcal{V}'$ and $\mathcal{V}''$ have dimension less than the dimension n of the subspace $\mathcal{V}$ we can first do an embedding. For example suppose $\mathcal{V}'$ has dimension $n' < n$. Then let $\mathcal{W}'$ be a space of dimension $n - n'$. Construct the subspace collection

$$\widetilde{\mathcal{K}}' = \widetilde{\mathcal{E}}' \oplus \mathcal{J}' = \widetilde{\mathcal{V}}' \oplus \mathcal{H}', \tag{7.136}$$

where

$$\widetilde{\mathcal{V}}' = \mathcal{V}' \oplus \mathcal{W}', \quad \widetilde{\mathcal{E}}' = \mathcal{E}' \oplus \mathcal{W}'. \tag{7.137}$$

Then given a field $\widetilde{\mathbf{E}}'_1 \in \widetilde{\mathcal{V}}'$ we can express it as a sum $\mathbf{E}'_1 + \mathbf{W}'$ with $\mathbf{E}'_1 \in \mathcal{V}'$ and $\mathbf{W}' \in \mathcal{W}'$. We write $\mathbf{E}'_1 = \Psi\widetilde{\mathbf{E}}'_1$ where Ψ is the projection onto $\mathcal{V}'$. Given this $\mathbf{E}'_1$ and solving the Y-problem associated with $\mathcal{K}'$ we obtain fields $\mathbf{E}'$ and $\mathbf{J}'$ satisfying

$$\begin{aligned} \mathbf{E}' &= \mathbf{E}'_1 + \mathbf{E}'_2 \in \mathcal{E}', \quad \mathbf{E}'_1 \in \mathcal{V}', \ \mathbf{E}'_2 \in \mathcal{H}', \\ \mathbf{J}' &= \mathbf{J}'_1 + \mathbf{J}'_2 \in \mathcal{J}', \quad \mathbf{J}'_1 \in \mathcal{V}', \ \mathbf{J}'_2 = \mathbf{L}\mathbf{E}_2 \in \mathcal{H}. \end{aligned} \tag{7.138}$$

It follows that the Y-problem in the space $\widetilde{\mathcal{K}}'$ is solved with fields

$$\widetilde{\mathbf{E}}' = \mathbf{W} + \mathbf{E}' = \mathbf{W} + \mathbf{E}'_1 + \mathbf{E}'_2, \quad \text{and} \quad \mathbf{J}' = \mathbf{J}'_1 + \mathbf{J}'_2 \ \text{ with } \ \mathbf{J}'_2 = \mathbf{L}\mathbf{E}_2, \tag{7.139}$$

implying that

$$\mathbf{J}'_1 = -\mathbf{Y}\mathbf{E}'_1 = -\mathbf{Y}\Psi\widetilde{\mathbf{E}}'_1. \tag{7.140}$$

We conclude that the new Y-problem has an operator $\widetilde{\mathbf{Y}} = \mathbf{Y}\Psi$, i.e., its range is not the whole space $\widetilde{\mathcal{V}}'$ but only at most the subspace $\mathcal{V}'$. After making such embeddings to ensure that $\mathcal{V}'$ and $\mathcal{V}''$ (or rather $\widetilde{\mathcal{V}}'$ and $\widetilde{\mathcal{V}}''$ have the same dimension as the dimension n of the subspace $\mathcal{V}$, we are then free to add them.

The additive zero is easy to find. Let us consider the degenerate subspace collection

$$\mathcal{K}'' = \mathcal{E}'' = \mathcal{V}''. \tag{7.141}$$

Clearly $\mathcal{H}''$ contains only the zero vector, and we are forced to choose $\mathbf{L}'' = 0$. Given $\mathbf{E}_1 \in \mathcal{V}''$. The Y-problem is solved with vectors

$$\mathbf{E}'' = \mathbf{E}_1, \quad \mathbf{E}_1 = \mathbf{J}_1 = \mathbf{J}_2 = \mathbf{J} = 0. \tag{7.142}$$

Implying the associated Y-operator $\mathbf{Y}$ is zero. Thus the subspace collection (7.141) is the additive zero. Note that this subspace collection does not satisfy the property $\mathcal{E}'' \cap \mathcal{V}'' = \{\mathbf{0}\}$ which is needed for the inverse of $\mathbf{Y}$ to exist, which is not surprising since $\mathbf{Y} = 0$ has no inverse.

Now suppose we have a subspace collection

$$\mathcal{K}' = \mathcal{E}' \oplus \mathcal{J}' = \mathcal{V}' \oplus \mathcal{H}' \quad \text{with } \ \mathcal{H}' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_j, \tag{7.143}$$

with associated operator $\mathbf{Y}'(z'_1, z'_2, \ldots, z'_n)$ when

$$\mathbf{L}' = \sum_{i=1}^{j} z'_i \mathbf{\Lambda}'_i. \tag{7.144}$$

It is clear that if we replace $\mathbf{L}'$ by

$$\mathbf{L}' = -\sum_{i=1}^{j} z_i' \mathbf{\Lambda}_i', \tag{7.145}$$

then the solution to the Y-problem will give the Y-operator

$$\mathbf{Y}(-z_1', -z_2', \ldots, -z_n') = -\mathbf{Y}(z_1', z_2', \ldots, z_n'), \tag{7.146}$$

where to obtain this last identity we have used the homogeneity of the function. Since adding (7.146) to the associated operator $\mathbf{Y}'(z_1', z_2', \ldots, z_n')$ we started with gives zero, it is tempting to conclude that we have found the additive inverse. However the function (7.146) is not the Y-operator-valued function of $z_1', z_2', \ldots, z_n'$ associated with the subspace collection (7.143), whose definition does not allow us to choose $\mathbf{L}'$ of the form (7.145). This is made more clear in the case where we have an orthogonal subspace collection since then the imaginary part of $(\mathbf{V}, \mathbf{Y}(z_1', z_2', \ldots, z_n')\mathbf{V})$ is generally positive when $z_1', z_2', \ldots, z_n'$ all have positive imaginary parts, and $-\mathbf{Y}(z_1', z_2', \ldots, z_n')$ then does not share this Herglotz property. So the additive inverse of an orthogonal subspace collection should typically not be an orthogonal subspace collection. We will find the proper additive inverse in section 7.12.

7.8 Substitution of subspace collections

Another familiar operation that we can do with rational functions is to make substitutions. Substitution of one subspace collection in another is similar to the way it can be done in electrical circuits. An example is shown in Figure 7.7. Thus if $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ is a $m \times m$ matrix-valued homogeneous function of degree one and $Z'(z_1', z_2', \ldots, z_p')$ is a scalar-valued function, say normalized with

$$Z'(1, 1, \ldots, 1) = 1, \tag{7.147}$$

then

$$\mathbf{Y}''(z_1', z_2', \ldots, z_p', z_2, \ldots, z_n) = \mathbf{Y}(Z(z_1', z_2', \ldots, z_p'), z_2, \ldots, z_n) \tag{7.148}$$

will be another $m \times m$ matrix-valued homogeneous function of degree one. What is the analogous operation on subspace collections? It is natural to expect there should be one, just as in a network of n types of resistors one can replace each resistor of type 1 with a network of p other resistors.

Extending the treatment given in Section 29.1 of Milton (2002) let us suppose that we are given a $Y(n)$-subspace collection

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.149}$$

and a $(3, p)$-subspace collection

$$\mathcal{H}' = \mathcal{U}' \oplus \mathcal{E}' \oplus \mathcal{J}' = \mathcal{P}_1' \oplus \mathcal{P}_2' \oplus \cdots \oplus \mathcal{P}_p', \tag{7.150}$$

in which $\mathcal{V}$ is m-dimensional and $\mathcal{U}'$ is one-dimensional. Let $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ and $Z'(z_1', z_2', \ldots, z_p')$ denote the functions associated with these subspace collections. We take as our new $(2, n+p)$-subspace collection,

$$\mathcal{K}'' = \mathcal{E}'' \oplus \mathcal{J}'' = \mathcal{V}'' \oplus \mathcal{P}_1'' \oplus \mathcal{P}_2'' \oplus \cdots \oplus \mathcal{P}_n'', \tag{7.151}$$

where

$$\mathcal{E}'' = (\mathcal{E} \otimes \mathcal{U}') \oplus (\mathcal{P}_1 \otimes \mathcal{E}'), \quad \mathcal{J}'' = (\mathcal{J} \otimes \mathcal{U}') \oplus (\mathcal{P}_1 \otimes \mathcal{J}'), \tag{7.152}$$

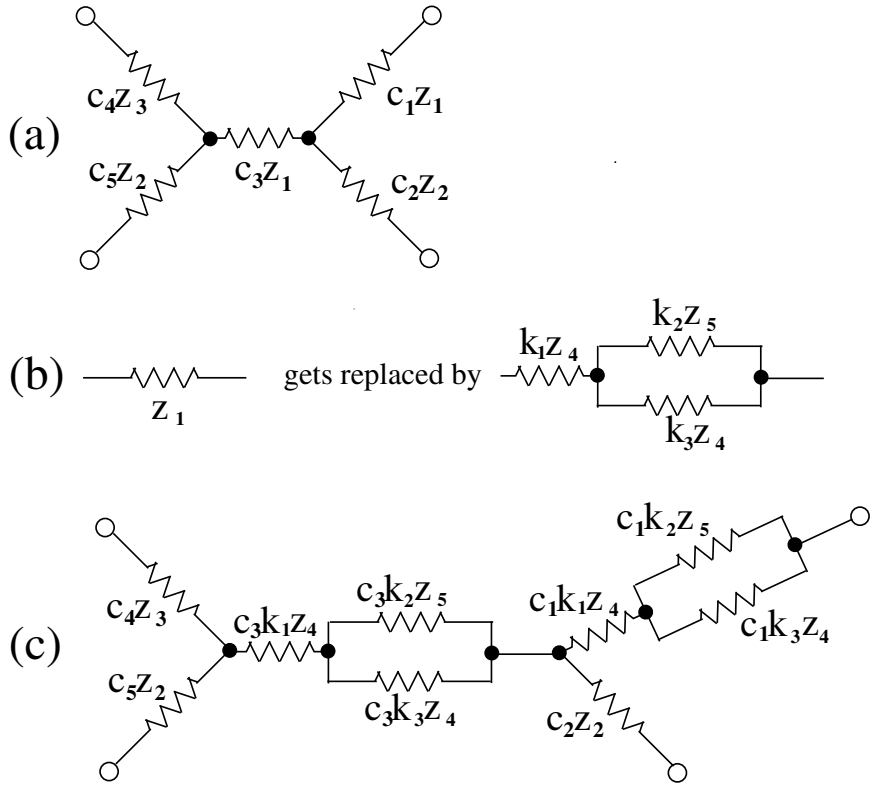


Figure 7.7: Substitution of Y- and Z-subspace collections is like replacing all resistors of one type by a compound network. If one takes a subspace collection, as, for example, represented by the 4-terminal network in (a) and replaces z_1 by the network in (b), where $k_1 + (1/k_2 + 1/k_3)^{-1} = 1$, to ensure this replacement does effect the resistance when $z_1 = z_4 = z_5 = 1$, one obtains the subspace collection as represented by the 4-terminal network in (c).

and

$$\begin{aligned} \mathcal{V}'' &= \mathcal{V} \otimes \mathcal{U}', \\ \mathcal{P}''_i &= \mathcal{P}_1 \otimes \mathcal{P}'_i \ \text{ for } \ 1 \le i \le p, \\ &= \mathcal{P}_{i+1-p} \otimes \mathcal{U}' \ \text{ for } \ p+1 \le i \le n+p-1, \end{aligned} \tag{7.153}$$

in which $\otimes$ denotes the operation of taking the tensor product of two subspaces. Vectors in the space

$$\mathcal{K}'' = \mathcal{E}'' \oplus \mathcal{J}'' = (\mathcal{K} \otimes \mathcal{U}') \oplus (\mathcal{P}_1 \otimes (\mathcal{E}' \oplus \mathcal{J}')), \tag{7.154}$$

spanned by these subspaces are represented as a pair $[\mathbf{P},\ \mathbf{u}']$ added to a linear combination of pairs of the form $[\mathbf{P}_1,\ \mathbf{P}']$, where $\mathbf{P} \in \mathcal{K}$, $\mathbf{u}' \in \mathcal{U}'$, $\mathbf{P}_1 \in \mathcal{P}_1$, and $\mathbf{P}' \in \mathcal{E}' \oplus \mathcal{J}'$.

Now define

$$\begin{aligned} \mathcal{H} &= \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \\ \mathcal{H}'' &= \mathcal{P}''_1 \oplus \mathcal{P}''_2 \oplus \cdots \oplus \mathcal{P}''_n, \end{aligned} \tag{7.155}$$

and suppose that we are given solutions to the equations

$$\begin{aligned} \mathbf{J}_2 &= \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i \mathbf{E}_2 \ \text{ with } \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}, \ \ \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}, \quad \mathbf{E}_1, \mathbf{J}_1 \in \mathcal{V}, \quad \mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}, \\ \mathbf{j}' + \mathbf{J}' &= \sum_{j=1}^{n} z'_j \mathbf{\Lambda}'_j (\mathbf{e}' + \mathbf{E}') \ \text{ with } \mathbf{e}', \mathbf{j}' \in \mathcal{U}', \ \ \mathbf{E}' \in \mathcal{E}', \ \ \mathbf{J}' \in \mathcal{J}', \end{aligned} \tag{7.156}$$

where

$$z_1 = Z(z'_1, z'_2, \ldots, z'_p), \tag{7.157}$$

while $\mathbf{\Lambda}_i$ and $\mathbf{\Lambda}'_j$ are the projections onto $\mathcal{P}_i$ and $\mathcal{P}'_j$. Let us introduce

$$\mathbf{P}_i = \mathbf{\Lambda}_i \mathbf{E}_2, \quad \mathbf{P}'_j = \mathbf{\Lambda}'_j (\mathbf{e}' + \mathbf{E}'), \tag{7.158}$$

and set

$$\begin{aligned} z''_i &= z'_i \ \text{ for } \ 1 \le i \le p, \\ &= z_{i+1-p} \ \text{ for } \ p+1 \le i \le n+p-1. \end{aligned} \tag{7.159}$$

Then, in the new subspace collection, the vectors

$$\begin{aligned} \mathbf{E}''_1 &= [\mathbf{E}_1, \ \mathbf{e}'] \in \mathcal{V}'', \quad \mathbf{E}''_2 = [\mathbf{E}_2, \ \mathbf{e}'] + [\mathbf{P}_1, \ \mathbf{E}'], \\ \mathbf{J}''_1 &= [\mathbf{J}_1, \ \mathbf{e}'] \in \mathcal{V}'', \quad \mathbf{J}''_2 = [\mathbf{J}_2, \ \mathbf{e}'] + [\mathbf{P}_1, \ \mathbf{J}'] \end{aligned} \tag{7.160}$$

satisfy

$$\mathbf{E}''_1 + \mathbf{E}''_2 \in \mathcal{E}'', \quad \mathbf{J}''_1 + \mathbf{J}''_2 \in \mathcal{J}''. \tag{7.161}$$

Additionally, we have

$$\mathbf{E}''_2 = [\sum_{i=1}^{n} \mathbf{P}_i, \ \mathbf{e}'] - [\mathbf{P}_1, \ \mathbf{e}'] + [\mathbf{P}_1, \ \sum_{i=1}^{p} \mathbf{P}'_i] = \sum_{i=1}^{n+p-1} \mathbf{P}''_i \in \mathcal{H}'', \tag{7.162}$$

where

$$\begin{aligned} \mathbf{P}''_i &= [\mathbf{P}_1, \ \mathbf{P}'_i] \ \text{ for } \ 1 \le i \le p, \\ &= [\mathbf{P}_{i+1-p}, \ \mathbf{e}'] \ \text{ for } \ p+1 \le i \le n+p-1 \end{aligned} \tag{7.163}$$

satisfies $\mathbf{P}''_i \in \mathcal{P}''_i$. Similarly, and using the fact implied by (7.157) that $\mathbf{j}' = Z\mathbf{e}' = z_1 \mathbf{e}'$, we have

$$\mathbf{J}''_2 = [\sum_{i=1}^{n} z_i \mathbf{P}_i, \ \mathbf{e}'] - [\mathbf{P}_1, \ \mathbf{j}'] + [\mathbf{P}_1, \sum_{i=1}^{p} z'_i \mathbf{P}'_i] = \sum_{i=1}^{n+p-1} z''_i \mathbf{P}''_i \in \mathcal{H}''. \tag{7.164}$$

Given a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ for $\mathcal{V}$ and a vector $\mathbf{u}'$ in $\mathcal{U}'$ it is natural to take $(\mathbf{v}_1, \mathbf{u}')$,$(\mathbf{v}_2, \mathbf{u}')$, …, $(\mathbf{v}_m, \mathbf{u}')$ as our basis for $\mathcal{V}''$. Choosing $\mathbf{e}'$ so that $\mathbf{e}' = \mathbf{u}'$, it is evident that $\mathbf{Y}(Z'(z'_1, z'_2, \ldots, z'_p), z_2, \ldots, z_n)$ is the matrix-valued function associated the new subspace collection, represented in these bases.

There is a similar subspace operation corresponding to substituting the Z-function $Z'(z'_1, z'_2, \ldots, z'_p)$ into another Z-function $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ to obtain

$$\mathbf{Z}''(z'_1, z'_2, \ldots, z'_p, z_2, \ldots, z_n) = \mathbf{Z}(Z(z'_1, z'_2, \ldots, z'_p), z_2, \ldots, z_n). \tag{7.165}$$

Given a $Z(n)$-subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.166}$$

and a $(3, p)$-subspace collection

$$\mathcal{H}' = \mathcal{U}' \oplus \mathcal{E}' \oplus \mathcal{J}' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_p, \tag{7.167}$$

in which $\mathcal{U}$ is m-dimensional and $\mathcal{U}'$ is one-dimensional, we take as our new $(3, n+p-1)$-subspace collection,

$$\mathcal{K}'' = \mathcal{U}'' \oplus \mathcal{E}'' \oplus \mathcal{J}'' = \mathcal{P}''_1 \oplus \mathcal{P}''_2 \oplus \cdots \oplus \mathcal{P}''_n, \tag{7.168}$$

where

$$\mathcal{U}'' = \mathcal{U} \otimes \mathcal{U}', \quad \mathcal{E}'' = (\mathcal{E} \otimes \mathcal{U}') \oplus (\mathcal{P}_1 \otimes \mathcal{E}'), \quad \mathcal{J}'' = (\mathcal{J} \otimes \mathcal{U}') \oplus (\mathcal{P}_1 \otimes \mathcal{J}'), \tag{7.169}$$

and

$$\begin{aligned} \mathcal{P}''_i &= \mathcal{P}_1 \otimes \mathcal{P}'_i \ \text{ for } \ 1 \le i \le p, \\ &= \mathcal{P}_{i+1-p} \otimes \mathcal{U}' \ \text{ for } \ p+1 \le i \le n+p-1. \end{aligned} \tag{7.170}$$

Suppose that we are given solutions to the equations

$$\begin{aligned} \mathbf{j} + \mathbf{J} &= \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i(\mathbf{e} + \mathbf{E}) \ \text{ with } \mathbf{e}, \mathbf{j} \in \mathcal{U}, \ \ \mathbf{E} \in \mathcal{E}, \ \ \mathbf{J} \in \mathcal{J}, \\ \mathbf{j}' + \mathbf{J}' &= \sum_{j=1}^{n} z'_j \mathbf{\Lambda}'_j(\mathbf{e}' + \mathbf{E}') \ \text{ with } \mathbf{e}', \mathbf{j}' \in \mathcal{U}', \ \ \mathbf{E}' \in \mathcal{E}', \ \ \mathbf{J}' \in \mathcal{J}', \end{aligned} \tag{7.171}$$

where $z_1 = Z(z'_1, z'_2, \ldots, z'_p)$, while $\mathbf{\Lambda}_i$ and $\mathbf{\Lambda}'_j$ are the projections onto $\mathcal{P}_i$ and $\mathcal{P}'_j$. Let us introduce

$$\mathbf{P}_i = \mathbf{\Lambda}_i(\mathbf{e} + \mathbf{E}), \quad \mathbf{P}'_j = \mathbf{\Lambda}'_j(\mathbf{e}' + \mathbf{E}'),$$

and define z''_i by (7.159), and $\mathbf{P}''_i \in \mathcal{P}''_i$ by (7.163). In the new subspace collection, the vectors

$$\begin{aligned} \mathbf{e}'' &= [\mathbf{e},\ \mathbf{e}'] \in \mathcal{U}'', \quad \mathbf{E}'' = [\mathbf{E},\ \mathbf{e}'] + [\mathbf{P}_1,\ \mathbf{E}'] \in \mathcal{E}'', \\ \mathbf{j}'' &= [\mathbf{j},\ \mathbf{e}'] \in \mathcal{U}'', \quad \mathbf{J}'' = [\mathbf{J},\ \mathbf{e}'] + [\mathbf{P}_1,\ \mathbf{J}'] \in \mathcal{J}'' \end{aligned} \tag{7.172}$$

satisfy

$$\begin{aligned}\mathbf{e}'' + \mathbf{E}'' &= [\sum_{i=1}^{n} \mathbf{P}_i,\ \mathbf{e}'] + [\mathbf{P}_1,\ \sum_{j=1}^{p} \mathbf{P}'_j] - [\mathbf{P}_1,\ \mathbf{e}'] \\ &= [\sum_{i=2}^{n} \mathbf{P}_i,\ \mathbf{e}'] + [\mathbf{P}_1,\ \sum_{j=1}^{p} \mathbf{P}'_j] \\ &= \sum_{i=1}^{n+p-1} \mathbf{P}''_i, \end{aligned} \tag{7.173}$$

and using (7.157) we get

$$\begin{aligned}\mathbf{j}'' + \mathbf{J}'' &= [\sum_{i=1}^{n} z_i\mathbf{P}_i,\ \mathbf{e}'] + [\mathbf{P}_1,\ \sum_{j=1}^{p} z'_i\mathbf{P}'_j] - [\mathbf{P}_1,\ \mathbf{j}'] \\ &= [\sum_{i=2}^{n} z_i\mathbf{P}_i,\ \mathbf{e}'] + [\mathbf{P}_1,\ \sum_{j=1}^{p} z'_i\mathbf{P}'_j] \\ &= \sum_{i=1}^{n+p-1} z''_i\mathbf{P}''_i. \end{aligned} \tag{7.174}$$

Given a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ for $\mathcal{U}$ and a vector $\mathbf{u}'$ in $\mathcal{U}'$ it is natural to take $(\mathbf{u}_1, \mathbf{u}')$,$(\mathbf{u}_2, \mathbf{u}')$, …, $(\mathbf{u}_m, \mathbf{u}')$ as our basis for $\mathcal{U}''$. Choosing $\mathbf{e}'$ so that $\mathbf{e}' = \mathbf{u}'$, it is evident from (7.172) that $\mathbf{Z}(Z'(z'_1, z'_2, \ldots, z'_p), z_2, \ldots, z_n)$ is the matrix-valued function associated the new subspace collection, represented in these bases.

7.9 Some other elementary operations on subspace collections

A further operation we can do on functions $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ while retaining the homogeneity of degree 1 in the variables $z_1, z_2, \ldots, z_n$ is to replace the function by $[\mathbf{Y}(1/z_1, 1/z_2, \ldots, 1/z_n)]^{-1}$. The analogous operation on the associated $Y(n)$-subspace collection is to interchange the subspaces $\mathcal{E}$ and $\mathcal{J}$. Similarly in a $Z(n)$ subspace collection, interchanging the subspaces $\mathcal{E}$ and $\mathcal{J}$ corresponds to replacing $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ by $[\mathbf{Z}(1/z_1, 1/z_2, \ldots, 1/z_n)]^{-1}$, as noted in Section 29.1 of Milton (2002). We call such a transformation a duality transformation. As a consequence of the duality transformation (7.60) immediately implies the formula

$$\mathbf{Z}^{-1} = \mathbf{\Gamma}_0[(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1)\mathbf{L}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1)]^{-1}\mathbf{\Gamma}_0. \tag{7.175}$$

One simple thing we can do in a function is set $z_j = z_k$: the analogous operation in a subspace collection is to replace $\mathcal{P}_j \oplus \mathcal{P}_k$ by a single subspace.

Another elementary operation we can do on a $Z(n)$ subspace collection is as follows. Let $\mathcal{U}$ be expressed as the direct sum

$$\mathcal{U} = \mathcal{U}' \oplus \mathcal{W}, \tag{7.176}$$

which defines the projection $\mathbf{\Phi}$ onto $\mathcal{U}'$. We now take as our subspace collection

$$\mathcal{H} = \mathcal{U}' \oplus \mathcal{E} \oplus \mathcal{J}' = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.177}$$

where

$$\mathcal{J}' = \mathcal{J} \oplus \mathcal{W}. \tag{7.178}$$

Then any solution to the Z-problem (7.27) with $\mathbf{e} \in \mathcal{U}'$ immediately generates a solution to the Z-problem associated with the subspace collection (7.177):

$$\mathbf{j}' \in \mathcal{U}', \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J}' \in \mathcal{J}', \quad \mathbf{j}' + \mathbf{J}' = \mathbf{L}(\mathbf{e} + \mathbf{E}), \tag{7.179}$$

where

$$\mathbf{L} = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i, \quad \mathbf{j}' = \mathbf{\Phi}\mathbf{j}, \quad \mathbf{J}' = \mathbf{J} + (\mathbf{I} - \mathbf{\Phi})\mathbf{j}, \tag{7.180}$$

which ensures that

$$\mathbf{j} + \mathbf{J} = \mathbf{j}' + \mathbf{J}' \quad \text{and} \quad (\mathbf{I} - \mathbf{\Phi})\mathbf{j} \in \mathcal{W}. \tag{7.181}$$

Hence the new subspace collection has a Z-operator

$$\mathbf{Z}' = \mathbf{\Phi}\mathbf{Z}, \tag{7.182}$$

when applied to fields in $\mathcal{U}'$.

7.10 Realizing any Y-matrix with elements that are rational functions of degree 1

Given any homogeneous rational function of degree 1,

$$Z(z_1, z_2, \ldots, z_n) = \frac{p(z_1, z_2, \ldots, z_n)}{q(z_1, z_2, \ldots, z_n)}, \tag{7.183}$$

satisfying the normalization $Z(1, 1, \ldots, 1) = 1$ where $p(z_1, z_2, \ldots, z_n)$ and $q(z_1, z_2, \ldots, z_n)$ are homogeneous polynomials of degree k and $k - 1$ respectively, where k is a positive integer, our first goal is to find a $Z(n)$ subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.184}$$

where $\mathcal{U}$ is one-dimensional which has $Z(z_1, z_2, \ldots, z_n)$ as its associated function. Without loss of generality we could set z_n =1, and then $p(z_1, z_2, \ldots, z_{n-1}, 1)$ and $q(z_1, z_2, \ldots, z_{n-1}, 1)$ are just arbitrary polynomials of the $n - 1$ variables $z_1, z_2, \ldots, z_{n-1}$. Also without loss of generality we can assume

$$p(1, 1, \ldots, 1) = q(1, 1, \ldots, 1) = 1. \tag{7.185}$$

The first step is to realize $Z(z_1, z_2, 1) = z_1 z_2$ as an associated Z-function. Note that (7.56) implies we can realize

$$Z(z_1, 1) = z_1^2, \tag{7.186}$$

and (7.57) implies we can realize

$$Z(z_1, z_2) = cz_1 + (1 - c)z_2, \tag{7.187}$$

for any constant c. Hence, by substitution we can realize

$$Z(z_1, z_2, 1) = 9(2z_1/3 + z_2/3)^2/8 - (2z_1 - z_2)^2/8 = z_1 z_2. \tag{7.188}$$

Making further substitutions, we can realize any product of the variables

$$Z(z_1, z_2, \ldots, z_{n-1}, 1) = z_1^{a_1} z_2^{a_2} \ldots z_{n-1}^{a_{n-1}}, \tag{7.189}$$

where the a_i are nonnegative integers. By repeated substitution in (7.187) we can realize any linear combination of such terms with coefficients summing to 1. Thus we can realize the polynomials $p(z_1, z_2, \ldots, z_{n-1}, 1)$ and $q(z_1, z_2, \ldots, z_{n-1}, 1)$.

Furthermore (7.56), with the roles of z_1 and z_2 interchanged, implies we can realize

$$Z(z_1, 1) = 1/z_1, \tag{7.190}$$

which by substitution into (7.188) implies we can realize

$$Z(z_1, z_2, 1) = z_2/z_1. \tag{7.191}$$

Substituting $p(z_1, z_2, \ldots, z_{n-1}, 1)$ for z_2 and $q(z_1, z_2, \ldots, z_{n-1}, 1)$ for z_1 we see we can find a subspace collection which realizes

$$Z(z_1, z_2, \ldots, z_{n-1}, 1) = \frac{p(z_1, z_2, \ldots, z_{n-1}, 1)}{q(z_1, z_2, \ldots, z_{n-1}, 1)} \tag{7.192}$$

as its associated Z-function when $z_n = 1$. When z_n is not 1, the subspace collection will by homogeneity realize the function (7.183).

Now from (7.44) we can realize

$$\mathbf{Y}(z_1) = \begin{pmatrix} a_{11} z_1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.193}$$

and realize

$$\mathbf{Y}(z_2) = \begin{pmatrix} 0 & a_{12} z_2 \\ 0 & 0 \end{pmatrix}. \tag{7.194}$$

By substitution of subspace collections, we can realize any Y-matrix where in the above formulae z_1 and z_2 are replaced by any normalized rational homogeneous functions of degree 1 (normalized in the sense that they take the value 1 when all variables take the value 1). Finally, by making suitable embeddings and adding subspace collections we can realize any Y-matrix with elements that are homogeneous rational functions of degree 1: (7.193) with the appropriate substitutions realizes each diagonal element, while (7.194) with the appropriate substitutions realizes each off-diagonal element.

7.11 Extension operations on subspace collections

Let us suppose we have a $Z(n)$ subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.195}$$

where $\mathcal{U}$ is m-dimensional. Let $\mathcal{V}$ be another m-dimensional space, and consider the space

$$\mathcal{K} = \mathcal{V} \oplus \mathcal{H}. \tag{7.196}$$

Suppose there is a nonsingular mapping $\mathbf{T}$ from $\mathcal{U}$ to $\mathcal{V}$. Define the subspace $\widetilde{\mathcal{E}}$ to consist of all vectors spanned by $\mathbf{u}+\mathbf{Tu}$ as $\mathbf{u}$ varies in $\mathcal{U}$. Define $\widetilde{\mathcal{J}}$ to consist of all vectors spanned by $\mathbf{u}-\mathbf{Tu}$ as $\mathbf{u}$ varies in $\mathcal{U}$. Clearly we have

$$\mathcal{V}\oplus\mathcal{U}=\widetilde{\mathcal{E}}\oplus\widetilde{\mathcal{J}}, \tag{7.197}$$

and consequently we obtain the $Y(n)$ subspace collection

$$\mathcal{K}=\mathcal{E}'\oplus\mathcal{J}'=\mathcal{V}\oplus\mathcal{P}_1\oplus\mathcal{P}_2\oplus\cdots\oplus\mathcal{P}_n, \tag{7.198}$$

in which

$$\mathcal{E}'=\widetilde{\mathcal{E}}\oplus\mathcal{E},\quad \mathcal{J}'=\widetilde{\mathcal{J}}\oplus\mathcal{J}. \tag{7.199}$$

Furthermore given vectors satisfying

$$\mathbf{j}+\mathbf{J}=\mathbf{L}(\mathbf{e}+\mathbf{E}),\quad \mathbf{E}\in\mathcal{E},\ \ \mathbf{J}\in\mathcal{J},\ \ \mathbf{e},\mathbf{j}\in\mathcal{U}, \tag{7.200}$$

where

$$\mathbf{j}=\mathbf{Ze},\quad \mathbf{L}=\sum_{\ell=1}^{n}z_i\mathbf{\Lambda}_i, \tag{7.201}$$

we can set

$$\mathbf{E}_2=\mathbf{e}+\mathbf{E}\in\mathcal{H},\quad \mathbf{E}_1=\mathbf{Te}\in\mathcal{V},\quad \mathbf{J}_2=\mathbf{j}+\mathbf{J}\in\mathcal{H},\quad \mathbf{J}_1=-\mathbf{Tj}. \tag{7.202}$$

Then we have

$$\mathbf{E}_1+\mathbf{E}_2=\mathbf{Te}+\mathbf{e}+\mathbf{E}\in\mathcal{E}',\quad \mathbf{J}_1+\mathbf{J}_2=-\mathbf{Tj}+\mathbf{j}+\mathbf{J}\in\mathcal{J}', \tag{7.203}$$

and

$$\mathbf{J}_1=-\mathbf{YE}_1,\quad \text{with}\ \ \mathbf{Y}=\mathbf{TZT}^{-1}. \tag{7.204}$$

Given a basis $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m$ for $\mathcal{U}$, with respect to which the matrix-valued function $\mathbf{Z}(z_1,z_2,\ldots,z_n)$ is defined, it is natural to take $\mathbf{Tu}_1,\mathbf{Tu}_2,\ldots,\mathbf{Tu}_m$ as our basis for $\mathcal{V}$. Then $\mathbf{T}$ is represented by the identity matrix, and the functions $\mathbf{Z}(z_1,z_2,\ldots,z_n)$ and $\mathbf{Y}(z_1,z_2,\ldots,z_n)$ are identical. We call the subspace collection (7.198) the extension of the subspace collection (7.195).

7.12 Reference transformations and additive inverses

Given the impedance network illustrated in **Figure 7.3** we are free the change the scaling constants c_i assigned to each bond to new constants c_i' and accordingly replace z_i with $z_i'=z_ic_i/c_i'$ without changing the overall electrical response of the network. Analogously, given a homogeneous rational function $\mathbf{Y}(z_1,z_2,\ldots,z_n)$ of degree one, an operation which preserves the homogeneity is obviously to multiply the variables by constants to obtain the function

$$\mathbf{Y}'(z_1',z_2',\ldots,z_n')=\mathbf{Y}(d_1z_1',d_2z_2',\ldots,d_nz_n'). \tag{7.205}$$

The associated operation on the $Y(n)$ subspace collection $(\mathcal{E},\mathcal{J})$ and $(\mathcal{V},\mathcal{P}_1,\mathcal{P}_2,\ldots,\mathcal{P}_n)$ is found by generalizing the analysis given after (29.3) in Milton (2002) and is as follows. Given nonzero (possibly complex) constants c_i^E and c_i^J, $i=1,\ldots,n$ we introduce the linear transformations

$$\psi^E(\mathbf{P})=\mathbf{\Pi}_1\mathbf{P}+\sum_{i=1}^{n}c_i^E\mathbf{\Lambda}_i\mathbf{P},\quad \psi^J(\mathbf{P})=\mathbf{\Pi}_1\mathbf{P}+\sum_{i=1}^{n}c_i^J\mathbf{\Lambda}_i\mathbf{P}, \tag{7.206}$$

on fields $\mathbf{P} \in \mathcal{K}$, where $\mathbf{\Lambda}_1$ is the projection onto $\mathcal{P}_1$. These transformations leave the subspaces $\mathcal{V}$ and $\mathcal{P}_i$ invariant. Define the spaces

$$\mathcal{E}' = \psi^E(\mathcal{E}) \text{ and } \mathcal{J}' = \psi^J(\mathcal{J}). \tag{7.207}$$

These will have the same dimension as $\mathcal{E}$ and $\mathcal{J}$ respectively. To see this, suppose $\psi^E(\mathbf{E}) = \psi^E(\mathbf{E}')$ for some $\mathbf{E}, \mathbf{E}' \in \mathcal{E}$. Then $\psi^E(\mathbf{E} - \mathbf{E}') = 0$ and since (7.206) implies $\psi^E(\mathbf{P}) = 0$ only when $\mathbf{P} = 0$ we conclude that $\mathbf{E} = \mathbf{E}'$. We need to make the technical assumption that

$$\psi^E(\mathbf{E}) \neq \psi^J(\mathbf{J}), \quad \text{for all nonzero } \mathbf{E} \in \mathcal{E}, \mathbf{J} \in \mathcal{J}, \tag{7.208}$$

to ensure $\mathcal{E}'$ and $\mathcal{J}'$ have no nonzero vector in common. A more insightful meaning to the condition (7.208) is given in the next section.

Let $(\mathcal{E}', \mathcal{J}')$ and $(\mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n)$ be our new subspace collection. Given a solution to the equations

$$\mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad (\mathbf{I} - \mathbf{\Pi}_1)\mathbf{J} = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i \mathbf{E}, \tag{7.209}$$

in the original subspace collection, in which $\mathbf{\Pi}_1$ is the projection onto $\mathcal{V}$, the fields $\mathbf{E}' = \psi^E(\mathbf{E})$ and $\mathbf{J}' = \psi^J(\mathbf{J})$ will be a solution to the equations

$$\mathbf{E}' \in \mathcal{E}', \quad \mathbf{J}' \in \mathcal{J}', \quad (\mathbf{I} - \mathbf{\Pi}_1)\mathbf{J}' = \sum_{i=1}^{n} z'_i \mathbf{\Lambda}_i \mathbf{E}', \tag{7.210}$$

in the new subspace collection with

$$z'_i = z_i c_i^J / c_i^E. \tag{7.211}$$

Since $\mathbf{\Pi}_1 \mathbf{E}' = \mathbf{\Pi}_1 \mathbf{E}$ and $\mathbf{\Pi}_1 \mathbf{J}' = \mathbf{\Pi}_1 \mathbf{J}$, it follows that the $\mathbf{Y}$-tensor functions of the two subspace collections are related by (7.205) where

$$d_i = c_i^E / c_i^J. \tag{7.212}$$

In particular, if we choose $c_i^E = -c_i^J$ for all i we obtain $d_i = -1$. Then using the homogeneity of the function $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ we see that

$$\mathbf{Y}'(z'_1, z'_2, \ldots, z'_n) = \mathbf{Y}(-z'_1, -z'_2, \ldots, -z'_n) = -\mathbf{Y}(z'_1, z'_2, \ldots, z'_n). \tag{7.213}$$

So if to another subspace collection, with an associated function $\mathbf{Y}''(z_1, z_2, \ldots, z_n)$, we add this new subspace collection according to the prescription given in Section 7.7, then it produces a subspace collection with an associated Y-function which is obtained by subtracting $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ from $\mathbf{Y}''(z_1, z_2, \ldots, z_n)$. In other words, when $c_i^E = -c_i^J$ for all i, the subspace collection with subspaces $(\mathcal{E}', \mathcal{J}')$ and $(\mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n)$ is the additive inverse of the original subspace collection, having subspaces $(\mathcal{E}, \mathcal{J})$ and $(\mathcal{V}, \mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n)$, where $(\mathcal{E}', \mathcal{J}')$ and $(\mathcal{E}, \mathcal{J})$ are linked by (7.207). If the technical condition (7.208) is not satisfied it appears that the subspace collection has no arithmetic inverse.

7.13 Operations on subspace collections leaving the associated function invariant

Note from (7.212) that if we choose $c_i^J = c_i^E$ for all i then the associated function remains invariant. More generally, if we are interested in leaving the associated function invariant, we could take

$$\mathcal{E}' = \mathbf{C}\mathcal{E}, \quad \mathcal{J}' = \mathbf{C}\mathcal{J}, \quad \mathcal{V}' = \mathbf{C}\mathcal{V}, \quad \mathcal{H}' = \mathbf{C}\mathcal{H}, \quad \mathcal{P}'_i = \mathbf{C}\mathcal{P}_i, \tag{7.214}$$

where $\mathbf{C}$ is a nonsingular linear operator which maps $\mathcal{K}$ to itself. Then the fields $\mathbf{E}' = \mathbf{C}\mathbf{E}$ and $\mathbf{J}' = \mathbf{C}\mathbf{J}$ will be a solution to the equations

$$\mathbf{E}' \in \mathcal{E}', \quad \mathbf{J}' \in \mathcal{J}', \quad (\mathbf{I} - \mathbf{\Pi}_1')\mathbf{J}' = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i' \mathbf{E}', \tag{7.215}$$

where

$$\mathbf{\Pi}_1' = \mathbf{C}\mathbf{\Pi}_1\mathbf{C}^{-1}, \quad \mathbf{\Lambda}_i' = \mathbf{C}\mathbf{\Lambda}_i\mathbf{C}^{-1} \tag{7.216}$$

are the projections onto $\mathcal{V}'$ and $\mathcal{P}_i'$. If $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_m$ is a basis for $\mathcal{V}$ then setting $\mathbf{v}_i' = \mathbf{C}\mathbf{v}_i$ we can take $\mathbf{v}_1'$,$\mathbf{v}_2'$,. . .,$\mathbf{v}_m'$ as a basis for $\mathcal{V}'$. Since multiplying by $\mathbf{C}$ is a linear operation the coefficients in the expansions

$$\mathbf{\Pi}_1'\mathbf{E}' = \sum_{i=1}^{m} E_i'\mathbf{u}_i', \quad \mathbf{\Pi}_1\mathbf{E} = \sum_{i=1}^{m} E_i\mathbf{u}_i, \quad \mathbf{\Pi}_1'\mathbf{J}' = \sum_{i=1}^{m} J_i'\mathbf{u}_i', \quad \mathbf{\Pi}_1\mathbf{J} = \sum_{i=1}^{m} J_i\mathbf{u}_i \tag{7.217}$$

can be equated:

$$E_i' = E_i, \quad J_i' = J_i, \tag{7.218}$$

and as a consequence the same matrix $\mathbf{Y}$ whose coefficients govern the relation

$$J_i = \sum_{i=1}^{k} Y_{ik} E_k, \tag{7.219}$$

also govern the relation

$$J_i' = \sum_{i=1}^{k} Y_{ik} E_k'. \tag{7.220}$$

Due to this equivalence it suffices in the preceding section to limit attention to the transformations (7.206) having $c_i^J = 1$ for all i: it is only the ratio $d_i = c_i^E/c_i^J$ that has any real significance. Then $\psi^J(\mathbf{P}) = \mathbf{P}$, and the technical condition (7.208) is violated only when there are nonzero vectors $\mathbf{E} \in \mathcal{E}$ and $\mathbf{J} \in \mathcal{J}$ such that

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2, \quad \mathbf{J} + \mathbf{J}_1 + \mathbf{J}_2, \quad \mathbf{E}_1 = \mathbf{J}_1 \in \mathcal{V}, \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2 \in \mathcal{H}, \quad \text{with } \mathbf{L} = \sum_{i=1}^{n} c_i^E \mathbf{\Lambda}_i. \tag{7.221}$$

Thus either $\mathbf{Y}(c_1^E, c_2^E, \ldots, c_n^E)$ has an eigenvalue of -1, or the Y-problem with $z_i = c_i^E$ for all i has a nonunique solution (with a nontrivial solution having $\mathbf{E}_2 \neq 0$ for the homogeneous problem with $\mathbf{E}_1 = 0$ and also $\mathbf{J}_1 = 0$, the latter not being needed for nonuniqueness but being needed if (7.221) holds). If we are looking for the arithmetic inverse we take $c_i^E = -1$ for all i, and the inverse exists except when $\mathbf{Y}(-1, -1, \ldots, -1) = -\mathbf{Y}(1, 1, \ldots, 1)$ has eigenvalue -1 or when the Y-problem with $z_i = 1$ for all i has a nonunique solution (with the homogeneous problem having a nontrivial solution with both $\mathbf{E}_1$ and $\mathbf{J}_1$ being zero).

There is a similar invariance of matrix functions associated with $Z(n)$ subspace collections under the linear transformations,

$$\mathcal{U}' = \mathbf{C}\mathcal{U}, \quad \mathcal{E}' = \mathbf{C}\mathcal{E}, \quad \mathcal{J}' = \mathbf{C}\mathcal{J}, \quad \mathcal{P}_i' = \mathbf{C}\mathcal{P}_i. \tag{7.222}$$

These invariances are quite natural, as they are isomorphic to changing the basis in the vector spaces $\mathcal{H}$ or $\mathcal{K}$. Thus, up to these trivial equivalences, the arithmetic inverse defined in the previous section is unique.

7.14 Multiplicative Inverses of superfunctions

To find the multiplicative inverse of a superfunction $(F^s)'$ we let $\mathcal{K}''$ be a vector space with the same dimension as $\mathcal{K}'$, and we take $\mathbf{C}$ as a nonsingular map from $\mathcal{K}'$ to $\mathcal{K}''$. We then let

$$\begin{aligned} \mathcal{J}'' &= \mathbf{C}(\mathcal{J}'), \quad \mathcal{H}'' = \mathbf{C}\mathcal{H}', \\ (\mathcal{V}^I)'' &= \mathbf{C}(\mathcal{V}^O)', \quad (\mathcal{V}^O)'' = \mathbf{C}(\mathcal{V}^I)', \quad \mathcal{P}_i'' = \mathbf{C}\mathcal{P}_i' \text{ for } i = 1, 2, \ldots, j. \end{aligned} \tag{7.223}$$

Introduce the transformation

$$\psi(\mathbf{P}) = \mathbf{\Pi}_1''\mathbf{P} - \mathbf{\Pi}_2'', \tag{7.224}$$

where $\mathbf{\Pi}_1''$ is the projection onto $(\mathcal{V}^I)'' \oplus (\mathcal{V}^O)''$ and $\mathbf{\Pi}_2''$ is the projection onto $\mathcal{H}''$. This is a special case of the transformations in (7.206). We let $\mathcal{E}'' = \psi(\mathbf{C}\mathcal{E}')$. Note that the output space $(\mathcal{V}^O)'$ gets mapped to the input space $(\mathcal{V}^I)''$, and the input space $(\mathcal{V}^I)'$ gets mapped to the output space $(\mathcal{V}^O)''$, and apart from these switchings we have essentially made an additive inverse in the Y-problem. We still require the technical condition mentioned in the last section, to ensure that this additive inverse exists: the operator $\mathbf{Y}(1, 1, \ldots, 1) - \mathbf{I}$ is nonsingular and the Y-problem with $z_i = 1$ for all i has a unique solution (or more precisely the homogeneous problem does not have a nontrivial solution with both $\mathbf{E}_1$ and $\mathbf{J}_1$ being zero).

Now suppose we are given a solution to the superfunction problem associated with $(F^s)'$,

$$\mathbf{E}' = (\mathbf{E}^I)' + (\mathbf{E}^O)' + \mathbf{E}_2' \in \mathcal{E}', \quad \mathbf{J}' = (\mathbf{J}^I)' + (\mathbf{J}^O)' + \mathbf{J}_2' \in \mathcal{J}', \quad \mathbf{J}_2' = \mathbf{L}'\mathbf{E}_2', \tag{7.225}$$

where

$$\mathbf{L}' = \sum_{i=1}^{j} z_i' \mathbf{\Lambda}_i', \tag{7.226}$$

in which $\mathbf{\Lambda}_i'$ is the projection onto $\mathcal{P}_i'$, and

$$(\mathbf{E}^I)', (\mathbf{J}^I)' \in (\mathcal{V}^I)', \quad (\mathbf{E}^O)', (\mathbf{J}^O)' \in (\mathcal{V}^O)', \quad \mathbf{E}_2', \mathbf{J}_2' \in \mathcal{H}'. \tag{7.227}$$

Now take vectors

$$\begin{aligned} \mathbf{E}'' &= \psi(\mathbf{C}\mathbf{E}'), \quad \mathbf{J}'' = -\mathbf{C}\mathbf{J}', \quad \mathbf{E}_2'' = -\mathbf{C}\mathbf{E}_2', \quad \mathbf{J}_2'' = -\mathbf{C}\mathbf{J}_2' \\ (\mathbf{E}^I)'' &= \mathbf{C}(\mathbf{E}^O)', \quad (\mathbf{E}^O)'' = \mathbf{C}(\mathbf{E}^I)', \quad (\mathbf{J}^I)'' = -\mathbf{C}(\mathbf{J}^O)', \quad (\mathbf{J}^O)'' = -\mathbf{C}(\mathbf{J}^I)'. \end{aligned} \tag{7.228}$$

These solve the superfunction problem associated with $(F^s)''$,

$$\mathbf{E}'' = (\mathbf{E}^I)'' + (\mathbf{E}^O)'' + \mathbf{E}_2'' \in \mathcal{E}'', \quad \mathbf{J}'' = (\mathbf{J}^I)'' + (\mathbf{J}^O)'' + \mathbf{J}_2'' \in \mathcal{J}'', \quad \mathbf{J}_2'' = \mathbf{L}''\mathbf{E}_2'', \tag{7.229}$$

where

$$\mathbf{L}'' = \sum_{i=1}^{j} z_i'' \mathbf{\Lambda}_i'', \quad z_i'' = z_i', \tag{7.230}$$

in which $\mathbf{\Lambda}_i''$ is the projection onto $\mathcal{P}_i''$, and

$$(\mathbf{E}^I)'', (\mathbf{J}^I)'' \in (\mathcal{V}^I)'', \quad (\mathbf{E}^O)'', (\mathbf{J}^O)'' \in (\mathcal{V}^O)'', \quad \mathbf{E}_2'', \mathbf{J}_2'' \in \mathcal{H}''. \tag{7.231}$$

Next let $\mathbf{M}_1$ denote the restriction of $\mathbf{C}$ to the subspace $(\mathcal{V}^O)'$, i.e., that operator mapping $(\mathcal{V}^O)'$ to $(\mathcal{V}^I)''$, such that $\mathbf{M}_1\mathbf{P} = \mathbf{C}\mathbf{P}$ for all $\mathbf{P} \in (\mathcal{V}^O)'$. Then from (7.228) we have $(\mathbf{E}^I)'' = \mathbf{M}_1(\mathbf{E}^O)'$ and $(\mathbf{J}^I)'' = -\mathbf{M}_1(\mathbf{J}^O)'$. To see that $(F^s)''$ is the inverse of the superfunction $(F^s)'$ when

$$\mathbf{L}' = \sum_{i=1}^{j} z_i' \mathbf{\Lambda}_i', \quad \mathbf{L}'' = \sum_{i=1}^{j} z_i' \mathbf{\Lambda}_i'', \tag{7.232}$$

we introduce the operator $\mathbf{M}_2$ which is the restriction of $\mathbf{C}^{-1}$ to the subspace $(\mathcal{V}^O)''$, i.e., that operator mapping $(\mathcal{V}^O)''$ to $(\mathcal{V}^I)'$, such that $\mathbf{M}_2\mathbf{P} = \mathbf{C}\mathbf{P}$ for all $\mathbf{P} \in (\mathcal{V}^I)'$. Then upon taking the product of the superfunctions (7.228) implies

$$\begin{pmatrix} (\mathbf{E}^O)'' \\ (\mathbf{J}^O)'' \end{pmatrix} = \mathbf{F} \begin{pmatrix} (\mathbf{E}^I)' \\ (\mathbf{J}^I)' \end{pmatrix}, \tag{7.233}$$

where

$$\mathbf{F} = \begin{pmatrix} (\mathbf{M}_2)^{-1} & 0 \\ 0 & -(\mathbf{M}_2)^{-1} \end{pmatrix} \tag{7.234}$$

is the multiplicative identity operator. From this analysis it looks like there are many multiplicative inverses, parameterized by $\mathbf{C}$, but in fact all are equivalent: this follows from the previous section.

7.15 Pruning the subspace collections

If an m terminal resistor network has a cluster of resistors which is not connected to the rest of the network, and that cluster does not have any terminals, only internal nodes, then clearly we can discard it without affecting the fields in the rest of the network and its response matrix. The analogous operation on subspace collections is called pruning.

When $\mathbf{L}$ is close to $z_0\mathbf{I}$ we can expand the inverses in (7.62) and (7.64) to obtain the series expansions

$$\mathbf{E} = \sum_{j=1}^{\infty} [\mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})/z_0]^j \mathbf{e}, \tag{7.235}$$

$$\mathbf{Z} = z_0\mathbf{\Gamma}_0 + \sum_{j=1}^{\infty} \mathbf{\Gamma}_0(\mathbf{L} - z_0\mathbf{I})[\mathbf{\Gamma}_1(\mathbf{L} - z_0\mathbf{I})/z_0]^j \mathbf{\Gamma}_0. \tag{7.236}$$

From these expansions it is evident that is only those fields in $\mathcal{H}$ that arise from products of the operators $\mathbf{\Gamma}_1$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, …, $\mathbf{\Lambda}_n$ applied to fields in $\mathcal{U}$ have any role in determining $\mathbf{E}$ and the associated function $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ (also $\mathbf{j}$ and $\mathbf{J}$): so we may as well prune away any other fields from the vector space $\mathcal{H}$. Thus we can redefine $\mathcal{H}$ as the smallest subspace containing $\mathcal{U}$ that is closed under the action of $\mathbf{\Gamma}_1$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, …, $\mathbf{\Lambda}_n$ and redefine

$$\mathcal{E} = \mathbf{\Gamma}_1\mathcal{H}, \quad \mathcal{J} = \mathbf{\Gamma}_2\mathcal{H}, \quad \mathcal{P}_j = \mathbf{\Lambda}_j\mathcal{H}, \quad j = 1, 2, \ldots, n. \tag{7.237}$$

This imposes constraints on the dimensions of these subspaces, as noted in Section 29.2 of Milton (2002) where the results are given in the case where $\mathcal{U}$ has dimension 1 and where the spaces are orthogonal. Let p_j

be the dimension of $\mathcal{P}_j$, $j = 1, 2, \ldots, n$, and let m, q_1 and q_2 represent the dimensions of $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$. The total dimension of the vector space $\mathcal{H}$ is therefore

$$h = m + q_1 + q_2 = p_1 + p_2 + \ldots + p_n. \tag{7.238}$$

Now the space

$$[\mathbf{\Lambda}_1(\mathcal{U} \oplus \mathcal{E})] \oplus [\mathbf{\Lambda}_2(\mathcal{U} \oplus \mathcal{E})] \oplus \ldots \oplus [\mathbf{\Lambda}_n(\mathcal{U} \oplus \mathcal{E})] \tag{7.239}$$

certainly contains $\mathcal{U}$, and is closed under $\mathbf{\Gamma}_1$ (because it contains $\mathcal{E}$) and is closed under $\mathbf{\Lambda}_j$ for each j. It therefore must be $\mathcal{H}$ and $\mathbf{\Lambda}_j(\mathcal{U} \oplus \mathcal{E})$ which has at most dimension $m + q_1$ must be $\mathcal{P}_j$. Therefore for each j we have the inequality

$$p_j \leq m + q_1, \tag{7.240}$$

and by summing these over j we see that

$$q_2 \leq (n-1)(m + q_1). \tag{7.241}$$

Similarly the subspace

$$[\mathbf{\Lambda}_1(\mathcal{U} \oplus \mathcal{J})] \oplus [\mathbf{\Lambda}_2(\mathcal{U} \oplus \mathcal{J})] \oplus \ldots \oplus [\mathbf{\Lambda}_n(\mathcal{U} \oplus \mathcal{J})] \tag{7.242}$$

can also be identified with $\mathcal{H}$ and we obtain the inequalities

$$p_j \leq m + q_2, \quad q_1 \leq (n-1)(m + q_2). \tag{7.243}$$

In the particular case when $n = 2$ the constraints (7.241) and (7.243) imply that the dimensions of the subspaces $\mathcal{E}$ and $\mathcal{J}$ can differ by at most m. Also in the case $n = 2$ we have

$$p_1 = (m + q_2 - p_2) + q_1 = (m + q_1 - p_2) + q_2 \geq \max\{q_1, q_2\}, \tag{7.244}$$

and similarly for p_2.

Likewise we can redefine $\mathcal{K}$ as the smallest subspace containing $\mathcal{V}$ that is closed under the action of $\mathbf{\Gamma}_1$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, …, $\mathbf{\Lambda}_n$ and redefine

$$\mathcal{E} = \mathbf{\Gamma}_1\mathcal{K}, \quad \mathcal{J} = \mathbf{\Gamma}_2\mathcal{K}, \quad \mathcal{P}_j = \mathbf{\Lambda}_j\mathcal{K}, \quad j = 1, 2, \ldots, n. \tag{7.245}$$

Let v be the dimension of $\mathcal{V}$, p_j be the dimension of $\mathcal{P}_j$, $j = 1, 2, \ldots, n$, and let q_1 and q_2 represent the dimensions of $\mathcal{E}$ and $\mathcal{J}$. The total dimension of the vector space $\mathcal{K}$ is therefore

$$h = q_1 + q_2 = v + p_1 + p_2 + \ldots + p_n. \tag{7.246}$$

The space

$$\mathcal{V} \oplus [\mathbf{\Lambda}_1(\mathcal{E})] \oplus [\mathbf{\Lambda}_2(\mathcal{E})] \oplus \ldots \oplus [\mathbf{\Lambda}_n(\mathcal{E})] \tag{7.247}$$

certainly contains $\mathcal{V}$, and is closed under $\mathbf{\Gamma}_1$ (because it contains $\mathcal{E}$) and is closed under $\mathbf{\Lambda}_j$ for each j. It therefore must be $\mathcal{K}$ and $\mathbf{\Lambda}_j(\mathcal{E})$ which has at most dimension q_1 must be $\mathcal{P}_j$. Thus for each j we have the inequality

$$p_j \leq q_1, \tag{7.248}$$

and summing these over j we obtain

$$q_2 \leq v + (n-1)q_1. \tag{7.249}$$

Similarly since

$$\mathcal{K} = \mathcal{V} \oplus [\mathbf{\Lambda}_1(\mathcal{J})] \oplus [\mathbf{\Lambda}_2(\mathcal{J})] \oplus \ldots \oplus [\mathbf{\Lambda}_n(\mathcal{J})], \tag{7.250}$$

we obtain the inequalities

$$p_j \leq q_2, \quad q_1 \leq v + (n-1)q_2. \tag{7.251}$$

When $n = 2$ the constraints (7.249) and (7.251) imply that the dimensions of the subspaces $\mathcal{E}$ and $\mathcal{J}$ can differ by at most v. Also in the case $n = 2$ we have

$$p_1 = (q_2 - p_2) + q_1 - v = (q_1 - p_2) + q_2 - v \geq \max\{q_1, q_2\} - v, \tag{7.252}$$

with a similar inequality for p_2.

7.16 Expressions for the numerator and denominator in the rational function

Assume that a $Z(n)$ subspace collection, with $m = 1$ has been pruned. Let $\mathbf{w}_1$,$\mathbf{w}_2$,...,$\mathbf{w}_{q_1+1}$ be a basis for $\mathcal{U} \oplus \mathcal{E}$ with $\mathbf{w}_1$ in $\mathcal{U}$ and $\mathbf{w}_2$,$\mathbf{w}_3$,...,$\mathbf{w}_{q_1+1}$ in $\mathcal{E}$. In this basis $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1)\mathbf{\Lambda}_i(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1)$ is represented by a $(q_1+1) \times (q_1+1)$ matrix $\mathbf{A}_i$, and since the $\mathbf{\Lambda}_i$ sum up to the identity operator it follows that

$$\sum_{i=1}^{n} \mathbf{A}_i = \mathbf{I}. \tag{7.253}$$

Also, because the subspace is pruned, $\mathbf{\Lambda}_i(\mathcal{U} \oplus \mathcal{E})$ can be identified with $\mathcal{P}_i$ which implies the matrix $\mathbf{A}_i$ must have at most rank p_i. It is exactly p_i if $\mathcal{P}_i \cap \mathcal{J} = \{\mathbf{0}\}$. The formula (7.175) for the Z-function implies

$$1/Z(z_1, z_2, \ldots, z_n) = \mathbf{e}_1 \cdot [\sum_{i=1}^{n} z_i \mathbf{A}_i]^{-1} \mathbf{e}_1, \tag{7.254}$$

where $\mathbf{e}_1$ is the $q_1 + 1$ component unit vector $[1, 0, 0, \ldots 0]^T$. Hence, following the argument given in Section 29.2 of Milton (2002), $Z(z_1, z_2, \ldots, z_n)$ can be expressed in the form (7.183) with numerator

$$p(z_1, z_2, \ldots, z_n) = \det[\sum_{i=1}^{n} z_i \mathbf{A}_i] = \sum_{a_1, a_2, \ldots, a_n} \alpha_{a_1 a_2 \ldots a_n} z_1^{a_1} z_2^{a_2} \ldots z_n^{a_n}, \tag{7.255}$$

of degree $1 + q_1$, in which the sum extends over all $a_1, a_2, \ldots, a_n$ with

$$\sum_{i=1}^{n} a_i = 1 + q_1, \quad 0 \leq a_i \leq p_i \quad \text{for } i = 1, 2, \ldots, n. \tag{7.256}$$

Typically one expects that the maximum power of z_i in this polynomial will be the rank of $\mathbf{A}_i$. However, for example, note that for the matrices

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad \mathbf{M}_2 = \mathbf{I} - \mathbf{M}_1, \tag{7.257}$$

the maximum power of z_1 in

$$\det[z_1\mathbf{M}_1 + z_2\mathbf{M}_2] = \det[(z_1 - z_2)\mathbf{M}_1 + z_2\mathbf{I}] = z_2[z_2^2 + 2z_2(z_1 - z_2)] \tag{7.258}$$

is 1 while $\mathbf{M}_1$ has rank 2.

Next let $\mathbf{w}_1$,$\mathbf{w}_{q_1+2}$,...,$\mathbf{w}_h$ be a basis for $\mathcal{U} \oplus \mathcal{J}$ with $\mathbf{w}_1$ in $\mathcal{U}$ and $\mathbf{w}_{q_1+2}$, $\mathbf{w}_{q_1+3}$...,$\mathbf{w}_h$ in $\mathcal{J}$. In this basis $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{\Lambda}_i(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ is represented by a $(q_2 + 1) \times (q_2 + 1)$ matrix $\mathbf{B}_i$, and since the $\mathbf{\Lambda}_i$ sum up to the identity operator it follows that

$$\sum_{i=1}^{n} \mathbf{B}_i = \mathbf{I}. \tag{7.259}$$

Also, because the subspace is pruned, $\mathbf{\Lambda}_i(\mathcal{U} \oplus \mathcal{J})$ can be identified with $\mathcal{P}_i$ which implies the matrix $\mathbf{B}_i$ must have rank at most p_i. It is exactly p_i if $\mathcal{P}_i \cap \mathcal{E} = \{\mathbf{0}\}$. The formula (7.60) for the Z-function implies

$$Z(z_1, z_2, \ldots, z_n) = \mathbf{e}_2 \cdot [\sum_{i=1}^{n} \mathbf{B}_i/z_i]^{-1}\mathbf{e}_2, \tag{7.260}$$

where $\mathbf{e}_2$ is the $q_2 + 1$ component unit vector $[1, 0, 0, \ldots 0]^T$. The denominator of this expression, as a polynomial in the variables $1/z_i$, is

$$\det[\sum_{i=1}^{n} \mathbf{B}_i/z_i] = \sum_{b_1,b_2,\ldots,b_n} \beta_{b_1 b_2 \ldots b_n}/z_1^{b_1} z_2^{b_2} \ldots z_n^{b_n}, \tag{7.261}$$

in which the sum extends over all $b_1, b_2, \ldots, b_n$ with

$$\sum_{i=1}^{n} b_i = 1 + q_2, \quad 0 \le b_i \le p_i \quad \text{for } i = 1, 2, \ldots, n. \tag{7.262}$$

Consequently, for the denominator in the expression (7.183) for $Z(z_1, z_2, \ldots, z_n)$, we can make the identification

$$q(z_1, z_2, \ldots, z_n) = \sum_{b_1,b_2,\ldots,b_n} \beta_{b_1 b_2 \ldots b_n} z_1^{p_1 - b_1} z_2^{p_2 - b_2} \ldots z_n^{p_n - b_n}, \tag{7.263}$$

which is a polynomial of degree $h - (1 + q_2) = q_1$. Furthermore the identities (7.253) and (7.259) imply the polynomial p and q satisfy the normalization (7.185), i.e.,

$$\sum_{a_1,a_2,\ldots,a_n} \alpha_{a_1 a_2 \ldots a_n} = 1, \qquad \sum_{b_1,b_2,\ldots,b_n} \beta_{b_1 b_2 \ldots b_n} = 1. \tag{7.264}$$

7.17 The correspondence between rational functions of one variable and $Z(2)$ subspace collections with $m = 1$

In the case $m = 1$ and $n = 2$ there are two cases to consider. When the dimension of h is even, $h = 2d$, then in order to satisfy the inequalities (7.240), (7.241) and (7.243) the subspaces $\mathcal{E}$ and $\mathcal{J}$ must have dimension d and $d - 1$ or vice versa and the subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ must have dimension d. Without loss of generality, by

making a duality transformation if necessary, let us suppose $\mathcal{E}$ has dimension $d-1$. Given $\mathbf{u} \in \mathcal{U}$ let us take as our basis for $\mathcal{H}$ the vectors

$$\mathbf{v}_{2j-1} = (\mathbf{\Gamma}_1\mathbf{\Lambda}_1)^{j-1}\mathbf{u}, \quad \mathbf{v}_{2j} = (\mathbf{\Lambda}_1\mathbf{\Gamma}_1)^{j-1}\mathbf{\Lambda}_1\mathbf{u}, \quad j = 1, 2, \ldots, d, \tag{7.265}$$

so that

$$\mathbf{v}_1 = \mathbf{u}, \quad \mathbf{v}_{2j} = \mathbf{\Lambda}_1\boldsymbol{\nu}_{2j-1}, \quad j = 1, 2, \ldots, d, \quad \mathbf{v}_{2j+1} = \mathbf{\Gamma}_1\boldsymbol{\nu}_{2j-1}, \quad j = 1, 2, \ldots, d-1. \tag{7.266}$$

These fields are independent since if they were not we could prune the subspace collection. The vectors $\mathbf{v}_{2j+1}, j = 1, 2, \ldots, d-1$, which number $d-1$, must form a basis for $\mathcal{E}$ and so it follows that

$$\mathbf{\Gamma}_1\mathbf{v}_{2d} = \sum_{i=1}^{d-1} \gamma_i \mathbf{v}_{2i+1}. \tag{7.267}$$

Also we have

$$\mathbf{\Gamma}_0\mathbf{v}_1 = \mathbf{v}_1, \quad \mathbf{\Gamma}_0\mathbf{v}_{2j} = \delta_j\mathbf{v}_1, \quad j = 1, 2, \ldots, d, \quad \mathbf{\Gamma}_0\mathbf{v}_{2j+1} = 0, \quad j = 1, 2, \ldots, d-1. \tag{7.268}$$

The $2d-1$ constants $\gamma_1, \ldots, \gamma_{d-1}$ and $\delta_1, \ldots, \delta_d$ characterize the geometry of the subspace collection. The field $\mathbf{e} + \mathbf{E}$ must have the expansion

$$\mathbf{e} + \mathbf{E} = \sum_{i=1}^{d} a_i \mathbf{v}_{2i-1}, \tag{7.269}$$

and consequently, setting $z_2 = 1$ we get

$$\mathbf{j} + \mathbf{J} = [\mathbf{I} + (z_1 - 1)\mathbf{\Lambda}_1](\mathbf{e} + \mathbf{E}) = \sum_{i=1}^{d} a_i \mathbf{v}_{2i-1} + (z_1 - 1)\sum_{i=1}^{d} a_i \mathbf{v}_{2i}. \tag{7.270}$$

Since $\mathbf{\Gamma}_1(\mathbf{j} + \mathbf{J}) = 0$ we arrive at the equations

$$\begin{aligned} 0 &= \sum_{i=2}^{d} a_i \mathbf{v}_{2i-1} + (z_1 - 1)\sum_{i=1}^{d-1} a_i \mathbf{v}_{2i+1} + (z_1 - 1)\sum_{i=1}^{d-1} a_d \gamma_i \mathbf{v}_{2i+1} \\ &= \sum_{i=1}^{d-1} [a_{i+1} + a_i(z_1 - 1) + \gamma_i a_d (z_1 - 1)]\mathbf{v}_{2i+1}, \end{aligned} \tag{7.271}$$

implying

$$a_{i+1} + a_i(z_1 - 1) + \gamma_i a_d (z_1 - 1) = 0, \quad i = 1, \ldots, d-1. \tag{7.272}$$

Choosing a normalization with $a_d = (1 - z_1)^{d-1}$ these equations are solved with

$$a_i = (1 - z_1)^{i-1} - \sum_{j=i}^{d-1} \gamma_{d-1+i-j}(1 - z_1)^j. \tag{7.273}$$

Since

$$\mathbf{\Gamma}_0(\mathbf{e} + \mathbf{E}) = a_1\mathbf{v}_1, \quad \mathbf{\Gamma}_0(\mathbf{j} + \mathbf{J}) = [a_1 + (z_1 - 1)\sum_{i=1}^{d} \delta_i a_i]\mathbf{v}_1, \tag{7.274}$$

we obtain

$$Z(z_1,1) = 1 + \frac{(z_1-1)\sum_{i=1}^{d}\delta_i a_i}{a_1}. \tag{7.275}$$

Conversely suppose we are given a rational function $Z(z_1,1)$ with a denominator of degree at most $d-1$ and a numerator of degree at most d satisfying $Z(1,1)=1$. It can be expressed in the form

$$Z(z_1,1) = \frac{p(z_1,1)}{q(z_1,1)} = 1 - \frac{\sum_{j=0}^{d-1} t_j(1-z_1)^{j+1}}{1-\sum_{j=1}^{d-1} s_j(1-z_1)^j}. \tag{7.276}$$

Comparing this with (7.275) we can make the identifications

$$\begin{aligned} 1-\sum_{j=1}^{d-1} s_j(1-z_1)^j &= a_1 = 1-\sum_{j=1}^{d-1}\gamma_{d-j}(1-z_1)^j, \\ -\sum_{j=0}^{d-1} t_j(1-z_1)^{j+1} &= (z_1-1)\sum_{i=1}^{d}\delta_i a_i \\ &= -\sum_{j=0}^{d-1}\delta_{j+1}(1-z_1)^{j+1} + \sum_{j=0}^{d-1}\sum_{i=1}^{j}\delta_i\gamma_{d-1+i-j}(1-z_1)^{j+1}, \end{aligned} \tag{7.277}$$

which imply

$$s_j = \gamma_{d-j}, \quad t_0 = \delta_1, \quad t_j = \delta_{j+1} - \sum_{i=1}^{j}\delta_i\gamma_{d-1+i-j}, \quad j=1,\dots,d-1. \tag{7.278}$$

Given the coefficients s and t we can inductively uniquely determine the coefficients γ and δ:

$$\gamma_j = s_{d-j}, \quad \delta_1 = t_0, \quad \delta_{j+1} = t_j + \sum_{i=1}^{j}\delta_i s_{1+j-i}, \quad j=1,\dots,d-1. \tag{7.279}$$

On the other hand when the dimension of h is odd, $h=2d-1$, then in order to satisfy the inequalities (7.240), (7.241) and (7.243) the subspaces $\mathcal{E}$ and $\mathcal{J}$ must have dimension $d-1$ and the subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ must have dimension $d-1$ and d or vice versa. Without loss of generality let us suppose $\mathcal{P}_1$ has dimension $d-1$. Given $\mathbf{u}\in\mathcal{U}$ let us take as our basis for $\mathcal{H}$ the vectors

$$\mathbf{v}_{2j-1} = (\mathbf{\Gamma}_1\mathbf{\Lambda}_1)^{j-1}\mathbf{u}, \quad j=1,2,\dots,d-1, \quad \mathbf{v}_{2j} = (\mathbf{\Lambda}_1\mathbf{\Gamma}_1)^{j-1}\mathbf{\Lambda}_1\mathbf{u}, \quad j=1,2,\dots,d, \tag{7.280}$$

which satisfy

$$\mathbf{v}_1 = \mathbf{u}, \quad \mathbf{v}_{2j} = \mathbf{\Lambda}_1\boldsymbol{\nu}_{2j-1}, \quad \mathbf{v}_{2j+1} = \mathbf{\Gamma}_1\boldsymbol{\nu}_{2j-1}, \quad j=1,2,\dots,d-1. \tag{7.281}$$

Again these fields are independent since if they were not we could prune the subspace collection. The vectors $\mathbf{v}_{2j}, j=1,2,\dots,d-1$, which number $d-1$, must form a basis for $\mathcal{P}_1$ and so it follows that

$$\mathbf{\Lambda}_1\mathbf{v}_{2d-1} = \sum_{i=1}^{d-1}\gamma_i\mathbf{v}_{2i}. \tag{7.282}$$

Also we have

$$\mathbf{\Gamma}_0\mathbf{v}_1=\mathbf{v}_1,\quad \mathbf{\Gamma}_0\mathbf{v}_{2j}=\delta_j\mathbf{v}_1,\quad j=1,2,\ldots,d-1,\quad \mathbf{\Gamma}_0\mathbf{v}_{2j+1}=0,\quad j=1,2,\ldots,d-1. \tag{7.283}$$

The $2d-2$ constants $\gamma_1,\ldots,\gamma_{d-1}$ and $\delta_1,\ldots,\delta_{d-1}$ characterize the geometry of the subspace collection. The field $\mathbf{e}+\mathbf{E}$ has the expansion (7.269) and so, with $z_2=1$,

$$\mathbf{j}+\mathbf{J}=[\mathbf{I}+(z_1-1)\mathbf{\Lambda}_1](\mathbf{e}+\mathbf{E})=\sum_{i=1}^{d}a_i\mathbf{v}_{2i-1}+(z_1-1)\sum_{i=1}^{d-1}a_i\mathbf{v}_{2i}+(z_1-1)\sum_{i=1}^{d-1}a_d\gamma_i\mathbf{v}_{2i}. \tag{7.284}$$

Since $\mathbf{\Gamma}_1(\mathbf{j}+\mathbf{J})=0$ we arrive at the equations

$$\begin{aligned}0&=\sum_{i=2}^{d}a_i\mathbf{v}_{2i-1}+(z_1-1)\sum_{i=1}^{d-1}a_i\mathbf{v}_{2i+1}+(z_1-1)\sum_{i=1}^{d-1}a_d\gamma_i\mathbf{v}_{2i+1}\\&=\sum_{i=1}^{d-1}[a_{i+1}+a_i(z_1-1)+\gamma_i a_d(z_1-1)]\mathbf{v}_{2i+1},\end{aligned} \tag{7.285}$$

implying (7.272) which has the solution (7.273). Since

$$\mathbf{\Gamma}_0(\mathbf{e}+\mathbf{E})=a_1\mathbf{v}_1,\quad \mathbf{\Gamma}_0(\mathbf{j}+\mathbf{J})=[a_1+(z_1-1)\sum_{i=1}^{d-1}\delta_i(a_i+a_d\gamma_i)]\mathbf{v}_1=[a_1-\sum_{i=1}^{d-1}\delta_i a_{i+1}]\mathbf{v}_1, \tag{7.286}$$

we obtain

$$Z(z_1,1)=1-\frac{\sum_{i=1}^{d-1}\delta_i a_{i+1}}{a_1}. \tag{7.287}$$

Conversely suppose we are given a rational function $Z(z_1,1)$ with a denominator of degree at most $d-1$ and a numerator of degree at most $d-1$. It can be expressed in the form

$$Z(z_1,1)=1-\frac{\sum_{j=1}^{d-1}t_j(1-z_1)^j}{1-\sum_{j=1}^{d-1}s_j(1-z_1)^j}. \tag{7.288}$$

Comparing this with (7.287) we can make the identifications

$$\begin{aligned}1-\sum_{j=1}^{d-1}s_j(1-z_1)^j&=a_1=1-\sum_{j=1}^{d-1}\gamma_{d-j}(1-z_1)^j,\\\sum_{j=1}^{d-1}t_j(1-z_1)^j&=\sum_{i=1}^{d-1}\delta_i a_{i+1}\\&=\sum_{j=1}^{d-1}\delta_j(1-z_1)^j-\sum_{j=2}^{d-1}\sum_{i=1}^{j-1}\delta_i\gamma_{d+i-j}(1-z_1)^j,\end{aligned} \tag{7.289}$$

which imply

$$s_j=\gamma_{d-j},\quad j=1,\ldots,d-1,\quad t_1=\delta_1,\quad t_j=\delta_j-\sum_{i=1}^{j-1}\delta_i\gamma_{d+i-j},\quad j=2,\ldots,d-1. \tag{7.290}$$

Given the coefficients s and t we can inductively uniquely determine the coefficients γ and δ:

$$\gamma_j = s_{d-j}, \quad j = 1, \ldots, d-1 \quad \delta_1 = t_1, \quad \delta_j = t_j + \sum_{i=1}^{j} \delta_i s_{j-i}, \quad j = 2, \ldots, d-1. \tag{7.291}$$

One can see from this analysis that there can be more than one pruned subspace collection associated with a rational function $Z(z_1, 1)$. It may happen that one pruned $Z(n)$ subspace collection gives rise to polynomials $p(z_1, 1) = f(z_1, 1)r(z_1, 1)$ and $q(z_1, 1) = g(z_1, 1)r(z_1, 1)$ while another pruned Z(n) subspace collection gives rise to polynomials $p'(z_1, 1) = f(z_1, 1)r'(z_1, 1)$ and $q'(z_1, 1) = t(z_1, 1)r'(z_1, 1)$, so that both give rise to the same function $Z(z_1, 1)$. However there is a one-to-one correspondence if the pruned subspace collection is such that the polynomials $p(z_1, z_2)$ and $q(z_1, z_2)$ have no factor in common, and this correspondence is given by the above algorithm

7.18 On the correspondence between certain rational functions of two variables and $Z(3)$ subspace collections with $m = 1$

In the case $m = 1$ and $n = 3$ can we uniquely recover a generic subspace collection (modulo the linear transformations (7.222)) from knowledge of the rational function $Z(z_1, z_2, 1)$? The answer is no, but let us first provide a counting argument which suggests that, at least in the generic case, we can recover the subspace collection up to a finite number of possibilities. The counting argument is similar to that given in Section 29.2 of Milton (2002) but here we do not assume that the subspaces are orthogonal.

How many independent coefficients $\alpha_{a_1 a_2 a_3}$ are there in a polynomial

$$p(z_1, z_2, 1) = \sum_{a_1, a_2, a_3} \alpha_{a_1 a_2 a_3} z_1^{a_1} z_2^{a_2}, \tag{7.292}$$

that satisfies

$$a_1 + a_2 + a_3 = 1 + q_1, \quad 0 \le a_i \le p_i \le 1 + q_1, \quad i = 1, 2, 3\,? \tag{7.293}$$

Without loss of generality, following Section 29.2 of Milton (2002), let us suppose that $p_1 \ge p_2 \ge p_3$. With a_1 fixed in the regime $0 \le a_1 < 1 + q_1 - p_2$, the constant a_2 can take integer values from $a_2 = 1 + q_1 - a_1 - p_3$ (where $a_3 = p_3$) to $a_2 = p_2$, that is, a total of $p_2 + p_3 + a_1 - q_1$ different values. With a_1 fixed in the regime $1 + q_1 - p_2 \le a_1 < 1 + q_1 - p_3$, the constant a_2 can take integer values from $a_2 = 1 + q_1 - a_1 - p_3$ (where $a_3 = p_3$) to $a_2 = 1 + q_1 - a_1$ (where $a_3 = 0$) that is, a total of $p_3 + 1$ different values. Finally, with a_1 fixed in the regime $1 + q_1 - p_3 \le a_1 \le p_1$, the constant a_2 can take integer values from $a_2 = 0$ to $a_2 = 1 + q_1 - a_1$ (where $a_3 = 0$), that is, a total of $2 + q_1 - a_1$ different values. Therefore the total number of coefficients in the polynomial is

$$\begin{aligned}
&\sum_{a_1=0}^{q_1-p_2} (p_2 + p_3 + a_1 - q_1) + \sum_{a_1=1+q_1-p_2}^{q_1-p_3} (p_3 + 1) + \sum_{a_1=1+q_1-p_3}^{p_1} (2 + q_1 - a_1) \\
&\quad = (q_1 - p_2 + 1)(p_2 + p_3 - q_1) + \frac{1}{2}(q_1 - p_2 + 1)(q_1 - p_2) + (p_2 - p_3)(p_3 + 1) \\
&\qquad + (p_1 + p_3 - q_1)(2 + p_3) - \frac{1}{2}((p_1 + p_3 - q_1)(p_1 + p_3 - q_1 + 1) \\
&\quad = k_1 + 1,
\end{aligned} \tag{7.294}$$

where

$$k_1 = [2(1+q_1)q_2 - p_1^2 - p_2^2 - p_3^2 + h]/2, \tag{7.295}$$

in which $h = p_1 + p_1 + p_3$ and $q_2 = h - 1 - q_1$. These coefficients are not all independent since, from (7.264) the $\alpha_{a_1 a_2 a_3}$ must sum to one. Subtracting this constraint gives k_1 independent coefficients.

Similarly in a polynomial

$$q(z_1, z_2, 1) = \sum_{b_1,b_2,b_3} \beta_{b_1 b_2 b_3} z_1^{p_1 - b_1} z_2^{p_2 - b_2}, \tag{7.296}$$

that satisfies

$$b_1 + b_2 + b_3 = 1 + q_2, \quad 0 \le a_i \le p_i \le 1 + q_2, \quad i = 1, 2, 3, \quad \sum_{b_1,b_2,b_3} \beta_{b_1 b_2 b_3} = 1, \tag{7.297}$$

there are a total of

$$k_2 = [2(1+q_2)q_1 - p_1^2 - p_2^2 - p_3^2 + h]/2 \tag{7.298}$$

independent coefficients. Hence the total number of independent coefficients in the rational function

$$Z(z_1, z_2, 1) = \frac{p(z_1, z_2, 1)}{q(z_1, z_2, 1)} \tag{7.299}$$

is

$$k_1 + k_2 = (1+q_1)q_2 + (1+q_2)q_1 - p_1^2 - p_2^2 - p_3^2 + h = h^2 - p_1^2 - p_2^2 - p_3^2 - q_1^2 - q_2^2. \tag{7.300}$$

Now how many parameters describe a $Z(n)$ subspace collection, when the spaces $\mathcal{U}$, $\mathcal{E}$, $\mathcal{J}$, $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$ have dimensions 1, q_1, q_2, p_1, p_2, and p_3, with $1 + q_1 + q_2 = p_1 + p_2 + p_3 = h$? Let $\mathbf{w}_1$,$\mathbf{w}_2$,…,$\mathbf{w}_h$ be a basis for $\mathcal{H}$ with $\mathbf{w}_1$ in $\mathcal{U}$, $\mathbf{w}_2$,$\mathbf{w}_3$,…,$\mathbf{w}_{q_1+1}$ in $\mathcal{E}$, and $\mathbf{w}_{q_1+2}$, $\mathbf{w}_{q_1+3}$…,$\mathbf{w}_h$ in $\mathcal{J}$. Recall that it requires $s(d-s)$ parameters to describe the orientation of a subspace of dimension s in a space of dimension d. Therefore, it requires

$$p_1(h - p_1) + (h - p_2)p_2 + (h - p_3)p_3 = h^2 - p_1^2 - p_2^2 - p_3^2 \tag{7.301}$$

parameters to describe the orientation of the subspaces $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$ with respect to this basis. However some of these subspace collections are equivalent, linked through transformations of the form (7.222). If respect to this basis $\mathbf{C}$ is represented by a matrix with block form

$$\mathbf{C} = \begin{pmatrix} c & 0 & 0 \\ 0 & \mathbf{C}_1 & 0 \\ 0 & 0 & \mathbf{C}_2 \end{pmatrix}, \tag{7.302}$$

where c is a scalar, while $\mathbf{C}_1$ and $\mathbf{C}_2$ are $q_1 \times q_1$ and $q_2 \times q_2$ matrices, then it will leave the subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ unchanged. The transformation $\mathbf{C} = a\mathbf{I}$ leaves all subspaces unchanged for any scalar $a \neq 0$, and so to factor out such trivial transformations we should choose $c = 1$. The number of remaining independent parameters in $\mathbf{C}$ is then $q_1^2 + q_2^2$. Subtracting these from (7.301) we see that the number of parameters describing the $Z(n)$ subspace collection is

$$h^2 - p_1^2 - p_2^2 - p_3^2 - q_1^2 - q_2^2 = k_1 + k_2. \tag{7.303}$$

The precise agreement between the number of coefficients in the rational function and the number of parameters describing the $Z(n)$ subspace collection is curious (since it holds for all q_1, q_2, p_1, p_2, and p_3, with $1 + q_1 + q_2 = p_1 + p_2 + p_3 = h$). Despite this coincidence we now show that it is not possible to uniquely recover a generic subspace collection (modulo the linear transformations (7.222)) from knowledge of the associated rational function $Z(z_1, z_2, 1)$.

Let us consider a subspace collection with $h = 5, q_1 = q_2 = 2, p_1 = p_2 = 1, p_3 = 3$ giving $k_1 + k_2 = 6$ according to the formula (7.300). Given $\mathbf{u} \in \mathcal{U}$ we choose as our basis the vectors

$$\mathbf{v}_0 = \mathbf{u}, \quad \mathbf{v}_1 = \mathbf{\Lambda}_1\mathbf{u}, \quad \mathbf{v}_2 = \mathbf{\Lambda}_2\mathbf{u}, \quad \mathbf{v}_3 = \mathbf{\Gamma}_1\mathbf{\Lambda}_1\mathbf{u}, \quad \mathbf{v}_4 = \mathbf{\Gamma}_1\mathbf{\Lambda}_2\mathbf{u}, \tag{7.304}$$

with the closure relations

$$\begin{aligned} \mathbf{\Lambda}_1\mathbf{v}_3 &= \gamma_1\mathbf{v}_1, \quad \mathbf{\Lambda}_2\mathbf{v}_3 = \gamma_2\mathbf{v}_2, \quad \mathbf{\Lambda}_1\mathbf{v}_4 = \gamma_3\mathbf{v}_1, \quad \mathbf{\Lambda}_2\mathbf{v}_4 = \gamma_4\mathbf{v}_1, \\ \mathbf{\Gamma}_0\mathbf{v}_1 &= \delta_1\mathbf{v}_0, \quad \mathbf{\Gamma}_0\mathbf{v}_2 = \delta_2\mathbf{v}_0, \end{aligned} \tag{7.305}$$

expressed in terms of the 6 parameters $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \delta_1$, and δ_2 which describe the subspace collection. The question is: can one uniquely recover these six parameters from $Z(z_1, z_2, 1)$? Although the following analysis extends easily to the case of arbitrary γ_1 and γ_4 let us assume, for simplicity, that $\gamma_1 = \gamma_4 = 0$ and ask whether one can recover the remaining four parameters. The field $\mathbf{e} + \mathbf{E}$ must have the expansion

$$\mathbf{e} + \mathbf{E} = a_0\mathbf{v}_0 + a_1\mathbf{v}_3 + a_2\mathbf{v}_4, \tag{7.306}$$

and consequently, setting $z_3 = 1$, we get

$$\begin{aligned} \mathbf{j} + \mathbf{J} &= [\mathbf{I} + (z_1 - 1)\mathbf{\Lambda}_1 + (z_2 - 1)\mathbf{\Lambda}_2](\mathbf{e} + \mathbf{E}) \\ &= a_0\mathbf{v}_0 + a_1\mathbf{v}_3 + a_2\mathbf{v}_4 + (z_1 - 1)(a_0 + a_2\gamma_3)\mathbf{v}_1 + (z_2 - 1)(a_0 + a_1\gamma_2)\mathbf{v}_2. \end{aligned} \tag{7.307}$$

Since $\mathbf{\Gamma}_1(\mathbf{j} + \mathbf{J}) = 0$ we arrive at the equations

$$0 = a_1\mathbf{v}_3 + a_2\mathbf{v}_4 + (z_1 - 1)(a_0 + a_2\gamma_3)\mathbf{v}_3 + (z_2 - 1)(a_0 + a_1\gamma_2)\mathbf{v}_4, \tag{7.308}$$

implying

$$a_1 + (z_1 - 1)(a_0 + a_2\gamma_3) = 0, \quad a_2 + (z_2 - 1)(a_0 + a_1\gamma_2) = 0. \tag{7.309}$$

These equations have as a solution,

$$\begin{aligned} a_0 &= 1 - (z_1 - 1)(z_2 - 1)\gamma_2\gamma_3, \\ a_1 &= \gamma_3(z_1 - 1)(z_2 - 1) - (z_1 - 1), \\ a_2 &= \gamma_2(z_1 - 1)(z_2 - 1) - (z_2 - 1). \end{aligned} \tag{7.310}$$

Since

$$\mathbf{\Gamma}_0(\mathbf{e} + \mathbf{E}) = a_0\mathbf{v}_0, \quad \mathbf{\Gamma}_0(\mathbf{j} + \mathbf{J}) = [a_0 + (z_1 - 1)(a_0 + a_2\gamma_3)\delta_1 + (z_2 - 1)(a_0 + a_1\gamma_2)\delta_2]\mathbf{v}_0, \tag{7.311}$$

we obtain

$$\begin{aligned} Z(z_1, z_2, 1) &= 1 + \frac{(z_1 - 1)(a_0 + a_2\gamma_3)\delta_1 + (z_2 - 1)(a_0 + a_1\gamma_2)\delta_2}{a_0} \\ &= 1 + \frac{\delta_1(z_1 - 1) - \gamma_3\delta_1(z_1 - 1)(z_2 - 1) + \delta_2(z_2 - 1) - \gamma_2\delta_2(z_1 - 1)(z_2 - 1)}{1 - (z_1 - 1)(z_2 - 1)\gamma_2\gamma_3}. \end{aligned} \tag{7.312}$$

Given this function we can uniquely determine δ_1 and δ_2 from the coefficients of $(z_1 - 1)$ and $(z_2 - 1)$ in the numerator. Also from the coefficients of $(z_1 - 1)(z_2 - 1)$ in the numerator and denominator we can uniquely determine

$$t_1 = \gamma_2\gamma_3, \quad t_2 = \gamma_3\delta_1 + \gamma_2\delta_2, \tag{7.313}$$

in terms of which there are two possible values of γ_2, namely

$$\gamma_2 = \frac{t_3 \pm \sqrt{t_3^2 - 3t_1\delta_1\delta_2}}{2\delta_1}. \tag{7.314}$$

Thus we cannot uniquely recover the subspace collection parameters from $Z(z_1, z_2, 1)$.

It remains an open question, raised at the end of Section 29.2 of Milton (2002), as to whether in general one can uniquely recover the subspace collection parameters when, with respect to some inner product, the subspaces $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are mutually orthogonal, and the subspaces $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ are mutually orthogonal. These orthogonality constraints overdetermine the system of equations needed to recover the subspace collection parameters which provides some hope that we can recover them. It would be useful if one could uniquely recover the subspace collection parameters (the weight and normalization matrices introduced in Milton, 1987a, 1987b) from say the effective conductivity $\sigma_*(\sigma_1, \sigma_2, \sigma_3)$ of an isotropic composite of three isotropic phases having conductivities σ_1, σ_2, and σ_3 as then one could obtain the effective response tensor for coupled field problems. We will see in **Chapter 9** that the effective response tensor just depends on the weight and normalization matrices for the uncoupled conductivity problem.

7.19 Visualizing the poles and zeros of functions associated with orthogonal $Z(3)$ subspace collections when $m = 1$

For scalar functions $Z(z_1, z_2, z_3)$, associated with orthogonal $Z(3)$ subspace collections, satisfying the homogeneity, Herglotz, and normalization properties, the trajectories of their poles and zeros in (z_1, z_2, z_3) space, with z_1, z_2, and z_3 taking real values, have a beautiful visualization as trajectories on three interlinked hexagons: To obtain this visualization we follow Appendix C in Nicorovici, McPhedran, and Milton (1993): see also **Figure 5** in that paper.

First note that if we set $z_3 = 1$, then the poles and zeros of $Z(z_1, z_2, 1)$ lie in one of the three quadrants:

- The quadrant $z_1 \le 0$, $z_2 \ge 0$;
- The quadrant $z_2 \le 0$, $z_1 \ge 0$;
- The quadrant $z_1 \le 0$, $z_2 \le 0$.

Of course we can visualize the pole and zero trajectories by plotting them in this plane, but this has the disadvantage that the three variables z_1, z_2 and z_3 are not treated in a symmetric way, and the disadvantage that its hard to see what is happening when z_1 and/or z_2 is large, and it is hard to see what is happening near the origin $z_1 = z_2 = 0$ since the trajectories can bunch up there. To get around this we map each of the three quadrants to a hexagon. Given a quadrant, the point $z_1 = z_2 = 0$ gets blown up to form one edge of the hexagon; the two edges of the quadrant where z_1 or z_2 is zero, but not the other, get mapped to two other edges of the hexagon; the two "boundaries" of the quadrant where $|z_1|$ or $|z_2|$ is infinite but other is finite get mapped to two more edges of the hexagon; finally $z_1 = z_2 = \infty$ gets mapped to the final sixth edge of the

hexagon. We remark that just as a pole trajectory can cross from one quadrant to another, so too can it jump from the boundary of one hexagon to the corresponding point on the boundary of another hexagon.

To be more precise, we introduce the three variables

$$t_1 = \frac{1}{1+|z_2/z_3|}, \quad t_2 = \frac{1}{1+|z_3/z_1|}, \quad t_3 = \frac{1}{1+|z_1/z_2|}. \tag{7.315}$$

Clearly (t_1, t_2, t_3) takes values in the unit cube. It is confined to a surface within the unit cube as the three ratios $|z_2|/|z_3|$, $|z_3|/|z_1|$ and $|z_1|/|z_2|$ are not independent, but have product 1. The next step is to map these three variables onto three variables s_1, s_2 and s_3 lying in the plane $s_1 + s_2 + s_3 = 0$ using the projection

$$s_1 = 2t_1 - t_2 - t_3, \quad s_2 = 2t_2 - t_3 - t_1, \quad s_3 = 2t_3 - t_1 - t_2. \tag{7.316}$$

Finally, we map these down to the x–y plane:

$$x = s_1, \quad y = (s_1 + 2s_2)/\sqrt{3}. \tag{7.317}$$

Some normalization is needed, so in the hexagon where z_1 is negative and z_2 and z_3 are positive, we plot $Z(z_1, z_2, z_3)/\sqrt{z_2 z_3}$; in the hexagon where z_2 is negative and z_1 and z_3 are positive, we plot $Z(z_1, z_2, z_3)/\sqrt{z_1 z_3}$; and in the hexagon where z_3 is negative and z_1 and z_2 are positive, we plot $Z(z_1, z_2, z_3)/\sqrt{z_1 z_2}$.

Figure 7.8 uses this approach to visualize the pole trajectory of a function $Z(z_1, z_2, z_3)$ associated with a $Z(3)$-subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \mathcal{P}_3, \tag{7.318}$$

where in this example $\mathcal{H}$ is 12-dimensional; $\mathcal{U}$ is one-dimensional; $\mathcal{P}_1$ is 3-dimensional; $\mathcal{P}_2$ is 6-dimensional; $\mathcal{P}_3$ is 3-dimensional. Note that as the subspace collection does not need pruning, the dimensions of $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$ can be immediately read off from the figure by simply counting the number of pole paths on each hexagon: figures (a), (b), and (c) have 3, 6 and 3 pole paths corresponding to the dimensions of $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$, respectively. To understand this, first recognize that when z_2 and z_3 are fixed, and real and positive, $Z(z_1, z_2, z_3)$ is a Herglotz function of z_1 taking real positive values when $z_1 > 0$. Thus all its poles must be simple and located on the negative real z_1-axis, i.e., on the hexagon (a). Also because the subspace is pruned $\mathbf{\Lambda}_1(\mathcal{U} \oplus \mathcal{J})$ can be identified with $\mathcal{P}_1$ (Section 7.16), and hence the matrix $\mathbf{C}_1$ representing $\mathbf{\Lambda}_1(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ has rank p_1. Then as $\mathbf{C}_1$ and $\mathbf{C}_1^T\mathbf{C}_1$ have equal rank (this well-known fact can easily be seen by showing that they have the same null-space), and as the subspace collection is orthogonal, it follows that the matrix $\mathbf{B}_1 = \mathbf{C}_1^T\mathbf{C}_1$ representing $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{\Lambda}_1(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ has exactly rank p_1. Similarly the matrix $\mathbf{A}_1$ representing $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{\Lambda}_1(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ has exactly rank p_1. Therefore the sum over a_1 in the numerator in (7.255), goes up to $a_1 = p_1$, while the sum in the denominator in (7.263), goes from 0 up to $b_1 = p_1$ or (when all the coefficients $\beta_{p_1 b_2 b_3}$ are zero) to $b_1 = p_1 - 1$: it cannot go only up to $b_1 = p_1 - 2$, since as a function of z_1, $Z(z_1, z_2, z_3)/\sqrt{z_2 z_3}$ with fixed $z_2 > 0$ and fixed $z_3 > 0$ can only have a simple pole at $z_1 = \infty$. When the sum over b_1 goes up to $b_1 = p_1$, there are clearly p poles of the function $Z(z_1, z_2, z_3)/\sqrt{z_2 z_3}$ on the hexagon as z_1 varies with fixed $z_2 > 0$ and fixed $z_3 > 0$. When the sum over b_1 goes up to $b_1 = p_1 - 1$, there are still p poles of the function $Z(z_1, z_2, z_3)/\sqrt{z_2 z_3}$ on the hexagon as z_1 varies with fixed $z_2 > 0$ and fixed $z_3 > 0$ provided we count the pole at $z_1 = \infty$.

The dimension q_2 of the subspace $\mathcal{J}$ can also generically be read off from the pole trajectories on the three hexagons. Consider the edge joining two of the hexagons that corresponds to the values $z_2 = 0$, and $z_3 = 1$

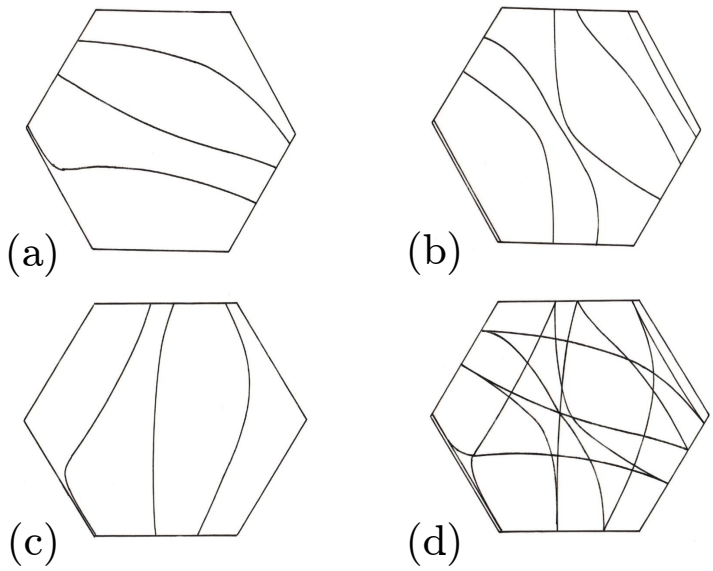


Figure 7.8: The pole trajectory of the function $Z(z_1, z_2, z_3)$ as visualized using the representation using three interlinked hexagons. The hexagon in (a) corresponds to real values of (z_1, z_2, z_3) where z_2 and z_3 have the same sign, but z_1 has the opposite sign. The hexagon in (b) corresponds to real values of (z_1, z_2, z_3) where z_1 and z_3 have the same sign, but z_2 has the opposite sign. The hexagon in (c) corresponds to real values of (z_1, z_2, z_3) where z_1 and z_2 have the same sign, but z_3 the opposite sign. By superimposing all three pictures one obtains (d) where the pole trajectory is like that of a billard ball bouncing around a hexagonal table, following curved paths. The zero trajectory is similiar, but for clarity we chose not to include it. Note that the dimensions 3, 6 and 3 of the subspaces $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$ can be immediately read off from the number of paths crossing the hexagons in (a), (b) and (c). These figures were obtained by Oscar Bruno in 1988, while he was working with me at the Courant Institute.

with $z_1 < 0$ varying. Then the only coefficients $\beta_{b_1 b_2 b_3}$ that can contribute to the denominator in (7.263) are those with $p_2 = b_2$. The first constraint in (7.262) then implies

$$b_1 + b_3 = 1 + q_2 - p_2. \tag{7.319}$$

So b_1 can only range from 0 up to the maximum of p_1 and $1 + q_2 - p_2 = p_1 + p_3 - q_1$. Note that according to the inequality (7.240), $q_1 \geq p_3 - 1$ so $1 + q_2 - p_2$ could be as large as $p_1 + 1$. If there are less than p_1 pole trajectories crossing this edge joining the hexagons, the number of these crossing pole trajectories should generically allow us to determine q_2 and hence q_1, assuming p_1, p_2 and p_3 have been determined from the number of pole trajectories on each hexagon. If there are exactly p_1 pole trajectories crossing the edge then q_2 could be p_3 or $p_3 + 1$. To determine which it is (or as an additional check on the value of q_2) we could look at pole trajectories, or zero trajectories, crossing other edges where the hexagons meet.

This visualization may be useful in finding other topological features of the trajectories, which hopefully

could be connected with topological features of the subspace collections.

7.20 Normalization operations on subspace collections

Rational functions of a single variable may be expanded in continued fractions, which incorporate successively higher and higher order terms in the series expansion of the function about a point. The analogous procedure with subspace collections is achieved through normalization and reduction operations, subject to some technical assumptions. The associated functions are then linked, and provided the technical assumptions hold at each level, these links provide continued fractions for multivariate functions $\mathbf{Z}(z_1, z_2, \ldots, z_n)$ and $\mathbf{Y}(z_1, z_2, \ldots, z_n)$ incorporating matrices of increasingly high dimension at each level in the continued fraction.

The normalization and reduction operations are discussed in this and the next section. For more insight, in the case where the subspaces in the direct sums are orthogonal (see Milton 1987a, 1987b and Sections 19.2, 20.6 and 29.5 in Milton 2002).

Normalization reverses extension. Given a subspace collection

$$\mathcal{K} = \mathcal{E}' \oplus \mathcal{J}' = \mathcal{V} \oplus \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.320}$$

define

$$\begin{aligned} \mathcal{H} &= \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \quad \mathcal{E} = \mathcal{E}' \cap \mathcal{H}, \quad \mathcal{J} = \mathcal{J}' \cap \mathcal{H}, \\ \mathcal{U} &= \boldsymbol{\Pi}_2 \boldsymbol{\Gamma}_1' \mathcal{V} = \boldsymbol{\Pi}_2 (\mathbf{I} - \boldsymbol{\Gamma}_2') \mathcal{V} = \boldsymbol{\Pi}_2 \boldsymbol{\Gamma}_2' \mathcal{V}, \qquad \widetilde{\mathcal{E}} = \boldsymbol{\Gamma}_1' \mathcal{V}, \quad \widetilde{\mathcal{J}} = \boldsymbol{\Gamma}_2' \mathcal{V}, \end{aligned} \tag{7.321}$$

where $\boldsymbol{\Gamma}_1'$ and $\boldsymbol{\Gamma}_2'$ are the projections onto $\mathcal{E}'$ and $\mathcal{J}'$, and $\boldsymbol{\Pi}_2$ is the projection onto $\mathcal{H}$.

We assume that the Y-problem has a unique solution when $\mathbf{L} = \mathbf{I}$ for $\mathbf{J}_1 \in \mathcal{V}$ given $\mathbf{E}_1 \in \mathcal{V}$. In other words, we assume that the equations

$$\begin{aligned} \mathbf{E}_1 + \mathbf{E}_2 &\in \mathcal{E}', \quad \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J}', \quad \mathbf{J}_2 = \mathbf{E}_2, \quad \mathbf{E}_1, \mathbf{J}_1 \in \mathcal{V}, \quad \mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}, \\ \mathbf{E}_1 + \underline{\mathbf{E}}_2 &\in \mathcal{E}', \quad \underline{\mathbf{J}}_1 + \underline{\mathbf{J}}_2 \in \mathcal{J}', \quad \underline{\mathbf{J}}_2 = \underline{\mathbf{E}}_2, \quad \underline{\mathbf{J}}_1 \in \mathcal{V}, \quad \underline{\mathbf{E}}_2, \underline{\mathbf{J}}_2 \in \mathcal{H}, \end{aligned} \tag{7.322}$$

imply $\underline{\mathbf{J}}_1 = \mathbf{J}_1$. Subtracting these equations we see that

$$\mathbf{E} \equiv \mathbf{E}_2 - \underline{\mathbf{E}}_2 \in \mathcal{E}', \quad \mathbf{J} \equiv \mathbf{J}_1 + \mathbf{J}_2 - \underline{\mathbf{J}}_1 - \underline{\mathbf{J}}_2 \in \mathcal{J}', \quad \mathbf{J}_2 - \underline{\mathbf{J}}_2 = \mathbf{E}. \tag{7.323}$$

These imply

$$\mathbf{E} \in \mathcal{H}, \quad \mathbf{E} = \mathbf{J} - \mathbf{v}, \quad \text{where} \quad \mathbf{v} = \mathbf{J}_1 - \underline{\mathbf{J}}_1. \tag{7.324}$$

The uniqueness assumption means that these equations imply $\mathbf{v} = 0$ (and if $\mathbf{v} = 0$ then necessarily $\mathbf{E} = \mathbf{J} = 0$ since $\mathcal{E}'$ and $\mathcal{J}'$ have no vector in common). The relation $\mathbf{E} = \mathbf{J} - \mathbf{v}$ with $\mathbf{E} \in \mathcal{E}' \cap \mathcal{H}$ implies

$$\mathbf{E} = -\boldsymbol{\Gamma}_1' \mathbf{v}, \tag{7.325}$$

which will only have the trivial solution $\mathbf{v} = 0$ if and only if

$$\mathcal{H} \cap \widetilde{\mathcal{E}} = \{\mathbf{0}\} \quad \text{and} \quad \mathcal{V} \cap \mathcal{J}' = \{\mathbf{0}\}, \tag{7.326}$$

where the latter guarantees that $\boldsymbol{\Gamma}_1' \mathbf{v} = 0$ implies $\mathbf{v} = 0$.

We also assume that the Y-problem has a unique solution when $\mathbf{L} = \mathbf{I}$ for $\mathbf{E}_1 \in \mathcal{V}$ given $\mathbf{J}_1 \in \mathcal{V}$. By similar analysis this is satisfied if and only if

$$\mathcal{H} \cap \widetilde{\mathcal{J}} = \{\mathbf{0}\} \quad \text{and} \quad \mathcal{V} \cap \mathcal{E}' = \{\mathbf{0}\}. \tag{7.327}$$

We now establish that

$$\mathcal{W} \equiv \widetilde{\mathcal{E}} \oplus \widetilde{\mathcal{J}} = \mathcal{V} \oplus \mathcal{U}. \tag{7.328}$$

First note that $\mathcal{V}$ and $\mathcal{U}$ have no vector in common since $\mathcal{U} \subset \mathcal{H}$, and similarly $\widetilde{\mathcal{E}}$ and $\widetilde{\mathcal{J}}$ have no vector in common since $\mathcal{E}' \cap \mathcal{J}' = \{\mathbf{0}\}$. Clearly $\mathcal{W}$ contains $\mathcal{V}$. To show it contains $\mathcal{U}$ notice that

$$\mathcal{U} = \mathbf{\Pi}_2\mathbf{\Gamma}_1'\mathcal{V} = (\mathbf{I} - \mathbf{\Pi}_1)\mathbf{\Gamma}_1'\mathcal{V} \subset \mathbf{\Gamma}_1'\mathcal{V} \oplus \mathbf{\Pi}_1\mathbf{\Gamma}_1'\mathcal{V} \subset \widetilde{\mathcal{E}} \oplus \mathcal{V} \subset \mathcal{W}. \tag{7.329}$$

Together these imply $\mathcal{V} \oplus \mathcal{U} \subset \mathcal{W}$. Finally we have

$$\widetilde{\mathcal{E}} = \mathbf{\Gamma}_1'\mathcal{V} = (\mathbf{\Pi}_1 + \mathbf{\Pi}_2)\mathbf{\Gamma}_1'\mathcal{V} \subset \mathbf{\Pi}_1\mathbf{\Gamma}_1'\mathcal{V} \oplus \mathbf{\Pi}_2\mathbf{\Gamma}_1'\mathcal{V} \subset \mathcal{V} \oplus \mathcal{U}, \tag{7.330}$$

and similarly $\widetilde{\mathcal{J}} \subset \mathcal{V} \oplus \mathcal{U}$. Together these imply $\mathcal{W} \subset \mathcal{V} \oplus \mathcal{U}$, establishing (7.328).

If $\mathcal{V}$ has dimension m then $\widetilde{\mathcal{E}}$ must also have dimension m since otherwise $\mathbf{\Gamma}_1'\mathbf{v} = 0$ for some nonzero $\mathbf{v} \in \mathcal{V}$, implying $\mathbf{v} = \mathbf{\Gamma}_2'\mathbf{v}$ which only has the solution $\mathbf{v} = 0$ since $\mathcal{V} \cap \mathcal{J}' = \{\mathbf{0}\}$. Similarly $\widetilde{\mathcal{J}}$ must have dimension m and (7.328) then implies $\mathcal{U}$ must have dimension m. The first condition in (7.326) implies

$$\mathcal{W} = \mathcal{U} \oplus \widetilde{\mathcal{E}}, \tag{7.331}$$

since $\mathcal{U} \subset \mathcal{H}$ and $\widetilde{\mathcal{E}}$ have no vector in common and are m-dimensional spaces contained in the $2m$-dimensional space $\mathcal{W}$. Now any vector $\mathbf{E}' \in \mathcal{E}'$ has the unique decomposition

$$\mathbf{E}' = \mathbf{E}_1' + \mathbf{P}, \quad \mathbf{E}_1' \in \mathcal{V}, \quad \mathbf{P} \in \mathcal{H}, \tag{7.332}$$

and according to (7.331) $\mathbf{E}_1'$ has the unique decomposition

$$\mathbf{E}_1' = -\mathbf{e} + \widetilde{\mathbf{E}}, \quad \mathbf{e} \in \mathcal{U}, \quad \widetilde{\mathbf{E}} \in \widetilde{\mathcal{E}}. \tag{7.333}$$

So we have the decomposition

$$\mathbf{E}' = \widetilde{\mathbf{E}} + \mathbf{E}, \tag{7.334}$$

where

$$\mathbf{E} = \mathbf{P} - \mathbf{e} = \mathbf{E}' - \widetilde{\mathbf{E}} \in \mathcal{E}' \cap \mathcal{H} = \mathcal{E}. \tag{7.335}$$

Also the first condition in (7.326) implies $\widetilde{\mathcal{E}}$ and $\mathcal{E} \subset \mathcal{H}$ have no vector in common, so the decomposition is unique. Therefore we conclude that

$$\mathcal{E}' = \widetilde{\mathcal{E}} \oplus \mathcal{E}, \tag{7.336}$$

and similarly the first condition in (7.327) implies

$$\mathcal{J}' = \widetilde{\mathcal{J}} \oplus \mathcal{J}. \tag{7.337}$$

These and (7.328) imply

$$\mathcal{K} = \mathcal{V} \oplus \mathcal{H} = \widetilde{\mathcal{E}} \oplus \mathcal{E} \oplus \widetilde{\mathcal{J}} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}, \tag{7.338}$$

and since $\mathcal{U}$, $\mathcal{E}$ and $\mathcal{J}$ are all contained in $\mathcal{H}$ we conclude that

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n. \tag{7.339}$$

Now a given $\mathbf{E}'_1 \in \mathcal{V}$ has the unique decomposition (7.333). This defines the nonsingular operator $\mathbf{K}$: $\mathcal{V} \to \mathcal{U}$ such that $\mathbf{e} = \mathbf{K}\mathbf{E}'_1$. (It is nonsingular because $\mathcal{V}$ and $\widetilde{\mathcal{E}} \subset \mathcal{E}'$ have no nonzero vector in common.) Now given $\mathbf{e}$, consider the solution to

$$\mathbf{e}, \mathbf{j} \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{j} + \mathbf{J} = \mathbf{L}(\mathbf{e} + \mathbf{E}), \quad \text{where } \mathbf{L} = \sum_{i=1}^{n} z_i \mathbf{\Lambda}_i, \tag{7.340}$$

where $\mathbf{\Lambda}_i$ is the projection onto $\mathcal{P}_i$, and from the definition of $\mathbf{Z}$, $\mathbf{j} = \mathbf{Z}\mathbf{e}$. Since the second condition in (7.326) implies $\mathcal{V}$ and $\widetilde{\mathcal{J}}$ have no vector in common we have

$$\mathcal{W} = \mathcal{V} \oplus \widetilde{\mathcal{J}}, \tag{7.341}$$

and consequently any $\mathbf{j} \in \mathcal{U}$ has the decomposition

$$\mathbf{j} = -\mathbf{J}'_1 + \widetilde{\mathbf{J}}, \quad \mathbf{J}'_1 \in \mathcal{V}, \quad \widetilde{\mathbf{J}} \in \widetilde{\mathcal{J}}, \tag{7.342}$$

which defines the nonsingular operator $\mathbf{M} : \mathcal{U} \to \mathcal{V}$ such that $\mathbf{J}'_1 = \mathbf{M}\mathbf{j}$. Defining

$$\mathbf{E}'_2 = \mathbf{e} + \mathbf{E}, \quad \mathbf{J}'_2 = \mathbf{j} + \mathbf{J}, \tag{7.343}$$

we have

$$\begin{aligned} \mathbf{E}'_1 + \mathbf{E}'_2 &= \mathbf{E}'_1 + \mathbf{e} + \mathbf{E} = \widetilde{\mathbf{E}} + \mathbf{E} \in \mathcal{E}', \\ \mathbf{J}'_1 + \mathbf{J}'_2 &= \mathbf{J}'_1 + \mathbf{j} + \mathbf{J} = \widetilde{\mathbf{J}} + \mathbf{J} \in \mathcal{J}', \end{aligned} \tag{7.344}$$

and

$$\mathbf{J}'_1 = \mathbf{M}\mathbf{j} = \mathbf{M}\mathbf{Z}\mathbf{e} = \mathbf{M}\mathbf{Z}\mathbf{K}\mathbf{E}'_1, \tag{7.345}$$

which by definition of the associated Y-function implies

$$\mathbf{Y}(z_1, z_2, \ldots, z_n) = \mathbf{M}\mathbf{Z}(z_1, z_2, \ldots, z_n)\mathbf{K}. \tag{7.346}$$

This is analogous to the relation (20.29) in Milton (2002) obtained in the case where the subspaces are mutually orthogonal.

In particular by letting $z_1 = z_2 = \ldots = z_n = 1$ we obtain

$$\mathbf{Y}(1, 1, \ldots, 1) = \mathbf{M}\mathbf{K}. \tag{7.347}$$

If $\mathbf{v}_1$, $\mathbf{v}_2$,... $\mathbf{v}_m$ are a basis for $\mathcal{V}$, and we choose $\mathbf{K}\mathbf{v}_1$, $\mathbf{K}\mathbf{v}_2$,... $\mathbf{K}\mathbf{v}_m$ as our basis for $\mathcal{U}$ then with these bases $\mathbf{K}$ is represented by the identity matrix $\mathbf{K} = \mathbf{I}$ and (7.346) and (7.347) imply

$$\mathbf{Y}(z_1, z_2, \ldots, z_n) = \mathbf{Y}(1, 1, \ldots, 1)\mathbf{Z}(z_1, z_2, \ldots, z_n). \tag{7.348}$$

7.21 Reduction operations on subspace collections

Extension is one way to go from a $Z(n)$ subspace collection to a $Y(n)$ subspace collection. Another way is through reduction, which has some features in common with normalization. Given a $Z(n)$ subspace collection

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \cdots \oplus \mathcal{P}_n, \tag{7.349}$$

let $\mathbf{\Gamma}_0$ be the projection onto $\mathcal{U}$, and let $\mathbf{\Lambda}_j$ be the projection onto $\mathcal{P}_j$. Define

$$\begin{aligned} \mathcal{K} = \mathcal{E} \oplus \mathcal{J}, \quad \mathcal{P}'_j = \mathcal{P}_j \cap \mathcal{K} \quad \text{for } j = 1, 2, \ldots, n, \\ \mathcal{V} = (\mathbf{I} - \mathbf{\Gamma}_0)[\mathbf{\Lambda}_1\mathcal{U} \oplus \mathbf{\Lambda}_2\mathcal{U} \oplus \cdots \oplus \mathbf{\Lambda}_n\mathcal{U}] \subset \mathcal{K}, \quad \widetilde{\mathcal{P}}_j = \mathbf{\Lambda}_j\mathcal{U}. \end{aligned} \tag{7.350}$$

We now establish that

$$\mathcal{W} \equiv \widetilde{\mathcal{P}}_1 \oplus \widetilde{\mathcal{P}}_2 \oplus \cdots \oplus \widetilde{\mathcal{P}}_n = \mathcal{U} \oplus \mathcal{V}. \tag{7.351}$$

First note that $\mathcal{V}$ and $\mathcal{U}$ have no vector in common since $\mathcal{V} \subset \mathcal{K}$, and similarly the subspaces $\widetilde{\mathcal{P}}_j$ have no vector in common since $\widetilde{\mathcal{P}}_j \subset \mathcal{P}_j$. Clearly $\mathcal{W}$ contains $\mathcal{U}$ since the projections $\mathbf{\Lambda}_j$ sum to the identity. To show it contains $\mathcal{V}$ note that

$$\mathcal{V} \subset \mathbf{\Lambda}_1\mathcal{U} \oplus \mathbf{\Lambda}_2\mathcal{U} \oplus \cdots \oplus \mathbf{\Lambda}_n\mathcal{U} + \mathbf{\Gamma}_0[\mathbf{\Lambda}_1\mathcal{U} \oplus \mathbf{\Lambda}_2\mathcal{U} \oplus \cdots \oplus \mathbf{\Lambda}_n\mathcal{U}] \subset \mathcal{W} + \mathcal{U} = \mathcal{W}. \tag{7.352}$$

Therefore we have that $\mathcal{U} \oplus \mathcal{V} \subset \mathcal{W}$. The converse inclusion that $\mathcal{W} \subset \mathcal{U} \oplus \mathcal{V}$ follows from the inclusion

$$\widetilde{\mathcal{P}}_j = [\mathbf{\Gamma}_0 + (\mathbf{I} - \mathbf{\Gamma}_0)]\mathbf{\Lambda}_j\mathcal{U} \subset \mathcal{U} \oplus \mathcal{V}, \tag{7.353}$$

which establishes (7.351). Next, to establish that for all j,

$$\mathcal{P}_j = \widetilde{\mathcal{P}}_j \oplus \mathcal{P}'_j, \tag{7.354}$$

we need to assume that for all j

$$\widetilde{\mathcal{P}}_j \cap \mathcal{K} = \{\mathbf{0}\}, \tag{7.355}$$

and that

$$\mathbf{\Lambda}_j\mathbf{u} = 0, \quad \mathbf{u} \in \mathcal{U} \tag{7.356}$$

only has the trivial solution $\mathbf{u} = 0$, i.e.,

$$\mathcal{U} \cap (\mathcal{P}_1 \oplus \mathcal{P}_2 \oplus \ldots \oplus \mathcal{P}_{j-1} \oplus \mathcal{P}_{j+1} \oplus \ldots \oplus \mathcal{P}_n) = \{\mathbf{0}\}. \tag{7.357}$$

These conditions imply that

$$\mathcal{U} = \mathbf{\Gamma}_0\mathbf{\Lambda}_j\mathcal{U}, \tag{7.358}$$

and hence that

$$\mathcal{U} \subset \mathbf{\Lambda}_j\mathcal{U} \oplus (\mathbf{I} - \mathbf{\Gamma}_0)\mathbf{\Lambda}_j\mathcal{U}, \tag{7.359}$$

which in turn implies that

$$\mathcal{U} \subset \widetilde{\mathcal{P}}_j + \mathcal{V}. \tag{7.360}$$

Then any vector $\mathbf{P} \in \mathcal{P}_j$ has the unique decomposition

$$\mathbf{P} = \mathbf{u} + \mathbf{K}, \ \text{ with } \ \mathbf{u} \in \mathcal{U}, \ \ \mathbf{K} \in \mathcal{K}, \tag{7.361}$$

and according to (7.360), $\mathbf{u}$ has the unique decomposition

$$\mathbf{u} = \mathbf{v} + \widetilde{\mathbf{P}} \text{ with } \mathbf{v} \in \mathcal{V}, \ \widetilde{\mathbf{P}} \in \widetilde{\mathcal{P}}_j, \tag{7.362}$$

which is unique because $\mathcal{V} \subset \mathcal{K}$ and $\widetilde{\mathcal{P}}_j$ have no nonzero vector in common. Therefore $\mathbf{P}$ has the unique decomposition

$$\mathbf{P} = \widetilde{\mathbf{P}} + \mathbf{P}', \tag{7.363}$$

where

$$\mathbf{P}' = \mathbf{v} + \mathbf{K} = \mathbf{P} - \widetilde{\mathbf{P}} \in \mathcal{P}_j \cap \mathcal{K} = \mathcal{P}'_j. \tag{7.364}$$

This decomposition and the fact that (7.355) implies $\widetilde{\mathcal{P}}_j$ and $\mathcal{P}'_j \subset \mathcal{K}$ have no vector in common establishes (7.354).

So we deduce that

$$\begin{aligned} \mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} &= \widetilde{\mathcal{P}}_1 \oplus \widetilde{\mathcal{P}}_2 \oplus \cdots \oplus \widetilde{\mathcal{P}}_n \oplus \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_n \\ &= \mathcal{U} \oplus \mathcal{V} \oplus \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_n, \end{aligned} \tag{7.365}$$

and since the $\mathcal{P}'_j$, $j = 1, 2, \ldots, n$ are all contained in $\mathcal{K}$ it follows that

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_n. \tag{7.366}$$

Now suppose that given $\mathbf{e} \in \mathcal{U}$ we can solve the equations

$$\mathbf{j} + \mathbf{J}_1 = \mathbf{L}(\mathbf{e} + \mathbf{E}_1), \quad \mathbf{J}_1 = -\mathbf{Y}\mathbf{E}_1, \quad \mathbf{e}, \mathbf{j} \in \mathcal{U}, \quad \mathbf{E}_1, \mathbf{J}_1 \in \mathcal{V}, \tag{7.367}$$

where $\mathbf{Y}$ is the Y-operator associated with the subspace collection (7.366). From the Y-problem we have

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \in \mathcal{E}, \quad \mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \in \mathcal{J} \quad \mathbf{J}_2 = \mathbf{L}\mathbf{E}_2, \quad \mathbf{E}_2, \mathbf{J}_2 \in \mathcal{H}', \tag{7.368}$$

where

$$\mathcal{H}' = \mathcal{P}'_1 \oplus \mathcal{P}'_2 \oplus \cdots \oplus \mathcal{P}'_n. \tag{7.369}$$

Since

$$\mathbf{j} + \mathbf{J}_1 + \mathbf{J}_2 = \mathbf{L}(\mathbf{e} + \mathbf{E}_1 + \mathbf{E}_2), \tag{7.370}$$

we see that these fields solve the Z-problem

$$\mathbf{e}, \mathbf{j} \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \quad \mathbf{j} + \mathbf{J} = \mathbf{L}(\mathbf{e} + \mathbf{E}), \tag{7.371}$$

and by definition $\mathbf{j} = \mathbf{Z}\mathbf{e}$. To solve (7.367) let $\mathbf{\Pi}_1$ be the projection onto $\mathcal{V}$. Then (7.367) implies

$$-\mathbf{Y}\mathbf{E}_1 = \mathbf{\Pi}_1\mathbf{L}(\mathbf{e} + \mathbf{\Pi}_1\mathbf{E}_1), \tag{7.372}$$

giving

$$\mathbf{E}_1 = -\mathbf{\Pi}_1(\mathbf{Y} + \mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{e}, \tag{7.373}$$

where the inverse is to be taken on the subspace $\mathcal{V}$. It follows that

$$\mathbf{j} + \mathbf{J}_1 = \mathbf{L}\mathbf{e} - \mathbf{L}\mathbf{\Pi}_1(\mathbf{Y} + \mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{e}, \tag{7.374}$$

implying

$$\mathbf{Z} = \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Pi}_1(\mathbf{Y} + \mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{\Gamma}_0. \tag{7.375}$$

This formula is analogous to that given in (29.12) of Milton (2002).

To obtain a more explicit way of writing (7.375) let us suppose we are given a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ of $\mathcal{U}$. Since (7.356) only has the trivial solution $\mathbf{u} = 0$ each space $\widetilde{\mathcal{P}}_j$ has dimension m. It then follows from (7.351) that $\mathcal{V}$ has dimension $m(n-1)$. Also, for $i = 1, 2, \ldots, n-1$, (7.351) implies $\mathbf{\Lambda}_i\mathbf{u}_j$ has the unique decomposition

$$\mathbf{\Lambda}_i\mathbf{u}_j = \sum_k w_{ijk}\mathbf{u}_k + \mathbf{v}_{ij}, \quad \mathbf{v}_{ij} \in \mathcal{V}, \tag{7.376}$$

for some set of constants w_{ijk}. To show that the vectors $\mathbf{v}_{ij}$, which number $m(n-1)$, are independent, let us suppose

$$0 = \sum_{i=1}^{n-1}\sum_{j=1}^{m} c_{ij}\mathbf{v}_{ij} = \sum_{i=1}^{n-1}\sum_{j=1}^{m} c_{ij}(\mathbf{\Lambda}_i\mathbf{u}_j - \sum_{k=1}^{m} w_{ijk}\mathbf{u}_k). \tag{7.377}$$

By letting $\mathbf{\Lambda}_n$ act on this equation and taking into account that (7.356) only has the trivial solution $\mathbf{u} = 0$ we see that

$$\sum_{i=1}^{n-1}\sum_{j=1}^{m}\sum_{k=1}^{m} c_{ij}w_{ijk}\mathbf{u}_k = 0. \tag{7.378}$$

Then substituting this in (7.377) and letting $\mathbf{\Lambda}_i$, $i \neq n$, act on (7.377) and again taking into account that (7.356) only has the trivial solution $\mathbf{u} = 0$ we obtain

$$\sum_{j=1}^{m} c_{ij}\mathbf{u}_j = 0, \tag{7.379}$$

which shows that all the c_{ij} must be zero. Therefore let us take the vectors $\mathbf{v}_{ij}$ as our basis for $\mathcal{V}$.

The identities

$$\mathbf{\Pi}_1\mathbf{\Lambda}_i\mathbf{\Gamma}_0\mathbf{u}_j = \mathbf{v}_{ij}, \quad \mathbf{\Gamma}_0\mathbf{\Lambda}_i\mathbf{\Gamma}_0\mathbf{u}_j = \sum_k w_{ijk}\mathbf{u}_k, \tag{7.380}$$

which follow from (7.376) then gives the matrix representations for $\mathbf{\Pi}_1\mathbf{\Lambda}_i\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_0\mathbf{\Lambda}_i\mathbf{\Gamma}_0$ in these bases, when $i \neq m$. Using the fact that $\mathbf{\Lambda}_n = \mathbf{I} - \sum_{i\neq n}\mathbf{\Lambda}_i$ we obtain

$$\mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 = z_n\mathbf{\Gamma}_0 + \sum_{i=1}^{n-1}(z_i - z_n)\mathbf{\Gamma}_0\mathbf{\Lambda}_i\mathbf{\Gamma}_0, \quad \mathbf{\Pi}_1\mathbf{L}\mathbf{\Gamma}_0 = \sum_{i=1}^{n-1}(z_i - z_n)\mathbf{\Pi}_1\mathbf{\Lambda}_i\mathbf{\Gamma}_0. \tag{7.381}$$

Now for $p \neq n$ (and $i \neq n$) (7.376) implies (no sum over p)

$$\begin{aligned}\mathbf{\Lambda}_p\mathbf{v}_{ij} &= \sum_k(\delta_{pi}\delta_{kj} - w_{ijk})\mathbf{\Lambda}_p\mathbf{u}_k \\ &= \sum_k(\delta_{pi}\delta_{kj} - w_{ijk})(\mathbf{v}_{pk} + \sum_q w_{pkq}\mathbf{u}_q).\end{aligned} \tag{7.382}$$

Thus we deduce

$$\begin{aligned}\mathbf{\Gamma}_0\mathbf{\Lambda}_p\mathbf{\Pi}_1\mathbf{v}_{ij} &= \sum_k(\delta_{pi}\delta_{kj}-w_{ijk})\sum_q w_{pkq}\mathbf{u}_q,\\ \mathbf{\Pi}_1\mathbf{\Lambda}_p\mathbf{\Pi}_1\mathbf{v}_{ij} &= \sum_k(\delta_{pi}\delta_{kj}-w_{ijk})\mathbf{v}_{pk},\end{aligned} \tag{7.383}$$

which gives the matrix representation for the operators $\mathbf{\Gamma}_0\mathbf{\Lambda}_p\mathbf{\Pi}_1$ and $\mathbf{\Pi}_1\mathbf{\Lambda}_p\mathbf{\Pi}_1$ in these bases ($p \neq n$), in terms of which we obtain the representation for the operators

$$\mathbf{\Gamma}_0\mathbf{L}\mathbf{\Pi}_1 = \sum_{p=1}^{n-1}(z_p-z_n)\mathbf{\Gamma}_0\mathbf{\Lambda}_p\mathbf{\Pi}_1, \quad \mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 = z_n\mathbf{\Pi}_1+\sum_{p=1}^{n-1}(z_p-z_n)\mathbf{\Pi}_1\mathbf{\Lambda}_p\mathbf{\Pi}_1. \tag{7.384}$$

Thus all the matrices representing the operators entering (7.375), aside from $\mathbf{Y}$, only depend on the parameters w_{ijk} and these parameters can be obtained from the representation in the basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ of $\mathbf{Z}$ when the differences $z_i - z_n$, $i = 1, 2, \ldots, n-1$ are small. To first order in these differences, (7.375), (7.381), and (7.384) imply

$$\mathbf{Z}\mathbf{u}_j \approx z_n\mathbf{u}_j + \sum_{i=1}^{n-1}(z_i-z_n)\sum_k w_{ijk}\mathbf{u}_k. \tag{7.385}$$

Thus knowing this expansion one can recover all the parameters w_{ijk}.

7.22 "Continued fraction expansions" of subspace collections

The idea to developing the continued fraction is that by a succession of reduction and normalization operations one obtains a series of recursion relations

$$\mathbf{Z} = \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Pi}_1(\mathbf{Y}+\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{\Gamma}_0, \tag{7.386}$$

$$\mathbf{Y} = \mathbf{M}^{(1)}\mathbf{Z}^{(1)}\mathbf{K}^{(1)}, \tag{7.387}$$

$$\mathbf{Z}^{(1)} = \mathbf{\Gamma}_0^{(1)}\mathbf{L}^{(1)}\mathbf{\Gamma}_0^{(1)} - \mathbf{\Gamma}_0^{(1)}\mathbf{L}^{(1)}\mathbf{\Pi}_1^{(1)}(\mathbf{Y}^{(1)}+\mathbf{\Pi}_1^{(1)}\mathbf{L}^{(1)}\mathbf{\Pi}_1^{(1)})^{-1}\mathbf{\Pi}_1^{(1)}\mathbf{L}^{(1)}\mathbf{\Gamma}_0^{(1)}, \tag{7.388}$$

$$\mathbf{Y}^{(1)} = \mathbf{M}^{(2)}\mathbf{Z}^{(2)}\mathbf{K}^{(2)}, \tag{7.389}$$

$$\mathbf{Z}^{(2)} = \mathbf{\Gamma}_0^{(2)}\mathbf{L}^{(2)}\mathbf{\Gamma}_0^{(2)} - \mathbf{\Gamma}_0^{(2)}\mathbf{L}^{(2)}\mathbf{\Pi}_1^{(2)}(\mathbf{Y}^{(2)}+\mathbf{\Pi}_1^{(2)}\mathbf{L}^{(2)}\mathbf{\Pi}_1^{(2)})^{-1}\mathbf{\Pi}_1^{(2)}\mathbf{L}^{(2)}\mathbf{\Gamma}_0^{(2)}, \tag{7.390}$$

and so forth, until the dimension of the remaining space goes to zero, or until one (or more) of the assumptions necessary to proceed with the normalization or reduction operation does not hold. By substituting (7.387) in (7.386), then substituting (7.388) in the resulting expression, and subsequently substituting (7.389) in this expression, and so on, one develops the continued fraction expansion for $\mathbf{Z}$ incorporating the variables z_1, z_2, ..., z_n and, as one goes down the continued fraction, information contained in the series expansion (7.235) at successively higher and higher levels of truncation. We do not address in this book whether one can go ahead with the continued fraction expansion (and if so how) when the assumptions made to proceed with the normalization or reduction operation do not hold. In the process of developing the continued fraction through reduction and normalization operations, one could at those steps where one is dealing with a Y-problem make any desired reference transformation as described in Section 7.12. In this way one incorporates information

at the subspace collection level that corresponds at the function level to known values of the function, and derivatives, at various points.

Such continued fraction expansions form the basis of the field equation recursion method for bounding the effective moduli of composites (Milton and Golden 1985; Milton 1987a, 1987b, 1991; Clark and Milton 1994; Clark 1997 and **Chapter 29** of (Milton 2002) in the abstract theory of composites as described in **Chapter 2** of this book: see also Section 9.10 and **Chapter 10** of this book). The basic idea, at least when we have an orthogonal subspace collection, is that crude estimates or bounds on the operator $\mathbf{Z}^{(j)}$ or $\mathbf{Y}^{(j)}$ at some intermediate level j give through the above recursion relations good approximations or tight bounds on $\mathbf{Z}$ or $\mathbf{Y}$ incorporating the parameters that enter the recursion relations at the different levels up to level j (obtained from series expansions up to a given order of the solutions of the Z-problem or Y-problem).

Acknowledgments

G.W. Milton thanks his husband John K. Patton for suggesting the name superfunction. The help of Oscar Bruno is gratefully acknowledged in producing **Figure 7.8**.

8 Substitution of subspace collections with nonorthogonal subspaces to accelerate Fast Fourier Transform methods applied to conducting composites

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

In this chapter we show the power of the algebra of subspace collections developed in the previous chapter. Specifically we accelerate the Fast Fourier Transform schemes of Moulinec and Suquet (1994, 1998) and Eyre and Milton (1999), for computing the fields and effective tensor in a conducting periodic medium by substituting a subspace collection with nonorthogonal subspaces inside one with orthogonal subspaces. This can be done when the effective conductivity as a function of the conductivity σ_1 of the inclusion phase (with the matrix phase conductivity set to 1) has its singularities confined to an interval $[-\beta, -\alpha]$ of the negative real σ_1 axis. Numerical results of Moulinec and Suquet show accelerated convergence for the model example of a square array of squares at 25% volume fraction. For other problems we show how Q_C^*-convex functions can be used to restrict the region where singularities of the effective tensor as a function of the component tensors might be found.

8.1 Introduction

Subspace collections with orthogonal subspaces obviously have a lot of relevance to physical problems. Here we give an example to show that subspace collections with nonorthogonal subspaces are an important tool in analysis, reaching beyond their connection with rational functions of several complex variables. This chapter is mostly self-contained. While it uses nonorthogonal subspace collections, it is not necessary for the reader to have read the previous chapter, **Chapter 7**. However, it is recommended for the reader to look at **Chapters 1** and **2** first.

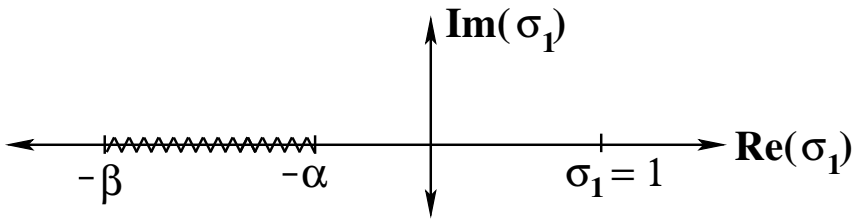


Figure 8.1: We assume that in the complex σ_1-plane, the function $\sigma_*(\sigma_1)$ has all its singularities on the negative real σ_1-axis, between $\sigma_1 = -\alpha$ and $\sigma_1 = -\beta$. Here the zig-zag line denotes a possible branch cut.

The specific problem we consider is a two-component conducting medium with conductivities σ_1 and σ_2, phase 1 being the inclusion phase. The dimension is $d = 2$ or $d = 3$. Without loss of generality we can take $\sigma_2 = 1$. Assume for simplicity that the composite has square $(d = 2)$ or cubic $(d = 3)$ symmetry so that the effective conductivity tensor is isotropic, taking the form $\sigma_*\mathbf{I}$. Bergman (1978) made the pioneering observation that the function $\sigma_*(\sigma_1)$ is an analytic function of σ_1 and has all its singularities on the negative real axis. He made some assumptions which were not valid (Milton 1979). These assumptions could be circumvented by approximating the composite by a large resistor network (Milton 1981a). Subsequently Golden and Papanicolaou (1983, 1985) gave a rigorous proof of the analytic properties. Here we make the additional assumption (only true for some geometries) that the function $\sigma_*(\sigma_1)$ has all its singularities confined to an interval $[-\beta, -\alpha]$ on the negative real axis, $\alpha > 0$ and $\beta > \alpha$ (see **Figure 8.1**). This information results in tighter bounds on the conductivity and complex conductivity, or equivalently the complex dielectric constant, (Bruno 1991a; Sawicz and Golden 1995; Golden 1998), and these bounds may be inverted to yield information about α and β from experimental measurements (Orum, Cherkaev, and Golden 2012). By contrast, here our goal is to utilize the information to improve the speed of convergence of Fast Fourier Transform methods for computing the fields in composites and their associated effective tensors. These methods were first introduced by Moulinec and Suquet (1994, 1998): see Moulinec and Silva (2014) for a recent review. One important application has been to viscoplastic polycrystals (Lebensohn 2001, Lee, Lebensohn, and Rollett 2011). For materials with a linear response, previous significant advances in the acceleration of these schemes were made by Eyre and Milton (1999) (see also the generalization in section 14.9 of Milton 2002, and in particular equation (14.38)), and by Willot, Abdallah, and Pellegrini (2014) and Willot (2015). The singularities in the interval $[-\beta, -\alpha]$ could be poles or could be a branch cut (if the inclusion has sharp corners, or if there is some randomness in the geometry). Note it is not only the effective conductivity which has this analytic form but also the electric field $\mathbf{E}(\mathbf{x}, \sigma_1)$ and current field $\mathbf{J}(\mathbf{x}, \sigma_1)$ at fixed $\mathbf{x}$, and fixed applied field, with σ_1 varying. If we didn't know anything about α and β there could potentially be singularities anywhere along the negative real σ_1 axis. Let's first look at the case $\alpha = 0$, $\beta = \infty$.

The original Fast Fourier Transform scheme of Moulinec and Suquet (1994, 1998) is based on series expansions, such as (2.35), which for conductivity takes the form

$$\boldsymbol{\sigma}_* = \sigma_0\mathbf{I} + \sum_{j=0}^{\infty} \boldsymbol{\Gamma}_0[\boldsymbol{\sigma}(\mathbf{x}) - \sigma_0\mathbf{I}][\boldsymbol{\Gamma}_1(\mathbf{I} - \boldsymbol{\sigma}/\sigma_0)]^j\boldsymbol{\Gamma}_0, \quad \mathbf{e} = \mathbf{e}_0 + \sum_{j=0}^{\infty}[\boldsymbol{\Gamma}_1(\mathbf{I} - \boldsymbol{\sigma}/\sigma_0)]^j\mathbf{e}_0, \tag{8.1}$$

for the effective conductivity $\boldsymbol{\sigma}_*$, and electric field $\mathbf{e}(\mathbf{x})$, with $\mathbf{e}_0$ being the applied (average) electric field. [To obtain (8.1) from (2.35) we use the fact that $\boldsymbol{\Gamma}_0(\sigma_0\mathbf{I})\boldsymbol{\Gamma}_1 = 0$.] Their key and beautiful idea was that since the

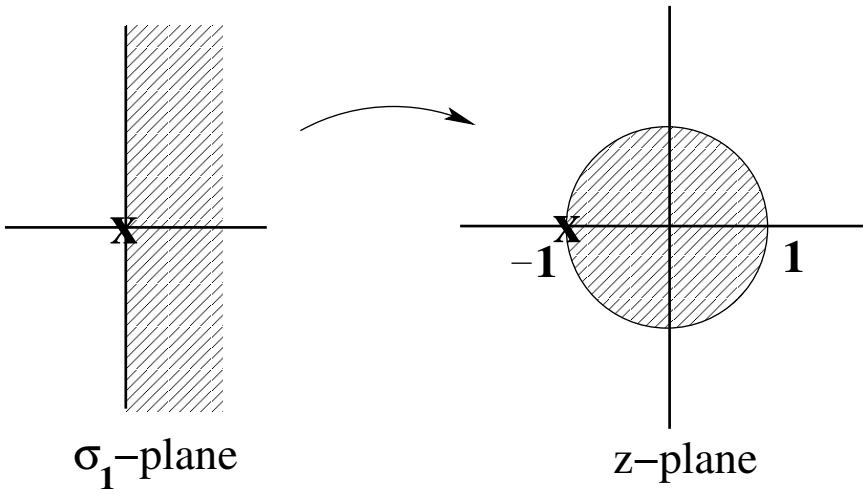


Figure 8.2: If we want a series expansion which converges in the entire right half of the σ_1-plane, when a singularity is located at the point $\mathbf{X}$, then we make a fractional linear transformation which takes the right half of the σ_1-plane to the unit disk in the z-plane, and find an series expansion in powers of z. The scheme of Moulinec and Suquet (1994, 1998) provides such an expansion.

action of $\boldsymbol{\Gamma}_1$ is readily evaluated in Fourier space, while the action of $(\mathbf{I} - \boldsymbol{\sigma}/\sigma_0)$ is readily evaluated in real space, both $\boldsymbol{\sigma}_*$ and $\mathbf{e}(\mathbf{x})$ can be readily calculated from these series by going back and forth between real and Fourier space, using Fast Fourier Transforms to do so.

In a two-phase medium with isotropic conductivities σ_1 and $\sigma_2 = 1$, and with the choice of $\sigma_0 = \frac{1}{2}(\sigma_1 + \sigma_2) = \frac{1}{2}(\sigma_1 + 1)$ used by Moulinec and Suquet, $(\mathbf{I} - \boldsymbol{\sigma}/\sigma_0)$ takes the value $(1 - \sigma_1)/(\sigma_1 + 1)$ in phase 1 and the value $-(1 - \sigma_1)/(\sigma_1 + 1)$ in phase 2. So it is clear that this scheme gives an expansion of the form

$$\sigma_*/\sigma_0 = 1 + \sum_{n=1}^{\infty} a_n \left(\frac{\sigma_1 - 1}{\sigma_1 + 1} \right)^n . \tag{8.2}$$

The transformation $z = (\sigma_1 - 1)/(\sigma_1 + 1)$ maps the right half of the complex plane to the unit disk (see **Figure 8.2**). So with $\alpha = 0$ the series will converge if $|z| < 1$, i.e., in the entire right half of the complex σ_1-plane, $\mathrm{Re}(\sigma_1) > 0$. (Recall that it is the distance from the origin to the nearest singularity in the $z-$plane which determines the radius of convergence in the $z-$plane, and hence the region of convergence in the σ_1-plane.)

With the accelerated scheme of Eyre and Milton (1999) (see also the generalization in section 14.9 of Milton 2002, and in particular equation (14.38)), one has the expansion

$$\sigma_*/\sqrt{\sigma_1} = 1 + \sum_{n=1}^{\infty} b_n \left(\frac{\sqrt{\sigma_1} - 1}{\sqrt{\sigma_1} + 1} \right)^n . \tag{8.3}$$

The transformation $z = (w-1)/(w+1)$ where $w = \sqrt{\sigma_1}$ maps the σ_1 complex plane minus the slit along the negative real axis to the unit disk $|z| = 1$ (see **Figure 8.3**). So with $\alpha = 0$ the series will converge if $|z| < 1$, i.e., in the entire complex plane minus the slit along the negative real axis. These arguments show that when

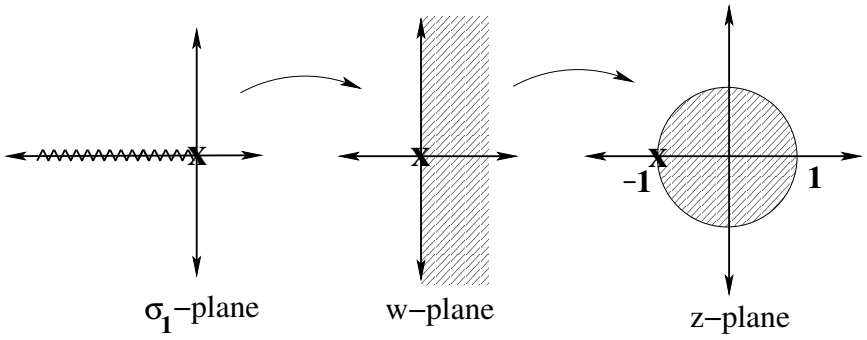


Figure 8.3: If we want a series expansion which converges in the entire the σ_1-plane minus the negative real σ_1-axis, then we first make a square root transformation which maps the cut complex σ_1-plane to the right half of the w-plane, followed by a fractional linear transformation which takes it to the unit disk in the z-plane, and find an expansion in powers of z. The scheme of Eyre and Milton (1999) provides such an expansion.

$\alpha = 0$ the accelerated scheme should have a larger region of convergence, and by the same line of reasoning a faster rate of convergence when both schemes converge. (Note, however, that if $\alpha > 0$ and $\beta < \infty$ the scheme of Moulinec and Suquet could outperform that of Eyre and Milton for small and very large values of σ_1, since the scheme of Moulinec and Suquet should then converge for sufficiently small or sufficiently large negative real values of σ_1.)

If we know α and β (or bounds for them) then it makes sense to use a transformation which maps the complex plane minus the slit $[-\beta, -\alpha]$ to the unit circle $|z| = 1$. Since the transformation

$$t = \frac{(\sigma_1 + \alpha)(1 + \beta)}{(\sigma_1 + \beta)(1 + \alpha)} = 1 + \frac{(\sigma_1 - 1)(\beta - \alpha)}{(\sigma_1 + \beta)(1 + \alpha)} \tag{8.4}$$

maps the interval $[-\beta, -\alpha]$ to $[-\infty, 0]$ and takes $\sigma_1 = 1$ to $t = 1$, it is clear that the transformation

$$z = \frac{\sqrt{t} - 1}{\sqrt{t} + 1} \quad \text{(with } t \text{ given above)} \tag{8.5}$$

maps the complex σ_1 plane minus the slit $[-\beta, -\alpha]$ to the unit disk $|z| = 1$ (see **Figure 8.4**). So we want to get a Fast Fourier Transform (FFT) method associated with an expansion

$$\frac{\sigma_*}{\sqrt{[(\sigma_1 + \alpha)(1 + \beta)]/[(\sigma_1 + \beta)(1 + \alpha)]}} = 1 + \sum_{n=1}^{\infty} c_n \left(\frac{\sqrt{\frac{(\sigma_1+\alpha)(1+\beta)}{(\sigma_1+\beta)(1+\alpha)}} - 1}{\sqrt{\frac{(\sigma_1+\alpha)(1+\beta)}{(\sigma_1+\beta)(1+\alpha)}} + 1} \right)^n . \tag{8.6}$$

It looks rather formidable but really it is just a matter of substituting

$$t = \frac{(\sigma_1 + \alpha)(1 + \beta)}{(\sigma_1 + \beta)(1 + \alpha)} \tag{8.7}$$

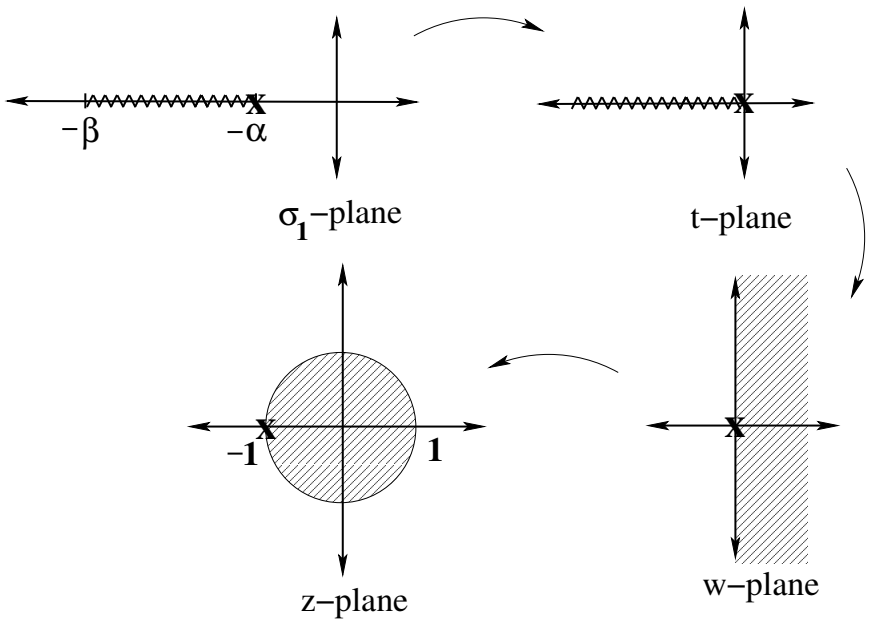


Figure 8.4: If we want a series expansion which converges in the entire the σ_1-plane minus the cut on σ_1-axis between $-\beta$ and $-\alpha$ then we first use a fractional linear transformation which maps this cut to the entire negative real axis in the t-plane, followed by a square root transformation mapping the cut complex t-plane to the right half of the w-plane, followed by a fractional linear transformation which takes it to the unit disk in the z-plane, and find an expansion in powers of z. The scheme developed in this chapter provides such an expansion.

in the Eyre and Milton (1999) scheme associated with

$$\sigma_*/\sqrt{t} = 1 + \sum_{n=1}^{\infty} c_n \left(\frac{\sqrt{t}-1}{\sqrt{t}+1} \right)^n . \tag{8.8}$$

But of course one wants to do this substitution at the level of the underlying Hilbert space, not just in the conductivity function.

8.2 Substitution at the level of the Hilbert space

Substitution at the level of the Hilbert space requires the substitution of subspace collections where the subspaces do not satisfy the orthogonality property. At a discrete level it is easy to get an idea of how this can be done. If we consider the composite as a resistor network, in phase 1 with resistors having resistance $R_1 = \delta/\sigma_1$, while in phase 2 with resistors having resistance $R_2 = \delta/\sigma_2$, so e.g., the composite is replaced by a network (see **Figure 8.5**(a)). Then at this level we could replace each resistor R_1 by the compound resistor

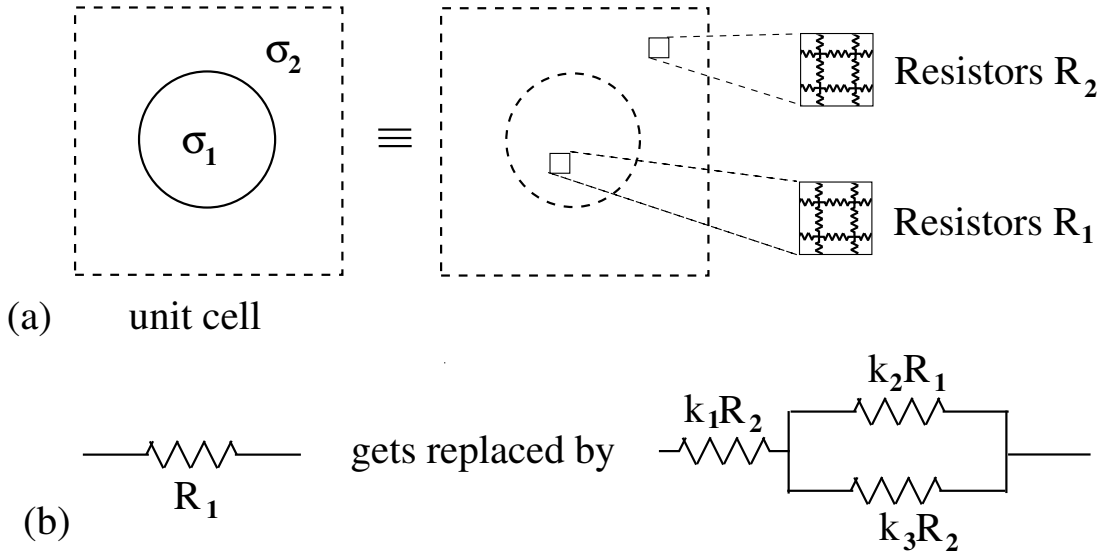


Figure 8.5: A composite containing two phases, as for example a periodic array of disks with a unit cell as shown in (a), can be approximated by a discrete network of resistors having resistors R_2 and R_1. At this discrete level we are free to replace each resistor having resistance R_1 by the combination of resistors in figure (b). This transformation at the Hilbert-space level, corresponds to a fractional linear transformation in the σ_1-plane.

(see **Figure 8.5**(b)) with k_1, k_2, k_3 real constants. That is the resistance $R_1 = \delta/\sigma_1$ gets replaced by

$$\begin{aligned} k_1 R_2 + \frac{1}{1/(k_2 R_1) + 1/(k_3 R_2)} &= \frac{k_1}{\sigma_2} + \frac{1}{\sigma_1/k_2 + \sigma_2/k_3} \\ &= \frac{k_1\sigma_1/k_2 + k_1/k_3 + 1}{\sigma_1/k_2 + 1/k_3}, \end{aligned} \tag{8.9}$$

when $\sigma_2 = 1$, and so σ_1 gets replaced by

$$\frac{(\sigma_1/k_2 + 1/k_3)\delta}{k_1\sigma_1/k_2 + k_1/k_3 + 1}, \tag{8.10}$$

which is a fractional linear transformation of σ_1. In other words, in phase 1 the number of fields is multiplied by 3 (replacing one resistor by 3). Note that with real positive values of k_1, k_2 and k_3 the transformation (8.10) maps the nonnegative real σ_1-axis onto an interval on the positive real axis, and the negative real σ_1-axis onto its complement, which cannot be the desired interval $[-\beta, -\alpha]$. Essentially, instead of making a transformation which has the effect of lengthening the branch cut as desired in the map at the top of **Figure 8.4**, we have a transformation which presumably (in some suitably defined metric) shortens the branch cut. This is why it is necessary to substitute nonorthogonal subspace collections, rather than orthogonal ones [see also the discussion below (8.25)], since, as we will see, they only act to lengthen the branch cut.

How to do this in general in the continuum case is described in section 29.1 of Milton (2002), bottom of

page 621 and page 622, for orthogonal subspace collections, and in the second half of Section 7.8 of this book, without assuming orthogonality. That is a bit abstract so let us go through it for the case in question.

8.3 The original subspace collection

Our starting point is the Hilbert space $\mathcal{H}$ of fields $\mathbf{P}(\mathbf{x})$ (real or complex $d-$dimensional vector fields), that are cell-periodic and square-integrable in the unit cell. We denote $\boldsymbol{\chi}$ as the projection onto all fields which are zero in phase 2, $\mathcal{U}$ as the space of constant vector fields, $\mathcal{E}$ as the space of gradients which derive from periodic potentials (i.e., $\mathbf{E} \in \mathcal{E}$ if $\nabla \times \mathbf{E} = 0$ and $\langle \mathbf{E} \rangle = 0$) and $\mathcal{J}$ as the space of divergence free fields with zero average value (i.e., $\mathbf{J} \in \mathcal{J}$ if $\nabla \cdot \mathbf{J} = 0$ and $\langle \mathbf{J} \rangle = 0$). In our Hilbert space the inner product is taken to be

$$(\mathbf{P}, \widetilde{\mathbf{P}}) = \int_{\text{unit cell}} \overline{\mathbf{P}(\mathbf{x})} \cdot \widetilde{\mathbf{P}}(\mathbf{x}), \tag{8.11}$$

where the overline denotes complex conjugation. With respect to this inner product the 3 spaces $\mathcal{U}, \mathcal{E}$ and $\mathcal{J}$ are orthogonal. The projection onto $\mathcal{E}$ we denote by $\boldsymbol{\Gamma}_1$. In Fourier space

$$\boldsymbol{\Gamma}_1 \widehat{\mathbf{P}}(\mathbf{k}) = \begin{cases} \frac{\mathbf{k} \otimes \mathbf{k}}{|\mathbf{k}|^2} \widehat{\mathbf{P}}(\mathbf{k}), & \mathbf{k} \neq 0, \\ 0, & \mathbf{k} = 0. \end{cases} \tag{8.12}$$

The field equations are solved once we have found electric fields $\mathbf{E}(\mathbf{x})$ and current fields $\mathbf{J}(\mathbf{x})$ such that

$$\mathbf{J} = [(\sigma_1 - \sigma_2)\boldsymbol{\chi} + \sigma_2]\mathbf{E}, \quad \mathbf{E} \in \mathcal{U} \oplus \mathcal{E}, \quad \mathbf{J} \in \mathcal{U} \oplus \mathcal{J}. \tag{8.13}$$

The effective tensor σ_* is then defined via

$$\boldsymbol{\Gamma}_0 \mathbf{J} = \sigma_* \boldsymbol{\Gamma}_0 \mathbf{E}, \tag{8.14}$$

where $\boldsymbol{\Gamma}_0$ is the projection onto $\mathcal{U}$. We assume the composite is isotropic so that σ_* is a scalar, although the method is easily generalized to anisotropic materials where the effective conductivity is a second order tensor.

8.4 The vector subspace collection we substitute into the original subspace collection

Now consider a $3-$dimensional subspace collection $\mathcal{H}'$ consisting of 3 component vectors $\mathbf{P} = [P_1, P_2, P_3]^T$ with inner product

$$(\mathbf{P}, \widetilde{\mathbf{P}}) = \sum_{i=1}^{3} \overline{P}_i \widetilde{P}_i, \tag{8.15}$$

where the overline denotes complex conjugation. The projection $\boldsymbol{\chi}' = p \otimes p$ projects onto the one dimensional space of fields proportional to the unit vector $\mathbf{p}$ where $\mathbf{p} = [p_1, p_2, p_3]^T$ and p_1, p_2, p_3 are given constants such that $p_1^2 + p_2^2 + p_3^2 = 1$. The p's could be complex but we *do not* mean $|p_1|^2 + |p_2|^2 + |p_3|^2 = 1$. Thus $\boldsymbol{\chi}'$ is

a projection but not an orthogonal projection when the p's are complex, as then $\boldsymbol{\chi}' = p \otimes p$ is not Hermitian. We take the following:

$$\begin{aligned} \mathcal{U}' &\text{ is the space of fields proportional to } (1,0,0)^T, \\ \mathcal{E}' &\text{ is the space of fields proportional to } (0,1,0)^T, \\ \mathcal{J}' &\text{ is the space of fields proportional to } (0,0,1)^T, \\ \mathcal{P}_1 &\text{ is the space of fields proportional to } (p_1,p_2,p_3)^T, \\ \mathcal{P}_2 &\text{ is the space of fields } (P_1,P_2,P_3)^T \\ &\text{such that } p_1P_1+p_2P_2+p_3P_3=0. \end{aligned} \tag{8.16}$$

The field equations become

$$\mathbf{J}' = [(t-\sigma_2)\boldsymbol{\chi}' + \sigma_2]\mathbf{E}', \quad \mathbf{E}' \in \mathcal{U}' \oplus \mathcal{E}', \quad \mathbf{J}' \in \mathcal{U}' \oplus \mathcal{J}', \tag{8.17}$$

where the constant t will be chosen so the associated "effective modulus" is σ_1. That is

$$\boldsymbol{\Gamma}_0\mathbf{J}' = \sigma_1\boldsymbol{\Gamma}_0\mathbf{E}', \tag{8.18}$$

where $\boldsymbol{\Gamma}_0$ is the projection onto $\mathcal{U}$, so that

$$J_1' = \sigma_1 E_1'. \tag{8.19}$$

Without loss of generality we can choose $E_1' = 1, J_1' = \sigma_1$ so the field equations become

$$\begin{pmatrix} J_1' \\ 0 \\ J_3' \end{pmatrix} = (t-\sigma_2) \underbrace{\begin{pmatrix} p_1^2 & p_1p_2 & p_1p_3 \\ p_1p_2 & p_2^2 & p_2p_3 \\ p_1p_3 & p_2p_3 & p_3^2 \end{pmatrix}}_{\boldsymbol{\chi}'} \begin{pmatrix} E_1' \\ E_2' \\ 0 \end{pmatrix} + \sigma_2 \begin{pmatrix} E_1' \\ E_2' \\ 0 \end{pmatrix}. \tag{8.20}$$

From the middle equation we get

$$(t-\sigma_2)p_1p_2E_1' + [(t-\sigma_2)p_2^2 + \sigma_2]E_2' = 0, \tag{8.21}$$

which with $E_1' = 1$ gives

$$E_2' = \frac{(\sigma_2 - t)p_1p_2}{(t-\sigma_2)p_2^2 + \sigma_2}. \tag{8.22}$$

So we have

$$\begin{aligned} \sigma_1 = J_1' &= p_1^2(t-\sigma_2) + \sigma_2 - \frac{(\sigma_2-t)^2p_1^2p_2^2}{(t-\sigma_2)p_2^2+\sigma_2} \\ &= \sigma_2 + \frac{p_1^2\sigma_2(t-\sigma_2)}{(t-\sigma_2)p_2^2+\sigma_2} \\ &= \sigma_2 + \frac{p_1^2\sigma_2}{p_2^2 + \sigma_2/(t-\sigma_2)}, \end{aligned} \tag{8.23}$$

which with $\sigma_2 = 1$ is satisfied with

$$t = 1 + \frac{\sigma_1 - 1}{p_1^2 - p_2^2(\sigma_1 - 1)} = 1 + \frac{(\sigma_1-1)(\beta-\alpha)}{(\sigma_1+\beta)(1+\alpha)}, \tag{8.24}$$

where

$$\begin{aligned} \alpha &= -1 - \frac{p_1^2}{p_2^2 - 1}, \\ \beta &= -1 - \frac{p_1^2}{p_2^2}. \end{aligned} \tag{8.25}$$

Note that $-\alpha$ (respectively $-\beta$) is obtained by substituting $t = 0$ (respectively $t = \infty$) in (8.24). Given real $\beta > \alpha > 0$ we need to choose p_1 and p_2 so that these equations are satisfied. This will necessitate complex solutions for p_1 and p_2 since otherwise β will be negative. Note that with p_1 and p_2 being complex, $\boldsymbol{\chi}'$ is no longer Hermitian, even though it is a (non self-adjoint) projection, and so we have a subspace collection which is not orthogonal: $\mathcal{P}_1$ is not orthogonal to $\mathcal{P}_2$. Also from the field equations with $E_1' = 1$ we get

$$J_3' = p_3(t - \sigma_2)(p_1 + p_2 E_2'), \tag{8.26}$$

i.e.,

$$J_3' = \frac{p_1 p_3 \sigma_2 (t - \sigma_2)}{(t - \sigma_2)p_2^2 + \sigma_2}. \tag{8.27}$$

8.5 The subspace collection after the substitution

Now consider the Hilbert space $\mathcal{H}''$ consisting of all periodic fields of the form

$$\mathbf{P}''(\mathbf{x}) = \underbrace{\begin{pmatrix} 0 \\ \mathbf{S}(\mathbf{x}) \\ \mathbf{T}(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x})}_{\in \mathcal{P}_1 \otimes (\mathcal{E}' \oplus \mathcal{J}')} + \underbrace{\begin{pmatrix} \mathbf{Q}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix}}_{\in \mathcal{H} \otimes \mathcal{U}'} = \begin{pmatrix} \mathbf{Q}(\mathbf{x}) \\ \boldsymbol{\chi}(\mathbf{x})\mathbf{S}(\mathbf{x}) \\ \boldsymbol{\chi}(\mathbf{x})\mathbf{T}(\mathbf{x}) \end{pmatrix}. \tag{8.28}$$

Fields in $\mathcal{U}''$ take the form

$$\mathbf{u}''(\mathbf{x}) = \begin{pmatrix} \mathbf{u}_0 \\ 0 \\ 0 \end{pmatrix} \in \mathcal{U} \otimes \mathcal{U}'. \tag{8.29}$$

Fields in $\mathcal{E}''$ take the form

$$\mathbf{E}''(\mathbf{x}) = \underbrace{\begin{pmatrix} 0 \\ \mathbf{S}(\mathbf{x}) \\ 0 \end{pmatrix} \boldsymbol{\chi}(\mathbf{x})}_{\in \mathcal{P}_1 \otimes \mathcal{E}'} + \underbrace{\begin{pmatrix} \widetilde{\mathbf{E}}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix}}_{\in \mathcal{E} \otimes \mathcal{U}'} = \begin{pmatrix} \widetilde{\mathbf{E}}(\mathbf{x}) \\ \mathbf{S}(\mathbf{x})\boldsymbol{\chi}(\mathbf{x}) \\ 0 \end{pmatrix}, \tag{8.30}$$

where $\widetilde{\mathbf{E}}(\mathbf{x}) \in \mathcal{E}$. Fields in $\mathcal{J}''$ take the form

$$\mathbf{J}''(\mathbf{x}) = \underbrace{\begin{pmatrix} 0 \\ 0 \\ \mathbf{T}(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x})}_{\in \mathcal{P}_1 \otimes \mathcal{J}'} + \underbrace{\begin{pmatrix} \widetilde{\mathbf{J}}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix}}_{\in \mathcal{J} \otimes \mathcal{U}'} = \begin{pmatrix} \widetilde{\mathbf{J}}(\mathbf{x}) \\ 0 \\ \mathbf{T}(\mathbf{x})\boldsymbol{\chi}(\mathbf{x}) \end{pmatrix}, \tag{8.31}$$

where $\widetilde{\mathbf{J}}(\mathbf{x}) \in \mathcal{J}$. The space $\mathcal{P}_1'$ consists of all vectors of the form

$$c \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}, \tag{8.32}$$

and $\mathcal{P}_1''$ consists of all fields $\mathbf{P}(\mathbf{x})$ of the form

$$\begin{pmatrix} p_1\mathbf{C}(\mathbf{x}) \\ p_2\mathbf{C}(\mathbf{x}) \\ p_3\mathbf{C}(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x}) \in \mathcal{P}_1 \otimes \mathcal{P}_1'. \tag{8.33}$$

Also $\mathcal{P}_2'$ consists of all vectors of the form

$$c \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \quad \text{where } p_1q_1 + p_2q_2 + p_3q_3 = 0, \tag{8.34}$$

and $\mathcal{P}_2''$ consists of all fields $\mathbf{P}(\mathbf{x})$ of the form

$$\underbrace{\begin{pmatrix} \mathbf{Q}_1(\mathbf{x}) \\ \mathbf{Q}_2(\mathbf{x}) \\ \mathbf{Q}_3(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x}) + (1 - \boldsymbol{\chi}(\mathbf{x})) \begin{pmatrix} \mathbf{R}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix}}_{\in(\mathcal{P}_1 \otimes \mathcal{P}_2' + \mathcal{P}_2 \otimes \mathcal{U}')} \quad \text{where } p_1\mathbf{Q}_1(\mathbf{x}) + p_2\mathbf{Q}_2(\mathbf{x}) + p_3\mathbf{Q}_3(\mathbf{x}) = 0. \tag{8.35}$$

The inner product on $\mathcal{H}''$ is defined to be

$$(\mathbf{P}, \widetilde{\mathbf{P}}) = \int_{\text{unit cell}} [\overline{\mathbf{S}(\mathbf{x})} \cdot \widetilde{\mathbf{S}}(\mathbf{x}) + \overline{\mathbf{T}(\mathbf{x})} \cdot \widetilde{\mathbf{T}}(\mathbf{x})]\boldsymbol{\chi}(\mathbf{x}) + \overline{\mathbf{Q}(\mathbf{x})} \cdot \widetilde{\mathbf{Q}}(\mathbf{x}), \tag{8.36}$$

where the overline denotes taking a complex conjugate. With this inner product, the subspaces $\mathcal{U}'', \mathcal{E}''$ and $\mathcal{J}''$ are mutually orthogonal. We define $\boldsymbol{\chi}'' = (\mathbf{p} \otimes \mathbf{p})\boldsymbol{\chi}$, i.e.,

$$\boldsymbol{\chi}'' \left\{ \begin{pmatrix} 0 \\ \mathbf{S}(\mathbf{x}) \\ \mathbf{T}(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x}) + \begin{pmatrix} \mathbf{Q}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix} \right\} = \begin{pmatrix} p_1^2\mathbf{I} & p_1p_2\mathbf{I} & p_1p_3\mathbf{I} \\ p_1p_2\mathbf{I} & p_2^2\mathbf{I} & p_2p_3\mathbf{I} \\ p_1p_3\mathbf{I} & p_2p_3\mathbf{I} & p_3^2\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{Q}(\mathbf{x}) \\ \mathbf{S}(\mathbf{x}) \\ \mathbf{T}(\mathbf{x}) \end{pmatrix} \boldsymbol{\chi}(\mathbf{x}), \tag{8.37}$$

where $\mathbf{I}$ is the $d \times d$ identity matrix which defines $\boldsymbol{\chi}''$ even if p_1, p_2 and p_3 are complex. Now the field equations are

$$\mathbf{J}'' = [(t - \sigma_2)\boldsymbol{\chi}'' + \sigma_2\mathbf{I}]\mathbf{E}'', \quad \mathbf{E}'' \in \mathcal{U}'' \oplus \mathcal{E}'', \quad \mathbf{J}'' \in \mathcal{U}'' \oplus \mathcal{J}''. \tag{8.38}$$

These are easy to solve given periodic solutions $\mathbf{J}(\mathbf{x})$ and $\mathbf{E}(\mathbf{x})$ to the equations in the Hilbert space $\mathcal{H}$, i.e.,

$$\mathbf{J} = [(\sigma_1 - \sigma_2)\boldsymbol{\chi} + \sigma_2]\mathbf{E}, \quad \nabla \cdot \mathbf{J} = 0, \quad \nabla \times \mathbf{E} = 0. \tag{8.39}$$

We take (with $E_1' = 1$)

$$\mathbf{E}'' = \begin{pmatrix} \mathbf{E}(\mathbf{x}) \\ E_2'\mathbf{E}(\mathbf{x})\boldsymbol{\chi}(\mathbf{x}) \\ 0 \end{pmatrix}, \quad \mathbf{J}'' = \begin{pmatrix} \mathbf{J}(\mathbf{x}) \\ 0 \\ J_3'\mathbf{J}(\mathbf{x})/\sigma_1 \end{pmatrix}. \tag{8.40}$$

Note that we have $\mathbf{E}'' \in \mathcal{U}'' \oplus \mathcal{E}''$ and $\mathbf{J}'' \in \mathcal{U}'' \oplus \mathcal{J}''$. Also, with $\sigma_2 = 1$, we have

$$\begin{aligned}((t-\sigma_2)\boldsymbol{\chi}'' + \sigma_2)\mathbf{E}'' &= (\mathbf{p} \otimes \mathbf{p}(t-\sigma_2) + \sigma_2])\begin{pmatrix} 1 \\ E_2'\mathbf{E}(\mathbf{x})\boldsymbol{\chi}(\mathbf{x}) \\ 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathbf{E}(\mathbf{x})(1-\boldsymbol{\chi}(\mathbf{x})) \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} J_1'\mathbf{E}(\mathbf{x}) \\ 0 \\ J_3'\mathbf{E}(\mathbf{x}) \end{pmatrix}\boldsymbol{\chi}(\mathbf{x}) + \begin{pmatrix} \mathbf{E}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix}(1-\boldsymbol{\chi}(\mathbf{x})) \\ &= \begin{pmatrix} \mathbf{J}(\mathbf{x}) \\ 0 \\ J_3'\mathbf{E}(\mathbf{x}) \end{pmatrix}\boldsymbol{\chi}(\mathbf{x}) + (1-\boldsymbol{\chi}(\mathbf{x}))\begin{pmatrix} \mathbf{J}(\mathbf{x}) \\ 0 \\ 0 \end{pmatrix} = \mathbf{J}''. \end{aligned} \tag{8.41}$$

Finally if $\boldsymbol{\Gamma}_0''$ is the projection onto $\mathcal{U}''$ we have

$$\begin{aligned}\boldsymbol{\Gamma}_0''\mathbf{E}'' &= \begin{pmatrix} \langle \mathbf{E} \rangle \\ 0 \\ 0 \end{pmatrix}\boldsymbol{\chi}(\mathbf{x}) + \begin{pmatrix} \langle \mathbf{E} \\ 0 \\ 0 \end{pmatrix}\rangle(1-\boldsymbol{\chi}(\mathbf{x})), \\ \boldsymbol{\Gamma}_0''\mathbf{J}'' &= \begin{pmatrix} \langle \mathbf{J} \rangle \\ 0 \\ 0 \end{pmatrix}\boldsymbol{\chi}(\mathbf{x}) + \begin{pmatrix} \langle \mathbf{J} \rangle \\ 0 \\ 0 \end{pmatrix}(1-\boldsymbol{\chi}(\mathbf{x})), \end{aligned} \tag{8.42}$$

and since $\langle \mathbf{J} \rangle = \sigma_* \langle \mathbf{E} \rangle$ we deduce that

$$\boldsymbol{\Gamma}_0''\mathbf{J}'' = \sigma_*\boldsymbol{\Gamma}_0''\mathbf{E}''. \tag{8.43}$$

That is, σ_* is still the effective tensor. Now the idea is to apply, either the basic Fast Fourier Transform scheme of Moulinec and Suquet (1994, 1998) to the Hilbert space $\mathcal{H}''$ that is associated with the expansion

$$\sigma_*/[(t+1)/2] = 1 + \sum_{n=1}^{\infty} d_n \left(\frac{t-1}{t+1}\right)^n, \tag{8.44}$$

or the accelerated Fast Fourier Transform method of Eyre and Milton (1999) as generalized in section 14.9 of Milton (2002) to the Hilbert space $\mathcal{H}''$ that is associated with the expansion

$$\sigma_*/\sqrt{t} = \sum_n c_n \left(\frac{\sqrt{t}-1}{\sqrt{t}+1}\right)^n. \tag{8.45}$$

The operator $\boldsymbol{\chi}''$ is easily evaluated in real space. The operator $\boldsymbol{\Gamma}_1''$ which projects onto $\mathcal{E}''$ is easily evaluated in Fourier space since

$$\boldsymbol{\Gamma}_1''\begin{pmatrix} \mathbf{Q}(\mathbf{x}) \\ \boldsymbol{\chi}(\mathbf{x})\mathbf{S}(\mathbf{x}) \\ \boldsymbol{\chi}(\mathbf{x})\mathbf{T}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}_1\mathbf{Q}(\mathbf{x}) \\ \boldsymbol{\chi}(\mathbf{x})\mathbf{S}(\mathbf{x}) \\ 0 \end{pmatrix}, \tag{8.46}$$

where in Fourier space

$$\boldsymbol{\Gamma}_1\widehat{\mathbf{Q}}(\mathbf{k}) = \begin{cases} \frac{\mathbf{k}\otimes\mathbf{k}\widehat{\mathbf{Q}}(\mathbf{k})}{|\mathbf{k}|^2}, & \mathbf{k}\neq 0, \\ 0, & \mathbf{k}=0. \end{cases} \tag{8.47}$$

Hence the accelerated Fast Fourier Transform method of Eyre and Milton can be directly applied in the Hilbert space $\mathcal{H}''$.

8.6 Proof of acceleration

Let us suppose $\sigma_1 \in \mathbb{R}$ is fixed such that $\sigma_1 > 1$. The rate of convergence will be determined by the magnitude of

$$|z| = \left|\frac{\sqrt{t}-1}{\sqrt{t}+1}\right|. \tag{8.48}$$

Since $t > 1$, when $\sigma_1 > 1$, the convergence will be quicker the smaller t is. Now as α increases,

$$t = 1 + \frac{(\sigma_1-1)(\beta-\alpha)}{(\sigma_1+\beta)(1+\alpha)} \tag{8.49}$$

decreases monotonically from the value $t = 1+((\sigma_1-1)\beta)/((\sigma_1+\beta))$ at $\alpha = 0$, to the value 1 at $\alpha = \beta$. So the larger the value of α, the faster the rate of convergence. Similarly as β decreases, t decreases monotonically from the value $t = 1+(\sigma_1-1)/(1+\alpha)$ at $\beta = \infty$ to the value of 1 at $\beta = \alpha$. So the smaller the value of β, the faster the rate of convergence. More generally when σ_1 is complex, to see the rate of convergence one should plot the contours $|z| = c$ in the complex σ_1-plane. It might be instructive to do this for particular values of α and β, and compare the contours with $\alpha = 0, \beta = \infty$.

8.7 The numerical example of Moulinec and Suquet

This numerical example is due to Hervé Moulinec and Pierre Suquet (to be published) and compares in a model example the speeds of convergence of results in the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}''$ for the Fast Fourier Transform scheme first proposed by Moulinec and Suquet (1994, 1998). Then the speeds of convergence of results are compared in the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}''$ for the accelerated scheme of Eyre and Milton (1999) [see also the generalization in section 14.9 of Milton (2002)]. In both cases the Fast Fourier Transform schemes converge substantially faster in the Hilbert space $\mathcal{H}''$.

It is to be emphasized that while these numerical results are only for the effective tensor, the real interest in the new algorithm is for obtaining results for the fields inside the body: accelerated rates of convergence for the effective conductivity alone could easily be obtained by using Padé approximants (Baker, Jr. and Graves-Morris 1981) or (almost equivalently) the associated bounds which use series expansion coefficients up to a given order (Milton 1981c; McPhedran and Milton 1981; Milton and McPhedran 1982: see also **Chapter 27** in Milton 2002 and **Chapter 10** in this book, and references therein). Using the new method I expect there will be a similar acceleration of the convergence rates for the fields. Indeed, bounds on the norm of the difference between the actual field and the field obtained by truncating the series expansion show these improved convergence rates. However this does not guarantee pointwise convergence, and in any case it needs to be numerically explored. Padé approximants methods could also be used for the fields, but this approach

has the disadvantage that one needs to simultaneously store a lot of information: not just the fields, but also the fields that appear up to a given order in the perturbation expansion for a nearly homogeneous medium.

The model example is a regular array of squares of conductivity σ_1 occupying a volume fraction of 25% in a matrix of conductivity $\sigma_2 = 1$, as illustrated in **Figure 8.6**.

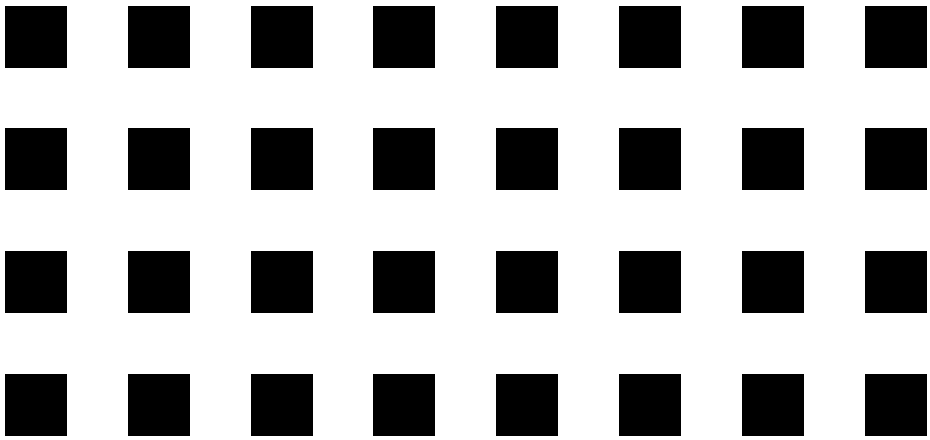


Figure 8.6: A periodic array of squares at a volume fraction of 25%, as illustrated here, provides a benchmark for testing the theory as, thanks to Obnosov (1999), there is an exact formula for its effective conductivity, given in (8.50).

An exact formula for the effective conductivity of this array (and for the interior fields) was discovered by Obnosov (1999),

$$\sigma_* = \sqrt{(1+3\sigma_1)/(3+\sigma_1)}, \tag{8.50}$$

and clearly has a branch cut between $\alpha = 1/3$ and $\beta = 3$. [Interestingly, for the four-phase checkerboard Mortola and Steffé (1985) conjectured a formula, that was later independently proved by Craster and Obnosov (2001) and Milton (2001).] Taking a wider estimate for the branch cut (with $\alpha = 1/4$ and $\beta = 4$), Moulinec and Suquet used the algorithm described in this chapter and found that the acceleration provided by the new method was generally substantially improved as shown in their **Figure 8.7** and **Figure 8.8**.

8.8 Estimating the parameters α and β

Bruno (1991a) has derived rigorous lower bounds on α and upper bounds on β, for suspensions of separated spheres in a medium. Actually we need only estimates on α and β. For instance if an estimate on α is slightly too large, then the transformations will look like those depicted in **Figure 8.9**. The series will still converge but only for $|z| < c$, where c will be close to 1 (but less than 1). However, the support of the measure will be highly dependent on small details of the microstructure: a tiny sharp cusp on the surface of an otherwise smooth inclusion will in general dramatically change the support of the measure: see Hetherington and Thorpe (1992), page 378 of Milton (2002), and equation (15) of Helsing, McPhedran, and Milton (2011). As remarked on page 378 of Milton (2002), this is related to the fact that these sharp corners behave as sinks of energy in the mathematically equivalent dielectric problem with the conductivities σ_1 and σ_2 being replaced by electrical

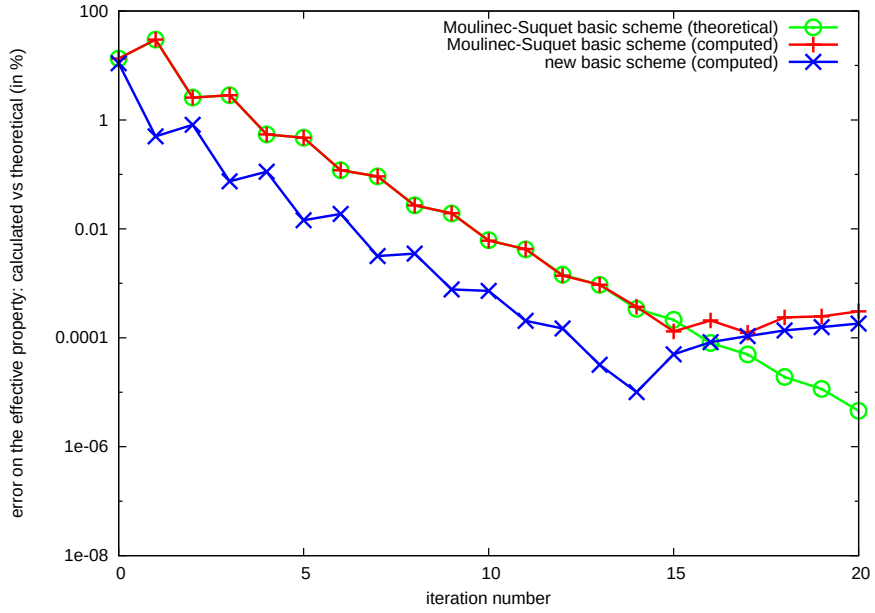


Figure 8.7: The results of Moulinec and Suquet (to be published) comparing the convergence of numerical results to the formula (8.50) for the effective conductivity σ_* when $\sigma_1 = 0$ and $\sigma_2 = 1$, for their basic scheme (Moulinec and Suquet 1994, 1998), and for the basic scheme applied in the Hilbert space $\mathcal{H}''$. Printed with permission of Hervé Moulinec and Pierre Suquet.

permittivities ε_1 and ε_2, where ε_2 is real and positive, while ε_1 is almost real and negative with a tiny positive imaginary part. Related observations and analysis include those of Qiu and Luk'yanchuk (2008), Pitkonen (2010), Estakhri and Alù (2013), and Bonnet-Ben Dhia, Chesnel, and Claeys (2013). Also there is a close connection with the essential spectrum of the Neumann–Poincaré operator on planar domains with corners (Helsing, Kang, and Lim 2016; Perfekt and Putinar 2016).

8.9 Bounds on the support of the measure using Q_C^*-convex functions

For many other equations in periodic composite materials robust things can be said about the measure which restrict its support. We assume we have a Hilbert space $\mathcal{H}$ of square integrable periodic fields $\mathbf{P}(\mathbf{x})$ with an inner product

$$(\mathbf{P}', \mathbf{P}) = \langle \overline{\mathbf{P}}' \cdot \mathbf{P} \rangle, \tag{8.51}$$

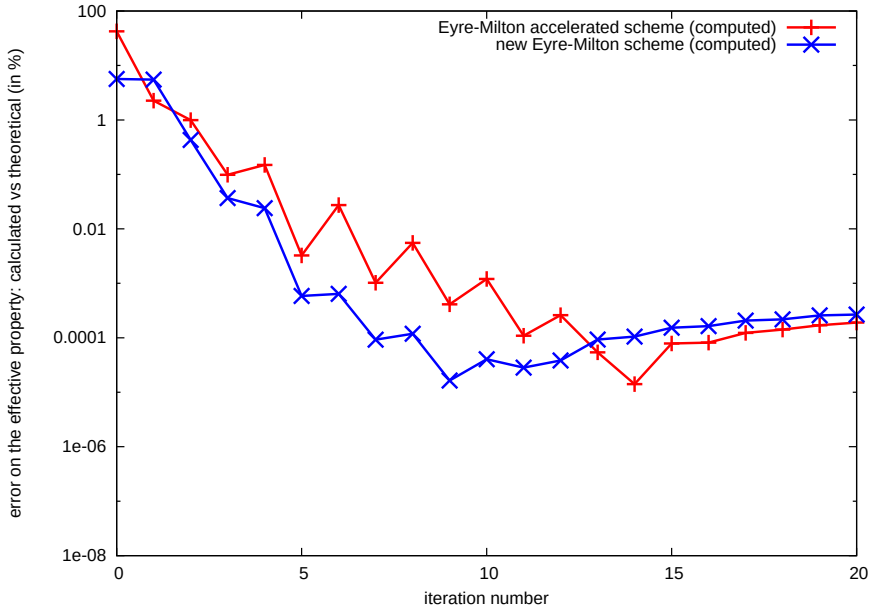


Figure 8.8: The results of Moulinec and Suquet (to be published) comparing the convergence of numerical results to the formula (8.50) for the effective conductivity σ_*, when $\sigma_1 = 0$ and $\sigma_2 = 1$ for the accelerated scheme of Eyre and Milton (1999) (see also the generalization in section 14.9 of Milton (2002)), and for the accelerated scheme applied in the Hilbert space $\mathcal{H}''$. Printed with permission of Hervé Moulinec and Pierre Suquet

where the angular brackets denote an average over the unit cell of periodicity. We also assume $\mathcal{H}$ has a decomposition into three orthogonal subspaces

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}. \tag{8.52}$$

We are interested in the equations

$$\mathbf{J}_0 + \mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})(\mathbf{E}_0 + \mathbf{E}(\mathbf{x})), \tag{8.53}$$

where

$$\mathbf{J}_0, \mathbf{E}_0 \in \mathcal{U}, \quad \mathbf{E} \in \mathcal{E}, \quad \mathbf{J} \in \mathcal{J}, \tag{8.54}$$

and where the local tensor field $\mathbf{L}(\mathbf{x})$ takes the form

$$\mathbf{L}(\mathbf{x}) = \mathbf{Q}(\mathbf{R}(\mathbf{x})) \left[\sum_{i=1}^{n} \chi_i(\mathbf{x}) \mathbf{L}_i \right] [\mathbf{Q}(\mathbf{R}(\mathbf{x}))]^T, \tag{8.55}$$

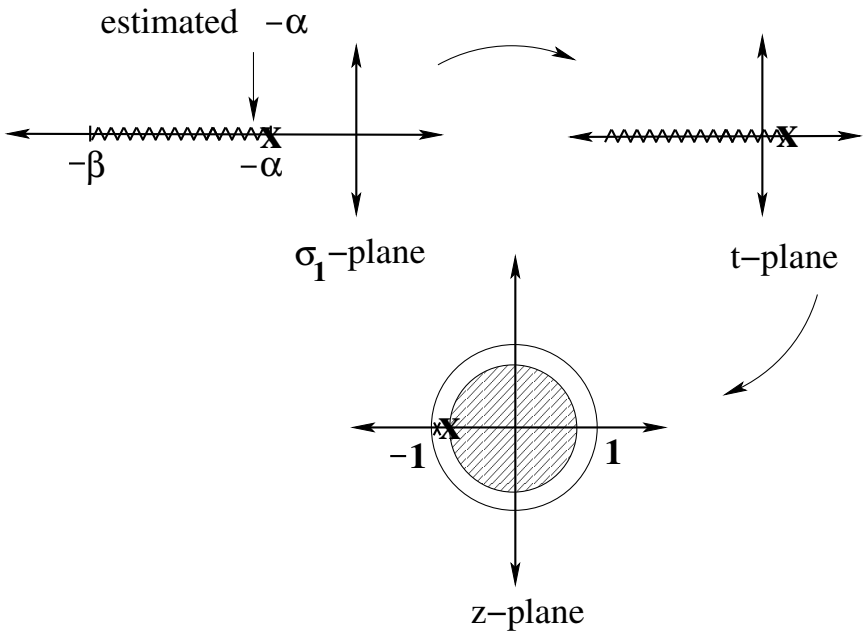


Figure 8.9: The new accelerated scheme should still work although not as effectively, even when we make a small error in the estimation of α (and/or β). With this error a singularity at $\mathbf{X}$, gets mapped to a point slightly greater than zero in the t-plane, and then gets mapped to a point slightly inside the unit disk in the z-plane. There is a consequent reduction in the radius of convergence, and an associated decrease in the rate of convergence, for the series in powers of z.

in which $\mathbf{Q}(\mathbf{R})$ is the orthogonal matrix (satisfying $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) associated with a rotation $\mathbf{R}$ acting on elements in the tensor space, while $\mathbf{R}(\mathbf{x})$ is a field of rotation matrices giving the local orientation of each phase, and $\chi_i(\mathbf{x})$ is the indicator function that is 1 in phase i and zero elsewhere.

First, following section 14.8 of Milton (2002) let us show that there are series expansions for the effective tensor that converge even when the local tensor $\mathbf{L}(\mathbf{x})$ is not everywhere positive definite. To obtain the series expansion we introduce a constant reference tensor $\mathbf{L}_0$ and an associated operator $\mathbf{\Gamma}$: we say

$$\mathbf{E}' = \mathbf{\Gamma}\mathbf{P} \quad \text{if and only if } \mathbf{E}' \in \mathcal{E} \text{ and } \mathbf{P} - \mathbf{L}_0\mathbf{E}' \in \mathcal{U} \oplus \mathcal{J}. \tag{8.56}$$

Explicitly, $\mathbf{\Gamma}$ is given by the operator

$$\mathbf{\Gamma} = \mathbf{\Gamma}_1(\mathbf{\Gamma}_1\mathbf{L}_0\mathbf{\Gamma}_1)^{-1}\mathbf{\Gamma}_1, \tag{8.57}$$

where the inverse is to be taken on the subspace $\mathcal{E}$. Introducing the polarization field

$$\mathbf{P} = (\mathbf{L} - \mathbf{L}_0)(\mathbf{E}_0 + \mathbf{E}) = \mathbf{J}_0 + \mathbf{J} - \mathbf{L}_0(\mathbf{E}_0 + \mathbf{E}), \tag{8.58}$$

we see it satisfies $\mathbf{\Gamma}\mathbf{P} = -\mathbf{E}$, and hence we have the identity

$$[\mathbf{I} + \mathbf{\Gamma}(\mathbf{L} - \mathbf{L}_0)](\mathbf{E}_0 + \mathbf{E}) = \mathbf{E}_0 + \mathbf{E} - \mathbf{E} = \mathbf{E}_0, \tag{8.59}$$

which gives

$$\mathbf{J}_0 + \mathbf{J} = \mathbf{L}(\mathbf{E}_0 + \mathbf{E}) = \mathbf{L}[\mathbf{I} + \boldsymbol{\Gamma}(\mathbf{L} - \mathbf{L}_0)]^{-1}\mathbf{E}_0. \tag{8.60}$$

Averaging both sides yields a formula for the effective tensor, and associated series expansion:

$$\mathbf{L}_* = \langle \mathbf{L}[\mathbf{I} + \boldsymbol{\Gamma}(\mathbf{L} - \mathbf{L}_0)]^{-1}\rangle = \sum_{j=0}^{\infty} \langle \mathbf{L}[\boldsymbol{\Gamma}(\mathbf{L}_0 - \mathbf{L})]^j\rangle, \tag{8.61}$$

where the angular brackets denote a volume average.

A sufficient condition for convergence is that the operator $\boldsymbol{\Gamma}(\mathbf{L}_0 - \mathbf{L})$ has norm less than 1. We take a reference tensor of the form

$$\mathbf{L}_0 = \sigma_0 \mathbf{I} + \mathbf{L}_0', \tag{8.62}$$

where the tensor $\mathbf{L}_0'$ is assumed to be bounded and self-adjoint, but need not be positive definite. We are interested in seeing whether the expansion for the effective tensor converges when σ_0 is very large. Expanding the operator $\boldsymbol{\Gamma}$ in powers of $1/\sigma_0$ gives

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1(\boldsymbol{\Gamma}_1\mathbf{L}_0\boldsymbol{\Gamma}_1)^{-1}\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_1/\sigma_0 - \boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1/\sigma_0^2 + \mathbf{R}_1, \tag{8.63}$$

with a remainder term

$$\mathbf{R}_1 = \sum_{j=2}^{\infty} \boldsymbol{\Gamma}_1(-1)^j(\mathbf{L}_0'\boldsymbol{\Gamma}_1)^j/\sigma_0^{j+1}. \tag{8.64}$$

When $\sigma_0 > \|\mathbf{L}_0'\|$ this remainder term has norm satisfying the bound

$$\|\mathbf{R}_1\| \leq \sum_{j=2}^{\infty} \|\mathbf{L}_0'\|^j/\sigma_0^{j+1} = \frac{\|\mathbf{L}_0'/\sigma_0\|^2}{\sigma_0 - \|\mathbf{L}_0'\|}. \tag{8.65}$$

This gives us a bound on the norm of $\boldsymbol{\Gamma}(\mathbf{L}_0 - \mathbf{L})$:

$$\begin{aligned} \|\boldsymbol{\Gamma}(\mathbf{L}_0 - \mathbf{L})\| &= \|(\boldsymbol{\Gamma}_1/\sigma_0 - \boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1/\sigma_0^2 + \mathbf{R}_1)(\sigma_0\mathbf{I} + \mathbf{L}_0' - \mathbf{L})\| \\ &= \|\boldsymbol{\Gamma}_1[\mathbf{I} + (\mathbf{L}_0' - \mathbf{L} - \mathbf{L}_0'\boldsymbol{\Gamma}_1)/\sigma_0] + \mathbf{R}_2\| \\ &\leq \|\mathbf{I} + (\mathbf{L}_0' - \mathbf{L} - \boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1)/\sigma_0\| + \|\mathbf{R}_2\|, \end{aligned} \tag{8.66}$$

where

$$\mathbf{R}_2 \equiv \mathbf{R}_1(\sigma_0\mathbf{I} + \mathbf{L}_0' - \mathbf{L}) - \boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1(\mathbf{L}_0' - \mathbf{L})/\sigma_0^2. \tag{8.67}$$

Now we can find a $\sigma^+ > \|\mathbf{L}_0'\|$ such that for $\sigma_0 > \sigma_0^+$,

$$\|(\sigma_0\mathbf{I} + \mathbf{L}_0' - \mathbf{L})\| < 2\sigma_0, \quad \sigma_0 - \|\mathbf{L}_0'\| > \sigma_0/2. \tag{8.68}$$

Hence there exists a constant $\beta_1 > 0$ (independent of σ_0) such that

$$\begin{aligned} \|\mathbf{R}_2\| &\leq \|\mathbf{R}_1\|\|(\sigma_0\mathbf{I} + \mathbf{L}_0' - \mathbf{L})\| + \|\boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1(\mathbf{L}_0' - \mathbf{L})\|/\sigma_0^2 \\ &\leq 4\|\mathbf{L}_0'/\sigma_0\|^2 + \|\boldsymbol{\Gamma}_1\mathbf{L}_0'\boldsymbol{\Gamma}_1(\mathbf{L}_0' - \mathbf{L})\|/\sigma_0^2 \leq \beta_1/\sigma_0^2. \end{aligned} \tag{8.69}$$

Now given any bounded operator $\mathbf{A}$ with adjoint $\mathbf{A}^\dagger$ and field $\mathbf{P} \in \mathcal{H}$ with $|\mathbf{P}|= 1$ we have

$$\begin{aligned} |(\mathbf{I} - \mathbf{A}/\sigma_0)\mathbf{P}|^2 &= (\mathbf{P}, (\mathbf{I} - \mathbf{A}/\sigma_0)^\dagger(\mathbf{I} - \mathbf{A}/\sigma_0)\mathbf{P} \\ &= 1 - (\mathbf{P}, (\mathbf{A} + \mathbf{A}^\dagger)\mathbf{P})/\sigma_0 + (\mathbf{P}, \mathbf{A}^\dagger\mathbf{A}\mathbf{P})/\sigma_0^2 \\ &\le 1 - (\mathbf{P}, (\mathbf{A} + \mathbf{A}^\dagger)\mathbf{P})/\sigma_0 + \|\mathbf{A}^\dagger\mathbf{A}\|/\sigma_0^2, \end{aligned} \tag{8.70}$$

and from the definition of the norm of an operator, this implies

$$\|\mathbf{I} - \mathbf{A}/\sigma_0)\| = 1 + \|\mathbf{A}^\dagger\mathbf{A}\|/\sigma_0^2 - \min_{\substack{\mathbf{P}\in\mathcal{H} \\ |\mathbf{P}|=1}} (\mathbf{P}, (\mathbf{A} + \mathbf{A}^\dagger)\mathbf{P})/\sigma_0. \tag{8.71}$$

Setting $\mathbf{A} = \mathbf{L} - \mathbf{L}_0' + \mathbf{L}_0'\mathbf{\Gamma}_1$ we see there exists a constant $\beta_2 > 0$ such that for $\sigma_0 > \sigma_0^+$,

$$\|\mathbf{\Gamma}(\mathbf{L}_0 - \mathbf{L})\| \le 1 + \beta_2/\sigma_0^2 - 2 \min_{\substack{\mathbf{P}\in\mathcal{H} \\ |\mathbf{P}|=1}} (\mathbf{P}, (\mathbf{L}_S - \mathbf{L}_0' + \mathbf{\Gamma}_1\mathbf{L}_0'\mathbf{\Gamma}_1)\mathbf{P})/\sigma_0, \tag{8.72}$$

where $\mathbf{L}_S = (\mathbf{L} + \mathbf{L}^\dagger)/2$. If the operator $\mathbf{L}_S - \mathbf{L}_0' + \mathbf{\Gamma}_1\mathbf{L}_0'\mathbf{\Gamma}_1$ is coercive in the sense that there exists a constant $\alpha' > 0$ such that

$$\mathbf{L}_S - \mathbf{L}_0' + \mathbf{\Gamma}_1\mathbf{L}_0'\mathbf{\Gamma}_1 > \alpha'\mathbf{I}, \tag{8.73}$$

then (8.72) implies

$$\|\mathbf{\Gamma}(\mathbf{L}_0 - \mathbf{L})\| < 1 + \beta_2/\sigma_0^2 - \alpha'/\sigma_0, \tag{8.74}$$

and this will surely be less than 1 for sufficiently large σ_0. Furthermore suppose the constant tensor $\mathbf{L}_0'$ is positive semi-definite on the subspace $\mathcal{E}$, i.e., the associated quadratic form

$$f(\mathbf{A}) = \overline{\mathbf{A}} \cdot \mathbf{L}_0'\mathbf{A} \tag{8.75}$$

is Q_C^*-convex, meaning it satisfies

$$\langle f(\mathbf{E})\rangle \ge 0, \quad \text{for all } \mathbf{E} \in \mathcal{E}, \tag{8.76}$$

where the angular brackets denote a volume average over the unit cell C of periodicity. Also suppose there exists a positive constant α' such that

$$\mathbf{\Gamma}_1\mathbf{L}_0'\mathbf{\Gamma}_1 \ge 0, \quad \mathbf{L}_S > \mathbf{L}_0' + \alpha'\mathbf{I}. \tag{8.77}$$

Then (8.73) will be satisfied. Now each term in the series expansion for $\mathbf{L}_*$ will be a polynomial in the elements of the matrices $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_n$ and hence in the domain where the series expansion converges, the effective tensor will be an analytic function of these matrix elements. Hence the support of the measure must lie outside the region where $\mathbf{L}_S > \mathbf{L}_0'$.

Note that the definition (8.76) is slightly different to the meaning of Q_C^*-convexity given in Milton (2013b) as here we allow for the space $\mathcal{E}$ to contain certain fields whose volume average is not zero (as happens when we treat the Schrödinger equation, with sources and with the energy having an imaginary part: see sections 13.6 and 13.7). When the fields in $\mathcal{E}$ satisfy homogeneous linear differential constraints, the condition (8.76) is easily reduced to an algebraic condition by taking Fourier transforms of (8.76). The projection onto the space $\mathcal{E}$ is given by projections $\mathbf{\Gamma}_1(\mathbf{k})$, that are local in Fourier space and (8.76) holds if an only if

$$\mathbf{\Gamma}_1(\mathbf{k})\mathbf{L}_0'\mathbf{\Gamma}_1(\mathbf{k}) \ge 0, \tag{8.78}$$

for all $\mathbf{k}$, including $\mathbf{k} = 0$ if $\mathcal{E}$ contains fields whose volume average is not zero, i.e., $\mathbf{\Gamma}_1(0) \neq 0$. As shown in Milton (2013b) this is basically the same procedure as followed by Murat and Tartar (Tartar 1979a; Murat and Tartar 1985; Tartar 1985) to obtain algebraic conditions for a quadratic function to be quasiconvex. For quadratic functions, quasiconvexity and Q^*-convexity are equivalent when the fields satisfy differential constraints involving derivatives of fixed order, but are not equivalent when the differential constraints have derivatives of mixed orders, as in time-harmonic wave-equations.

Acknowledgments

G.W. Milton wishes to especially thank Hervé Moulinec and Pierre Suquet for helping simplify his notation for the fields in the space $\mathcal{H}''$, and for doing the numerical experiments reported in section 8.7, thus providing evidence that the method suggested here works in practice, and not only in theory. Additionally G.W. Milton is grateful to the National Science Foundation for support through grant DMS-1211359.

9 The response of linear inhomogeneous systems to coupled fields: Bounds and perturbation expansions

Mordehai Milgrom* **and Graeme W. Milton****

*Department of Particle Physics and Astrophysics, Weizmann Institute of Science, 76100 Rehovot, Israel
**Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA
emails: moti.milgrom@weizmann.ac.il, milton@math.utah.edu

Abstract

We consider the response of a multicomponent body to n fields, such as electric fields, magnetic fields, temperature gradients, concentration gradients, etc., where each component, which is possibly anisotropic, may cross couple the various fields with different fluxes, such as electrical currents, electrical displacement currents, magnetic induction fields, energy fluxes, particle fluxes, etc. We obtain the form of the perturbation expansions of the fields and response tensor in powers of matrices which measure the difference between each component tensor and a homogeneous reference tensor $\mathbf{L}_0$. For the case of a statistically homogeneous or periodic composite the expansion coefficients can be expressed in terms of positive semidefinite normalization matrices alternating with positive semidefinite weight matrices, which at each given level sum to the identity matrix. In an appropriate basis the projection operators onto the relevant subspaces can be expressed in block tridiagonal form, where the blocks are functions of these weight and normalization matrices. This leads to continued fraction expansions for the effective tensor, and by truncating the continued fraction at successive levels one obtains a nested sequence of bounds on the effective tensor incorporating successively more weight and normalization matrices. The weight matrices and normalization matrices can be calculated from the series expansions of the fields which solve the conductivity problem alone, without any couplings to other fields, and then they can be used to obtain the solution for the fields and effective tensor in coupled field problems in composites.

9.1 Introduction

This chapter is concerned with the response of coupled fields and fluxes in a three-dimensional body Ω to potentials prescribed at the boundary of the body and with how this response depends on the material constants of the body. The effective tensor of a statistically homogeneous or periodic composite, with coupling between

the fields, is a special case which we will study in more depth. The set of fields $\vec{\boldsymbol{E}} = (\vec{E}_1, \vec{E}_2, ..., \vec{E}_n)$ which are each curl-free, may include electric fields, magnetic fields, temperature gradients, or concentration gradients and the associated fluxes $\vec{\boldsymbol{J}} = (\vec{J}_1, \vec{J}_2, ..., \vec{J}_n)$ may include electrical currents, electrical displacement currents, magnetic induction fields, energy fluxes and particle fluxes. We assume there are no sources inside the body, so each of these fluxes is divergence-free. Assuming a simply connected topology of the body, each of the curl-free fields derive from a potential $\vec{E}_j = -\vec{\nabla}\phi^j$. At each point within the body or medium we assume a linear constitutive relation $\vec{\boldsymbol{J}} = \overleftrightarrow{\boldsymbol{L}}\vec{\boldsymbol{E}}$ between the fluxes and fields through a position dependent symmetric positive-definite tensor $\overleftrightarrow{\boldsymbol{L}}(\vec{r})$ of material constants. The tensor may have off-diagonal couplings which cause a single driving field, such as a temperature gradient, to induce fluxes of all types.

The body is assumed to be an aggregate of grains (possibly infinite in number) comprised of a finite number M of components (phases) that have at least orthorhombic symmetry with the crystal orientation varying from grain to grain, thus $\overleftrightarrow{\boldsymbol{L}}(\vec{r})$ is assumed to be piecewise constant. Let $\boldsymbol{L}^{l,\alpha},\ l = 1, 2, \ldots M,\ \alpha = 1, 2, 3$ be the $n \times n$ principal response matrices of the l-th component, defined more precisely in the next section.

We investigate the response of the set of fluxes, $\vec{\boldsymbol{J}}(\vec{r})$, measured at a given position $\vec{r}$ within the body, to the potentials $\boldsymbol{\phi}(\vec{r}) = [\phi^1(\vec{r}), ..., \phi^n(\vec{r})]$ prescribed at the boundary of the body. Without loss of generality (see Milgrom (1990) for a discussion of this point) it is assumed that the prescribed potentials are all in proportion to a fixed scalar function $f(\vec{r})$ defined at points $\vec{r}$ on the surface of the body, i.e., $\phi^j(\vec{r}) = \phi_0^j f(\vec{r})$ for all $\vec{r} \in \partial\Omega$, and we consider how the set of fluxes $\vec{\boldsymbol{J}}(\vec{r})$ at $\vec{r}$ vary with the choice of the vector $\boldsymbol{\phi}_0 = (\phi_0^1, ..., \phi_0^n)$ of proportionality constants: since this relation is linear it is governed by a response tensor $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ giving $\vec{\boldsymbol{J}}(\vec{r}) = \vec{\boldsymbol{\mathcal{L}}}(\vec{r})\boldsymbol{\phi}_0$. This tensor $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ is the object of our analysis. Specifically we examine the dependence of $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ on the set of crystal moduli $\boldsymbol{L}^{l,\alpha}\ \ (l = 1, 2, ..., M, \alpha = 1, 2, 3)$ when each is close to a constant tensor $\boldsymbol{L}_0$, i.e., when the material constants of the body are close to being homogeneous and isotropic. To simplify notations these crystal moduli are relabeled as $\boldsymbol{L}_a\ \ (a = 1, 2, ..., p)$, avoiding repetitions in the original set of crystal moduli due to crystal symmetries of isotropy or uniaxiality: thus, when there are no symmetries (other than orthorhombic symmetry) a represents the pair (l, α) and $p = 3M$, but p could be less than $3M$ if some of the phases are isotropic or uniaxial.

A formal expression is obtained for the coefficients appearing in the series expansion of $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ in powers of the differences $\boldsymbol{\epsilon_a} = \boldsymbol{L}_a - \boldsymbol{L}_0\ (a = 1, 2, .., p)$. We say formal because these coefficients are difficult to evaluate and because their (nonlinear and nonlocal) dependence on the overall shape of the body, on the division of the body into grains and on the orientation of the crystals in each grain is complicated. What is interesting is the explicit form of the expansion. This is a non trivial issue since the set of matrices $\boldsymbol{L}_a\ (a = 1, 2, ..., p)$ do not necessarily commute. The issue has been addressed in part by Milgrom (1990) from general analytic considerations. Milgrom noted that the functional dependence of $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ on the $\boldsymbol{L}_a$ must satisfy two constraints:

(i) *Covariance*, the property that for any real, nonsingular, n by n matrix $\boldsymbol{W}$ with transpose $\boldsymbol{W}^T$ acting only on the field indices, the response tensor $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ transforms to $\boldsymbol{W}\vec{\boldsymbol{\mathcal{L}}}(\vec{r})\boldsymbol{W}^T$ when all of the crystal moduli $\boldsymbol{L}_a$ are replaced by the moduli $\boldsymbol{W}\boldsymbol{L}_a\boldsymbol{W}^T$. Covariance follows from the observation that we are free to define a new set of (curl-free) fields $\vec{\boldsymbol{E}}\,' = (\boldsymbol{W}^T)^{-1}\vec{\boldsymbol{E}}$ and a new set of (divergence-free) fluxes $\vec{\boldsymbol{J}}' = \boldsymbol{W}\vec{\boldsymbol{J}}$ by taking linear combinations of the old set of fields and fluxes while preserving at the same time the self-adjointness of the tensor $\overleftrightarrow{\boldsymbol{L}}(\vec{r})' = \boldsymbol{W}\overleftrightarrow{\boldsymbol{L}}(\vec{r})\boldsymbol{W}^T$ in the constitutive relation $\vec{\boldsymbol{J}}\,' = \overleftrightarrow{\boldsymbol{L}}'\vec{\boldsymbol{E}}\,'$. Clearly $\boldsymbol{W}\vec{\boldsymbol{\mathcal{L}}}(\vec{r})\boldsymbol{W'}$ is simply the old response tensor $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ expressed in terms of the new fields.

(ii) *Disjunction*, the property that when the matrices $\boldsymbol{L}_a$ are block diagonal of the same form then so must $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ have a similar block diagonal form in the field indices, and furthermore the elements of $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ within each block only depend on the elements of the $\boldsymbol{L}_a$'s in the corresponding blocks. Disjunction follows from

the observation that if a subset of fields is decoupled from another subset of fields then the effective response tensor must reflect this decoupling.

These analytic considerations alone eliminate from consideration many candidates for the terms in the series expansion, such as for example $\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon}_{\boldsymbol{a_1}}\boldsymbol{\epsilon}_{\boldsymbol{a_2}}$, and leave terms such as $\boldsymbol{\epsilon}_{\boldsymbol{a_1}}\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon}_{\boldsymbol{a_2}}\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon}_{\boldsymbol{a_3}}$ (that in fact do occur in the series expansion) as natural candidates.

The technique we employ in the present chapter is a simple generalization of an approach used in the theory of composite materials to derive series expansions for the effective conductivity or elasticity tensor of a nearly homogeneous multiphase material. A lot of the progress that has been made on series expansions and associated bounds on effective tensors is summarized in the review articles of Willis (1981a) and Hashin (1983) and the books of Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009). Brown, Jr. (1955), in a pioneering paper, obtained the series expansion of the effective conductivity σ^* of an isotropic composite of two isotropic components with nearly equal conductivities σ_1 and σ_2, and found that the coefficient of $(\sigma_1 - \sigma_2)^n$ in this expansion depends on the n-point correlation function giving the probability that a fixed configuration of n-points lands with all points in component 1 when placed randomly in the composite. Subsequently many other series expansions were derived for the effective conductivity tensor or elasticity tensor of nearly homogeneous composites: see for example, Herring (1960), Prager (1960), Beran and Molyneux (1963), Beran (1968),Beran and McCoy (1970)), Fokin and Shermergor (1969), Dederichs and Zeller (1973), Hori (1973), Zeller and Dederichs (1973), Gubernatis and Krumhansl (1975), Kröner (1977), Willis (1981a), Milton and Phan-Thien (1982), Phan-Thien and Milton (1982), Sen and Torquato (1989), Torquato (1997), Tartar (1989, 1990), and Bruno (1991b). Our analysis closely follows that of Willis (1981a) and Phan–Thien and Milton (1982, 1983).

Our analysis gives, as a simple corollary, a series expansion for the effective tensor $\overleftrightarrow{\boldsymbol{L}}^*$ that governs the constitutive relation between the local average of $\vec{\boldsymbol{J}}$ and the local average of $\vec{\boldsymbol{E}}$ in a statistically homogeneous or periodic composite material. (These averages are taken over a length scale much larger than the microstructure, yet smaller than any macroscopic lengths associated with variations in the applied fields.) This expansion is derived in Section 9.4 where the body is assumed to be filled with such a composite material, with microstructure much smaller than the dimensions of the body. From the response tensor $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ associated with linear potentials specified on the boundary, i.e., with $f(\vec{r}) = -\vec{r}\cdot\vec{v}_0$ on $\partial\Omega$, where $\vec{v}_0$ gives the direction of the applied field, we directly obtain the effective tensor $\overleftrightarrow{\boldsymbol{L}}^*$ of the composite.

The coefficients in the series expansion of $\overleftrightarrow{\boldsymbol{L}}^*$ in powers of the $\boldsymbol{\epsilon}_{\boldsymbol{a}}$ $(a = 1, 2, .., p)$ are useful for obtaining bounds on $\overleftrightarrow{\boldsymbol{L}}^*$. In particular they likely contain sufficient information to determine the weight and normalization matrices that were introduced by Milton (1987a, 1987b), following the introduction of scalar-valued weights and normalization factors by Milton and Golden (1985). Thus these parameters are seen to have a natural significance in the context of coupled field problems. In any case the weight matrices and normalization matrices can be calculated from the series expansions of the fields. It is noteworthy that they can be calculated from the series expansions of the fields which solve the conductivity problem alone, without any couplings to other fields, and then they can be used to obtain the solution for the fields and effective tensor $\overleftrightarrow{\boldsymbol{L}}^*$ in coupled field problems.

With these geometric parameters we show how one can compute, for coupled field problems, the Wiener–Beran and Hashin–Shtrikman type bounds of any order: these bounds, derived for the effective conductivity by Milton (1981c) and Milton and McPhedran (1982) (see also McPhedran and Milton 1981) and extended here to bounds on $\overleftrightarrow{\boldsymbol{L}}^*$, generalize the bounds of Wiener (1912), Hashin and Shtrikman (1962), Beran (1965), Willis (1977), Phan-Thien and Milton (1982), and Sen and Torquato (1989). They do not, however, encompass the

optimal two-dimensional, two-phase bounds of Cherkaev and Gibiansky (1992) and Clark and Milton (1995) which couple effective tensors using additional information about the differential constraints on the fields, or duality relations satisfied by the effective tensor as a function of the component moduli. Again many of the existing bounds are summarized in the review articles of Willis (1981a) and Hashin (1983) and the books of Cherkaev (2000), Milton (2002), Allaire (2002), Torquato (2002), Tartar (2009).

We also point out that in the case $p = 2$, as observed by Milgrom and Shtrikman (1989b; 1989a) following related work by Straley (1981), that when $\boldsymbol{\epsilon_1}$ and $\boldsymbol{\epsilon_2}$ are symmetric one can simultaneously diagonalize them using congruence transformations. Equivalently, one can define a new set of potentials and a new set of fluxes, which are linear combinations of the old ones, so that the equations are decoupled [see also the developments in Milgrom (1990, 1997) and Chen (1995, 1997), **Chapter 6** of Milton (2002), and Section 1.5 of this book.] Then, if we are considering the response of a composite, one can apply the bounds that are relevant in this case, such as those derived by Prager (1969), Bergman (1976, 1978) and Milton (1981c) [that were numerically tested by McPhedran and Milton (1981)] and by Cherkaev and Gibiansky (1992) and Clark and Milton (1995). In a wider context many of these bounds are closely related to the bounds on Stieltjes functions derived by Baker, Jr. (1969), in what he calls the fitting problem, and bounds on Herglotz functions that are associated with Nevanlinna–Pick interpolation (Pick 1915; Nevanlinna 1919, 1929; Delsarte, Genin, and Kamp 1981; Theorem 3.1 of **Chapter 5** of Kreĭn and Nudel'man 1974; Fedčina 1972; Delsarte, Genin, and Kamp 1979); Chen and Koç 1994, 1995; Theorem 2.2 in **Chapter VIII** of Foiaş and Frazho 1990; Alpay 2001; Agler and McCarthy 2002). When $p = 2$ and n is large, an alternative approach suggested by Day, Grant, Sievers, and Thorpe (2000) is to construct approximations to the measure which enters the integral representation formula for the response (Day and Thorpe 1999; Day, Grant, Sievers, and Thorpe 2000; Cherkaev 2001; Zhang and Cherkaev 2009; Cherkaev and Bonifasi-Lista 2011) and then use this to predict the response when n is increased. Generalizing the ideas of McPhedran, McKenzie, and Milton (1982), an alternative hybrid approach is to use the fact that the bounds imply rational functions satisfying the required analytic constraints pass through the exact data for a given value of n if the rational functions are of the appropriate degree. Then, if there are errors in the measurements (with known error bars), one can generate a family of such rational functions that are compatible with the data and in this way now only obtain predictions of the response when n is increased, but also obtain an idea of the probable error associated with these predictions. It could happen that there is no rational function compatible with the data. This could be a sign that the errors have been underestimated, or that there is some physics involved which has not been captured by the underlying equations.

This chapter is mostly self-contained, although it may be helpful for the reader to study **Chapters 1** and **2** first.

9.2 Setting of the problem and equations for the fields

We consider the problem of linear response to n coupled fields derivable from potentials $\phi^k,\ k = 1, ..., n$. The problem is described, in detail, by Milgrom (1990), and we give a succinct description here. The body consists of a space domain Ω within which the position-dependent response tensor is $L_{\alpha i\beta k}(\vec{r})$, where i, k are field indices and α, β are space indices. The αth component of the ith flux is given by the constitutive relation

$$J^i_\alpha(\vec{r}) = -\sum_{k=1}^{n}\sum_{\beta=1}^{3} L_{\alpha i\beta k}(\vec{r})\partial_\beta\phi^k(\vec{r}), \tag{9.1}$$

or, suppressing the indices:

$$\vec{\boldsymbol{J}} = -\overleftrightarrow{\boldsymbol{L}}\vec{\nabla}\boldsymbol{\phi}. \tag{9.2}$$

We shall be using boldface letters for quantities that are vectors or tensors in the field indices. Also, a $\rightarrow$ above a character indicates a vector in the space indices and a $\leftrightarrow$ above a character indicates a matrix in the space indices. So, for example, $\vec{\boldsymbol{J}}$, $\vec{\boldsymbol{E}}$, and $\vec{\boldsymbol{e}}$ are vectors in both space and field indices; $\overleftrightarrow{\boldsymbol{L}}$, and $\overleftrightarrow{\boldsymbol{\epsilon}}$ are second rank tensors in both types of indices; $\boldsymbol{L}$ is a matrix in the field indices; $\boldsymbol{\phi}$ and $\boldsymbol{\phi}_0$ are vectors in the field indices; $\vec{r}$ is a vector in the space indices; and $\overleftrightarrow{\Gamma}_1$ is a matrix in the space indices.

The equation

$$\vec{\nabla} \cdot \vec{\boldsymbol{J}} = 0, \tag{9.3}$$

determines the fields $\boldsymbol{\phi}$ within Ω, given the boundary conditions.

The response we consider is the field vector of n fluxes, $\vec{\boldsymbol{J}}(\vec{r})$, measured at a given position $\vec{r}$ within Ω, and is taken to respond to the boundary conditions dictated on the surface, $\partial\Omega$, of Ω. As explained in Milgrom (1990), we may, without loss of generality, restrict ourselves to boundary conditions of the form

$$\boldsymbol{\phi}(\vec{r}) = \boldsymbol{\phi}_0 f(\vec{r}), \quad \vec{r} \in \partial\Omega. \tag{9.4}$$

We then define the response matrix $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ such that

$$\vec{\boldsymbol{J}}(\vec{r}) = \vec{\boldsymbol{\mathcal{L}}}(\vec{r})\boldsymbol{\phi}_0. \tag{9.5}$$

We shall be interested in a piecewise-homogeneous system, so Ω is divided into a (possibly infinite) number of domains, as in Fig. 1, each of which is filled with one of M (possibly anisotropic) components, with an arbitrary orientation of its axes. We restrict ourselves to components that have, at least, an orthorhombic symmetry. The response matrix, $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$, depends, then, on the shape of Ω, on the choice of $f(\vec{r})$, on the division of Ω into sub-domains, on the orientations of the different components within these homogeneous sub-domains, and on the response properties of the individual components. In the principal axes of the lth component we can write

$$L^l_{\alpha k \beta m} = L^{l,\alpha}_{km}\delta_{\alpha\beta}, \tag{9.6}$$

where there is no summation over α and the $\boldsymbol{L}^{l,\alpha}$ $(\alpha = 1, 2, 3)$ are the principal response matrices of component l, $l = 1, 2, \ldots, M$. Let p be the total number of such principal matrices characterizing all the components. So, there is only one such matrix for an isotropic component, two for a component with uniaxial symmetry, and three for a component with orthorhombic symmetry. We shall use a single index notation with $\boldsymbol{L_a}$, $a = 1, ..., p$ instead of the doubly indexed $\boldsymbol{L}^{l,\alpha}$. (Depending on the symmetry, α here takes one, two, or three values.)

In the isotropic and homogeneous case we have

$$\boldsymbol{L_a} = \boldsymbol{L}_0, \tag{9.7}$$

for all a. When there are departures from isotropy and homogeneity we write

$$\boldsymbol{L_a} = \boldsymbol{L}_0 + \boldsymbol{\epsilon_a}, \tag{9.8}$$

and seek to expand $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ in the elements ϵ^a_{ik} of the $\boldsymbol{\epsilon_a}$'s.

To this end we first derive a formal expression for the driving field, $\vec{\boldsymbol{E}} = -\vec{\nabla}\boldsymbol{\phi}$ produced within Ω by the boundary conditions $\boldsymbol{\phi}_0$. Let $\vec{\boldsymbol{E}}_0$ be the driving field that is produced by these same boundary conditions *in the homogeneous, isotropic* case. We can write

$$\vec{\boldsymbol{E}}_0(\vec{r}) = \boldsymbol{\phi}_0 \vec{v}_0(\vec{r}), \tag{9.9}$$

where $\vec{v}_0 = -\vec{\nabla}\psi_0$, and ψ_0 is the single-field solution of the Laplace equation, in Ω, with boundary condition $\psi(\vec{r}) = f(\vec{r})$ on $\partial\Omega$; thus $\vec{\nabla} \cdot \vec{\boldsymbol{E}}_0 = 0$. The difference field

$$\vec{\boldsymbol{e}} \equiv \vec{\boldsymbol{E}} - \vec{\boldsymbol{E}}_0 = -\vec{\nabla}\boldsymbol{\psi}, \tag{9.10}$$

is derivable from a potential $\boldsymbol{\psi}$, that vanishes on $\partial\Omega$. Now introduce

$$\overleftrightarrow{\boldsymbol{\epsilon}}(\vec{r}) = \overleftrightarrow{\boldsymbol{L}}(\vec{r}) - \boldsymbol{L}_0 \overleftrightarrow{I}, \tag{9.11}$$

where we use $\overleftrightarrow{I}$ for the unit matrix in space indices; $\overleftrightarrow{\boldsymbol{I}}$ for the identity in both space and field indices; and I for the identity operator which when acting on a function leaves it invariant. Then the flux field,

$$\vec{\boldsymbol{J}}(\vec{r}) = \overleftrightarrow{\boldsymbol{L}}(\vec{r})\vec{\boldsymbol{E}}(\vec{r}), \tag{9.12}$$

can thus be written as

$$\vec{\boldsymbol{J}}(\vec{r}) = [\boldsymbol{L}_0 \overleftrightarrow{I} + \overleftrightarrow{\boldsymbol{\epsilon}}(\vec{r})](\vec{\boldsymbol{E}}_0 + \vec{\boldsymbol{e}})(\vec{r}). \tag{9.13}$$

Taking the divergence of (9.13), and remembering that $\vec{\boldsymbol{J}}$ and $\boldsymbol{L}_0\vec{\boldsymbol{E}}_0$ are divergence-free, we obtain

$$\boldsymbol{L}_0\vec{\nabla} \cdot \vec{\boldsymbol{e}} + \vec{\nabla} \cdot (\overleftrightarrow{\boldsymbol{\epsilon}}\,\vec{\boldsymbol{E}}) = 0, \tag{9.14}$$

or equivalently,

$$\Delta\boldsymbol{\psi} = \vec{\nabla} \cdot (\boldsymbol{L}_0^{-1}\overleftrightarrow{\boldsymbol{\epsilon}}\,\vec{\boldsymbol{E}}). \tag{9.15}$$

Define, now, the inverse Laplacian, Δ^{-1}, as the nonlocal operator which, acting on a density function $\rho(\vec{r})$, defined in Ω, gives the potential φ that solves the Poisson's equation $\Delta\varphi = -\rho$, and vanishes on the surface, $\partial\Omega$. Then, from (9.15) and (9.10) we can write

$$\vec{\boldsymbol{e}} = -\overleftrightarrow{\Gamma}_1\boldsymbol{L}_0^{-1}\overleftrightarrow{\boldsymbol{\epsilon}}\,\vec{\boldsymbol{E}}, \tag{9.16}$$

where

$$\overleftrightarrow{\Gamma}_1 \equiv \vec{\nabla}\Delta^{-1}\vec{\nabla}\cdot, \tag{9.17}$$

is nonlocal, with kernel $\overleftrightarrow{\Gamma}_1(\vec{r}, \vec{r}\,')$, and acts on a vector field $\vec{u}(\vec{r}\,')$ to give the vector field

$$\vec{v}(\vec{r}) = \int_\Omega d\vec{r}\,' \overleftrightarrow{\Gamma}_1(\vec{r}\,, \vec{r}\,')\vec{u}(\vec{r}\,'), \tag{9.18}$$

that has the same divergence as $\vec{u}$, and is derivable from a potential that vanishes on $\partial\Omega$. Clearly, $\overleftrightarrow{\Gamma}_1$ is a projection operator:

$$\overleftrightarrow{\Gamma}_1\overleftrightarrow{\Gamma}_1 = \overleftrightarrow{\Gamma}_1, \tag{9.19}$$

implying its kernel satisfies

$$\overleftrightarrow{\Gamma}_1(\vec{r},\vec{r}\,') = \int_\Omega d\vec{r}\,''\ \overleftrightarrow{\Gamma}_1(\vec{r},\vec{r}\,'')\overleftrightarrow{\Gamma}_1(\vec{r''},\vec{r}\,'). \tag{9.20}$$

In addition, because $\overleftrightarrow{\Gamma}_1$ gives zero when it acts on a uniform vector field and always produces a vector field with zero integral over Ω, we have

$$\int_\Omega d\vec{r}\,'\ \overleftrightarrow{\Gamma}_1(\vec{r},\vec{r}\,') = 0, \qquad \int_\Omega d\vec{r}\ \overleftrightarrow{\Gamma}_1(\vec{r},\vec{r}\,') = 0. \tag{9.21}$$

The operator $\overleftrightarrow{\Gamma}_1$ is also self-adjoint, i.e.,

$$\overleftrightarrow{\Gamma}_1(\vec{r},\vec{r}\,') = [\overleftrightarrow{\Gamma}_1(\vec{r}\,',\vec{r})]^T, \tag{9.22}$$

where T denotes the transpose. To see this, suppose one is given vector fields $\vec{u}(\vec{r})$ and $\vec{v}(\vec{r})$. Let $\varphi(\vec{r})$ and $\psi(\vec{r})$ be potentials that vanish on the boundary $\partial\Omega$ such that

$$\vec{u} = \vec{\nabla}\varphi + \vec{\nabla}\times\mathbf{A}, \quad \vec{v} = \vec{\nabla}\psi + \vec{\nabla}\times\mathbf{B}, \tag{9.23}$$

for some vector potentials $\mathbf{A}(\vec{r})$ and $\mathbf{B}(\vec{r})$. Then the definition of $\overleftrightarrow{\Gamma}_1$ implies $\overleftrightarrow{\Gamma}_1\vec{u} = \vec{\nabla}\varphi$ and $\overleftrightarrow{\Gamma}_1\vec{v} = \vec{\nabla}\psi$. So we have

$$\int_\Omega \vec{v}\cdot(\overleftrightarrow{\Gamma}_1\vec{u}) = \int_\Omega \vec{\nabla}\psi\cdot\vec{\nabla}\varphi + \int_\Omega(\vec{\nabla}\times\mathbf{B})\cdot\vec{\nabla}\varphi. \tag{9.24}$$

Using the divergence theorem, the last integral vanishes,

$$\int_\Omega(\vec{\nabla}\times\mathbf{B})\cdot\vec{\nabla}\varphi = \int_\Omega\vec{\nabla}\cdot[\varphi(\vec{\nabla}\times\mathbf{B})] = \int_{\partial\Omega}\varphi\mathbf{n}\cdot(\vec{\nabla}\times\mathbf{B}) = 0, \tag{9.25}$$

where $\mathbf{n}$ is the outwards normal to $\partial\Omega$, and we have used the fact that $\varphi = 0$ on $\partial\Omega$. Switching the roles of $\vec{v}$ and $\vec{u}$ in (9.24) gives the same result, and so we obtain

$$\int_\Omega \vec{v}\cdot(\overleftrightarrow{\Gamma}_1\vec{u}) = \int_\Omega \vec{u}\cdot(\overleftrightarrow{\Gamma}_1\vec{v}), \tag{9.26}$$

which means $\overleftrightarrow{\Gamma}_1$ is self-adjoint.

Adding $\vec{\boldsymbol{E}}_0$ to both sides of (9.16), we can write

$$(I + \overleftrightarrow{\Gamma}_1\boldsymbol{L}_0^{-1}\overleftrightarrow{\boldsymbol{\epsilon}})\vec{\boldsymbol{E}} = \vec{\boldsymbol{E}}_0, \tag{9.27}$$

or

$$\vec{\boldsymbol{E}} = (I + \overleftrightarrow{\Gamma}_1\boldsymbol{L}_0^{-1}\overleftrightarrow{\boldsymbol{\epsilon}})^{-1}\vec{\boldsymbol{E}}_0. \tag{9.28}$$

Thus, from the definition of the response matrix $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$, equation (9.5), from relation (9.9) between $\vec{\boldsymbol{E}}_0$ and $\boldsymbol{\phi}_0$, and from relation (9.12) between $\vec{\boldsymbol{J}}$ and $\vec{\boldsymbol{E}}$, we get

$$\vec{\boldsymbol{\mathcal{L}}}(\vec{r}) = \int d\vec{r}\,'\ \overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')\vec{v}_0(\vec{r}\,'), \tag{9.29}$$

where $\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')$ is the kernel of a nonlocal operator $\overleftrightarrow{\boldsymbol{\mathcal{S}}}$ [acting on the field $\vec{v}_0$], given by

$$\overleftrightarrow{\boldsymbol{\mathcal{S}}} = (\boldsymbol{L}_0 \overleftrightarrow{I} + \overleftrightarrow{\boldsymbol{\epsilon}})(I + \overleftrightarrow{\Gamma}_1 \boldsymbol{L}_0^{-1} \overleftrightarrow{\boldsymbol{\epsilon}})^{-1}. \tag{9.30}$$

The vector field $\vec{v}_0$ only carries the information on the exact form of the boundary conditions $[f(\vec{r}_0)]$; it is $\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')$ that plays the role of the response tensor of the system.

9.3 The expansion of the response tensor

We now use (9.30) to develop a series expansion for the response tensor $\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')$. Specializing to the piecewise homogeneous case, we express $\overleftrightarrow{\boldsymbol{\epsilon}}(\vec{r})$ in terms of the $\boldsymbol{\epsilon_a}$'s defined in equation (9.8). Defining the indicator function, $\chi_l(\vec{r})$, such that $\chi_l(\vec{r}) = 1$ in a subregion occupied by component l, $l = 1, 2, \ldots, M$ and $\chi_l(\vec{r}) = 0$ otherwise, we can write for the $\alpha\beta$ element of $\overleftrightarrow{\boldsymbol{\epsilon}}$

$$\boldsymbol{\epsilon}_{\alpha\beta}(\vec{r}) = \sum_{l=1}^{M} \sum_{\eta=1}^{3} \chi_l(\vec{r}) R_{\alpha\eta}(\vec{r}) \boldsymbol{\epsilon}^{l,\eta} R^T_{\eta\beta}(\vec{r}), \tag{9.31}$$

where

$$\boldsymbol{\epsilon}^{l,\eta} = \boldsymbol{L}^{l,\eta} - \boldsymbol{L}_0, \tag{9.32}$$

and $\boldsymbol{L}^{l,\eta}$ are the principal response matrices of component l, $R(\vec{r})$ is the rotation matrix from the principal axes to the orientation the component has at position $\vec{r}$, and $R^T(\vec{r})$ is its transpose (inverse).

Equation (9.31) can be cast in the form

$$\boldsymbol{\epsilon}_{\alpha\beta}(\vec{r}) = \sum_{a=1}^{p} \Lambda^a_{\alpha\beta}(\vec{r}) \boldsymbol{\epsilon_a}, \tag{9.33}$$

where the elements, $\Lambda^a_{\alpha\beta}(\vec{r})$, of $\overleftrightarrow{\Lambda}_a$ are defined as follows: For an orthorhombic component, there are three $\overleftrightarrow{\Lambda}_a$'s, where a replaces the double index l, η, and

$$\Lambda^{l,\eta}_{\alpha\beta}(\vec{r}) = \chi_l(\vec{r}) R_{\alpha\eta}(\vec{r}) R^T_{\eta\beta}(\vec{r}) \tag{9.34}$$

(with no summation over η). When the component l is isotropic, it contributes only one $\overleftrightarrow{\Lambda}_a$, with

$$\Lambda^a_{\alpha\beta}(\vec{r}) = \chi_l(\vec{r}) \sum_{\eta=1}^{3} R_{\alpha\eta}(\vec{r}) R^T_{\eta\beta}(\vec{r}) = \chi_l(\vec{r}) \delta_{\alpha\beta}. \tag{9.35}$$

Similarly, for a uniaxial component there are two matrices $\overleftrightarrow{\Lambda}_a(\vec{r})$. It is easy to ascertain that

$$\overleftrightarrow{\Lambda}_a(\vec{r}) \overleftrightarrow{\Lambda}_b(\vec{r}) = \delta_{ab} \overleftrightarrow{\Lambda}_a(\vec{r}), \tag{9.36}$$

and we also have

$$\sum_{a=1}^{p} \overleftrightarrow{\Lambda}_a(\vec{r}) = \overleftrightarrow{I}\,. \tag{9.37}$$

Now, substituting (9.33) in expression (9.30) for $\overleftrightarrow{\boldsymbol{\mathcal{S}}}$, and expanding, we get

$$\overleftrightarrow{\boldsymbol{\mathcal{S}}} = \boldsymbol{L}_0 \overleftrightarrow{I} + \sum_{s=1}^{\infty} \sum_{a_1,\ldots,a_s=1}^{p} (-1)^{s+1} \overleftrightarrow{K}_{a_1\ldots a_s} \boldsymbol{\epsilon_{a_1}} \boldsymbol{L}_0^{-1} \boldsymbol{\epsilon_{a_2}} \boldsymbol{L}_0^{-1} \ldots \boldsymbol{L}_0^{-1} \boldsymbol{\epsilon_{a_s}}, \tag{9.38}$$

where the reduced operator $\overleftrightarrow{K}_{a_1\ldots a_s}$ is given by

$$\overleftrightarrow{K}_{a_1\ldots a_s} = (I - \overleftrightarrow{\Gamma}_1)\overleftrightarrow{\Lambda}_{a_1} \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_2} \overleftrightarrow{\Gamma}_1 \ldots \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_s}. \tag{9.39}$$

Note that each operator $\overleftrightarrow{\Gamma}_1$ in the above relation acts on the whole expression to its right including the field on which $\overleftrightarrow{K}_{a_1\ldots a_s}$ acts: it does not just act on the adjacent $\overleftrightarrow{\Lambda}_a(\vec{r})$ factor. The reduced operators, which are matrices in the space indices, are purely geometrical. They depend on the geometry of the region Ω, on its division into homogeneous sub-regions, and on the orientation of the components within these sub-regions. They do not depend on the form of the boundary condition $f(\vec{r}_0)$, which enter through $\vec{v}_0(\vec{r})$ (on which the $\overleftrightarrow{K}$'s act); they also do not depend on the response coefficients of the components, which enter through the field-matrix terms in (9.38).

Using (9.29), the corresponding reduced, expansion coefficients of the response $\vec{\boldsymbol{\mathcal{L}}}(\vec{r})$ are

$$\vec{\kappa}_{a_1\ldots a_s}(\vec{r}) = [\overleftrightarrow{K}_{a_1\ldots a_s}\ \vec{v}_0](\vec{r}) = \int_\Omega d\vec{r}\,'\ \overleftrightarrow{K}_{a_1\ldots a_s}(\vec{r},\vec{r}\,')\vec{v}_0(\vec{r}\,'), \tag{9.40}$$

where $\overleftrightarrow{K}_{a_1\ldots a_s}(\vec{r},\vec{r}\,')$ is the kernel of the operator $\overleftrightarrow{K}_{a_1\ldots a_s}$. Note that the reduced coefficients are not independent: Summing over the last index gives

$$\sum_{a_s=1}^{p} \vec{\kappa}_{a_1\ldots a_s} = 0, \tag{9.41}$$

from (9.37), and the fact that $\overleftrightarrow{\Gamma}_1$ acting on a divergence-free vector field (such as $\vec{v_0}$) gives 0. Summing over the first index we also have

$$\sum_{a_1=1}^{p} \vec{\kappa}_{a_1\ldots a_s} = 0, \tag{9.42}$$

because $(I - \overleftrightarrow{\Gamma}_1)\overleftrightarrow{\Gamma}_1 = 0$. Summing over any, but the last, or first, index gives a reduced coefficient with one less index:

$$\sum_{a_i=1}^{p} \vec{\kappa}_{a_1\ldots a_s} = \vec{\kappa}_{a_1\ldots a_{i-1}a_{i+1}\ldots a_s}. \tag{9.43}$$

These follow directly from (9.39), and stem from the fact that we could arbitrarily redefine $\boldsymbol{L}_0$ by adding to it a constant matrix, and subtract that matrix from the $\boldsymbol{\epsilon_a}$'s, without affecting $\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')$. So there are really only $(p-1)^s$ independent s-th order coefficients, not p^s.

9.4 The expansion of the effective tensor of a composite

We now focus attention on an important subclass of inhomogeneous bodies: those filled with a statistically homogeneous or periodic composite material with microstructure much smaller than the dimensions of the body. It is well-known and can be rigorously proved (see for example Golden and Papanicolaou (1983)) that if there exists an intermediate length scale λ much larger than the homogeneities yet much smaller than the length scales associated with the dimensions of Ω and with variations in the applied potentials, then $\overleftrightarrow{\boldsymbol{L}}(\vec{r})$ can be replaced by a constant effective tensor $\boldsymbol{L}^*$ without disturbing the macroscopic response of the body. At distances from the boundary $\partial\Omega$, inside the body, sufficiently greater than λ this effective tensor $\overleftrightarrow{\boldsymbol{L}}^*$ governs the relation between the fields

$$< \vec{\boldsymbol{J}} >_{\Theta(\vec{r})} = \frac{1}{|\Theta(\vec{r})|}\int_{\Theta(\vec{r})} d\vec{r}\,'\, \vec{\boldsymbol{J}}(\vec{r}\,'), \qquad < \vec{\boldsymbol{E}} >_{\Theta(\vec{r})} = \frac{1}{|\Theta(\vec{r})|}\int_{\Theta(\vec{r})} d\vec{r}\,'\, \vec{\boldsymbol{E}}(\vec{r}\,'), \tag{9.44}$$

obtained by averaging $\vec{\boldsymbol{J}}(\vec{r}\,')$ and $\vec{\boldsymbol{E}}(\vec{r}\,')$ over a sphere $\Theta(\vec{r})$ of volume $|\Theta(\vec{r})|$, centered at $\vec{r}$, with radius λ, through the constitutive relation

$$< \vec{\boldsymbol{J}} >_{\Theta(\vec{r})} = \overleftrightarrow{\boldsymbol{L}}^* < \vec{\boldsymbol{E}} >_{\Theta(\vec{r})} . \tag{9.45}$$

Another tensor of interest is the microscopic response tensor $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,', \vec{r})$ which governs the linear relation between $\vec{\boldsymbol{J}}(\vec{r}\,')$, for points $\vec{r}\,'$ in $\Theta(\vec{r})$, and $< \vec{\boldsymbol{E}} >_{\Theta(\vec{r})}$:

$$\vec{\boldsymbol{J}}(\vec{r}) = \overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,', \vec{r}) < \vec{\boldsymbol{E}} >_{\Theta(\vec{r})} . \tag{9.46}$$

This tensor $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,', \vec{r})$ is only well-defined if there is a sufficient separation of length scales so that homogenization theory (see the many references in the introduction in **Chapter 2** and in particular Bensoussan, Lions, and Papanicolaou 1978 and Kozlov 1978) applies. Then $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,', \vec{r})$ is independent of the choice of $f(\vec{r}_0)$ (subject to it being smooth and only varying on the macroscopic scale), on the choice of $\boldsymbol{\phi}_0$, and (assuming statistical homogeneity) on the value of $\vec{r}$. Then we may vary $f(\vec{r}_0)$ and $\boldsymbol{\phi}_0$ to change $< \vec{\boldsymbol{E}} >_{\Theta(\vec{r})}$ and thus determine $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,', \vec{r}) = \overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,')$ through (9.46). For materials that are periodic inside Ω, with periodic cell much smaller than the size of Ω, $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,')$ can be obtained from the fields that solve the homogenization cell problem. i.e., with $\vec{\boldsymbol{J}}(\vec{r})$ and $\vec{\boldsymbol{E}}(\vec{r})$ having the same periodicity as the material, and the cell average of $\vec{\boldsymbol{E}}(\vec{r})$ having any value we desire.

By assumption $< \vec{\boldsymbol{E}} >_{\Theta(\vec{r})}$ has a smooth dependence on $\vec{r}$ and so by taking the average of (9.46) over points $\vec{r}\,'$ in the sphere $\Theta(\vec{r})$ we can identify $\overleftrightarrow{\boldsymbol{L}}^*$ with the average of $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,')$,

$$\overleftrightarrow{\boldsymbol{L}}^* = \; <\overleftrightarrow{\boldsymbol{\mathcal{L}}}>_{\Theta(\vec{r})} = \frac{1}{|\Theta(\vec{r})|}\int_{\Theta(\vec{r})} d\vec{r}\,'\; \overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,'). \tag{9.47}$$

To determine $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r})$ and hence $\overleftrightarrow{\boldsymbol{L}}^*$ it suffices to prescribe linear potentials on the boundary $\partial\Omega$ of Ω, i.e., to suppose $f(\vec{r}_0)$ takes the form

$$f(\vec{r}_0) = -\vec{r}_0 \cdot \vec{v}_0, \tag{9.48}$$

where $\vec{v}_0$ is a constant vector. Then the fields $\vec{\boldsymbol{E}}_0$ and $<\vec{\boldsymbol{E}}>_{\Theta(\vec{r})}$ which solve the constitutive equations in a homogeneous body are uniform,

$$\vec{\boldsymbol{E}}_0 =<\vec{\boldsymbol{E}}>_{\Theta(\vec{r})}= \vec{v}_0\boldsymbol{\phi}_0. \tag{9.49}$$

Consequently for the purpose of determining both $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,')$ and $\overleftrightarrow{\boldsymbol{L}}^*$ the averages $<\quad>_{\Theta(\vec{r})}$ over each sphere $\Theta(\vec{r})$ can be replaced by averages $<\quad>_\Omega$ over the entire body Ω. Also to simplify subsequent formula let us select our dimensions of length so that the body has unit volume,

$$|\Omega|= 1. \tag{9.50}$$

Then, averages over Ω can be equated with integrals over Ω. From (9.49) and the relations (9.5) and (9.46) of $\vec{\boldsymbol{\mathcal{L}}}(\vec{r}\,')$ and $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}\,')$ we have

$$\vec{\boldsymbol{\mathcal{L}}}(\vec{r}) =\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r})\vec{v}_0. \tag{9.51}$$

where we have relabeled $\vec{r}\,'$ as $\vec{r}$ to avoid confusion in the subsequent formulae. This, in conjunction with (9.29) and (9.47), leads directly to the expressions

$$\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}) = \int_\Omega d\vec{r}\,'\,\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,'), \tag{9.52}$$

$$\overleftrightarrow{\boldsymbol{L}}^*= \int_\Omega d\vec{r}\int_\Omega d\vec{r}\,'\,\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,'), \tag{9.53}$$

for the microscopic response tensor $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r})$ and the effective tensor $\overleftrightarrow{\boldsymbol{L}}^*$. Substitution of the series expansion for $\overleftrightarrow{\boldsymbol{\mathcal{S}}}(\vec{r},\vec{r}\,')$ into these expressions gives the desired series expansions

$$\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r}) = \boldsymbol{L}_0\,\overleftrightarrow{I} +\sum_{s=1}^{\infty}\sum_{a_1,\ldots,a_s=1}^{p}(-1)^{s+1}\,\overleftrightarrow{A}_{a_1\ldots a_s}\,\boldsymbol{\epsilon_{a_1}}\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon_{a_2}}\boldsymbol{L}_0^{-1}\ldots\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon_{a_s}}, \tag{9.54}$$

$$\overleftrightarrow{\boldsymbol{L}}^*= \boldsymbol{L}_0\,\overleftrightarrow{I} +\sum_{s=1}^{\infty}\sum_{a_1,\ldots,a_s=1}^{p}(-1)^{s+1}\,\overleftrightarrow{\alpha}_{a_1\ldots a_s}\,\boldsymbol{\epsilon_{a_1}}\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon_{a_2}}\boldsymbol{L}_0^{-1}\ldots\boldsymbol{L}_0^{-1}\boldsymbol{\epsilon_{a_s}}, \tag{9.55}$$

for $\overleftrightarrow{\boldsymbol{\mathcal{L}}}\,(\vec{r})$ and $\overleftrightarrow{\boldsymbol{L}}^*$ in powers of the $\boldsymbol{\epsilon_a}$'s with coefficients

$$\begin{aligned}\overleftrightarrow{A}_{a_1\ldots a_s}\,(\vec{r}) &= \int_\Omega d\vec{r}\,'\;\overleftrightarrow{K}_{a_1\ldots a_s}\,(\vec{r},\vec{r}\,')\\ &= \int_\Omega d\vec{r}\,'\,[(I-\overleftrightarrow{\Gamma}_1)\overleftrightarrow{\Lambda}_{a_1}\overleftrightarrow{\Gamma}_1\overleftrightarrow{\Lambda}_{a_2}\overleftrightarrow{\Gamma}_1\cdots\overleftrightarrow{\Gamma}_1\overleftrightarrow{\Lambda}_{a_s}](\vec{r},\vec{r}\,'),\end{aligned} \tag{9.56}$$

$$\begin{aligned}\overleftrightarrow{\alpha}_{a_1\ldots a_s}&= \int_\Omega d\vec{r}\;\overleftrightarrow{A}_{a_1\ldots a_s}\,(\vec{r})\\ &= \int_\Omega d\vec{r}\int_\Omega d\vec{r}\,'\,[\overleftrightarrow{\Lambda}_{a_1}\overleftrightarrow{\Gamma}_1\overleftrightarrow{\Lambda}_{a_2}\overleftrightarrow{\Gamma}_1\cdots\overleftrightarrow{\Gamma}_1\overleftrightarrow{\Lambda}_{a_s}](\vec{r},\vec{r}\,'),\end{aligned} \tag{9.57}$$

where the prefactor of $(I - \overleftrightarrow{\Gamma}_1)$ has been dropped from the last equation because $\overleftrightarrow{\Gamma}_1$ acting upon any field produces a field with zero integral over Ω: see (9.21).

As a consequence of (9.19), (9.21) and (9.37) the coefficients $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$ when $s > 1$ satisfy

$$\sum_{a_1=1}^{p} \overleftrightarrow{\alpha}_{a_1 \ldots a_s} = 0, \tag{9.58}$$

$$\sum_{a_s=1}^{p} \overleftrightarrow{\alpha}_{a_1 \ldots a_s} = 0, \tag{9.59}$$

$$\sum_{a_i=1}^{p} \overleftrightarrow{\alpha}_{a_1 \ldots a_s} = \overleftrightarrow{\alpha}_{a_1 \ldots a_{i-1} a_{i+1} \ldots a_s} \ . \tag{9.60}$$

In the special case $s = 1$ (9.37) implies

$$\sum_{a=1}^{p} \overleftrightarrow{\alpha}_a = \overleftrightarrow{I} \ . \tag{9.61}$$

Due to these identities it suffices, for any choice of reference index $q \in \{1, 2, ...p\}$, to consider the subset of coefficients $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}, s = 1, 2, ...$ generated as the indices a_i range over the reduced set $\{1, 2, ..q-1, q+1, ...p\}$ skipping the reference index q. The remaining coefficients $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$ where at least one index $a_i = q$ can then be recovered using (9.58)-(9.61). In addition, recall from (9.26) that the operator $\overleftrightarrow{\Gamma}_1$ is self-adjoint (this is also evident from (9.65) below). Also $\overleftrightarrow{\Lambda}_a$ is obviously self-adjoint. So (9.57) implies that the matrix $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$ is transformed to its transpose under reversal of the ordering of its subscripts:

$$\overleftrightarrow{\alpha}_{a_s a_{s-1} \ldots a_2 a_1} = (\overleftrightarrow{\alpha}_{a_1 a_2 \ldots a_{s-1} a_s})^T. \tag{9.62}$$

There are further identities satisfied by the coefficients $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$. In particular, the first order coefficients satisfy

$$\mathrm{Tr}(\overleftrightarrow{\alpha}_a) = \mathrm{Tr}(\overleftrightarrow{\alpha}^{l\beta}) = \int_\Omega d\vec{r} \ \mathrm{Tr}(\overleftrightarrow{\Lambda}^{l\beta}) = m_l f_l, \tag{9.63}$$

where f_l denotes the volume fraction occupied by component l and m_l takes values 1,2 or 3 according to whether the component l has orthorhombic symmetry, uniaxial symmetry, or isotropic symmetry. The last identity in (9.63) follows immediately for orthorhombic components by taking the trace in (9.34) (i.e., $\mathrm{Tr}[\overleftrightarrow{\Lambda}^{l\beta}(\vec{r})] = \chi_l(\vec{r})$), and for isotropic components by taking the trace in (9.35) (i.e., $\mathrm{Tr}[\overleftrightarrow{\Lambda}^{l\beta}(\vec{r})] = 3\chi_l(\vec{r})$).

The trace of the second order coefficient $\overleftrightarrow{\alpha}_{ab}$ can also be easily evaluated when the components are isotropic. To see this let us, for simplicity, suppose that the composite material is periodic with periodicity h much smaller than the dimensions of Ω. The action of $\overleftrightarrow{\Gamma}_1$ on any h-periodic vector field $\vec{u}(\vec{r}\,')$ is local in Fourier space and produces a vector field $\vec{v}(\vec{r})$ given by (9.18) with Fourier components

$$\vec{v}(\vec{k}) = \overleftrightarrow{\Gamma}_1(\vec{k})\vec{u}(\vec{k}), \tag{9.64}$$

in which $\vec{u}(\vec{k})$ denotes the Fourier component of $\vec{u}(\vec{r})$ and where the matrix $\overleftrightarrow{\Gamma}_1(\vec{k})$ has elements

$$\begin{aligned} \{\Gamma_1\}_{ij}(\vec{k}) &= k_i k_j / |\vec{k}|^2 \qquad \vec{k} \neq 0, \\ &= 0 \qquad\qquad\ \ \vec{k} = 0. \end{aligned} \tag{9.65}$$

Clearly (9.65) implies

$$\begin{aligned} \mathrm{Tr}(\overleftrightarrow{\Gamma}_1(\vec{k})) &= 1 \quad \vec{k} \neq 0, \\ &= 0 \quad \vec{k} = 0, \end{aligned} \tag{9.66}$$

and it follows that the operator

$$\Gamma(\vec{r}, \vec{r}\,') \equiv \mathrm{Tr}(\overleftrightarrow{\Gamma}_1(\vec{r}, \vec{r}\,')) \tag{9.67}$$

acts on any h-periodic scalar field $u(\vec{r})$ to produce the scalar field

$$v(\vec{r}) = \int_\Omega d\vec{r}\,' \, \Gamma(\vec{r}, \vec{r}\,')u(\vec{r}) = u(\vec{r}) \; - \; \int_\Omega d\vec{r}\,' \, u(\vec{r}\,'). \tag{9.68}$$

When the components are isotropic $\overleftrightarrow{\Lambda}_a(\vec{r}) = \Lambda_a(\vec{r}) \; \overleftrightarrow{I}$ and we have

$$\begin{aligned} \mathrm{Tr}(\overleftrightarrow{\alpha}_{ab}) &= \int_\Omega d\vec{r} \int \; d\vec{r}\,' \, [\Lambda_a \Gamma \Lambda_b](\vec{r}, \vec{r}\,') \\ &= \int_\Omega d\vec{r} \Lambda_a(\vec{r}) \Lambda_b(\vec{r}) - \int_\Omega d\vec{r} \Lambda_a(\vec{r}) \; \int_\Omega d\vec{r}\,' \Lambda_b(\vec{r}\,') \\ &= \delta_{ab} f_a \; - \; f_a f_b, \end{aligned} \tag{9.69}$$

where again f_a and f_b are the volume fractions of the components a and b.

In two-dimensional composites (9.69) is a simple corollary of one of an infinite set of identities satisfied by the coefficients $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$. These follow from the simple duality observation (see, for example, Keller (1964), Dykhne (1970) and Mendelson (1975)) that a 90° rotation, $\overleftrightarrow{R}_\perp$ acting on a curl-free field produces a divergence-free field and vice versa. Equivalently, from (9.65) we see immediately that

$$\overleftrightarrow{R}_\perp \overleftrightarrow{\Gamma}_1 (\overleftrightarrow{R}_\perp)^T = I - \overleftrightarrow{\Gamma}_1 - \overleftrightarrow{\Gamma}_0, \tag{9.70}$$

or alternatively,

$$\overleftrightarrow{R}_\perp \overleftrightarrow{\Gamma}_1 = (I - \overleftrightarrow{\Gamma}_1 - \overleftrightarrow{\Gamma}_0) \overleftrightarrow{R}_\perp, \tag{9.71}$$

where $\overleftrightarrow{\Gamma}_0(\vec{r}, \vec{r}\,')$ is the operator which simply acts to average the field: $\overleftrightarrow{\Gamma}_0(\vec{r}, \vec{r}\,')$ acting on a field $\vec{u}(\vec{r}\,')$ produces the uniform field

$$\vec{v}(\vec{r}) = \int_\Omega d\vec{r}\,' \, \overleftrightarrow{\Gamma}_0(\vec{r}, \vec{r}\,') \vec{u}(\vec{r}\,') = \int_\Omega d\vec{r}\,' \, \vec{u}(\vec{r}\,'), \tag{9.72}$$

and $\overleftrightarrow{R}_\perp$ is the operator which acts locally upon a field $\vec{u}(\vec{r})$ rotating it by 90° to produce the field $\vec{v}(\vec{r})$ with elements

$$v_\alpha = \sum_{\beta=1}^{2} R^\perp_{\alpha\beta} u_\beta, \tag{9.73}$$

where $R^\perp_{\alpha\beta}$ are in turn the elements of the matrix

$$\mathbf{R}_\perp = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{9.74}$$

for a 90° rotation. Accordingly we can use (9.70) to express $\overleftrightarrow{R}_\perp \overleftrightarrow{\alpha}_{a_1\ldots a_s}(\overleftrightarrow{R}_\perp)^T$ as a linear combination of the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_m}$ with $m \le s$. For example, if the components are isotropic $\overleftrightarrow{R}_\perp$ commutes with $\overleftrightarrow{\Lambda}_a$ and we have

$$\begin{aligned}\overleftrightarrow{R}_\perp\ \overleftrightarrow{\alpha}_{a_1a_2}\ (\overleftrightarrow{R}_\perp)^T &= \int_\Omega d\vec{r}\int_\Omega d\vec{r}\,'\,[\overleftrightarrow{\Lambda}_{a_1}(I-\overleftrightarrow{\Gamma}_1-\overleftrightarrow{\Gamma}_0)\overleftrightarrow{\Lambda}_{a_2}](\vec{r},\vec{r}\,')\\ &= -\overleftrightarrow{\alpha}_{a_1a_2}+\delta_{a_1a_2}f_{a_1}\overleftrightarrow{I}-f_{a_1}f_{a_2}\overleftrightarrow{I},\end{aligned} \tag{9.75}$$

$$\begin{aligned}\overleftrightarrow{R}_\perp\ \overleftrightarrow{\alpha}_{a_1a_2a_3}\ (\overleftrightarrow{R}_\perp)^T &= \int_\Omega d\vec{r}\int_\Omega d\vec{r}\,'\,[\overleftrightarrow{\Lambda}_{a_1}(I-\overleftrightarrow{\Gamma}_1-\overleftrightarrow{\Gamma}_0)\overleftrightarrow{\Lambda}_{a_2}(I-\overleftrightarrow{\Gamma}_1-\overleftrightarrow{\Gamma}_0)\overleftrightarrow{\Lambda}_{a_3}](\vec{r},\vec{r}\,')\\ &=\overleftrightarrow{\alpha}_{a_1a_2a_3}-\delta_{a_1a_2}\overleftrightarrow{\alpha}_{a_2a_3}-\overleftrightarrow{\alpha}_{a_1a_2}\delta_{a_2a_3}+f_{a_1}\overleftrightarrow{\alpha}_{a_2a_3}+\overleftrightarrow{\alpha}_{a_1a_2}f_{a_3}\\ &+\delta_{a_1a_2}\delta_{a_2a_3}f_{a_1}\overleftrightarrow{I}-\delta_{a_1a_2}f_{a_2}f_{a_3}\overleftrightarrow{I}-f_{a_1}f_{a_2}\delta_{a_2a_3}\overleftrightarrow{I}+f_{a_1}f_{a_2}f_{a_3}\overleftrightarrow{I}.\end{aligned} \tag{9.76}$$

The identity (9.69) is easily seen to follow from (9.75) by taking the trace of that equation. We now return to considering three dimensional composite materials.

When only one field is present, i.e., $n=1$, then the knowledge of the series expansion of $\overleftrightarrow{\boldsymbol{L}}^*$ in powers of the $\boldsymbol{\epsilon_a}$ up to a given order s is insufficient to determine the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ when $p \ge 3$ and $s \ge 3$. For example, consider the problem of electrical conductivity,

$$\vec{\nabla}\cdot J(\vec{r})=0,\ \ \vec{\nabla}\times E(\vec{r})=0,\ \ J(\vec{r})=\sigma(\vec{r})E(\vec{r}),\ \ \sigma(\vec{r})=\sum_{a=1}^{p}\sigma_a\overleftrightarrow{\Lambda}_a, \tag{9.77}$$

in a nearly homogeneous, nearly isotropic material with small values of the conductivity differences

$$\varepsilon_a=\sigma_a-\sigma_o. \tag{9.78}$$

Since the scalar quantities σ_o and ε_a commute, (9.55) reduces to the well-known series expansion for the effective conductivity

$$\overleftrightarrow{\sigma}^*=\sigma_o\overleftrightarrow{I}+\sum_{s=1}^{\infty}\sum_{a_1,\ldots,a_s=1}^{p}(-1)^{s+1}\overleftrightarrow{\beta}_{a_1\ldots a_s}\,\varepsilon_{a_1}\varepsilon_{a_2}\ldots\varepsilon_{a_s}/(\sigma_o)^{s-1}, \tag{9.79}$$

with coefficients

$$\overleftrightarrow{\beta}_{a_1\ldots a_s}=[\overleftrightarrow{\alpha}_{a_1\ldots a_s}]_{\rm sym}\equiv\frac{1}{s!}\sum_{\text{permutations}}\overleftrightarrow{\alpha}_{p(a_1\ldots a_s)}, \tag{9.80}$$

where the brackets $[\ ]_{\rm sym}$ denote a symmetrization over all $s!$ permutations $p(a_1\ldots a_s)$ of the field indices $a_1\ldots a_s$, excluding the space indices. In view of (9.62) we have, for example,

$$\begin{aligned}\overleftrightarrow{\beta}_{a_1}&=\overleftrightarrow{\alpha}_{a_1},\qquad \overleftrightarrow{\beta}_{a_1a_2}=\frac{1}{2}[\overleftrightarrow{\alpha}_{a_1a_2}+(\overleftrightarrow{\alpha}_{a_1a_2})^T],\\ \overleftrightarrow{\beta}_{a_1a_2a_3}&=\frac{1}{6}[\overleftrightarrow{\alpha}_{a_1a_2a_3}+\overleftrightarrow{\alpha}_{a_2a_3a_1}+\overleftrightarrow{\alpha}_{a_3a_1a_2}+(\overleftrightarrow{\alpha}_{a_1a_2a_3}+\overleftrightarrow{\alpha}_{a_2a_3a_1}+\overleftrightarrow{\alpha}_{a_3a_1a_2})^T].\end{aligned} \tag{9.81}$$

If the coefficients $\overleftrightarrow{\beta}_{a_1...a_j}$ are known for all $j \leq m$ then it is clearly impossible to recover all the coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ for $s \leq m$: one can only recover the linear combinations given by (9.80). However this does not eliminate the possibility that the coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ could be recovered from knowledge of the entire infinite set of coefficients $\overleftrightarrow{\beta}_{a_1...a_j}$. As we will see in Section 9.7, the value that $\overleftrightarrow{\alpha}_{a_1...a_s}$ can take is nonlinearly correlated with the coefficients $\overleftrightarrow{\beta}_{a_1...a_j}$ with $j \leq m$ through a set of matrix inequalities and it is conceivable that these matrix inequalities are sufficiently stringent to uniquely determine a given coefficient $\overleftrightarrow{\alpha}_{a_1...a_s}$ as m tends to infinity.

9.5 The weights and normalization matrices and a stratification of the Hilbert space

Suppose the coefficients $\overleftrightarrow{A}_{a_1...a_s}(\vec{r})$ of the microscopic response tensor $\overleftrightarrow{\boldsymbol{\mathcal{L}}}(\vec{r})$ are known as functions of $\vec{r}$, for all s up to a given order m, and for all combinations of indices a_i taken from the set $\{1, 2, ..., p\}$. In light of (9.57) one might think that this information would only be sufficient to determine the coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ for $s \leq m$. However this does not take into account the relations

$$\begin{aligned}\int_\Omega d\vec{r}\,[\overleftrightarrow{A}_{a_i a_{i-1}...a_1}(\vec{r})]^T\,\overleftrightarrow{A}_{a_{i+1}a_{i+2}...a_s}(\vec{r}) \\ = \int_\Omega d\vec{r}\int_\Omega d\vec{r}\,'\,[\overleftrightarrow{\Lambda}_{a_1}\overleftrightarrow{\Gamma}_1...\overleftrightarrow{\Lambda}_{a_i}(I-\overleftrightarrow{\Gamma}_1)\overleftrightarrow{\Lambda}_{a_{i+1}}\overleftrightarrow{\Gamma}_1...\overleftrightarrow{\Lambda}_{a_s}](\vec{r},\vec{r}\,') \\ = \delta_{a_i a_{i+1}}\,\overleftrightarrow{\alpha}_{a_1 a_2...a_{i-1}a_{i+1}...a_s} - \overleftrightarrow{\alpha}_{a_1...a_s},\end{aligned} \tag{9.82}$$

implied by (9.56), (9.19) and (9.36). These relations, which hold for all $i \in \{1, 2, ..., s-1\}$, allow the coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ to be determined for $s \leq 2m$ from knowledge of the functions $\overleftrightarrow{A}_{a_1...a_s}(\vec{r})$ for all $s \leq m$.

Now note that (9.57) and (9.82) imply inequalities such as the positive semidefiniteness of the tensors $\overleftrightarrow{\alpha}_{aa}$ and $\overleftrightarrow{\alpha}_a - \overleftrightarrow{\alpha}_{aa}$, $a = 1, 2...p$. The question of what other inequalities apply to the coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ has been analyzed in depth by Milton (1987a, 1987b). Briefly, and as proved later in Section 9.7, the set of coefficients $\overleftrightarrow{\alpha}_{a_1...a_s}$ for $s \leq 2m$ derives from, and in turn uniquely determine, a set of normalization matrices $\overleftrightarrow{N}^j$, $j = 1, 2, ...m$, and weight matrices $\overleftrightarrow{W}^j_a$, $a = 1, 2, ...p$, $j = 0, 1, 2, ...m-1$ that are real and symmetric and satisfy

$$\overleftrightarrow{N}^j \geq 0, \quad \overleftrightarrow{W}^j_a \geq 0, \quad \sum_{a=1}^{p} \overleftrightarrow{W}^j_a = \overleftrightarrow{I}^j, \tag{9.83}$$

where $\overleftrightarrow{I}^j$ denotes the k-dimensional identity matrix, where in a space of 3 dimensions, $k = 3(p-1)^j$. These matrices have elements $N^j_{\tau,\mu}$, $W^j_{a,\tau,\mu}$ and

$$I^j_{\tau,\mu} = \delta_{\tau\mu}, \tag{9.84}$$

labeled by strings $\tau = a_1 a_2 ... a_j \alpha$ and $\mu = b_1 b_2 ... b_j \beta$ of integers a_i or b_i, $i = 1, 2, ..j$ chosen from the set $\{1, 2, ...q-1, q+1...p\}$ (skipping the reference index q) terminated by a single space index α or β chosen from the set $\{1, 2, 3\}$. Thus each matrix has dimension $3(p-1)^j$ dependent on j, for $p > 2$.

Conversely, if a set of $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ derive from any sequence of $3(p-1)^j$-dimensional symmetric real matrices $\overleftrightarrow{N}^j, j = 1, 2, ..$ and $\overleftrightarrow{W}^j_a, a = 1, 2, ...p$, $j = 0, 1, 2, ...$ satisfying (9.83) then there always exists a set of commuting projection operators, $\overleftrightarrow{\Lambda}_a, a = 1, 2, ..p$, satisfying (9.36) and (9.37), and another noncommuting projection operator $\overleftrightarrow{\Gamma}_1$, satisfying (9.19) such that $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ is given by (9.57): we will see in Section 9.7 that, with a suitable choice of basis, the operators $\overleftrightarrow{\Lambda}_a$ only depend on the weight matrices, while $\overleftrightarrow{\Gamma}_1$ in this representation only depends on the normalization matrices. However not every sequence of normalization and weight matrices corresponds to a composite: there are additional subtle restrictions on the operators $\overleftrightarrow{\Lambda}_a$ and $\overleftrightarrow{\Gamma}_1$ in a composite which lead to nontrivial restrictions on the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$. In particular, as noticed by Zhikov, Kozlov, and Oleinik (1994), when all the components are isotropic a theorem of Meyers (1963) implies that, in the limit as the volume fraction f_a of component a tends to zero, $\overleftrightarrow{\boldsymbol{L}}^*$ cannot depend on $\boldsymbol{L_a}$ unless of course $\boldsymbol{L_a}$ has infinite or zero eigenvalues. In other words there exist inequalities which force any coefficient $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$, with $a_i = a$ for some $i \in \{1, 2, ...s\}$, to approach zero as $f_a = \mathrm{Tr}(\overleftrightarrow{\alpha}_a)$ tends to zero.

It remains to link the expansion coefficients with the weight and normalization matrices and to derive suitable representations for the operators $\overleftrightarrow{\Lambda}_a$ and $\overleftrightarrow{\Gamma}_1$. In the rest of the chapter, lower-case greek letters, other than α or β will always be used to denote strings of indices, where each index except the last is an element of the set $\{1, 2, ...q-1, q+1...p\}$ and where the final space index takes values from the set $\{1, 2, 3\}$. The length j of a string will refer to the number of indices in the string excluding the final space index. Also we use commas to separate strings of indices that label the elements of a matrix. Finally, a $\leftrightarrow$ above a character accompanied by a superscript j will indicate a $3(p-1)^j$ dimensional matrix in the string indices,with strings of length j.

First consider the sequence of fields obtained in the following fashion. We begin with a set of three or two uniform fields $\vec{x}_\alpha$, $(\alpha = 1, 2, 3)$ each aligned with its corresponding coordinate axis. [The notation is somewhat bad as $\vec{x}_\alpha$ should not be confused with a variable or spatial coordinate, but it follows the notation given in appendix 1 of Milton (1987a).] Then we set

$$\vec{p}_{a_1 a_2 \ldots a_k \alpha}(\vec{r}) = \overleftrightarrow{\Lambda}_{a_1} \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_2} \overleftrightarrow{\Gamma}_1 \cdots \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_k} \vec{x}_\alpha, \tag{9.85}$$

$$\vec{e}_{a_1 a_2 \ldots a_k \alpha}(\vec{r}) = \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_1} \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_2} \overleftrightarrow{\Gamma}_1 \cdots \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_k} \vec{x}_\alpha. \tag{9.86}$$

Note that the response coefficients $\overleftrightarrow{A}_{a_1\ldots a_s}(\vec{r})$ derive from these fields: from (9.56) we have

$$\overleftrightarrow{A}_{a_1\ldots a_s}(\vec{r})\vec{x}_\alpha = \vec{p}_{a_1\ldots a_s\alpha}(\vec{r}) - \vec{e}_{a_1\ldots a_s\alpha}(\vec{r}). \tag{9.87}$$

Introducing the standard inner product,

$$(\vec{u}, \vec{v}) = \int_\Omega d\vec{r}\, \overline{\vec{u}(\vec{r})} \cdot \vec{v}(\vec{r}), \tag{9.88}$$

between any two real fields $\vec{u}(\vec{r})$ and $\vec{v}(\vec{r})$, where the overline denotes complex conjugation, it is clear (see also (9.82)) that the inner product between any pair of the above fields can be written in terms of the elements of the coefficient matrix $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$: we have

$$(\vec{e}_\tau, \vec{e}_\eta) = \alpha_{\bar{\tau}\eta}, \tag{9.89}$$

$$(\vec{e}_\tau, \vec{p}_\eta) = (\vec{p}_\tau, \vec{e}_\eta) = \alpha_{\bar{\tau}\eta}, \tag{9.90}$$

$$(\vec{p}_{a\tau}, \vec{p}_{b\eta}) = \delta_{ab}\alpha_{\bar{\tau}a\eta}, \tag{9.91}$$

where τ and η represent strings of indices of lengths j and k respectively, and $\bar{\tau}$ is obtained from τ by reversing the sequence of indices in the string.

The space spanned by these fields has a natural stratification into a sequence of orthogonal subspaces $\mathcal{X}^0, \mathcal{Y}^1, \mathcal{X}^1, \mathcal{Y}^2, \mathcal{X}^2, ...$ The subspace $\mathcal{X}^0$ is defined as the subspace spanned by the uniform fields $\vec{x}_\alpha,\ \alpha = 1, 2, 3$. Let $\mathcal{F}^j$ denote the subspace spanned by the fields $\vec{x}_\alpha, \vec{p}_\eta(\vec{r})$ and $\vec{e}_\eta(\vec{r})$ as η ranges over all strings of length j. Also let $\mathcal{G}^j$ denote the closure of $\mathcal{F}^{j-1}$ under the action of the set of operators $\overleftrightarrow{\Lambda}_a\ a = 1, 2, ..p$: this is the space spanned by $\mathcal{F}^{j-1}$ and fields $\vec{p}_\tau$ as τ ranges over strings of length j. Note that $\mathcal{F}^j$ in turn is the closure of $\mathcal{G}^j$ under the action of $\overleftrightarrow{\Gamma}_1$. These subspaces satisfy the inclusion relations

$$\mathcal{X}^0 = \mathcal{F}^0 \subset \mathcal{G}^1 \subset \mathcal{F}^1 \subset \mathcal{G}^2 \subset \mathcal{F}^2 \subset \mathcal{G}^3 \cdots. \tag{9.92}$$

Accordingly we define $\mathcal{Y}^j, j = 1, 2, ...$ as the subspace of $\mathcal{G}^j$ which is the orthogonal complement of $\mathcal{F}^{j-1}$, and $\mathcal{X}^j, j = 1, 2, ...$ as the subspace of $\mathcal{F}^j$ which is the orthogonal complement of $\mathcal{G}^j$.

The weights and normalization matricesare obtained through the introduction of an orthonormal basis set of fields, comprised of fields $\vec{x}_\eta(\vec{r})$, denoted as type x, and fields $\vec{y}_\eta(\vec{r})$, denoted as type y, generated by a special version of Gram–Schmidt orthogonalization applied to the sequence of fields $\vec{p}_\tau(\vec{r})$ and $\vec{e}_\tau(\vec{r})$. These basis fields $\vec{x}_\eta(\vec{r})$ and $\vec{y}_\eta(\vec{r})$ will be called fields of order j if the string η has length j. Any linear combination of type x (or type y) basis fields of order j will also be called a type x (or type y) field of order j and we will establish that these type x (or type y) fields of order j are precisely the fields in the subspace $\mathcal{X}^j$ (or $\mathcal{Y}^j$).

9.6 Construction of the basis fields and weights and normalization factors

Those readers not interested in the details of the construction of the basis fields and weight and normalization matrices can skip to Section 9.9. We follow the construction procedure outlined in Appendix 1 of Milton (1987a). Recall that the uniform fields $\vec{x}_\alpha$ are already defined. Let us therefore suppose, for some $j \geq 1$, that all type x basis fields of order $j-1$ have been introduced. The weight matrices $\overleftrightarrow{W}_a^{j-1}$ are then defined via

$$W^{j-1}_{a,\omega,\rho} \equiv (\vec{x}_\omega, \overleftrightarrow{\Lambda}_a \vec{x}_\rho), \tag{9.93}$$

where ω and ρ are strings of length $j-1$. Next we introduce the first set of auxiliary fields

$$\vec{a}_{a\omega}(\vec{r}) \equiv \overleftrightarrow{\Lambda}_a(\vec{r})\vec{x}_\omega(\vec{r}) - \sum_\zeta W^{j-1}_{a,\omega,\zeta}\vec{x}_\zeta(\vec{r}), \tag{9.94}$$

which are defined in this way to ensure orthogonality to the previous set of type x fields of order $j-1$. Also from (9.37) it is evident that

$$\sum_{a=1}^{p} \vec{a}_{a\omega}(\vec{r}) = 0, \tag{9.95}$$

and consequently it suffices to consider the subset of fields $\vec{a}_{a\omega}(\vec{r})$ as the index a ranges over the reduced set $\{1, 2, .., q-1, q+1, ...p\}$. The inner products between the fields in this subset are given by

$$(\vec{a}_{a\omega}, \vec{a}_{b\rho}) = Y^j_{a\omega,b\rho}, \tag{9.96}$$

where

$$Y^j_{a\omega,b\rho} \equiv \delta_{ab} W^{j-1}_{a,\omega,\rho} - \sum_\zeta W^{j-1}_{a,\omega,\zeta} W^{j-1}_{b,\zeta,\rho}, \tag{9.97}$$

and the indices a and b belong to the reduced set (as does any other index in the strings ω and ρ apart from the terminating index). We normalize these fields to obtain the desired family of type y basis fields of order j,

$$\begin{aligned} \vec{y}_{b\rho} &\equiv \sum_{a\neq q} \sum_\omega C^j_{b\rho,a\omega} \vec{a}_{a\omega} \\ &= \sum_{a\neq q} \sum_\omega C^j_{b\rho,a\omega} (\overleftrightarrow{\Lambda}_a \vec{x}_\omega - \sum_\zeta W^{j-1}_{a,\omega,\zeta} \vec{x}_\zeta), \end{aligned} \tag{9.98}$$

where

$$\overleftrightarrow{C}^j \equiv (\overleftrightarrow{Y}^j)^{-1/2}. \tag{9.99}$$

Similarly, starting from these fields, let us introduce the commuting pair of matrices

$$U^j_{\tau,\phi} \equiv (\vec{y}_\tau, \overleftrightarrow{\Gamma}_1 \vec{y}_\phi), \tag{9.100}$$

$$V^j_{\tau,\phi} \equiv (\vec{y}_\tau, (I - \overleftrightarrow{\Gamma}_1)\vec{y}_\phi) = \delta_{\tau\phi} - U^j_{\tau,\phi}, \tag{9.101}$$

where the string indices τ and ϕ are now of length j. In terms of these matrices the normalization matrix is defined via

$$\overleftrightarrow{N}^j \equiv (\overleftrightarrow{U}^j)^{-1} - \overleftrightarrow{I}^j, \tag{9.102}$$

implying

$$\overleftrightarrow{U}^j = (\overleftrightarrow{I}^j + \overleftrightarrow{N}^j)^{-1}, \qquad \overleftrightarrow{V}^j = \{(\overleftrightarrow{I}^j + (\overleftrightarrow{N}^j)^{-1}\}^{-1}. \tag{9.103}$$

Next we generate the second set of auxiliary fields

$$\vec{b}_\tau(\vec{r}) \equiv \int_\Omega d\vec{r}\,' \overleftrightarrow{\Gamma}_1(\vec{r}, \vec{r}\,') \vec{y}_\tau(\vec{r}\,') \; - \; \sum_\nu U^j_{\tau,\nu} \vec{y}_\nu(\vec{r}), \tag{9.104}$$

which are orthogonal to the fields $\vec{y}_\phi$, and have inner products

$$(\vec{b}_\tau, \vec{b}_\phi) = \sum_\nu U^j_{\tau,\nu} V^j_{\nu,\phi}. \tag{9.105}$$

Normalizing these fields then produces the next orthonormal set of type x basis fields of order j:

$$\vec{x}_\phi \equiv \sum_\tau D^j_{\phi,\tau} \vec{b}_\tau = \sum_\tau D^j_{\phi,\tau} (\overleftrightarrow{\Gamma}_1 \vec{y}_\tau \; - \; \sum_\nu U^j_{\tau,\nu} \vec{y}_\nu), \tag{9.106}$$

where

$$\overleftrightarrow{D}^j \equiv (\overleftrightarrow{U}^j \overleftrightarrow{V}^j)^{-1/2} = (\overleftrightarrow{N}^j)^{1/2} + (\overleftrightarrow{N}^j)^{-1/2}. \tag{9.107}$$

By induction this completes the definition of the basis fields, and weight and normalization matrices.

From the definitions (9.93),(9.97), (9.100) and (9.101) it is clear that the matrices $\overleftrightarrow{W}^{j-1}_a, \overleftrightarrow{Y}^j, \overleftrightarrow{U}^j$ and $\overleftrightarrow{V}^j$ are positive semidefinite. Furthermore from (9.103) and from the orthonormality of the sets of fields, $\vec{x}_\omega$

and $\vec{y}_\tau$ it follows that the weights and normalization matrices satisfy (9.83). We avoid considering the rather special limiting case where the matrices $\overleftrightarrow{W}_a^{j-1}, \overleftrightarrow{U}^j$ and $\overleftrightarrow{V}^j$ have zero eigenvalues. In this event the matrices $\overleftrightarrow{Y}^j$ and $\overleftrightarrow{U}^j\overleftrightarrow{V}^j$ become singular and technical difficulties arise in the above construction procedure because the inverses needed in (9.99) and (9.107) do not exist.

The set of normalization and weight matrices obtained in this way clearly depend on the choice of reference component q. However the subspace spanned by type x (or type y) fields of order j remains invariant: it is only the basis within each subspace that changes when the choice of reference component is changed. Consequently the eigenvalues of the weight and normalization matrices do not depend on the choice of reference media.

Observe from (9.98) and (9.106) that for $a \neq q$

$$\overleftrightarrow{\Lambda}_a \vec{x}_\omega = \sum_\zeta W^{j-1}_{a,\omega,\zeta}\vec{x}_\zeta + \sum_{b\neq q}\sum_\rho M^j_{a\omega,b\rho}\vec{y}_{b\rho}, \tag{9.108}$$

$$\overleftrightarrow{\Gamma}_1 \vec{y}_\tau = \sum_\nu U^j_{\tau,\nu}\vec{y}_\nu + \sum_\phi X^j_{\tau,\phi}\vec{x}_\phi, \tag{9.109}$$

where

$$\overleftrightarrow{X}^j \equiv (\overleftrightarrow{U}^j\overleftrightarrow{V}^j)^{1/2} = \{(\overleftrightarrow{N}^j)^{1/2} + (\overleftrightarrow{N}^j)^{-1/2}\}^{-1}. \tag{9.110}$$

$$\overleftrightarrow{M}^j \equiv (\overleftrightarrow{Y}^j)^{1/2} = (\overleftrightarrow{C}^j)^{-1}, \tag{9.111}$$

and $\overleftrightarrow{Y}^j$ in turn is given by (9.97).

Applying $\overleftrightarrow{\Lambda}_c$, with $c \neq q$ to both sides of this first equation and $\overleftrightarrow{\Gamma}_1$ to both sides of the second equation gives

$$\sum_{b\neq q}\sum_\rho M^j_{a\omega,b\rho}\overleftrightarrow{\Lambda}_c\vec{y}_{b\rho} = \sum_\zeta(\delta_{ac}\delta_{\omega\zeta} - W^{j-1}_{a,\omega,\zeta})\overleftrightarrow{\Lambda}_c\vec{x}_\zeta, \tag{9.112}$$

$$\sum_\phi X^j_{\tau,\phi}\overleftrightarrow{\Gamma}_1\vec{x}_\phi = \sum_\nu V^j_{\tau,\nu}\overleftrightarrow{\Gamma}_1\vec{y}_\nu. \tag{9.113}$$

Substituting (9.108) and (9.109) back into these expressions produces after some algebraic manipulation,

$$\overleftrightarrow{\Lambda}_c\vec{y}_{b\rho} = \sum_{a\neq q}\sum_\zeta Q^j_{c,b\rho,a\zeta}\vec{y}_{a\zeta} + \sum_\zeta M^j_{b\rho,c\zeta}\vec{x}_\zeta, \tag{9.114}$$

$$\overleftrightarrow{\Gamma}_1\vec{x}_\nu = \sum_\phi V^j_{\nu,\phi}\vec{x}_\phi + \sum_\phi X^j_{\nu,\phi}\vec{y}_\phi, \tag{9.115}$$

where $\overleftrightarrow{Q}^j_c$ is the matrix,

$$\overleftrightarrow{Q}^j_c \equiv \overleftrightarrow{M}^j_c(\overleftrightarrow{W}^{j-1}_c)^{-1}(\overleftrightarrow{M}^j_c)^T, \tag{9.116}$$

and $\overleftrightarrow{M}^j_a$, with transpose $(\overleftrightarrow{M}^j_a)^T$, is the rectangular submatrix of the square matrix $\overleftrightarrow{M}^j$ defined in (9.111) with elements $M^j_{a\tau,\lambda}$ labeled by the strings τ and λ.

So $\overleftrightarrow{\Lambda}_c$ acting upon any basis field produces a linear combination of two fields: one field of the same order and type as the basis field and the other field of adjacent order and opposite type. By contrast $\overleftrightarrow{\Gamma}_1$ acting on any basis field produces a field of the same order but mixed type.

By construction the basis fields of a given order form an orthonormal set. To establish the orthonormality of the entire basis set we still need to show that the basis fields of order j are orthogonal to the subspace spanned by the fields of order at most $j-1$. Note that this subspace can also be identified with the subspace $\mathcal{F}^{j-1}$ spanned by the fields x_α, $\vec{p}_\eta(\vec{r})$, and $\vec{e}_\eta(\vec{r})$ as η ranges over strings of length $k \leq j-1$. We argue by induction and begin by assuming that the collection of fields $\vec{x}_\eta$, and $\vec{y}_\eta$ of order at most $j-1$ forms an orthonormal basis of $\mathcal{F}^{j-1}$: this is clearly true when $j=1$ because then $\mathcal{F}^0$ is the three dimensional space spanned by the fields $\vec{x}_\alpha$. In particular the assumption implies that within $\mathcal{F}^{j-1}$ basis fields of different types or different orders are orthogonal. Since $\overleftrightarrow{\Lambda}_a$ is self-adjoint (9.94) implies that

$$(\vec{a}_{a\omega}, \vec{x}_\eta) = (\vec{x}_\omega, \overleftrightarrow{\Lambda}_a \vec{x}_\eta), \qquad (\vec{a}_{a\omega}, \vec{y}_\eta) = (\vec{x}_\omega, \overleftrightarrow{\Lambda}_a \vec{y}_\eta), \tag{9.117}$$

where the string ω has length $j-1$. The choice of auxiliary fields guarantees that the first inner product is zero when the length k of the string η equals $j-1$. It is also zero when $k < j-1$ because then (9.108) implies $\overleftrightarrow{\Lambda}_a \vec{x}_\eta \in \mathcal{G}^{j-1}$ where $\mathcal{G}^{j-1}$ can now be identified with the space spanned by fields in $\mathcal{F}^{j-2}$ and type y fields of order $j-1$. Similarly the second inner product is zero because (9.114) implies $\overleftrightarrow{\Lambda}_a \vec{y}_\eta \in \mathcal{G}^{j-1}$. Since these inner products are zero we conclude that the auxiliary fields $\vec{a}_{a\omega}$ are orthogonal to $\mathcal{F}^{j-1}$. The type y fields of order j are linear combinations of these auxiliary fields and so must also be orthogonal to the space $\mathcal{F}^{j-1}$. Analogous considerations show that the inner products

$$(\vec{b}_\tau, \vec{x}_\eta) = (\vec{y}_\tau, \overleftrightarrow{\Gamma}_1 \vec{x}_\eta), \qquad (\vec{b}_\tau, \vec{y}_\eta) = (\vec{y}_\tau, \overleftrightarrow{\Gamma}_1 \vec{y}_\eta), \tag{9.118}$$

implied by (9.104) are zero when the string τ has length j. We deduce that the type x fields of order j are also orthogonal to the space $\mathcal{F}^{j-1}$. This completes the proof of orthonormality of the basis. As a corollary, it follows that $\mathcal{X}^j$ and $\mathcal{Y}^j$ represent respectively the type x fields and type y fields of order j.

9.7 Representation of the projection operators and recovery of weight and normalization matrices from series expansion coefficients

Clearly (9.108) and (9.114) determine the action of $\overleftrightarrow{\Lambda}_a$ on the basis fields while (9.109) and (9.115) determine the action of $\overleftrightarrow{\Gamma}_1$. It immediately follows that the projection operators $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$ for $a \neq q$ are represented in

this basis by the block tridiagonal infinite matrices

$$\overleftrightarrow{\Lambda}_a = \begin{bmatrix} \overleftrightarrow{W}^0_a & \overleftrightarrow{M}^1_a & & & \\ (\overleftrightarrow{M}^1_a)^T & \overleftrightarrow{Q}^1_a & 0 & & \\ & & \overleftrightarrow{W}^1_a & \overleftrightarrow{M}^2_a & \\ 0 & & (\overleftrightarrow{M}^2_a)^T & \overleftrightarrow{Q}^2_a & \\ & & & & \ddots \end{bmatrix},$$

$$\overleftrightarrow{\Gamma}_1 = \begin{bmatrix} 0 & 0 & 0 & & \\ & \overleftrightarrow{U}^1 \quad \overleftrightarrow{X}^1 & & & \\ 0 & & 0 & & \\ & \overleftrightarrow{X}^1 \quad \overleftrightarrow{V}^1 & & & \\ & & \overleftrightarrow{U}^2 & \overleftrightarrow{X}^2 & \\ 0 & 0 & \overleftrightarrow{X}^2 & \overleftrightarrow{V}^2 & \\ & & & & \ddots \end{bmatrix}. \tag{9.119}$$

The blocks in these matrices act upon fields of the order indicated by the block superscript, with the exception of the rectangular blocks $(\overleftrightarrow{M}^j_a)^T$ which act on fields of order $j-1$. The blocks going across a given row act on fields alternating between type x and type y, beginning with type x. The tridiagonal form of the matrices representing $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$ reflects the fact that the procedure for constructing the basis fields is similar to the procedure used in the Lanczos algorithm for tridiagonalization of symmetric matrices (see, for example, pages 414–419 of Strang 1986). The operator

$$\overleftrightarrow{\Lambda}_q = I - \sum_{a \neq q} \overleftrightarrow{\Lambda}_a \tag{9.120}$$

also can be represented by the matrix in (9.119) with $a = q$ provided we define $\overleftrightarrow{M}^j_q$ via

$$\overleftrightarrow{M}^j_q \equiv - \sum_{a \neq q} \overleftrightarrow{M}^j_a, \tag{9.121}$$

and $\overleftrightarrow{Q}^j_q$ via (9.116).

The matrix representing $\overleftrightarrow{\Lambda}_{a_1} \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_2} \overleftrightarrow{\Gamma}_1 \cdots \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_s}$ is generated by taking products of the matrices in (9.119). Also for any operator $\overleftrightarrow{B}(\vec{r}, \vec{r}\,')$ with elements $B_{\alpha\beta}(\vec{r}, \vec{r}\,')$ we have

$$\int_\Omega d\vec{r} \int_\Omega d\vec{r}\,' \; B_{\alpha\beta}(\vec{r}, \vec{r}\,') = (\vec{x}_\alpha, \overleftrightarrow{B}\, \vec{x}_\beta). \tag{9.122}$$

In particular then $\overleftrightarrow{\alpha}_{a_1 \ldots a_s}$ is the first block which appears in the matrix representing $\overleftrightarrow{\Lambda}_{a_1} \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_2} \overleftrightarrow{\Gamma}_1 \cdots \overleftrightarrow{\Gamma}_1 \overleftrightarrow{\Lambda}_{a_s}$. In this way we obtain expressions, such as

$$\overleftrightarrow{\alpha}_{a_1} = \overleftrightarrow{W}^0_{a_1}, \tag{9.123}$$

$$\overleftrightarrow{\alpha}_{a_1 a_2} = \overleftrightarrow{M}^1_{a_1} \overleftrightarrow{U}^1 (\overleftrightarrow{M}^1_{a_2})^T, \tag{9.124}$$

$$\overleftrightarrow{\alpha}_{a_1a_2a_3} = \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{U}^1\overleftrightarrow{Q}^1_{a_2}\overleftrightarrow{U}^1(\overleftrightarrow{M}^1_{a_3})^T + \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{X}^1\overleftrightarrow{W}^1_{a_2}\overleftrightarrow{X}^1(\overleftrightarrow{M}^1_{a_3})^T, \tag{9.125}$$

$$\begin{aligned}\overleftrightarrow{\alpha}_{a_1a_2a_3a_4} = {} & \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{U}^1\overleftrightarrow{Q}^1_{a_2}\overleftrightarrow{U}^1\overleftrightarrow{Q}^1_{a_3}\overleftrightarrow{U}^1(\overleftrightarrow{M}^1_{a_4})^T + \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{U}^1\overleftrightarrow{Q}^1_{a_2}\overleftrightarrow{X}^1\overleftrightarrow{W}^1_{a_3}\overleftrightarrow{X}^1(\overleftrightarrow{M}^1_{a_4})^T \\ & + \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{X}^1\overleftrightarrow{W}^1_{a_2}\overleftrightarrow{X}^1\overleftrightarrow{Q}^1_{a_3}\overleftrightarrow{U}^1(\overleftrightarrow{M}^1_{a_4})^T + \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{X}^1\overleftrightarrow{W}^1_{a_2}\overleftrightarrow{V}^1\overleftrightarrow{W}^1_{a_3}\overleftrightarrow{X}^1(\overleftrightarrow{M}^1_{a_4})^T \\ & + \overleftrightarrow{M}^1_{a_1}\overleftrightarrow{X}^1\overleftrightarrow{M}^2_{a_2}\overleftrightarrow{U}^2(\overleftrightarrow{M}^2_{a_3})^T\overleftrightarrow{X}^1(\overleftrightarrow{M}^1_{a_4})^T,\end{aligned} \tag{9.126}$$

for the $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ in terms of the normalization and weight matrices. Conversely, if the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ are known then (9.123) gives $\overleftrightarrow{W}^0_{a_1}$, and (9.124),(9.125), and (9.126) can be solved successively for $\overleftrightarrow{U}^1$, $\overleftrightarrow{W}^1_{a_2}$, and $\overleftrightarrow{U}^2$. These enter the equations linearly. Prior to solving each one of these equations it is necessary to determine the remaining matrices, $\overleftrightarrow{M}^1_{a_1}, \overleftrightarrow{Q}^1_{a_1}, \overleftrightarrow{X}^1, \overleftrightarrow{V}^1$ or $\overleftrightarrow{M}^2_{a_2}$ that also enter the equation in question: these are obtained from their definitions (9.111), (9.116), (9.110), and (9.101), which give them in terms of the matrices $\overleftrightarrow{W}^0_{a_1}, \overleftrightarrow{U}^1$, or $\overleftrightarrow{W}^1_{a_2}$, found from solving the previous equations.

In general the linear equation for the remaining unknown matrix $\overleftrightarrow{U}^j$ or $\overleftrightarrow{W}^j_a$, encountered at respectively the $2j$th stage or $(2j+1)$th stage, will be sandwiched between products of the matrices $\overleftrightarrow{M}^i$ and $\overleftrightarrow{X}^i$. These linear equations have a solution if we assume, as before, that the positive semidefinite matrices $\overleftrightarrow{M}^i$ and $\overleftrightarrow{X}^i$ are nonsingular for all $i \leq j$.

It can be checked through matrix multiplication that the set of matrices $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$ defined via (9.119) are projection operators satisfying (9.19), (9.36), and (9.37) for any choice of normalization and weight matrices satisfying (9.83). Consequently any further restrictions on the set of possible normalization and weight matrices must come from additional information about the operators $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$, such as the identity (9.70) which holds for two-dimensional composites.

Note that we have only shown that the weights and normalization matrices can be recovered from the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$. A separate question, which we do not address, is whether these coefficients can be recovered from the series expansion (9.55) in powers of the elements of the matrices $\boldsymbol{\epsilon_a}$, $a = 1, 2, \ldots, p$. Since the matrix $\boldsymbol{\epsilon_a}\boldsymbol{L}_0^{-1}$ does not generally commute with $\boldsymbol{\epsilon_b}\boldsymbol{L}_0^{-1}$, when $b \neq a$ it seems likely that one should be able to recover the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ if p was sufficiently large. But without a proof the most we can say is what we said in the introduction: that the series probably contains sufficient information to determine the weight and normalization matrices.

9.8 Simplification for two-dimensional, isotropic composites

It can be checked through matrix multiplication that the set of matrices $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$ defined via (9.119) are projection operators satisfying (9.19), (9.36), and (9.37) for any choice of normalization and weight matrices satisfying (9.83). Consequently any further restrictions on the set of possible normalization and weight matrices must come from additional information about the operators $\overleftrightarrow{\Gamma}_1$ and $\overleftrightarrow{\Lambda}_a$, such as the identity (9.70) which holds for two-dimensional composites.

In particular, if the composite is two-dimensional, statistically isotropic and has isotropic components then (9.70) implies that each normalization matrix is simply the identity matrix. Indeed, the isotropy of the composite implies $\overleftrightarrow{\boldsymbol{L}}^* = \overleftrightarrow{I}\ \boldsymbol{L}^*$ for all choices of moduli $\boldsymbol{L_a}$ and consequently all the coefficients $\overleftrightarrow{\alpha}_{a_1\ldots a_s}$ are

also proportional to $\overleftrightarrow{I}$. It follows that the weights and normalization matrices are also proportional to $\overleftrightarrow{I}$ in their space indices:

$$W^j_{c,a_1\ldots a_s\alpha,b_1\ldots b_s\beta} = w^j_{c,a_1\ldots a_s,b_1\ldots b_s}\delta_{\alpha\beta},$$
$$N^j_{a_1\ldots a_s\alpha,b_1\ldots b_s\beta} = n^j_{a_1\ldots a_s,b_1\ldots b_s}\delta_{\alpha\beta}, \tag{9.127}$$

and hence commute with $\overleftrightarrow{R}_\perp$. The isotropy of the components implies $\overleftrightarrow{R}_\perp$ also commutes with the operators $\overleftrightarrow{\Lambda}_a$. We next need to establish that

$$\overleftrightarrow{R}_\perp \vec{x}_{a_1\ldots a_s\alpha} = (-1)^s \sum_{\beta=1}^{2} R^\perp_{\alpha\beta}\vec{x}_{a_1\ldots a_s\beta}, \tag{9.128}$$

$$\overleftrightarrow{R}_\perp \vec{y}_{a_1\ldots a_s\alpha} = (-1)^{s+1} \sum_{\beta=1}^{2} R^\perp_{\alpha\beta}\vec{y}_{a_1\ldots a_s\beta}. \tag{9.129}$$

To see this first observe that (9.71) implies that $\mathcal{F}^j$ and $\mathcal{G}^j$ are each closed under the action of $\overleftrightarrow{R}_\perp$, and as a consequence so are the spaces $\mathcal{X}^j$ and $\mathcal{Y}^j$. Now we proceed by supposing there exists an j such that (9.128) holds true for all $s \leq j-1$ and for all permutations of indices: this is clearly true when $j = 0$. Now for any strings ρ and ϕ of length $j-1 \geq 0$, (9.98) and (9.106) imply

$$\vec{y}_{b\rho} - \sum_{a\neq q}\sum_{\omega} C^j_{b\rho,a\omega}(\overleftrightarrow{\Lambda}_a\vec{x}_\omega) \in \mathcal{F}^{j-1}, \tag{9.130}$$

$$\vec{x}_\phi - \sum_{\tau} D^j_{\phi,\tau}(\overleftrightarrow{\Gamma}_1\vec{y}_\tau) \in \mathcal{G}^j, \tag{9.131}$$

where ω has length $j-1$. By our supposition we can use to (9.128) to compute the action of $\overleftrightarrow{R}_\perp$ on $\vec{x}_\omega$. Applying $\overleftrightarrow{R}_\perp$ to (9.130) and using (9.127) and (9.71) brings one to the conclusion that

$$\overleftrightarrow{R}_\perp\vec{y}_{a_1\ldots a_j\alpha} + (-1)^j \sum_{\beta=1}^{2} R^\perp_{\alpha\beta}\vec{y}_{a_1\ldots a_j\beta} \in \mathcal{F}^{j-1}, \tag{9.132}$$

for all combinations of indices. But $\mathcal{Y}^j$ is closed under the action of $\overleftrightarrow{R}_\perp$ and since $\mathcal{Y}^j$ is orthogonal to $\mathcal{F}^{j-1}$ we infer that the field in (9.132) is zero, i.e., that (9.129) holds for $s = m$. Applying $\overleftrightarrow{R}_\perp$ to (9.131) and using a similar argument establishes that (9.128) holds when $s = m+1$. By induction this completes the proof of (9.128) and (9.129). In turn these imply via (9.70) that

$$U^j_{\tau,\phi} = (\vec{y}_\tau, \overleftrightarrow{\Gamma}_1\vec{y}_\phi) = (\vec{y}_\tau, (\overleftrightarrow{R}_\perp)^T\overleftrightarrow{\Gamma}_1\overleftrightarrow{R}_\perp\vec{y}_\phi) = (\vec{y}_\tau, (I - \overleftrightarrow{\Gamma}_1)\vec{y}_\phi) = \delta_{\tau\phi} - U^j_{\tau,\phi}. \tag{9.133}$$

From the definition (9.102) it follows that

$$\overleftrightarrow{N}^j = \overleftrightarrow{I}^j, \tag{9.134}$$

and consequently the operator $\overset{\leftrightarrow}{\Gamma}_1$ is represented by the matrix

$$\overset{\leftrightarrow}{\Gamma}_1 = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & & \\ & \overset{\leftrightarrow}{I}^1 \quad \overset{\leftrightarrow}{I}^1 & & & \\ 0 & \overset{\leftrightarrow}{I}^1 \quad \overset{\leftrightarrow}{I}^1 & 0 & & \\ & & \overset{\leftrightarrow}{I}^2 \quad \overset{\leftrightarrow}{I}^2 & & \\ 0 & 0 & \overset{\leftrightarrow}{I}^2 \quad \overset{\leftrightarrow}{I}^2 & & \\ & & & & \ddots \end{bmatrix}. \tag{9.135}$$

Note also from (9.128) and (9.129) that $\overset{\leftrightarrow}{R}_\perp$ has the representation,

$$\overset{\leftrightarrow}{R}_\perp = \begin{bmatrix} \overset{\leftrightarrow}{R}^0_\perp & & & & \\ & -\overset{\leftrightarrow}{R}^1_\perp & 0 & & \\ & 0 & \overset{\leftrightarrow}{R}^1_\perp & & \\ & & & -\overset{\leftrightarrow}{R}^2_\perp & \\ & & & & \ddots \end{bmatrix}, \tag{9.136}$$

where $\overset{\leftrightarrow}{R}^j_\perp$ is the rotation matrix with elements

$$R^{j\perp}_{a_1\ldots a_s\alpha,b_1\ldots b_s\beta} = R^\perp_{\alpha\beta}\prod_{i=1}^{s}\delta_{a_ib_i}. \tag{9.137}$$

When $p = 2$ and the composite is two-dimensional but possibly anisotropic, the set of all possible sequences of weight and normalization matrices has been completely characterized, and furthermore microgeometries have been identified which correspond to every such sequence. This was accomplished by Milton (1986b) for composites of two isotropic phases and by Clark and Milton (1994) for a polycrystal built from a single anisotropic crystal. In both cases the microgeometries that can simulate any sequence were found to be a hierarchical laminate. These two-dimensional microstructures can mimic the entire behavior of $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ as a function of the component moduli while keeping the microstructure fixed.

9.9 Bounds and methods for bounding the effective tensor

Bounds on the effective tensor $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ follow directly from the variational principles,

$$\vec{\boldsymbol{E}}_0\cdot\overset{\leftrightarrow}{\boldsymbol{L}}^*\vec{\boldsymbol{E}}_0 = \min_{\vec{\boldsymbol{e}}(\vec{r})}\int_\Omega d\vec{r}\,(\vec{\boldsymbol{E}}_0+\vec{\boldsymbol{e}}(\vec{r}))\cdot\overset{\leftrightarrow}{\boldsymbol{L}}(\vec{r})(\vec{\boldsymbol{E}}_0+\vec{\boldsymbol{e}}(\vec{r})), \tag{9.138}$$

$$\vec{\boldsymbol{J}}_0\cdot(\overset{\leftrightarrow}{\boldsymbol{L}}^*)^{-1}\vec{\boldsymbol{J}}_0 = \min_{\vec{\boldsymbol{j}}(\vec{r})}\int_\Omega d\vec{r}\,(\vec{\boldsymbol{J}}_0+\vec{\boldsymbol{j}}(\vec{r}))\cdot(\overset{\leftrightarrow}{\boldsymbol{L}}(\vec{r}))^{-1}(\vec{\boldsymbol{J}}_0+\vec{\boldsymbol{j}}(\vec{r})), \tag{9.139}$$

where $\vec{\boldsymbol{E}}_0$ and $\vec{\boldsymbol{J}}_0$ are uniform fields, and the minimization extends over statistically homogeneous or periodic fields $\vec{\boldsymbol{e}}(\vec{r})$ and $\vec{\boldsymbol{j}}(\vec{r})$ satisfying

$$\vec{\nabla} \times \vec{\boldsymbol{e}}(\vec{r}) = 0, \qquad \int_\Omega d\vec{r}\, \vec{\boldsymbol{e}}(\vec{r}) = 0, \tag{9.140}$$

$$\vec{\nabla} \cdot \vec{\boldsymbol{j}}(\vec{r}) = 0, \qquad \int_\Omega d\vec{r}\, \vec{\boldsymbol{j}}(\vec{r}) = 0. \tag{9.141}$$

Substitution of the trial fields $\vec{\boldsymbol{e}}(\vec{r}) = 0$ and $\vec{\boldsymbol{j}}(\vec{r}) = 0$ gives the arithmetic and harmonic mean bounds,

$$[\sum_{a=1}^{p} \overset{\leftrightarrow}{\alpha}_a\, (\boldsymbol{L_a})^{-1}]^{-1} \le \overset{\leftrightarrow}{\boldsymbol{L}}^{*j-1} \le \sum_{a=1}^{p} \overset{\leftrightarrow}{\alpha}_a\, \boldsymbol{L_a}. \tag{9.142}$$

Better bounds result from a more judicious choice of trial fields. For example, to derive improved upper bounds one can follow the approach of Beran (1965, 1966) and choose a trial field of the form

$$\vec{\boldsymbol{e}}(\vec{r}) = \sum_{s=1}^{j} \sum_{a_1,..,a_s=1}^{p} \sum_{\alpha=1}^{3} \boldsymbol{c}_{a_1 a_2 \ldots a_s \alpha} \vec{e}_{a_1 a_2 \ldots a_s \alpha}(\vec{r}), \tag{9.143}$$

where the fields $\vec{e}_{a_1 a_2 \ldots a_s \alpha}(\vec{r})$ are given by (9.86), and then minimize (9.138) to find the best choice of the coefficients $\boldsymbol{c}_{a_1 a_2 \ldots a_s \alpha}$, which are vectors in the field indices. The bound generated by this procedure when expanded in a power series agrees with the terms in the series (9.55) for all s up to and including $s = 2j+1$, and for this reason is called the Wiener–Beran type upper bound of order $2j+1$: a bound is said to be of order m if the series expansion of the bound and the series expansion of $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ agree for all s up to and including $s = m$. An analogous choice of trial field $\vec{\boldsymbol{j}}(\vec{r})$ generates the Wiener–Beran type lower bound of order $2j+1$ through the variational principle (9.139). Bounds of even order are generated by substituting an appropriate choice of trial polarization field into the Hashin–Shtrikman variational principles (Hashin and Shtrikman 1962). yielding Hashin–Shtrikman type bounds.

These bounds on $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ are naturally expressed in terms of the normalization and weight matrices. For this purpose it is useful to expand $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ as a continued fraction rather than as a power series. A direct extension of the analysis of Milton (1987a,1987b) gives a continued fraction expansion for the effective tensor

$$\overset{\leftrightarrow}{\boldsymbol{L}}^* \equiv \overset{\leftrightarrow}{\boldsymbol{L}}^{*0}, \tag{9.144}$$

generated by setting

$$\boldsymbol{L}_0 = \boldsymbol{L}_q, \tag{9.145}$$

and eliminating the tensors $\overset{\leftrightarrow}{\boldsymbol{L}}^{*j}$ for $j \ge 1$ from the recursion relations

$$\begin{aligned} \overset{\leftrightarrow}{\boldsymbol{L}}^{*j-1} = & \sum_{a=1}^{p} \overset{\leftrightarrow}{W}{}_a^{j-1} \boldsymbol{L_a} \\ - \sum_{a,b \ne q} \boldsymbol{\epsilon_a} \overset{\leftrightarrow}{M}{}_a^j \{ & \overset{\leftrightarrow}{I}{}^j \boldsymbol{L}_0 + \sum_{c \ne q} \overset{\leftrightarrow}{Q}{}_c^j \boldsymbol{\epsilon_c} + (\overset{\leftrightarrow}{N}{}^j)^{1/2} \overset{\leftrightarrow}{\boldsymbol{L}}^{*j} (\overset{\leftrightarrow}{N}{}^j)^{1/2} \}^{-1} (\overset{\leftrightarrow}{M}{}_b^j)^T \boldsymbol{\epsilon_b}, \end{aligned} \tag{9.146}$$

where, in accordance with our previous definitions,

$$\overleftrightarrow{Q}^j_c = \overleftrightarrow{M}^j_c(\overleftrightarrow{W}^{j-1}_c)^{-1}(\overleftrightarrow{M}^j_c)^T, \quad \overleftrightarrow{M}^j = (\overleftrightarrow{Y}^j)^{1/2}, \quad Y^j_{a\omega,b\rho} = \delta_{ab}W^{j-1}_{a,\omega,\rho} - \sum_\zeta W^{j-1}_{a,\omega,\zeta}W^{j-1}_{b,\zeta,\rho}, \tag{9.147}$$

and $\overleftrightarrow{M}^j_a$, with transpose $(\overleftrightarrow{M}^j_a)^T$, is the rectangular submatrix of the square matrix $\overleftrightarrow{M}^j$ with elements $M^j_{a\tau,\lambda}$ labeled by the strings τ and λ. Note that $\overleftrightarrow{\boldsymbol{L}}^{*j}$ has elements $\boldsymbol{L}^{*j}_{\tau k\eta m}$ labeled by field indices $k, m \in \{1, 2..p\}$ and string indices $\tau = a_1a_2..a_j\alpha$, $\mu = b_1b_2..b_j\beta$ with a_i and $b_i \in \{1, 2, ..q-1, q+1, ..p\}$, and α and $\beta \in \{1, 2, 3\}$. Also note that $\overleftrightarrow{W}^j_a$, $\overleftrightarrow{M}^j_a$ and $\overleftrightarrow{N}^j$ act on the string indices, not on the field indices.

There are other equivalent ways of expressing $\overleftrightarrow{\boldsymbol{L}}^{*j-1}$ in terms of $\overleftrightarrow{\boldsymbol{L}}^{*j}$ (Milton 1987a). For example (9.146) can be replaced by its dual form

$$\begin{aligned}(\overleftrightarrow{\boldsymbol{L}}^{*j-1})^{-1} &= \sum_{a=1}^p \overleftrightarrow{W}^{j-1}_a(\boldsymbol{L_a})^{-1} \\ - \sum_{a,b\neq q} \boldsymbol{\eta_a}\overleftrightarrow{M}^j_a\{\overleftrightarrow{I}^j\boldsymbol{L}_0^{-1} + \sum_{c\neq q}\overleftrightarrow{Q}^j_c\boldsymbol{\eta_c} &+ (\overleftrightarrow{N}^j)^{-1/2}(\overleftrightarrow{\boldsymbol{L}}^{*j})^{-1}(\overleftrightarrow{N}^j)^{-1/2}\}^{-1}(\overleftrightarrow{M}^j_b)^T\boldsymbol{\eta_b},\end{aligned} \tag{9.148}$$

where

$$\boldsymbol{\eta_a} \equiv (\boldsymbol{L_a})^{-1} - \boldsymbol{L}_0^{-1}. \tag{9.149}$$

Eliminating the matrices $\overleftrightarrow{\boldsymbol{L}}^{*j}$ from this recursion relation generates an alternative continued fraction expansion of $\overleftrightarrow{\boldsymbol{L}}^*$.

The tensors $\overleftrightarrow{\boldsymbol{L}}^{*j}, j = 0, 1, 2, ...$ have an interpretation in the context of the solution $\vec{\boldsymbol{J}}(\vec{r})$ for any given field $\vec{\boldsymbol{E}}_0 \in \mathcal{X}^j$ (the space $\mathcal{X}^j$ now plays the role that was played by the uniform fields) to the equations

$$\overleftrightarrow{\Gamma}^j_1\vec{\boldsymbol{J}} = 0, \quad \vec{\boldsymbol{J}}(\vec{r}) = \overleftrightarrow{\boldsymbol{L}}(\vec{r})(\vec{\boldsymbol{E}}_0(\vec{r}) + \vec{\boldsymbol{e}}(\vec{r})), \quad \overleftrightarrow{\Gamma}^j_1\vec{\boldsymbol{e}} = \vec{\boldsymbol{e}}, \tag{9.150}$$

where $\overleftrightarrow{\Gamma}^j_1$ is the nonlocal operator,

$$\overleftrightarrow{\Gamma}^j_1 = \overleftrightarrow{\Gamma}_1 - \overleftrightarrow{\Upsilon}^j, \tag{9.151}$$

and $\overleftrightarrow{\Upsilon}^j$ (which commutes with $\overleftrightarrow{\Gamma}_1$) is the projection onto the space

$$\mathcal{E}^j \equiv \{\vec{u}(\vec{r}) \in \mathcal{X}^j \oplus \mathcal{Y}^j | \overleftrightarrow{\Gamma}_1\vec{u} = \vec{u}\} \tag{9.152}$$

of order j fields which are curl-free and have zero average value. In the representation (9.119) $\overleftrightarrow{\Gamma}^j_1$ is obtained from $\overleftrightarrow{\Gamma}_1$ by setting the blocks $\overleftrightarrow{U}^j, \overleftrightarrow{V}^j$ and $\overleftrightarrow{X}^j$ to zero. Note that $\overleftrightarrow{\Gamma}^j_1$ is a projection and acts upon any field to produce a curl-free field with zero average value. So in particular $\vec{\boldsymbol{e}}(\vec{r})$ (but not $\vec{\boldsymbol{E}}_0(\vec{r})$) is the gradient of a potential.

A simple application of the Lax–Milgram theorem (see, for example, Section 5.8 of Gilbarg and Trudinger 1983) shows that these equations always have a unique solution for $\vec{\boldsymbol{J}}(\vec{r})$, for any choice of field $\vec{\boldsymbol{E}}_0 \in \mathcal{X}^j$,

provided that the set of tensors $\overset{\leftrightarrow}{\Lambda}_a$ are positive definite and bounded. Let us define $\overset{\leftrightarrow}{\Gamma}_0^j$ as the projection onto the subspace $\mathcal{X}^j$ and $\vec{\boldsymbol{J}}_0$ as the component

$$\vec{\boldsymbol{J}}_0 = \overset{\leftrightarrow}{\Gamma}_0^j \boldsymbol{J}, \tag{9.153}$$

of the field $\vec{\boldsymbol{J}}(\vec{r})$ which lies in the subspace $\mathcal{X}^j$. Since the relation between $\vec{\boldsymbol{J}}_0$ and $\vec{\boldsymbol{E}}_0$ is linear we can write

$$\vec{\boldsymbol{J}}_0 = \overset{\leftrightarrow}{\boldsymbol{L}}^{*j} \vec{\boldsymbol{E}}_0. \tag{9.154}$$

This linear relation serves to define $\overset{\leftrightarrow}{\boldsymbol{L}}^{*j}$: it is a linear map from the space $\mathcal{X}^j$ to itself. When $j = 0$ these equations reduce to the previous set (9.3), (9.10), and (9.12) and so we can make the identification (9.144) between $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ and $\overset{\leftrightarrow}{\boldsymbol{L}}^{*0}$.

From the matrix representation (9.119) of the operators $\overset{\leftrightarrow}{\Lambda}_a$ and $\overset{\leftrightarrow}{\Gamma}_1$ it is clear that $\overset{\leftrightarrow}{\Lambda}_a$ and $\overset{\leftrightarrow}{\Gamma}_1^j$ do not couple $\vec{\boldsymbol{E}}_0$ with fields in the space $\mathcal{G}^j$. Thus the fields in $\mathcal{G}^j$ play no role in the solutions of the equations (9.150). Consequently we can now eliminate from our basis those fields $\vec{x}_\tau, \vec{y}_\tau \in \mathcal{G}^j$. In the remaining reduced basis the operators $\overset{\leftrightarrow}{\Lambda}_a$ and $\overset{\leftrightarrow}{\Gamma}_1^j$ have the representation

$$\overset{\leftrightarrow}{\Lambda}_a = \begin{bmatrix} \overset{\leftrightarrow}{W}_a^j & \overset{\leftrightarrow}{M}_a^{j+1} & & & \\ (\overset{\leftrightarrow}{M}_a^{j+1})^T & \overset{\leftrightarrow}{Q}_a^{j+1} & 0 & & \\ & & \overset{\leftrightarrow}{W}_a^{j+1} & \overset{\leftrightarrow}{M}_a^{j+2} & \\ 0 & & (\overset{\leftrightarrow}{M}_a^{j+2})^T & \overset{\leftrightarrow}{Q}_a^{j+2} & \\ & & & & \ddots \end{bmatrix},$$

$$\overset{\leftrightarrow}{\Gamma}_1^j = \begin{bmatrix} 0 & 0 & 0 & & \\ & \overset{\leftrightarrow}{U}^{j+1} \quad \overset{\leftrightarrow}{X}^{j+1} & & & \\ 0 & \overset{\leftrightarrow}{X}^{j+1} \quad \overset{\leftrightarrow}{V}^{j+1} & 0 & & \\ & & \overset{\leftrightarrow}{U}^{j+2} & \overset{\leftrightarrow}{X}^{j+2} & \\ 0 & 0 & \overset{\leftrightarrow}{X}^{j+2} & \overset{\leftrightarrow}{V}^{j+2} & \\ & & & & \ddots \end{bmatrix}. \tag{9.155}$$

The similarity with (9.119) makes it evident that whatever role the sequence $\overset{\leftrightarrow}{W}_a^0$,$\overset{\leftrightarrow}{N}^1$,$\overset{\leftrightarrow}{W}_a^1$,$\overset{\leftrightarrow}{N}^2$,$\overset{\leftrightarrow}{W}_a^2$,... of weight and normalization matrices plays in determining $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ is played in an identical way by the sequence $\overset{\leftrightarrow}{W}_a^j$,$\overset{\leftrightarrow}{N}^{j+1}$,$\overset{\leftrightarrow}{W}_a^{j+1}$,$\overset{\leftrightarrow}{N}^{j+2}$,$\overset{\leftrightarrow}{W}_a^{j+2}$,... in determining $\overset{\leftrightarrow}{\boldsymbol{L}}^{*j}$. This self-similarity is also evident from the continued fraction expansions for $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ and $\overset{\leftrightarrow}{\boldsymbol{L}}^{*j}$ implied by (9.146).

If the entire set of normalization and weight matrices is known then these continued fractions expansions allow the effective tensor $\overset{\leftrightarrow}{\boldsymbol{L}}^*$ to be computed to an arbitrarily high degree of accuracy. For example we could truncate the continued fraction at some stage m by setting

$$\overset{\leftrightarrow}{\boldsymbol{L}}^{*m} = \overset{\leftrightarrow}{I}^m \boldsymbol{L}_q, \tag{9.156}$$

which is a natural choice, corresponding to replacing the set of weights $\overset{\leftrightarrow}{W}{}_a^m$ by the weights

$$\overset{\leftrightarrow}{W}{}_q^m = \overset{\leftrightarrow}{I}{}^m, \quad \overset{\leftrightarrow}{W}{}_a^m = 0, \ \ \forall a \neq q, \tag{9.157}$$

consistent with the constraints (9.83). Then the tensor $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*0}$ obtained from the recursion relations (9.146) is an m-th order rational approximate to $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$, and it can be proved that this approximate converges to $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ as m tends to infinity, for any positive definite bounded set of moduli $\boldsymbol{L_a}, a = 1, .., p$ (Milton 1987b). The approximates also converge when the moduli are complex, provided the tensors $\boldsymbol{L_a}$ are symmetric and bounded and such that there exists a phase angle θ for which

$$\mathrm{Re}(e^{i\theta}\boldsymbol{L_a}) > 0, \quad \forall a, \tag{9.158}$$

where $\mathrm{Re}(A)$ denotes the real part of the quantity A. Such complex moduli have a physical interpretation. When the fields $\vec{\boldsymbol{J}}$ and $\vec{\boldsymbol{E}}$ oscillate sinusoidally in time t with frequency ω then they can be expressed as the real part of complex fields $\vec{\boldsymbol{J}}_c(\vec{r})$ and $\vec{\boldsymbol{E}}_c(\vec{r})$,

$$\vec{\boldsymbol{J}}(\vec{r},\omega) = \mathrm{Re}(e^{iwt}\vec{\boldsymbol{J}}_c(\vec{r})), \quad \vec{\boldsymbol{E}}(\vec{r},\omega) = \mathrm{Re}(e^{iwt}\vec{\boldsymbol{E}}_c(\vec{r})). \tag{9.159}$$

Provided the wavelength of this oscillation is sufficiently large compared with the microstructure these complex fields satisfy the quasistatic equations,

$$\vec{\nabla}\cdot\vec{\boldsymbol{J}}_c(\vec{r}) = 0, \quad \vec{\nabla}\times\vec{\boldsymbol{E}}_c(\vec{r}) = 0, \quad \vec{\boldsymbol{J}}_c(\vec{r}) = \overset{\leftrightarrow}{\boldsymbol{L}}(\vec{r})\vec{\boldsymbol{E}}_c(\vec{r}), \tag{9.160}$$

with a complex tensor $\overset{\leftrightarrow}{\boldsymbol{L}}(\vec{r})$ given by

$$\overset{\leftrightarrow}{\boldsymbol{L}}(\vec{r}) = \sum_{a=1}^{p} \overset{\leftrightarrow}{\Lambda}_a \boldsymbol{L_a}, \tag{9.161}$$

where the moduli $\boldsymbol{L_a}$ are complex and frequency dependent. The thermodynamic requirement that dissipation of power into entropy be positive ensures that (9.158) holds when $\theta = 0$. Each rational approximate satisfies the properties of covariance and disjunction, discussed in the introduction, and has the additional required analytic property that

$$\mathrm{Re}(e^{i\theta}\ \overset{\leftrightarrow}{\boldsymbol{L}}{}^*) > 0, \tag{9.162}$$

for any set of tensors $\boldsymbol{L_a}$ satisfying (9.158).

Bounds on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ follow from elementary bounds on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$. In particular, the inequalities

$$0 \leq \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} \leq \infty \overset{\leftrightarrow}{I}{}^j, \tag{9.163}$$

or equivalently the inequalities

$$[\sum_{a=1}^{p} \overset{\leftrightarrow}{W}{}_a^{j-1}(\boldsymbol{L_a})^{-1}]^{-1} \leq \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1} \leq \sum_{a=1}^{p} \overset{\leftrightarrow}{W}{}_a^{j-1}\boldsymbol{L_a}, \tag{9.164}$$

when substituted in the recursion relations (9.146) or (9.149) produce the Wiener–Beran type bounds on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ of order $2j-1$, while the inequalities

$$\boldsymbol{L}^- \overset{\leftrightarrow}{I}{}^j \leq \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} \leq \boldsymbol{L}^+ \overset{\leftrightarrow}{I}{}^j, \tag{9.165}$$

which hold for all tensors $\boldsymbol{L}^-$ and $\boldsymbol{L}^+$ such that

$$\boldsymbol{L}^- \le \boldsymbol{L}_{\boldsymbol{a}} \le \boldsymbol{L}^+, \quad 1 \le a \le p, \tag{9.166}$$

when substituted in (9.146) or (9.149) produce the Hashin–Shtrikman type bounds on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ of order $2j$. By substitution we mean precisely that an upper bound on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ is obtained by setting $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} = \infty \overset{\leftrightarrow}{I}{}^j$ or $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} = \boldsymbol{L}^+$ and solving the recursion relations for $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*0}$ and that a lower bound on $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ is obtained by setting $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} = 0$ or $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} = \boldsymbol{L}^-$ and solving for $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*0}$.

9.10 Bounds using the field-equation recursion method

The inequalities (9.164) and (9.165) can be easily derived without reference to variational principles using the field recursion method for bounding effective tensors. This approach utilizes the recursive structure of the equations (9.146) and the inequalities (9.83) on the normalization and weight matrices. The first step in the method is to conjecture a set of restrictions that might apply to $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ irrespective of what values the weights and normalization matrices take, subject only to the constraints (9.83)-or perhaps additional constraints if these are known. This conjecture need not be very restrictive, and could be guided by the form of the recursion relations (9.146). For example let us conjecture that $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ is positive semidefinite. The next step is to first check that the tensor $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*m}$ given by (9.156) satisfies the conjecture, and indeed it does. Then the remaining task is to assume the conjecture is true for some j and show this implies $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1}$ also satisfies the conjecture, for any choice of the weight matrices $\overset{\leftrightarrow}{W}{}_a^{j-1}$ and normalization matrices $\overset{\leftrightarrow}{N}{}^j$ satisfying (9.83): it obviously does since from the recursion relations (9.146) and (9.148) it follows that (9.163) implies (9.164) which in turn implies $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1}$ is positive semidefinite. By induction any rational approximate for $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ generated by choosing $m > j$ and making the substitution (9.156) satisfies the conjecture, and since these approximates converge to $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ as m tends to infinity, we conclude that $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ itself must be positive semidefinite. The conjecture is proved and it clearly implies both (9.164) and (9.165). The recursion method has the advantage that it also works when the moduli $\boldsymbol{L}_{\boldsymbol{a}}$ are complex (Milton, 1987a; 1987b)

In the special case of a composite with $p = 2$ the strings of indices merely consist of a repeated string of either $2's$ or $1's$ (according to whether $q = 1$ or $q = 2$) terminated by a space index. Let us drop this redundant information and allow the elements of the weight and normalization matrices to be addressed only by the space indices. Also when $p = 2$ the matrices $\overset{\leftrightarrow}{W}{}_1^j$ and $\overset{\leftrightarrow}{W}{}_2^j = \overset{\leftrightarrow}{I}{}^j - \overset{\leftrightarrow}{W}{}_1^j$ commute and so we have

$$\overset{\leftrightarrow}{Y}{}^j = \overset{\leftrightarrow}{W}{}_1^j \overset{\leftrightarrow}{W}{}_2^j, \qquad \overset{\leftrightarrow}{M}{}^j = (\overset{\leftrightarrow}{W}{}_1^j \overset{\leftrightarrow}{W}{}_2^j)^{1/2}, \qquad \overset{\leftrightarrow}{Q}{}_1^j = \overset{\leftrightarrow}{W}{}_2^j, \qquad \overset{\leftrightarrow}{Q}{}_2^j = \overset{\leftrightarrow}{W}{}_1^j. \tag{9.167}$$

Without loss of generality we take $q = 2$, and correspondingly $\boldsymbol{L}_0 = \boldsymbol{L}_2$. Then the recursion relation (9.146) simplifies to

$$\begin{aligned} \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1} &= \overset{\leftrightarrow}{W}{}_1^{j-1} \boldsymbol{L}_1 + \overset{\leftrightarrow}{W}{}_2^{j-1} \boldsymbol{L}_2 \\ -(\boldsymbol{L}_1 - \boldsymbol{L}_2) \overset{\leftrightarrow}{M}{}^j \{ \overset{\leftrightarrow}{W}{}_1^{j-1} \boldsymbol{L}_2 + \overset{\leftrightarrow}{W}{}_2^{j-1} \boldsymbol{L}_1 + (\overset{\leftrightarrow}{N}{}^j)^{1/2} \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} (\overset{\leftrightarrow}{N}{}^j)^{1/2} \}^{-1} \overset{\leftrightarrow}{M}{}^j (\boldsymbol{L}_1 - \boldsymbol{L}_2), \end{aligned} \tag{9.168}$$

which for $\boldsymbol{L}_1 \neq \boldsymbol{L}_2$ can be inverted to give $\overleftrightarrow{\boldsymbol{L}}^{*j}$ in terms of $\overleftrightarrow{\boldsymbol{L}}^{*j-1}$:

$$\overleftrightarrow{\boldsymbol{L}}^{*j} = (\overleftrightarrow{N}^j)^{-1/2}\{-\overleftrightarrow{W}_1^{j-1}\boldsymbol{L}_2 - \overleftrightarrow{W}_2^{j-1}\boldsymbol{L}_1 + (\boldsymbol{L}_1 - \boldsymbol{L}_2)\overleftrightarrow{M}^j(\overleftrightarrow{W}_1^{j-1}\boldsymbol{L}_1 + \overleftrightarrow{W}_2^{j-1}\boldsymbol{L}_2 - \overleftrightarrow{\boldsymbol{L}}^{*j-1})^{-1}\overleftrightarrow{M}^j(\boldsymbol{L}_1 - \boldsymbol{L}_2)\}(\overleftrightarrow{N}^j)^{-1/2}. \tag{9.169}$$

Supposing that the components are isotropic phases occupying volume fractions f_1 and f_2, (9.63) implies

$$W^0_{1,\alpha,\beta} = f_1\delta_{\alpha\beta}, \qquad W^0_{2,\alpha,\beta} = f_2\delta_{\alpha\beta}, \tag{9.170}$$

and consequently when $j = 1$ (9.169) takes the form

$$\overleftrightarrow{\boldsymbol{L}}^{*1} = (\overleftrightarrow{N}^1)^{-1/2}\overleftrightarrow{\boldsymbol{Y}}^*(\overleftrightarrow{N}^1)^{-1/2}, \tag{9.171}$$

where $\overleftrightarrow{\boldsymbol{Y}}^*$, not to be confused with the matrix $\overleftrightarrow{Y}^j$, is given by

$$\overleftrightarrow{\boldsymbol{Y}}^* = -f_1\ \overleftrightarrow{I}\ \boldsymbol{L}_2 - f_2\ \overleftrightarrow{I}\ \boldsymbol{L}_1 + \ f_1f_2(\boldsymbol{L}_1 - \boldsymbol{L}_2)(f_1\ \overleftrightarrow{I}\ \boldsymbol{L}_1 + f_2\ \overleftrightarrow{I}\ \boldsymbol{L}_2 - \overleftrightarrow{\boldsymbol{L}}^*)^{-1}(\boldsymbol{L}_1 - \boldsymbol{L}_2). \tag{9.172}$$

9.11 Bounds using the translation method

It turns out that bounds on $\overleftrightarrow{\boldsymbol{L}}^*$ derived via the translation method follow from elementary bounds on this tensor $\overleftrightarrow{\boldsymbol{Y}}^*$. This method was discovered independently by Murat and Tartar (1979b;1985;1985) and by Lurie and Cherkaev (1982;1984) and applied to generate bounds that characterize for $n = 1$ the region in tensor space filled by the range of values $\overleftrightarrow{\boldsymbol{L}}^*$ takes as the microstructure varies over all configurations while keeping the moduli $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ and the volume fraction f_1 fixed. Subsequently it was noted that the corresponding region filled by the possible values of $\overleftrightarrow{\boldsymbol{Y}}^*$ did not depend on the choice of volume fraction f_1 (Milton 1986a). Cherkaev and Gibiansky (1992) extended the characterization to $n = 2$, assuming a two-dimensional geometry. Subsequently Clark and Milton (1995) obtained the characterization for arbitrary n, using fractional linear transformations which preserve the analytic properties as functions of the component moduli.

To explain the translation method let us focus on bounding $\overleftrightarrow{\boldsymbol{L}}^*$ from below. Then one needs to find a suitable translation tensor $T_{\alpha i\beta k}$, where i, k are field indices and α, β are space indices, satisfying

$$\overleftrightarrow{I}\ \boldsymbol{L_a}\ \geq\ \overleftrightarrow{\boldsymbol{T}}, \quad a = 1, 2, \tag{9.173}$$

and with the additional property that

$$\int_\Omega d\vec{r}\,\vec{\nabla}\boldsymbol{\psi}\cdot\overleftrightarrow{\boldsymbol{T}}\,\vec{\nabla}\boldsymbol{\psi} \geq 0, \tag{9.174}$$

for all periodic potentials $\boldsymbol{\psi}$ with elements $\psi^k(\vec{r}), k = 1, 2, ..n$. Any positive semidefinite tensor satisfies this last constraint. However the converse is not true, and in fact the interesting applications to bounds come from

translations $\overleftrightarrow{\boldsymbol{T}}$ which are not positive semidefinite. The key idea in the method is to consider a comparison composite with its moduli translated from $\overleftrightarrow{\boldsymbol{L}}(\vec{r})$ to the moduli

$$\overleftrightarrow{\boldsymbol{L'}}(\vec{r}) \equiv \overleftrightarrow{\boldsymbol{L}}(\vec{r}) - \overleftrightarrow{\boldsymbol{T}}, \tag{9.175}$$

which are positive semidefinite as a consequence of (9.173). From (9.174) and from the variational definition (9.138) applied to the effective tensor $\overleftrightarrow{\boldsymbol{L'}}^*$ of the comparison composite we have, for all uniform fields $\vec{\boldsymbol{E}}_0$,

$$\begin{aligned}
\vec{\boldsymbol{E}}_0 \cdot \overleftrightarrow{\boldsymbol{L'}}^* \vec{\boldsymbol{E}}_0 &= \min_{\boldsymbol{\psi}(\vec{r})} \{ \int_\Omega d\vec{r}\, (\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r})) \cdot (\overleftrightarrow{\boldsymbol{L'}}(\vec{r}))(\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r}))\} \\
&= \min_{\boldsymbol{\psi}(\vec{r})} \{ \int_\Omega d\vec{r}\, (\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r})) \cdot (\overleftrightarrow{\boldsymbol{L}}(\vec{r}))(\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r})) \\
&\qquad - \int_\Omega d\vec{r}\, (\vec{\nabla}\boldsymbol{\psi}(\vec{r})) \cdot (\overleftrightarrow{\boldsymbol{T}})(\vec{\nabla}\boldsymbol{\psi}(\vec{r})) \; - \vec{\boldsymbol{E}}_0 \cdot \overleftrightarrow{\boldsymbol{T}}\, \vec{\boldsymbol{E}}_0 \} \\
&\leq \min_{\boldsymbol{\psi}(\vec{r})} \{ \int_\Omega d\vec{r}\, (\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r})) \cdot (\overleftrightarrow{\boldsymbol{L}}(\vec{r}))(\vec{\boldsymbol{E}}_0 - \vec{\nabla}\boldsymbol{\psi}(\vec{r}))\} \; - \vec{\boldsymbol{E}}_0 \cdot \overleftrightarrow{\boldsymbol{T}}\, \vec{\boldsymbol{E}}_0 \\
&= \vec{\boldsymbol{E}}_0 \cdot (\overleftrightarrow{\boldsymbol{L}}^* - \overleftrightarrow{\boldsymbol{T}}) \vec{\boldsymbol{E}}_0,
\end{aligned} \tag{9.176}$$

which is equivalent to the tensor inequality

$$\overleftrightarrow{\boldsymbol{L'}}^* \;\leq\; \overleftrightarrow{\boldsymbol{L}}^* - \overleftrightarrow{\boldsymbol{T}}\,. \tag{9.177}$$

Substituting this in the harmonic mean bounds on $\overleftrightarrow{\boldsymbol{L}}^*$,

$$(\overleftrightarrow{\boldsymbol{L'}}^*)^{-1} \;\leq\; \int_\Omega d\vec{r}\, (\overleftrightarrow{\boldsymbol{L'}}(\vec{r}))^{-1}, \tag{9.178}$$

yields the translation bounds,

$$(\overleftrightarrow{\boldsymbol{L}}^* - \overleftrightarrow{\boldsymbol{T}})^{-1} \;\leq\; \int_\Omega d\vec{r}\, (\overleftrightarrow{\boldsymbol{L}}(\vec{r}) - \overleftrightarrow{\boldsymbol{T}})^{-1}, \tag{9.179}$$

which for composites of two isotropic materials reduces to

$$(\overleftrightarrow{\boldsymbol{L}}^* - \overleftrightarrow{\boldsymbol{T}})^{-1} \;\leq\; f_1 (\overleftrightarrow{I}\, \boldsymbol{L}_1 - \overleftrightarrow{\boldsymbol{T}})^{-1} + f_2 (\overleftrightarrow{I}\, \boldsymbol{L}_2 - \overleftrightarrow{\boldsymbol{T}})^{-1}. \tag{9.180}$$

Cherkaev and Gibiansky (1992) noticed through algebraic manipulation, that these bounds when expressed in terms of $\overleftrightarrow{\boldsymbol{Y}}^*$ simplify to

$$\overleftrightarrow{\boldsymbol{Y}}^* + \overleftrightarrow{\boldsymbol{T}} \;\geq\; 0. \tag{9.181}$$

In their proof they assumed that $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ commute. Later this assumption was found unnecessary and moreover a direct and simple proof of (9.181) was found from a variational expression for $\overleftrightarrow{\boldsymbol{Y}}^*$ (Milton 1991). An interesting feature of the translation method is that the sharpest bounds are usually obtained from translations $\overleftrightarrow{\boldsymbol{T}}$ with couplings between the fields, even when $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$, and hence $\overleftrightarrow{\boldsymbol{L}}^*$, have no such couplings.

When $p > 2$ the transformation (9.146) cannot simply be inverted because the matrices $\overset{\leftrightarrow}{M}{}_a^j$ are rectangular and have no unique inverse. Also it is clear that the tensor $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ is larger than the tensor $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1}$ and so contains more information. However if more than one field was present, i.e., if $n \geq 2$, and if $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ was known as a function of the $\boldsymbol{L_a}$, then in principle one could expand $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ in a power series, possibly extract the coefficients $\overset{\leftrightarrow}{\alpha}_{a_1 \ldots a_s}$ and subsequently find the weights and normalization matrices. By this means one could recover both $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ and

$$\overset{\leftrightarrow}{\boldsymbol{Y}}{}^{*j} \equiv (\overset{\leftrightarrow}{N}{}^j)^{-1/2} \overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j} (\overset{\leftrightarrow}{N}{}^j)^{-1/2} \tag{9.182}$$

as a function of the $\boldsymbol{L_a}$ through the continued fraction formula for $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j}$ implied by (9.146). Naturally we expect that there exists a more direct way of recovering the function $\overset{\leftrightarrow}{\boldsymbol{Y}}{}^{*j}(\boldsymbol{L}_1, \boldsymbol{L}_2, .., \boldsymbol{L}_p)$ from the function $\overset{\leftrightarrow}{\boldsymbol{L}}{}^{*j-1}(\boldsymbol{L}_1, \boldsymbol{L}_2, .., \boldsymbol{L}_p)$. One intriguing question is whether this direct recovery process, whatever it is, works when $n = 1$. If it does then the sequence of matrices $\overset{\leftrightarrow}{N}{}^j$ and $\overset{\leftrightarrow}{W}{}_a^j$ could be recovered by expanding each function $\overset{\leftrightarrow}{\boldsymbol{Y}}{}^{*j-1}(\boldsymbol{L}_1, \boldsymbol{L}_2, .., \boldsymbol{L}_p)$ to first order, and consequently $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ could be calculated even when more than one field is present. In other words, knowledge of the conductivity function $\overset{\leftrightarrow}{\sigma}{}^*(\sigma_1, \sigma_2, ...\sigma_p)$ without couplings would be sufficient to uniquely determine the effective tensor $\overset{\leftrightarrow}{\boldsymbol{L}}{}^*$ with couplings present.

Acknowledgments

This manuscript was largely complete in 1992, and at that time Graeme W. Milton was an associate professor at the Courant Institute, and a recipient of a Packard Fellowship, and Mordehai Milgrom was a visitor to the Courant Institute. Conversely Graeme W. Milton benefited from a visit to the Weizmann Institute. The Courant Institute, the Packard Foundation, the University of Utah, and the Weizmann Institute are gratefully thanked for their support.

10 A rigorous approach to the field recursion method for two-component composites with isotropic phases

Maxence Cassier*, Aaron Welters and Graeme W. Milton***

*Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA
**Department of Mathematical Sciences, Florida Institute of Technology, Melbourne, FL 32901, USA
Emails: cassier@math.utah.edu, awelters@fit.edu, milton@math.utah.edu

Abstract

In this chapter we give a rigorous derivation of the field equation recursion method in the abstract theory of composites to two-component composites with isotropic phases. This method is of great interest since it has proven to be a powerful tool in developing sharp bounds for the effective tensor of a composite material. The reason is that the effective tensor $\mathbf{L}_*$ can be interpreted in the general framework of the abstract theory of composites as the Z-operator on a certain orthogonal $Z(2)$ subspace collection. The base case of the recursion starts with an orthogonal $Z(2)$ subspace collection on a Hilbert space $\mathcal{H}$, the Z-problem, and the associated Y-problem. We provide some new conditions for the solvability of both the Z-problem and the associated Y-problem. We also give explicit representations of the associated Z-operator and Y-operator and study their analytical properties. An iteration method is then developed from a hierarchy of subspace collections and their associated operators which leads to a continued fraction representation of the initial effective tensor $\mathbf{L}_*$.

Key words: field recursion method, abstract theory of composites, effective tensors, subspace collections, Z-operators, Y-operators, analytic properties

10.1 Introduction

In this chapter we give a rigorous derivation of the field equation recursion method (Milton and Golden 1985; Milton 1987a, 1987b, 1991; Clark and Milton 1994; Clark 1997 and **Chapter 29** of Milton 2002 in the abstract theory of composites as described in **Chapter 2** of this book). The field equation recursion method utilizes information obtained from series expansions to derive bounds on the effective response. The series expansion information is encoded in a series of positive semidefinite weight and normalization matrices. The response could be that of a composite material, resistor network, spring network, or as shown in **Chapter 3**, that of a multicomponent body with inclusions having a size comparable to that of the body. More generally, it is applicable to any problem that can be phrased, in the abstract theory of composites, as a Y-problem or Z-problem, involving subspace collections with orthogonal subspaces: see **Chapter 2**. Thus it is applicable to the many equations discussed in **Chapter 1**, where the key identity is used to establish orthogonality of subspaces see Section 1.19 of **Chapter 1**, and **Chapter 3**. In particular, as follows from Section 3.12 in **Chapter 3**, it is applicable to the electromagnetism in multiphase bodies, and thus provides an alternative to bounds based on variational principles or on the analyticity, established in **Chapter 3** and more rigorously in **Chapter 4**, of the Dirichlet-to-Neumann map as functions of the frequency and component moduli. As shown in **Chapter 4**, the bounds can be used in an inverse way to give information on the interior structure inside the body.

For composite materials the n-th order series expansion terms, in the expansion of the effective moduli and fields in powers of the contrast between the materials, can in theory be derived from information derived from n-point correlation functions. These incorporate, for example in the case of two-component composites, the information contained in the probability that a polyhedron lands with all vertices in phase 1 when dropped randomly in the composite. The series expansions were first derived by Brown, Jr. (1955) for two phase conducting composites and were subsequently extended by others (Herring 1960; Prager 1960; Beran and Molyneux 1963; Beran 1968; Beran and McCoy 1970; Fokin and Shermergor 1969; Dederichs and Zeller 1973; Hori 1973; Zeller and Dederichs 1973; Gubernatis and Krumhansl 1975; Kröner 1977; Willis 1981a; Milton and Phan-Thien 1982; Phan-Thien and Milton 1982; Sen and Torquato 1989; Torquato 1997; Tartar 1989, 1990; and Bruno 1991b). See also the book of Torquato (2002), **Chapters 14** and **15** in (Milton 2002), and **Chapter 9** in this book, where series expansions are derived for coupled field problems. In practice these multipoint correlation functions are difficult to obtain: only three-point correlations can be directly obtained from cross-sectional photographs of three-dimensional materials, since in general four points have one point lying outside the plane containing the other 3 points. A far more practical way of obtaining the series expansion information, when the geometry is known is to use Fast Fourier Transform methods (Moulinec and Suquet 1994, 1998; Eyre and Milton (1999); Willot, Abdallah, and Pellegrini (2014); and **Chapter 8**: when the composite is not periodic one can take a cubic representative volume element of it, and periodically extend it. For the response of bodies, rather than composites, as discussed in **Chapter 3**, the computation of the series expansion is more difficult, as computation of the action of the nonlocal operators $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ requires one to know the Green's function associated with the body, with appropriate boundary conditions, and is not simply calculated through Fourier transforms.

For two-component composites the bounds calculated from the field expansion recursion method can alternatively be obtained from variational principles (Milton and McPhedran 1982), or via the analytic method (Bergman 1978; Milton 1981c; Bergman 1993). The rapid convergence of the bounds to the actual effective conductivity as successively more series expansion coefficients are incorporated in the bounds has been shown in numerical studies (McPhedran and Milton 1981; Milton and McPhedran 1982). In a wider context these bounds mostly correspond to bounds on Stieltjes functions, and are related to Padé approximants (Gragg

1968; Baker, Jr. 1969; Field 1976; Baker, Jr. and Graves-Morris 1981). The real power of the method lies in its application to multiphase composites, where it allows one to obtain bounds on complex effective moduli that have not been obtained by any other method (see Milton 1987b and section 29.6 of Milton 2002). In fact many of the bounds are sharp in the context of the abstract theory of composites: there is a subspace collection with an associated response function that saturates the bounds. For real composite materials the differential constraints on the fields typically impose further restrictions on the subspace collections which make many bounds nonoptimal in this context. For example, for two-dimensional isotropic composites of two isotropic conducting phases, the differential constraints imply that the normalization matrices are simply identity matrices (and using this fact the bounds can be improved) while for other problems (even three dimensional conductivity, with two isotropic phases) the complete set of constraints on the normalization matrices and weight matrices is unknown.

The differential constraints are incorporated in the translation method for obtaining bounds (Tartar 1979b, particularly theorem 8; Murat and Tartar 1985; Tartar 1985; Lurie and Cherkaev 1982, 1984 - see also the books of the books of Allaire 2002; Cherkaev 2000; Milton 2002; Tartar 2009; and Torquato 2002 and references therein), which however usually has the disadvantage that there are so many "translations" to choose from that it is hard to know which ones will produce the best bounds. The field equation recursion method has the advantage of being applicable to any problem that can be formulated in the abstract theory of composites, and the advantage that it produces bounds in a systematic way.

We have in mind effective tensors $\mathbf{L}_* = \mathbf{L}_*(\mathbf{L}_1, \ldots, \mathbf{L}_n)$ with tensors $\mathbf{L}_j$ as variables, but for simplicity we will just consider in this chapter the case $n = 2$ in which the 2-variables are just scalars. Then in the abstract setting an effective tensor is just the Z-operator associated to a Z-problem on a $Z(n)$ subspace collection (see **Chapter 29** in Milton 2002 and **Chapters 2** and **7** of this book) and provides a solution to the Z-problem. The field recursion method is a way to systematically develop, via fractional linear transformations at each level of the recursion, the continued fraction expansions of the effective tensor $\mathbf{L}_*$, i.e., the initial Z-operator in the recursion. This method is powerful for developing sharp bounds on the effective tensor analogous to the sharp bounds of Stieltjes functions via its continued fraction representation using Padé approximants (Gragg 1968; Baker, Jr. 1969; Field 1976; Baker, Jr. and Graves-Morris 1981). We caution, however, that the approach using fractional linear transformations is similar in many ways to the Schur (1917a,1917b) algorithm (see also section 1.1 in Agler and McCarthy 2002) for successively reducing in the number of interpolation points in the Nevanlinna–Pick interpolation problem. This algorithm has the disadvantage that computational or experimental errors can lead to problems even for a modest number of iterations (Section 1.1 in Agler and McCarthy 2002; Foiaş, Jolly, and Li 2002).

The rest of this chapter is organized as follows. In section 10.2 we introduce, in the Hilbert space setting for the abstract theory of composites, the effective tensor $\mathbf{L}_*(L_1, L_2)$ and its properties as a function of the composite phases L_1 and L_2. This tensor is defined as the Z-operator associated to an orthogonal $Z(2)$ subspace collection in a Hilbert space $\mathcal{H}$. Next, in section 10.3.1 we generate the $\mathbf{Y}_*(L_1, L_2)$ tensor on the associated orthogonal $Y(2)$ subspace collection and develop its functional properties as well as prove the important linear fractional transformations that connect $\mathbf{L}_*$ and $\mathbf{Y}_*$ together. We then introduce a new effective tensor ${\mathbf{L}_*}^{(1)}(L_1, L_2)$ on a new orthogonal $Z(2)$ subspace collection in a Hilbert space $\mathcal{H}^{(1)}$ which is a subspace of the initial Hilbert space $\mathcal{H}$. The base case of the recursion is then concluded by showing that these two tensors $\mathbf{Y}_*$ and ${\mathbf{L}_*}^{(1)}$ are congruent (i.e., there exists an invertible operator $\mathbf{K}$ such that $\mathbf{Y}_* = \mathbf{K}{\mathbf{L}_*}^{(1)}\mathbf{K}^\dagger$). Finally, in section 10.3.2 we prove via induction (under additional assumptions) that we can repeat this process. In other words, we can define an effective tensor ${\mathbf{L}_*}^{(i)}(L_1, L_2)$ on some $Z(2)$ subspace collection in a Hilbert space $\mathcal{H}^{(i)}$ which is a subspace of $\mathcal{H}^{(i-1)}$. This method provides a continued fraction representation of the effective tensor $\mathbf{L}_*$ in terms of the ${\mathbf{L}_*}^{(i)}$ and ${\mathbf{Y}_*}^{(i)}$. Thus, if one can provide bounds at each level of this

representation then one can obtain very tight bounds on $\mathbf{L}_*$, which is one of the most important applications of this field recursion method (Milton 1987b; see also **Chapter 9** of this book).

A major achievement in this chapter is the use of the theory of Fredholm operators which allows us to significantly reduce, at each level of the induction, the necessary assumptions required to define the operators ${\mathbf{L}_*}^{(i)}$ and ${\mathbf{Y}_*}^{(i)}$. And although our approach here is abstract, which simplifies mathematically many aspects of the field equation recursion method, we want to keep in mind that its development is motivated by its applications in the theory of composites. As such, we use throughout this chapter the quintessential example of the effective conductivity tensor for composite materials in quasistatic electromagnetism [see, for instance, Milton and Golden 1985; Milton 1991; Clark and Milton 1994; Clark and Milton 1995; Clark 1997; Milton 2002] in order to illustrate the theory.

This chapter assumes the reader is familiar with **Chapters 1** and **2**. **Chapter 9** is closely related to the subject matter of the chapter, although need not be read beforehand. Like **Chapter 4** it is written in a rigorous mathematical style and care has been taken to explain most technical definitions to ensure it is accessible to non-mathematicians.

Before we proceed, let us introduce some notation and definitions:

- We denote by $\mathcal{B}(\mathcal{H})$ the Banach space of all bounded linear operators on a Hilbert space $\mathcal{H}$ and endow this space with the operator norm.
- If $\mathbf{T} \in \mathcal{B}(\mathcal{H})$, we denote respectively by $\operatorname{Ran}\mathbf{T}$, $\operatorname{Ker}\mathbf{T}$ and $\mathbf{T}^\dagger$ its range, kernel (i.e., nullspace) and adjoint operator.
- If $\mathbf{T} \in \mathcal{B}(\mathcal{H})$, $\operatorname{Re}\mathbf{T}$ and $\operatorname{Im}\mathbf{T}$ stand for the bounded operators

$$\operatorname{Re}\mathbf{T} = \frac{\mathbf{T}+\mathbf{T}^\dagger}{2} \text{ and } \operatorname{Im}\mathbf{T} = \frac{\mathbf{T}-\mathbf{T}^\dagger}{2i}.$$

- An operator $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ is said to be positive definite (i.e., $\mathbf{T} > 0$) if

$$\forall \mathbf{u} \in \mathcal{H} \setminus \{\mathbf{0}\},\ (\mathbf{u}, \mathbf{T}\mathbf{u}) > 0, \tag{10.1}$$

  and negative definite if $-\mathbf{T}$ is positive definite. If the inequality "$>$" in (10.1) is instead replaced by "$\geq$", then one says that $\mathbf{T}$ is positive semidefinite.
- An operator $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ is said to be coercive (or uniformly positive) if

$$\exists \alpha > 0 \mid\ \forall \mathbf{u} \in \mathcal{H},\ (\mathbf{u}, \mathbf{T}\mathbf{u}) \geq \alpha\,(\mathbf{u}, \mathbf{u}),$$

  and in that case we write $\mathbf{T} \geq \alpha\mathbb{I}$. An operator is said uniformly negative if $-\mathbf{T}$ is coercive.
- We denote by $\mathcal{E}^\perp$ the orthogonal complement of a subspace $\mathcal{E}$ of $\mathcal{H}$,
- $\operatorname{cl} S$ stands for the topological closure of a set S in a Hilbert space $\mathcal{H}$ in the metric topology generated by the inner product norm.

Definition 28. *(multivariate analyticity) Let U be an open set of $\mathbb{C}^n$ and F a complex Banach space. A function $f : U \to \mathbf{F}$ is said to be analytic if it is differentiable on $\mathcal{U}$. If we denote by $\mathbf{Z} = (z_1, \cdots, z_n)$ the points of U, then by Hartogs' theorem (see Theorem 36.8 in chapter VIII, section 36, p. 271 of Mujica 1986), f is analytic in U if and only if it is a differentiable function in each variable z_i separately.*

Definition 29. *Let $\mathbf{T} \in \mathcal{B}(\mathcal{H})$, we say that the operator $\mathbf{T}$ is Fredholm if $\operatorname{Ran}\mathbf{T}$ is a closed subspace of $\mathcal{H}$ and if $\operatorname{Ker}\mathbf{T}$ and $\operatorname{Ker}\mathbf{T}^\dagger$ are finite dimensional spaces. In that case, the index $\operatorname{ind}\mathbf{T}$ of $\mathbf{T}$ is defined by*

$$\operatorname{ind}\mathbf{T} = \dim\operatorname{Ker}\mathbf{T} - \dim\operatorname{Ker}\mathbf{T}^\dagger.$$

Remark 30. *Using the well-known identity $\operatorname{cl}\operatorname{Ran}\mathbf{T} = (\operatorname{Ker}\mathbf{T}^\dagger)^\perp$, it follows that a Fredholm operator of index 0 is invertible if and only if $\operatorname{Ker}\mathbf{T} = 0$.*

We will use the following propositions:

Proposition 31. *An operator $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ is a Fredholm operator of index 0 if and only if can be written as $\mathbf{T} = \mathbf{A} + \mathbf{K}$ with $\mathbf{A}$ an invertible operator and $\mathbf{K}$ a finite-rank operator, i.e., $\dim\operatorname{Ran}\mathbf{K} < \infty$.*

The proof of this result can be found in chapter XV, section 15.2, p. 350, Corollary 2.4 in Gohberg, Goldberg, and Kaashoek 2003.

Proposition 32. *Suppose $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ is an orthogonal decomposition of a Hilbert space $\mathcal{H}$ with $\dim\mathcal{H}_1 < \infty$. Let $\mathbf{T} \in \mathcal{B}(\mathcal{H})$ be written, with respect to this decomposition, as a 2×2 block operator matrix form*

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}. \tag{10.2}$$

If $\mathbf{T}$ is a Fredholm operator of index n then $\mathbf{T}_{22} \in \mathcal{B}(\mathcal{H}_2)$ is a Fredholm operator of index n. In particular, if $\mathbf{T}$ is invertible then $\mathbf{T}_{22}$ has index 0.

Proof. Suppose that $\mathbf{T}$ is a Fredholm operator of index n then since $\dim\mathcal{H}_1 < \infty$, the operator

$$\begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & 0 \end{bmatrix} \tag{10.3}$$

is a finite-rank operator in $\mathcal{B}(\mathcal{H})$ which implies by the Proposition 31 that

$$\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{T}_{22} \end{bmatrix} = \mathbf{T} - \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & 0 \end{bmatrix} \tag{10.4}$$

is a Fredholm operator in $\mathcal{B}(\mathcal{H})$ of index n. Thus,

$$\mathcal{H}_1 \oplus \operatorname{Ker}\mathbf{T}_{22}, \quad \mathcal{H}_1 \oplus \operatorname{Ker}\mathbf{T}_{22}^\dagger, \tag{10.5}$$

are finite-dimensional, $\operatorname{Ran}\mathbf{T}_{22}$ is closed, and

$$n = \dim(\mathcal{H}_1 \oplus \operatorname{Ker}\mathbf{T}_{22}) - \dim(\mathcal{H}_1 \oplus \operatorname{Ker}\mathbf{T}_{22}^\dagger) = \dim(\operatorname{Ker}\mathbf{T}_{22}) - \dim(\operatorname{Ker}\mathbf{T}_{22}^\dagger).$$

Therefore, $\mathbf{T}_{22}$ is a Fredholm operator of index n. In particular, if $\mathbf{T}$ is invertible then it is a Fredholm operator of index 0 and so is $\mathbf{T}_{22}$. This completes the proof. □

10.2 Formulation of the problem for two-component composites

The setting from **Chapter 2** is a Hilbert space $\mathcal{H}$ which has an inner product $(\mathbf{P}_1, \mathbf{P}_2)$ defined for all $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{H}$ having the usual properties that

$$(\mathbf{P}_1, \mathbf{P}_2) = \overline{(\mathbf{P}_2, \mathbf{P}_1)}, \quad (\mathbf{P}_1, \mathbf{P}_1) > 0 \ \text{ for all } \mathbf{P}_1 \neq 0, \tag{10.6}$$

with the convention, for this chapter and **Chapter 4** only, that it is linear in the second component and antilinear in the first (which is opposite to that used in the rest of the book). This Hilbert space is assumed to have the decomposition

$$\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J} = \mathcal{P}_1 \oplus \mathcal{P}_2, \tag{10.7}$$

where the subspaces $\mathcal{U}, \mathcal{E}$, and $\mathcal{J}$ are mutually orthogonal with respect to this inner product as are the subspaces $\mathcal{P}_1, \mathcal{P}_2$. Moreover, we assume $\mathcal{U}$ is a finite-dimensional subspace and that $\mathcal{P}_1$ and $\mathcal{P}_2$ are nonzero subspaces. Under these assumptions, the decomposition in (10.7) is called an orthogonal $Z(2)$ subspace collection (see **Chapter 7**).

We denote by $\mathbf{\Gamma}_0$, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ the orthogonal projections on $\mathcal{U}$, $\mathcal{E}$, and $\mathcal{J}$, respectively, and denote by $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ the orthogonal projections onto $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. We want to emphasize that we don't assume that the operators $\mathbf{\Lambda}_a$ and $\mathbf{\Gamma}_i$ commute. In this abstract setting, the Hilbert space $\mathcal{H}$ can be infinite-dimensional and as such it may not be clear that the subspaces introduced above must all be closed subspaces so that their corresponding orthogonal projections on each subspace exist. These facts though follow immediately from the following classical proposition which we prove here so that this chapter is self-contained.

Proposition 33. *Let $\mathcal{H}$ be a Hilbert space with inner product (10.6). If $\mathcal{M}_1$ and $\mathcal{M}_2$ are subspaces of $\mathcal{H}$ which are mutually orthogonal and $\mathcal{H} = \mathcal{M}_1 \oplus \mathcal{M}_2$ then $\mathcal{M}_1$ and $\mathcal{M}_2$ are closed sets in the norm topology on $\mathcal{H}$ and, for $i, j = 1, 2$ with $i \neq j$, the projection $\mathbf{Q}_i : \mathcal{H} \to \mathcal{H}$ onto $\mathcal{M}_i$ along $\mathcal{M}_j$, which is uniquely defined by $\mathbf{Q}_i\mathbf{P} = \mathbf{P}_i$ if $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ for some vectors $\mathbf{P}_1 \in \mathcal{M}_1$ and $\mathbf{P}_2 \in \mathcal{M}_2$, is an orthogonal projection.*

Proof. Suppose $\mathcal{M}_1$ and $\mathcal{M}_2$ are subspaces of $\mathcal{H}$ which are mutually orthogonal and $\mathcal{H} = \mathcal{M}_1 \oplus \mathcal{M}_2$. Then every $\mathbf{P} \in \mathcal{H}$ can be written as $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ for some unique $\mathbf{P}_1 \in \mathcal{M}_1$, $\mathbf{P}_2 \in \mathcal{M}_2$ and there exists unique linear operators $\mathbf{Q}_i : \mathcal{H} \to \mathcal{H}$ defined by $\mathbf{Q}_i\mathbf{P} = \mathbf{P}_i$, $i = 1, 2$. In particular, this implies that they are projections, i.e., $\mathbf{Q}_i^2 = \mathbf{Q}_i$, $i = 1, 2$, with the property that they sum to the identity operator, i.e., $\mathbf{Q}_1 + \mathbf{Q}_2 = I_{\mathcal{H}}$. It now follows from the fact that the spaces $\mathcal{M}_1, \mathcal{M}_2$ are mutually orthogonal that for any $\mathbf{P} \in \mathcal{H}$ with $\mathbf{P} = \mathbf{P}_1 + \mathbf{P}_2$ for some $\mathbf{P}_1 \in \mathcal{M}_1$, $\mathbf{P}_2 \in \mathcal{M}_2$ we have $||\mathbf{Q}_i\mathbf{P}|| = ||\mathbf{P}_i|| \leq ||\mathbf{P}_1|| + ||\mathbf{P}_2|| = ||\mathbf{P}||$ and $||\mathbf{Q}_i\mathbf{P}_i|| = ||\mathbf{P}_i||$ for $i = 1, 2$. Thus, since $\mathcal{H} = \mathcal{M}_1 \oplus \mathcal{M}_1$, this implies that in the operator norm $||\mathbf{Q}_i|| = 1$ or $\mathbf{Q}_i = 0$ and hence $\mathbf{Q}_i \in \mathcal{B}(\mathcal{H})$ for $i = 1, 2$. Thus, since $\mathcal{B}(\mathcal{H})$ is the space of all continuous linear operator from $\mathcal{H}$ into $\mathcal{H}$ in the norm topology on $\mathcal{H}$ then this implies $\mathbf{Q}_i^{-1}(\{\mathbf{0}\}) = \mathcal{M}_i$ is a closed set. Finally, since the spaces $\mathcal{M}_1$ and $\mathcal{M}_2$ are mutually orthogonal then for any $\mathbf{P}_i, \mathbf{R}_i \in \mathcal{M}_i$, $i = 1, 2$ we have $(\mathbf{Q}_i(\mathbf{P}_1 + \mathbf{P}_2), \mathbf{R}_1 + \mathbf{R}_2) = (\mathbf{P}_i, \mathbf{R}_i) = (\mathbf{P}_1 + \mathbf{P}_2, \mathbf{Q}_i(\mathbf{R}_1 + \mathbf{R}_2))$ which implies that $\mathbf{Q}_i^\dagger = \mathbf{Q}_i$, i.e., the projection operator $\mathbf{Q}_i$ is self-adjoint, that is, it is an orthogonal projection. This completes the proof. □

We now define the bounded linear operator-valued function on $\mathcal{H}$ by

$$\mathbf{L} = \mathbf{L}(L_1, L_2) = L_1\mathbf{\Lambda}_1 + L_2\mathbf{\Lambda}_2. \tag{10.8}$$

It is a function of the complex variables L_1, L_2, and $\mathbf{L} : \mathbb{C}^2 \to \mathcal{B}(\mathcal{H})$ is an analytic $\mathcal{B}(\mathcal{H})$-valued function in the sense of the Definition 28.

Remark 34. *Defining this functional framework for the conductivity equation of a two-component periodic composite defined by a unit cell $\mathcal{D}$, which is an open simply connected set with a Lipschitz continuous boundary, and composed of isotropic materials characterized by their complex conductivity σ_1 and σ_2 leads to:*

- $\mathcal{H} = \mathbf{L}^2(\mathcal{D}) = [L^2(\mathcal{D})]^d$ *the space of square-integrable, d-dimensional, vector-field-valued functions on the unit cells $\mathcal{D}$ endowed with the inner product*

$$(\mathbf{E}, \mathbf{E}^{'}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \mathbf{E}^{\top} \overline{\mathbf{E}^{'}} \, \mathrm{d}\mathbf{x}, \ \forall \mathbf{E}, \mathbf{E}^{'} \in \mathcal{M}_{d,1}(\mathbb{C}),$$

  where $|\mathcal{D}|$ denotes the volume of $\mathcal{D}$,

- $\mathcal{U}$ *is the d-dimensional space defined by*

$$\mathcal{U} = \{\mathbf{U} \in \mathcal{H} \mid \mathbf{U} \equiv C, \textit{ for some } C \in \mathcal{M}_{d,1}(\mathbb{C})\},$$

- $\mathcal{E} = \left\{\mathbf{E} \in \mathcal{H} \mid \operatorname{curl} \mathbf{E} = 0 \textit{ in } \mathcal{D} \textit{ and } \int_{\mathcal{D}} \mathbf{E} \, \mathrm{d}\mathbf{x} = 0 \textit{ (zero average condition)}\right\}$,

- $\mathcal{J} = \left\{\mathbf{J} \in \mathcal{H} \mid \operatorname{div} \mathbf{J} = 0 \textit{ in } \mathcal{D} \textit{ and } \int_{\mathcal{D}} \mathbf{J} \, \mathrm{d}\mathbf{x} = 0\right\}$.

For a proof of the orthogonal decomposition $\mathbf{L}^2(\mathcal{D}) = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}$, we refer to ***Chapter 12*** *of Milton (2002). Although the definition (10.8) of the operator $\mathbf{L}$ may seem abstract, one sees it is natural in the theory of composites (see Milton 2002) where $\mathbf{L}$ is the conductivity tensor*

$$\mathbf{L}(\sigma_1, \sigma_2) = \sigma_1 \chi_1 \mathbb{I} + \sigma_2 \chi_2 \mathbb{I}. \tag{10.9}$$

In this context, χ_i is the indicator function of the domain occupied by phase i and the scalars L_1 and L_2 in (10.8) are respectively the complex conductivities σ_1 and σ_2. Thus, $\mathbf{\Lambda}_1 = \chi_1 \mathbb{I}$ and $\mathbf{\Lambda}_2 = \chi_2 \mathbb{I}$ represent respectively the projections onto the spaces $\mathcal{P}_1$ and $\mathcal{P}_2$ of $\mathbf{L}^2(\mathcal{D})$ fields which are nonzero only inside phase 1 or 2. Finally, it should be emphasized that the restriction of $\mathbf{L}$ on the subspace $\mathcal{U} \oplus \mathcal{E}$ represents the constitutive law of the composite which links an electric field in $\mathcal{U} \oplus \mathcal{E}$ to a current density in $\mathcal{U} \oplus \mathcal{J}$.

Following **Chapter 7**, associated to this $Z(2)$ subspace collection (10.7) is a linear operator-valued function $\mathbf{L}_*(L_1, L_2)$ (i.e., the associated Z-operator) acting on the subspace $\mathcal{U}$. To obtain this function we begin by solving the following problem: for a given $\mathbf{e} \in \mathcal{U}$, find a unique triplet of vectors $\mathbf{j} \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$, and $\mathbf{J} \in \mathcal{J}$ that satisfy

$$\mathbf{j} + \mathbf{J} = \mathbf{L}(\mathbf{e} + \mathbf{E}), \tag{10.10}$$

also known as the Z-problem (see **Chapter 7**). The associated operator $\mathbf{L}_*$, by definition, governs the linear relation

$$\mathbf{j} = \mathbf{L}_* \mathbf{e}. \tag{10.11}$$

To obtain a representation of $\mathbf{L}_*$ using $\mathbf{L}$, we follow the approach of **Chapter 12** of Milton (2002) which consists of expressing $\mathbf{L}$ as a 3×3 block operator matrix

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{00} & \mathbf{L}_{01} & \mathbf{L}_{02} \\ \mathbf{L}_{10} & \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{20} & \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \tag{10.12}$$

with respect to the decomposition $\mathcal{H} = \mathcal{U} \oplus \mathcal{E} \oplus \mathcal{J}$. Solving the Z-problem (10.10) for a given $\mathbf{e} \in \mathcal{U}$ is then equivalent to proving that the system

$$\begin{cases} \mathbf{L}_{00}\mathbf{e} + \mathbf{L}_{01}\mathbf{E} = \mathbf{j}, \\ \mathbf{L}_{10}\mathbf{e} + \mathbf{L}_{11}\mathbf{E} = 0, \\ \mathbf{L}_{20}\mathbf{e} + \mathbf{L}_{21}\mathbf{E} = \mathbf{J}, \end{cases} \tag{10.13}$$

admits a unique solution $(\mathbf{j}, \mathbf{E}, \mathbf{J}) \in \mathcal{U} \times \mathcal{E} \times \mathcal{J}$. Under the assumption that the block operator $\mathbf{L}_{11}$ (which represent the restriction of the operator $\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_1$ on the subspace $\mathcal{E}$) is invertible, the solution of the system (10.13) is unique and given by

$$\begin{cases} \mathbf{j} = (\mathbf{L}_{00} - \mathbf{L}_{01}\mathbf{L}_{11}^{-1}\mathbf{L}_{10})\,\mathbf{e}, \\ \mathbf{E} = -\mathbf{L}_{11}^{-1}\,\mathbf{L}_{10}\,\mathbf{e}, \\ \mathbf{J} = \mathbf{L}_{20}\,\mathbf{e} + \mathbf{L}_{21}\,\mathbf{E}, \end{cases} \tag{10.14}$$

which defines the Z-operator $\mathbf{L}_*$ via a Schur complement by:

$$\mathbf{L}_* = (\mathbf{L}_{00} - \mathbf{L}_{01}\mathbf{L}_{11}^{-1}\mathbf{L}_{10}) = \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_1(\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_1)^{-1}\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_0, \tag{10.15}$$

where the second equality in the last relation has to be restricted to the subspace $\mathcal{U}$ of $\mathcal{H}$.

Another representation formula of the operator $\mathbf{L}_*$:

$$\mathbf{L}_* = \mathbf{\Gamma}_0[(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)]^{-1}\mathbf{\Gamma}_0, \tag{10.16}$$

is proved in Section 12.8 of Milton (2002), equation (12.59) (see also Section 7.4, equation (7.60) of this book) under the assumption that the inverse of $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ exists on the subspace $\mathcal{U} \oplus \mathcal{J}$ (which requires in particular that $\mathbf{L}$ is invertible on $\mathcal{H}$).

Remark 35. *For instance, if* $\operatorname{Im}\mathbf{L} \geq \alpha\mathbb{I}$ *or* $\operatorname{Re}\mathbf{L} \geq \alpha\mathbb{I}$ *for some* $\alpha > 0$ *then the inverse of* $\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_1$ *on* $\mathcal{E}$ *and the inverse of* $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ *on* $\mathcal{U} \oplus \mathcal{J}$ *exists and thus the* $Z-$ *problem (10.10) is well-defined and both representation formulas (10.15) and (10.16) of* $\mathbf{L}_*$ *hold.*

Our next proposition shows that these two representations of $\mathbf{L}_*$ are both well-defined together or neither is when $\mathbf{L}^{-1}$ exists.

Proposition 36. *If* $\mathbf{L}(L_1, L_2)$ *is invertible then the operator* $\mathbf{A} := \mathbf{\Gamma}_1\mathbf{L}(L_1, L_2)\mathbf{\Gamma}_1 : \mathcal{E} \to \mathcal{E}$ *is a Fredholm operator of index* n *if and only if the operator* $\mathbf{B} := (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}(L_1, L_2)^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2) : \mathcal{U} \oplus \mathcal{J} \to \mathcal{U} \oplus \mathcal{J}$ *is a Fredholm operator of index* n*. Furthermore, the operator* $\mathbf{A}$ *is invertible if and only if* $\mathbf{B}$ *is invertible, and in that case we have*

$$\mathbf{A}^{-1} = \mathbf{\Gamma}_1\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Gamma}_1 - \mathbf{\Gamma}_1\mathbf{L}(L_1, L_2)^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{B}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Gamma}_1, \tag{10.17}$$

$$\mathbf{B}^{-1} = (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}(L_1, L_2)(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2) - (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}(L_1, L_2)\mathbf{\Gamma}_1\mathbf{A}^{-1}\mathbf{\Gamma}_1\mathbf{L}(L_1, L_2)(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2). \tag{10.18}$$

Proof. Suppose $\mathbf{L}(L_1, L_2)$ is invertible. Then with respect to the orthogonal decomposition of the Hilbert space $\mathcal{H} = \mathcal{E} \oplus (\mathcal{U} \oplus \mathcal{J})$, with the corresponding orthogonal projections $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2$, the operator $\mathbf{L}(L_1, L_2)$ and its inverse $\mathbf{L}(L_1, L_2)^{-1}$ can be written as the 2×2 block operator matrices

$$\mathbf{L}(L_1, L_2) = \begin{bmatrix} \mathbf{A} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{L}(L_1, L_2)^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B} \end{bmatrix}. \tag{10.19}$$

This means that $\mathbf{A}$ and $\mathbf{B}$ are matricially coupled operators and therefore the proof of this proposition now follows immediately from Corollary 4.3, pp. 46–47, section III.4 of Gohberg, Goldberg, and Kaashoek 1990. □

Remark 37. *The invertibility condition of* $\mathbf{L}(L_1, L_2)$ *is straightforward to check. Indeed, we have*

$$\mathbf{L}(L_1, L_2) \text{ is invertible if and only if } L_1 \neq 0 \text{ and } L_2 \neq 0.$$

Moreover, the operator $\mathbf{L}(L_1, L_2)^{-1}$ *is given by the following formula:*

$$\mathbf{L}(L_1, L_2)^{-1} = \frac{1}{L_1}\Lambda_1 + \frac{1}{L_2}\Lambda_2.$$

Nevertheless, we emphasize that the invertibility of $\mathbf{L}$ *on* $\mathcal{H}$ *does not imply in general the invertibility of* $\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_1$ *on the subspace* $\mathcal{E}$.

It follows from the formulas (10.15) or (10.16) that the analytic $\mathcal{B}(\mathcal{U})$-valued function $\mathbf{L}_*(L_1, L_2)$ (see Definition 28) satisfies the following homogeneity, normalization, and Herglotz properties:

The homogeneity property:

$$\mathbf{L}_*(L_1, L_2) = \frac{1}{c}\mathbf{L}_*(cL_1, cL_2), \tag{10.20}$$

for all choices of constants $c \neq 0$.

The normalization property:

$$\mathbf{L}_*(1, 1) = \mathbb{I}_{\mathcal{H}}. \tag{10.21}$$

The Herglotz property:

$$\operatorname{Im}\mathbf{L}_*(L_1, L_2) > 0, \quad \text{when } \operatorname{Im}(L_1) > 0 \text{ and } \operatorname{Im}(L_2) > 0. \tag{10.22}$$

The Herglotz property of $\mathbf{L}_*$ may not be obvious and so we provide two proofs. The first one depends on the representation formula (10.16) of $\mathbf{L}_*$. The second one is independent of the representation formulas (10.15) or (10.16) (which may not exist) and just assumes that $\mathbf{L}_*$ is well-defined by the Z-problem (10.10). Nevertheless, as it was mentioned in Remark 35, this coercivity assumption has the convenience to be a sufficient condition of the well-posedness of the Z-problem and to justify the existence of the both representation formulas (10.15) and (10.16) of $\mathbf{L}_*$.

Proof. By the definition (10.8) of the operator $\mathbf{L}$, it is straightforward that the conditions $\operatorname{Im}(L_1) > 0$ and $\operatorname{Im}(L_2) > 0$ imply that $\operatorname{Im}(\mathbf{L}) \geq \beta\mathbb{I}$ with $\beta = \min(\operatorname{Im}(L_1), \operatorname{Im}(L_2)) > 0$. Thus, by the Lax–Milgram theorem one deduces immediately that $\mathbf{L}$ is invertible on $\mathcal{H}$ and by virtue of the identities:

$$\operatorname{Im}(\mathbf{T}^{-1}) = -(\mathbf{T}^{-1})^\dagger \operatorname{Im}(\mathbf{T})\mathbf{T}^{-1}, \quad \operatorname{Im}(\mathbf{M}^\dagger\mathbf{A}\mathbf{M}) = \mathbf{M}^\dagger \operatorname{Im}(\mathbf{A})\mathbf{M}, \tag{10.23}$$

which hold for every invertible operator $\mathbf{T}$ of $\mathcal{B}(H)$ and every $\mathbf{A}, \mathbf{M} \in \mathcal{B}(H)$, we get immediately that $\operatorname{Im}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ is a uniformly negative operator on the subspace $\mathcal{U} \oplus \mathcal{J}$ and thus (by the Lax–Milgram theorem) $(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)$ is invertible on $\mathcal{U} \oplus \mathcal{J}$ and hence the operator $\mathbf{L}_*$ is well-defined by (10.16) if $\operatorname{Im}(L_1) > 0$ and $\operatorname{Im}(L_2) > 0$. Moreover, by using again (10.23), one concludes that $\operatorname{Im}\mathbf{L}_* = \operatorname{Im}\mathbf{\Gamma}_0[(\mathbf{\Gamma}_0+\mathbf{\Gamma}_2)\mathbf{L}^{-1}(\mathbf{\Gamma}_0+\mathbf{\Gamma}_2)]^{-1}\mathbf{\Gamma}_0$ is coercive on $\mathcal{U}$ (in particular this implies that $\mathbf{L}_*$ is invertible).

The second version of this proof is based on the definition (10.10) and (10.11) of the Z-problem, and follows the treatment given at the end of Section 2.6 (see also the end of Section 12.10 in Milton 2002). Assume that the $\mathbf{L}_*(L_1, L_2)$ is well-defined by the Z-problem. Let $\mathbf{e} \in \mathcal{U}$ then there exists a $\mathbf{E} \in \mathcal{E}$, $\mathbf{J} \in \mathcal{J}$, and $\mathbf{j} \in \mathcal{U}$ such that $\mathbf{L}_*(L_1, L_2)(\mathbf{e} + \mathbf{E}) = \mathbf{j} + \mathbf{J}$. Hence we obtain

$$\begin{aligned}((\mathbf{e}+\mathbf{E}), \operatorname{Im}\mathbf{L}(\mathbf{e}+\mathbf{E})) &= \operatorname{Im}(\mathbf{e}+\mathbf{E}, \mathbf{L}(\mathbf{e}+\mathbf{E})) \\ &= \operatorname{Im}(\mathbf{e}+\mathbf{E}, \mathbf{j}+\mathbf{J}) \\ &= \operatorname{Im}(\mathbf{e}, \mathbf{L}_*\mathbf{e}) \\ &= (\mathbf{e}, \operatorname{Im}\mathbf{L}_*\mathbf{e}). \end{aligned} \tag{10.24}$$

Thus, the positive definiteness of $\operatorname{Im}\mathbf{L}(L_1, L_2)$ under this hypothesis $\operatorname{Im}(L_1) > 0$ and $\operatorname{Im}(L_2) > 0$ implies immediately the coercivity of $\operatorname{Im}[\mathbf{L}_*(L_1, L_2)]$ as $\mathcal{U}$ is finite-dimensional, in particular, it also implies that $\mathbf{L}_*(L_1, L_2)$ is invertible. □

10.3 Field equation recursion method for two-component composites

10.3.1 The base case

In the field equation recursion method, the first step is to generate, from the orthogonal $Z(2)$ subspace collection (10.7), an orthogonal $Y(2)$ subspace collection (see **Chapter 9**) namely, the Hilbert space

$$\mathcal{K} = \mathcal{E} \oplus \mathcal{J} = \mathcal{V} \oplus \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}, \tag{10.25}$$

where $\mathcal{K}$ is the orthogonal complement of $\mathcal{U}$ in $\mathcal{H}$, i.e.,

$$\mathcal{K} = \mathcal{H} \ominus \mathcal{U}. \tag{10.26}$$

Then one generates another orthogonal $Z(2)$ subspace collection, namely, the Hilbert space

$$\mathcal{H}^{(1)} = \mathcal{U}^{(1)} \oplus \mathcal{E}^{(1)} \oplus \mathcal{J}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}, \tag{10.27}$$

with a suitable choice of the spaces, satisfying the inclusion relations

$$\mathcal{E}^{(1)} \subseteq \mathcal{E}, \;\; \mathcal{J}^{(1)} \subseteq \mathcal{J}, \;\; \mathcal{P}_1^{(1)} \subseteq \mathcal{P}_1, \;\; \mathcal{P}_2^{(1)} \subseteq \mathcal{P}_2. \tag{10.28}$$

We have already defined $\mathcal{K} = \mathcal{E} \oplus \mathcal{J}$, so our goal is to define the spaces $\mathcal{P}_1^{(1)}$, $\mathcal{P}_2^{(1)}$, $\mathcal{V}$, $\mathcal{E}_1^{(1)}$, $\mathcal{E}_2^{(1)}$, $\mathcal{J}_1^{(1)}$, $\mathcal{J}_2^{(1)}$, $\mathcal{H}^{(1)}$, and $\mathcal{U}^{(1)}$. The spaces $\mathcal{P}_1^{(1)}$ and $\mathcal{P}_2^{(1)}$ are defined as

$$\mathcal{P}_1^{(1)} = \mathcal{P}_1 \cap \mathcal{K}, \;\; \mathcal{P}_2^{(1)} = \mathcal{P}_2 \cap \mathcal{K}. \tag{10.29}$$

$\mathcal{V}$ is the orthogonal complement of $\mathcal{H}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}$ in $\mathcal{K}$, i.e.,

$$\mathcal{V} = \mathcal{K} \ominus \mathcal{H}^{(1)}, \tag{10.30}$$

the spaces $\mathcal{E}^{(1)}$ and $\mathcal{J}^{(1)}$ are defined as

$$\mathcal{E}^{(1)} = \mathcal{E} \cap \mathcal{H}^{(1)}, \;\; \mathcal{J}^{(1)} = \mathcal{J} \cap \mathcal{H}^{(1)}, \tag{10.31}$$

and finally $\mathcal{U}^{(1)}$ is defined as the orthogonal complement of $\mathcal{E}^{(1)} \oplus \mathcal{J}^{(1)}$ in the space $\mathcal{H}^{(1)}$, i.e.,

$$\mathcal{U}^{(1)} = \mathcal{H}^{(1)} \ominus [\mathcal{E}^{(1)} \oplus \mathcal{J}^{(1)}]. \tag{10.32}$$

The following projections $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$, $\mathbf{\Lambda}_1^{(1)}$, and $\mathbf{\Lambda}_2^{(1)}$ play a key role and are respectively defined as the orthogonal projections onto $\mathcal{E}$, $\mathcal{J}$, $\mathcal{V}$, $\mathcal{H}^{(1)}$, $\mathcal{P}_1^{(1)}$, and $\mathcal{P}_2^{(1)}$.

Remark 38. *By the subspace inclusion (10.28) we have*

$$\mathbf{L} = L_1\mathbf{\Lambda}_1 + L_2\mathbf{\Lambda}_2 = L_1\mathbf{\Lambda}_1^{(1)} + L_2\mathbf{\Lambda}_2^{(1)} \text{ on } \mathcal{H}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)} \tag{10.33}$$

$$\mathbf{L} = L_2\mathbf{\Lambda}_2^{(1)} \text{ on } \mathcal{P}_2^{(1)}, \ \ \mathbf{L} = L_2\mathbf{\Lambda}_1^{(1)} \text{ on } \mathcal{P}_1^{(1)}. \tag{10.34}$$

Remark 39. *The interpretation of the functional spaces which appear in the two subspace collections $Y(2)$ and $Z(2)$ in (10.25) and (10.27) in the case of the conductivity equation for a two-component composite are:*

- $\mathcal{P}_i^{(1)} = \mathcal{P}_i \cap \mathcal{K}$ *for* $i = 1, 2$ *is the subspace of* $\mathbf{L}^2(\mathcal{D})$ *of all fields whose support is included in the phase* i *and have zero average in the unit cell* $\mathcal{D}$*,*
- $\mathcal{H}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}$ *is the subspace of* $\mathbf{L}^2(\mathcal{D})$ *of all fields which have zero average in each phase,*
- $\mathcal{V}$ *is the subspace of* $\mathbf{L}^2(\mathcal{D})$ *of all fields which are constant in each phase and have zero average in the unit cell* $\mathcal{D}$*,*
- $\mathcal{E}^{(1)} = \left\{\mathbf{E} \in \mathcal{H}^{(1)} \mid \operatorname{curl} \mathbf{E} = 0 \text{ in } \mathcal{D}\right\}$, $\mathcal{J}^{(1)} = \left\{\mathbf{J} \in \mathcal{H}^{(1)} \mid \operatorname{div} \mathbf{J} = 0 \text{ in } \mathcal{D}\right\}$.

The definition (10.32) of the space $\mathcal{U}^{(1)}$ is more complicated to interpret, we will see in the Corollary 46 that it can also be defined as

$$\mathcal{U}^{(1)} = (\mathbf{\Gamma}_1\mathcal{V} \oplus \mathbf{\Gamma}_2\mathcal{V}) \ominus \mathcal{V},$$

in other words, as the orthogonal complement of $\mathcal{V}$ in $\mathbf{\Gamma}_1\mathcal{V} \oplus \mathbf{\Gamma}_2\mathcal{V}$.

The Y-operator of the associated Y-problem in the $Y(2)$ subspace collection (10.25) will be the Y-tensor function $\mathbf{Y}_*(L_1, L_2)$ and the effective tensor function of the $Z(2)$-subspace collection (10.27) will be $\mathbf{L}_*{}^{(1)}(L_1, L_2)$, i.e., the Z-operator of the associated Z-problem in the $Z(2)$ subspace collection (10.27). The purpose of this subsection is to define these two operators and prove that they are congruent operators.

Following Section 7.2, associated to this $Y(2)$ subspace collection (10.25) is the Y-tensor function (i.e., the associated Y-operator) which is a linear operator-valued function $\mathbf{Y}_*(L_1, L_2)$ acting on the space $\mathcal{V}$. To obtain this function we begin by solving the following problem: for a given $\mathbf{E}_1 \in \mathcal{V}$, find a unique vector pair: $\mathbf{E} \in \mathcal{E}$, $\mathbf{J} \in \mathcal{J}$ that satisfy

$$\mathbf{J}_2 = \mathbf{L}\mathbf{E}_2, \ \ \mathbf{E}_2 = \mathbf{\Pi}_2\mathbf{E}, \ \ \mathbf{J}_2 = \mathbf{\Pi}_2\mathbf{J}, \ \ \mathbf{E}_1 = \mathbf{\Pi}_1\mathbf{E}, \tag{10.35}$$

also known as the Y-problem for the $Y(2)$ subspace collection (10.25) (see **Chapter 7**). The associated operator $\mathbf{Y}_*$, by definition, governs the linear relation

$$\mathbf{J}_1 = -\mathbf{Y}_*\mathbf{E}_1, \tag{10.36}$$

where $\mathbf{J}_1 = \mathbf{\Pi}_1\mathbf{J}$. The formula for the operator $\mathbf{Y}_*$, as given in equation (19.29) of Milton (2002) [see also Section 7.4, equation (7.68), of this book] is

$$\mathbf{Y}_* = \mathbf{\Pi}_1\mathbf{\Gamma}_2(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}\mathbf{\Gamma}_2\mathbf{\Pi}_1, \tag{10.37}$$

where the inverse $(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}$, if it exists, is to be taken on the subspace $\mathcal{J}$. For instance, if $\mathrm{Im}\,\mathbf{L} \geq \alpha\mathbb{I}$ or $\mathrm{Re}\,\mathbf{L} \geq \alpha\mathbb{I}$ for some $\alpha > 0$ and Assumption 41 below is satisfied, then the inverse of $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2$ exists on $\mathcal{J}$.

Remark 40. *It follows from Remark 38 that*

$$\mathbf{\Pi}_2\mathbf{L}\mathbf{\Pi}_2 = \mathbf{\Pi}_2\mathbf{L} = \mathbf{L}\mathbf{\Pi}_2 \ \text{ on } \mathcal{K}. \tag{10.38}$$

which implies that if $\mathbf{L}^{-1}$ *exists then*

$$\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2 = \mathbf{\Pi}_2\mathbf{L}^{-1} = \mathbf{L}^{-1}\mathbf{\Pi}_2 \ \text{ on } \mathcal{K}. \tag{10.39}$$

Assumption 41. *We will assume that*

$$\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2 \geq \beta\mathbf{\Gamma}_2 \ \text{ for some } \beta > 0. \tag{10.40}$$

Proposition 42. *A necessary and sufficient condition for Assumption 41 to be true is*

$$\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}. \tag{10.41}$$

Proof. For any $\mathbf{P} \in \mathcal{V} \cap \mathcal{J}$ we have $\mathbf{\Gamma}_2\mathbf{P} = \mathbf{P}$ and $\mathbf{\Pi}_2\mathbf{P} = 0$ so that $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2\mathbf{P} = 0$. This proves that a necessary condition for Assumption 41 to be true is $\mathcal{V}\cap\mathcal{J} = \{\mathbf{0}\}$. We will now prove that it is also a sufficient condition. The key observations are that $\mathbb{I}_\mathcal{K} = \mathbf{\Gamma}_1 + \mathbf{\Gamma}_2 = \mathbf{\Pi}_1 + \mathbf{\Pi}_2$ which implies

$$0 \leq \mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2 = \mathbf{\Gamma}_2(\mathbb{I}_\mathcal{K} - \mathbf{\Pi}_1)\mathbf{\Gamma}_2 = \mathbf{\Gamma}_2 - \mathbf{\Gamma}_2\mathbf{\Pi}_1\mathbf{\Gamma}_2, \tag{10.42}$$

and $\mathbf{\Pi}_1$ is a finite-rank operator since its range is $\mathcal{V}$ which is a finite-dimensional subspace by Proposition 45. This implies that

$$0 \leq \mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J} = \mathbb{I}_\mathcal{J} - \mathbf{\Gamma}_2\mathbf{\Pi}_1\mathbf{\Gamma}_2|_\mathcal{J}, \tag{10.43}$$

where $\mathbf{\Gamma}_2\mathbf{\Pi}_1\mathbf{\Gamma}_2|_\mathcal{J}$ is a finite-rank operator on $\mathcal{J}$, the range of $\mathbf{\Gamma}_2$. This implies by Proposition 31 that $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is an index 0 Fredholm operator on $\mathcal{J}$. Thus, either $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}\mathbf{P} = 0$, $\mathbf{P} \in \mathcal{J}$ has a nontrivial solution or $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is invertible on $\mathcal{J}$. But $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}\mathbf{P} = 0$, $\mathbf{P} \in \mathcal{J}$ holds if and only if $\mathbf{P} \in \mathcal{V} \cap \mathcal{J}$. This implies $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is invertible if and only if $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$. This proves that $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2 > 0$ on $\mathcal{J}$ if and only if $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$. We now want to prove that we actually have the stronger result, namely, if $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$ then Assumption 41 is true. This follows from immediately from the fact $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is a positive semidefinite self-adjoint bounded operator (its norm is bounded by 1) which is invertible. Hence, the spectrum $\sigma(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J})$ of $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is contained in the closure of its numerical range which is a real convex compact set:

$$[\alpha, \beta] = \mathrm{cl}\,\{(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2\,\mathbf{u}, \mathbf{u}) \mid \mathbf{u} \in \mathcal{J}, \ \ \|\mathbf{u}\| = 1\},$$

and this interval $[\alpha, \beta] \subseteq [0, 1]$ has its endpoints α and β belonging to $\sigma(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J})$. Thus, as the operator $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J}$ is invertible, α must be positive and therefore we get by the definition of the numerical range that $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2|_\mathcal{J} \geq \alpha\,\mathbb{I}_\mathcal{J}$. □

Corollary 43. *Define the operator-valued function* $\mathbf{F}(L_1, L_2) : \mathcal{J} \to \mathcal{J}$ *on* $\mathbb{C}^2$ *by*

$$\mathbf{F}(L_1, L_2) := \mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2.$$

If $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$ *then* $\mathrm{Ker}[\mathbf{F}(L_1, L_2)] = \mathrm{Ker}[\mathbf{F}(L_1, L_2)^\dagger] = \{\mathbf{0}\}$ *and the following three statements are equivalent:*

(i) $\mathbf{Y}_*(L_1, L_2)$ *is well-defined by the formula (10.37);*

(ii) $\mathrm{Ran}[\mathbf{F}(L_1, L_2)] = \mathcal{J}$*;*

(iii) $\mathbf{F}(L_1, L_2)$ *is a Fredholm operator of index* 0*.*

Proof. Suppose $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$. We will now show that $\mathrm{Ker}[\mathbf{F}(L_1, L_2)] = \{\mathbf{0}\}$. Let $\mathbf{J} \in \mathrm{Ker}[\mathbf{F}(L_1, L_2)]$. Then $\mathbf{J} \in \mathcal{J}$ and $\mathbf{0} = \mathbf{F}(L_1, L_2)\mathbf{J} = \mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{J}$ implying that $\mathbf{0} = \mathbf{\Pi}_2\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{J}$. But $\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Gamma}_2$ is invertible on $\mathcal{J}$ (as we showed in the proof of Proposition 42) and $\mathbf{\Pi}_2$ commutes with $\mathbf{L}(L_1, L_2)$ and hence with $\mathbf{L}(L_1, L_2)^{-1}$ implying $\mathbf{0} = \mathbf{\Pi}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{J} = \mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{J}$. Therefore, we conclude that $\mathbf{\Pi}_2\mathbf{J} = \mathbf{0}$ which implies $\mathbf{J} \in \mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$ and hence $\mathrm{Ker}[\mathbf{F}(L_1, L_2)] = \{\mathbf{0}\}$. As $\mathbf{F}(L_1, L_2)^\dagger = \mathbf{F}(\overline{L_1}, \overline{L_2})$, this implies $\mathrm{Ker}[\mathbf{F}(L_1, L_2)^\dagger] = \{\mathbf{0}\}$. The proof of the rest of the statement now follows immediately from these facts and the fact that since $\mathcal{J}$ is a Hilbert space then $\mathrm{Ker}[\mathbf{F}(L_1, L_2)^\dagger]^\perp = \mathrm{cl}\{\mathrm{Ran}[\mathbf{F}(L_1, L_2)]\}$. □

Using this corollary we can now give, in the next proposition, a condition on $\mathbf{L}$ that implies $\mathbf{Y}_*$ is well-defined by the formula (10.36).

Proposition 44. *If* $(L_1, L_2) \in \mathbb{C}^2$ *is such that* $\mathbf{L}(L_1, L_2)$ *is invertible and the operator* $\mathbf{G}(L_1, L_2) := (\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2)\mathbf{L}(L_1, L_2)^{-1}(\mathbf{\Gamma}_0 + \mathbf{\Gamma}_2) : \mathcal{U} \oplus \mathcal{J} \to \mathcal{U} \oplus \mathcal{J}$ *is also invertible then the operator* $\mathbf{F}(L_1, L_2) : \mathcal{J} \to \mathcal{J}$ *(defined in Corollary 43) is a Fredholm operator of index* 0*. Moreover, if* $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$ *then* $\mathbf{F}(L_1, L_2)$ *is invertible and* $\mathbf{Y}_*(L_1, L_2)$ *is well-defined by formula (10.37).*

Proof. Suppose that $(L_1, L_2) \in \mathbb{C}^2$ such that $\mathbf{L}(L_1, L_2)$ is invertible and the operator $\mathbf{G}(L_1, L_2) : \mathcal{U} \oplus \mathcal{J} \to \mathcal{U} \oplus \mathcal{J}$ defined above is also invertible. Then with respect to the orthogonal decomposition of the Hilbert space $\mathcal{U} \oplus \mathcal{J}$, with the corresponding orthogonal projections $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_2$, we can write this bounded linear operator in the 2×2 block matrix form

$$\mathbf{G}(L_1, L_2) = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}, \tag{10.44}$$

where $\mathbf{G}_{22} : \mathcal{J} \to \mathcal{J}$ is the operator

$$\mathbf{G}_{22} = \mathbf{\Gamma}_2\mathbf{G}(L_1, L_2)\mathbf{\Gamma}_2 = \mathbf{\Gamma}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Gamma}_2. \tag{10.45}$$

It follows immediately from Proposition 32 that $\mathbf{G}_{22}$ is a Fredholm operator of index 0. Using the relation (10.39) and the decomposition $\mathbf{\Gamma}_2 = \mathbf{\Pi}_1\mathbf{\Gamma}_2 + \mathbf{\Pi}_2\mathbf{\Gamma}_2$, we can link $\mathbf{G}_{22}$ and $\mathbf{F}(L_1, L_2)$ by the following relation:

$$\mathbf{F}(L_1, L_2) = \mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2 \tag{10.46}$$

$$= \mathbf{G}_{22} - \mathbf{\Gamma}_2\mathbf{L}(L_1, L_2)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_2. \tag{10.47}$$

Thus, as $\mathrm{Ran}\,\mathbf{\Pi}_1 = \mathcal{V}$ is a finite-dimensional space (see Corollary 46), by the invariance of the Fredholm index under perturbation by a compact operator (see Theorem 4.1, p. 355, section 15.4 of Gohberg, Goldberg, and Kaashoek 2003), we obtain that $\mathbf{F}(L_1, L_2)$ is a Fredholm operator of index 0. The rest of the statements of this proposition now follow immediately from these facts and Corollary 43. This completes the proof. □

We will now show that just as $\mathbf{L}_*$ satisfies the homogeneity and Herglotz property so too does $\mathbf{Y}_*$ under the Assumption 41, that is, we prove the following:

The homogeneity property:

$$\mathbf{Y}_*(L_1, L_2) = \frac{1}{c}\mathbf{Y}_*(cL_1, cL_2) \tag{10.48}$$

is satisfied for all choices of constants $c \neq 0$.

The Herglotz property:

$$\operatorname{Im} \mathbf{Y}_*(L_1, L_2) > 0, \quad \text{when } \operatorname{Im}(L_1) > 0 \text{ and } \operatorname{Im}(L_2) > 0. \tag{10.49}$$

Proof. We will prove these two properties under Assumption 41. First, the homogeneity property follows immediately from formula (10.37) and the homogeneity property of $\mathbf{L}_*$. Next, the Herglotz property of $\mathbf{Y}_*$ will be proven in the same as for $\mathbf{L}_*$ by using formula (10.37) and the identities (10.23). [Alternatively, the Herglotz property of $\mathbf{Y}_*$ follows directly from (2.82) in the same way as the Herglotz property of $\mathbf{L}_*$ follows from (2.68) or (10.24).] First, we obtain that under the Herglotz conditions $\operatorname{Im}(L_1) > 0$ and $\operatorname{Im}(L_2) > 0$, that $\operatorname{Im}(\mathbf{L}^{-1})$ is uniformly negative. Then by Assumption 41, it follows that the operator $\operatorname{Im}(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)$ is uniformly negative on $\mathcal{J}$ and that $\operatorname{Im}(\mathbf{\Gamma}_2(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}\mathbf{\Gamma}_2)$ is coercive on $\mathcal{J}$. From these facts we conclude that

$$\operatorname{Im}(\mathbf{\Pi}_1\mathbf{\Gamma}_2(\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{L}^{-1}\mathbf{\Pi}_2\mathbf{\Gamma}_2)^{-1}\mathbf{\Gamma}_2\mathbf{\Pi}_1) > 0 \text{ on } \mathcal{V}.$$

□

The relationship between $\mathbf{L}_*$ and $\mathbf{Y}_*$, proven in Section 19.1 of Milton (2002) and for two-component composites and in the abstract setting in Section 7.21, is given by

$$\mathbf{L}_* = \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Pi}_1[\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 + \mathbf{Y}_*]^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{\Gamma}_0. \tag{10.50}$$

Proof. For every $\mathbf{e} \in \mathcal{U}$ which has a solution to the Z-problem, there exists $\mathbf{j} \in \mathcal{U}$, $\mathbf{E} \in \mathcal{E}$, and $\mathbf{J} \in \mathcal{J}$ such that

$$\mathbf{j} + \mathbf{J} = \mathbf{L}(\mathbf{e} + \mathbf{E}), \quad \mathbf{L}_*\mathbf{e} = \mathbf{j}, \tag{10.51}$$

the latter by definition of $\mathbf{L}_*$. It follows from Remark 40 that $\mathbf{E}_1 = \mathbf{\Pi}_1\mathbf{E} \in \mathcal{V}$ is a solution to the Y-problem (10.35), i.e.,

$$\mathbf{J}_2 = \mathbf{L}\mathbf{E}_2, \quad \mathbf{E}_2 = \mathbf{\Pi}_2\mathbf{E}, \quad \mathbf{J}_2 = \mathbf{\Pi}_2\mathbf{J}, \quad \mathbf{E}_1 = \mathbf{\Pi}_1\mathbf{E}, \tag{10.52}$$

and, by definition we have

$$\mathbf{J}_1 = -\mathbf{Y}_*\mathbf{E}_1, \quad \mathbf{J}_1 = \mathbf{\Pi}_1\mathbf{J}. \tag{10.53}$$

Thus we can write

$$\mathbf{j} + \mathbf{J}_1 + \mathbf{J}_2 = \mathbf{L}(\mathbf{e} + \mathbf{E}_1 + \mathbf{E}_2). \tag{10.54}$$

We will now prove from these facts that the relationship (10.50) holds. First, we solve for $\mathbf{E}_1$ in terms of $\mathbf{e}$ from the identity

$$-\mathbf{Y}_*\mathbf{E}_1 = \mathbf{J}_1 = \mathbf{\Pi}_1\mathbf{L}(\mathbf{e} + \mathbf{E}_1) = \mathbf{\Pi}_1\mathbf{L}(\mathbf{e} + \mathbf{\Pi}_1\mathbf{E}_1) \tag{10.55}$$

and we find that

$$\mathbf{E}_1 = -\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 + \mathbf{Y}_*)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{e}, \tag{10.56}$$

where the inverse is taken on the subspace $\mathcal{V}$. It then follows that

$$\mathbf{j} + \mathbf{J}_1 = \mathbf{L}(\mathbf{e} + \mathbf{E}_1) = \mathbf{L}\mathbf{e} + \mathbf{L}\mathbf{E}_1 = \mathbf{L}\mathbf{e} - \mathbf{L}\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 + \mathbf{Y}_*)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{e}. \tag{10.57}$$

Hence we have

$$\mathbf{L}_*\mathbf{e} = \mathbf{j} = \mathbf{\Gamma}_0(\mathbf{j} + \mathbf{J}_1) = [\mathbf{\Gamma}_0\mathbf{L}\mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{L}\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 + \mathbf{Y}_*)^{-1}\mathbf{\Pi}_1\mathbf{L}\mathbf{\Gamma}_0]\mathbf{e}. \tag{10.58}$$

Therefore, it follows that the relationship (10.50) is true if $\mathbf{L}_*$, $\mathbf{Y}_*$, and $(\mathbf{\Pi}_1\mathbf{L}\mathbf{\Pi}_1 + \mathbf{Y}_*)^{-1}$ are well-defined operators (which is true, for instance, when $\mathrm{Im}(L_1) > 0$, $\mathrm{Im}(L_2) > 0$, and Assumption 41 holds). □

Now consider the Hilbert space $\mathcal{H}^{(1)}$, i.e., the orthogonal $Z(2)$ subspace collection (10.27). First, we emphasize here that the operator $\mathbf{L}$ commutes with the orthogonal projection $\mathbf{\Pi}_2$ whose range is $\mathcal{H}^{(1)}$ and hence by Remark 38 it follows that

$$\mathbf{L} = \mathbf{\Pi}_2\mathbf{L}\mathbf{\Pi}_2 = L_1\mathbf{\Lambda}_1^{(1)} + L_2\mathbf{\Lambda}_2^{(1)} \text{ on } \mathcal{H}^{(1)}, \tag{10.59}$$

where now $\mathbf{\Lambda}_1^{(1)}$ and $\mathbf{\Lambda}_2^{(1)}$ on $\mathcal{H}^{(1)}$ are the orthogonal projections onto $\mathcal{P}_1^{(1)}$ and $\mathcal{P}_2^{(1)}$, respectively. Now associated with the operator $\mathbf{L}$ on $\mathcal{H}^{(1)}$ and this orthogonal $Z(2)$ subspace collection is the linear operator-valued function ${\mathbf{L}_*}^{(1)}(L_1, L_2)$ (i.e., the associated Z-operator) acting on the subspace $\mathcal{U}^{(1)}$ which is the solution of the corresponding Z-problem in this new $Z(2)$ subspace collection. The results in section 10.2 now apply to this operator ${\mathbf{L}_*}^{(1)}(L_1, L_2)$ which, just like $\mathbf{L}_*$, satisfies the homogeneity, Herglotz, and normalization properties [i.e., the normalization property in this case is ${\mathbf{L}_*}^{(1)}(1, 1) = \mathbb{I}_{\mathcal{U}^{(1)}}$]. In particular, ${\mathbf{L}_*}^{(1)}$ admits the two following representation formulas (when they are well-defined) similar to the formulas and (10.15) (10.16) but now on the orthogonal $Z(2)$ subspace collection (10.27):

$${\mathbf{L}_*}^{(1)} = \mathbf{\Gamma}_0^{(1)}\mathbf{L}\mathbf{\Gamma}_0^{(1)} - \mathbf{\Gamma}_0^{(1)}\mathbf{L}\mathbf{\Gamma}_1^{(1)}(\mathbf{\Gamma}_1^{(1)}\mathbf{L}\mathbf{\Gamma}_1^{(1)})^{-1}\mathbf{\Gamma}_1^{(1)}\mathbf{L}\mathbf{\Gamma}_0^{(1)} \tag{10.60}$$

and

$${\mathbf{L}_*}^{(1)} = \mathbf{\Gamma}_0^{(1)}[(\mathbf{\Gamma}_0^{(1)} + \mathbf{\Gamma}_2^{(1)})\mathbf{L}^{-1}(\mathbf{\Gamma}_0^{(1)} + \mathbf{\Gamma}_2^{(1)})]^{-1}\mathbf{\Gamma}_0^{(1)}, \tag{10.61}$$

where $\mathbf{\Gamma}_0^{(1)}, \mathbf{\Gamma}_1^{(1)}$ and $\mathbf{\Gamma}_2^{(2)}$ are respectively the orthogonal projections onto $\mathcal{U}^{(1)}$, $\mathcal{E}^{(1)}$, $\mathcal{J}^{(1)}$, which are related to the orthogonal projection $\mathbf{\Pi}_2$ onto $\mathcal{H}^{(1)}$ by $\mathbf{\Pi}_2 = \mathbf{\Gamma}_0^{(1)} + \mathbf{\Gamma}_1^{(2)} + \mathbf{\Gamma}_2^{(2)}$. Then, can one link the operators $\mathbf{Y}_*$ associated with $\mathcal{V}$ and ${\mathbf{L}_*}^{(1)}$ associated with $\mathcal{U}^{(1)}$? The answer is yes. The first step in the proof is to show that these two operators act on finite-dimensional spaces of the same dimension, in other words that $\dim\mathcal{U}^{(1)} = \dim(\mathcal{V})$. This is a consequence of the following proposition and corollary.

Proposition 45. *Let $\mathcal{L}$ be a Hilbert space with two orthogonal projections $\mathcal{Q}_1$, $\mathcal{Q}_2$ which satisfy $\mathcal{Q}_1 + \mathcal{Q}_2 = \mathbb{I}_{\mathcal{L}}$. Assume that $\mathcal{L}$ has an orthogonal decomposition*

$$\mathcal{L} = \mathcal{M} \oplus \mathcal{N}. \tag{10.62}$$

Define the subspaces $\mathcal{N}_1, \mathcal{N}_2$ by

$$\mathcal{N}_1 = (\boldsymbol{\mathcal{Q}}_1\mathcal{L} \cap \mathcal{N}) \oplus (\boldsymbol{\mathcal{Q}}_2\mathcal{L} \cap \mathcal{N}), \;\; \mathcal{N}_2 = \mathcal{N} \ominus \mathcal{N}_1, \tag{10.63}$$

where the latter denotes the orthogonal complement of $\mathcal{N}_1$ in $\mathcal{N}$. Then we have

$$\mathcal{M} \oplus \mathcal{N}_2 = \boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M}. \tag{10.64}$$

In particular, if $\mathcal{M}$ is finite-dimensional then $\mathcal{N}_2$ is finite-dimensional with

$$\dim(\mathcal{N}_2) = \dim(\boldsymbol{\mathcal{Q}}_1\mathcal{M}) + \dim(\boldsymbol{\mathcal{Q}}_2\mathcal{M}) - \dim(\mathcal{M}) \leq \dim(\mathcal{M}). \tag{10.65}$$

Moreover, $\dim(\mathcal{N}_2) = \dim(\mathcal{M})$ if and only if $\boldsymbol{\mathcal{Q}}_1\mathcal{L} \cap \mathcal{M} = \boldsymbol{\mathcal{Q}}_2\mathcal{L} \cap \mathcal{M} = \{\mathbf{0}\}$.

Proof. First, $\mathcal{N}_1 \subseteq (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M})^\perp$ holds since $\boldsymbol{\mathcal{Q}}_1, \boldsymbol{\mathcal{Q}}_2$ are invariant on $\mathcal{N}_1$ and $\mathcal{M}$ is orthogonal to $\mathcal{N}$ which contains $\mathcal{N}_1$. This implies $\mathcal{M} \subseteq \boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M} \subseteq (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M})^{\perp\perp} \subseteq \mathcal{N}_1^\perp = \mathcal{M} \oplus \mathcal{N}_2$. Let $\mathbf{P} \in \mathcal{M} \oplus \mathcal{N}_2 \ominus (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M})$. Then $\mathbf{P} \in (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M})^\perp \subseteq \mathcal{M}^\perp = \mathcal{N}_1 \oplus \mathcal{N}_2$ implying $\mathbf{P} \in \mathcal{N}_2$. But by invariance of $\boldsymbol{\mathcal{Q}}_1, \boldsymbol{\mathcal{Q}}_2$ on $\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M}$ we must also have $\boldsymbol{\mathcal{Q}}_1\mathbf{P}, \boldsymbol{\mathcal{Q}}_2\mathbf{P} \in (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M})^\perp \subseteq \mathcal{N}$ and thus $\boldsymbol{\mathcal{Q}}_1\mathbf{P}, \boldsymbol{\mathcal{Q}}_2\mathbf{P} \in \mathcal{N}_1$. This implies $\mathbf{P} = \boldsymbol{\mathcal{Q}}_1\mathbf{P} + \boldsymbol{\mathcal{Q}}_2\mathbf{P} \in \mathcal{N}_1 \cap \mathcal{N}_2 = \{\mathbf{0}\}$ and hence $\mathbf{P} = \mathbf{0}$. Therefore, $\mathcal{M} \oplus \mathcal{N}_2 \ominus (\boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M}) = \{\mathbf{0}\}$ implying the relation (10.64): $\mathcal{M} \oplus \mathcal{N}_2 = \boldsymbol{\mathcal{Q}}_1\mathcal{M} \oplus \boldsymbol{\mathcal{Q}}_2\mathcal{M}$, as desired. Now assume that $\mathcal{M}$ is a finite-dimensional space, then the relation (10.65) follows immediately from the decomposition (10.64). Next, the equality $\dim(\mathcal{N}_2) = \dim(\mathcal{M})$ is equivalent to $\dim(\mathbf{Q}_1\mathcal{M}) = \dim(\mathbf{Q}_2\mathcal{M}) = \dim(\mathcal{M})$ and, by the rank theorem, it is equivalent to the injectivity of the restriction of $\mathbf{Q}_1$ and $\mathbf{Q}_2$ on $\mathcal{M}$. Using the fact $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are two projections which satisfy $\boldsymbol{\mathcal{Q}}_1 + \boldsymbol{\mathcal{Q}}_2 = \mathbb{I}_\mathcal{L}$, one can easily show that this injectivity condition is equivalent to $\boldsymbol{\mathcal{Q}}_1\mathcal{L} \cap \mathcal{M} = \boldsymbol{\mathcal{Q}}_2\mathcal{L} \cap \mathcal{M} = \{\mathbf{0}\}$. This completes the proof. □

Corollary 46. *The spaces $\mathcal{V}$ and $\mathcal{U}^{(1)}$ are finite-dimensional and orthogonal to $\mathcal{U}$ and $\mathcal{V}$, respectively. Furthermore, we have*

$$\mathcal{U} \oplus \mathcal{V} = \boldsymbol{\Lambda}_1\mathcal{U} \oplus \boldsymbol{\Lambda}_2\mathcal{U}, \tag{10.66}$$

$$\mathcal{V} \oplus \mathcal{U}^{(1)} = \boldsymbol{\Gamma}_1\mathcal{V} \oplus \boldsymbol{\Gamma}_2\mathcal{V}, \tag{10.67}$$

and

$$\dim(\mathcal{V}) = \dim(\boldsymbol{\Lambda}_1\mathcal{U}) + \dim(\boldsymbol{\Lambda}_2\mathcal{U}) - \dim(\mathcal{U}) \leq \dim(\mathcal{U}), \tag{10.68}$$

$$\dim(\mathcal{U}^{(1)}) = \dim(\boldsymbol{\Gamma}_1\mathcal{V}) + \dim(\boldsymbol{\Gamma}_2\mathcal{V}) - \dim(\mathcal{V}) \leq \dim(\mathcal{V}). \tag{10.69}$$

Moreover, the following two statements are true:
(i) $\dim(\mathcal{V}) = \dim(\mathcal{U})$ if and only if $\mathcal{P}_1 \cap \mathcal{U} = \mathcal{P}_2 \cap \mathcal{U} = \{\mathbf{0}\}$;
(ii) $\dim(\mathcal{U}^{(1)}) = \dim(\mathcal{V})$ if and only if $\mathcal{E} \cap \mathcal{V} = \mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$.

Proof. These results follow immediately from Proposition 45. Indeed, in the Hilbert space $\mathcal{H}$ we have the orthogonal decomposition $\mathcal{H} = \mathcal{U} \oplus \mathcal{K}$ and two orthogonal projections $\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2$ satisfying $\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 = \mathbb{I}_\mathcal{H}$. The spaces $\mathcal{L}, \mathcal{M}, \mathcal{N}, \mathcal{N}_1, \mathcal{N}_2$ in the above proposition are $\mathcal{H}, \mathcal{U}, \mathcal{K}, \mathcal{H}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}, \mathcal{V}$, respectively, since $\boldsymbol{\Lambda}_1\mathcal{H} = \mathcal{P}_1, \boldsymbol{\Lambda}_2\mathcal{H} = \mathcal{P}_2$. Similarly, on the Hilbert space $\mathcal{K}$ we have $\mathcal{K} = \mathcal{V} \oplus \mathcal{H}^{(1)}$ and there are two orthogonal projections $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2$ satisfying $\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 = \mathbb{I}_\mathcal{K}$. The spaces $\mathcal{L}, \mathcal{M}, \mathcal{N}, \mathcal{N}_1, \mathcal{N}_2$ in the above proposition are $\mathcal{K}, \mathcal{V}, \mathcal{H}^{(1)}, \mathcal{H}^{(1)} \ominus \mathcal{U}^{(1)}, \mathcal{U}^{(1)}$, respectively, since $\boldsymbol{\Gamma}_1\mathcal{K} = \mathcal{E}, \boldsymbol{\Gamma}_2\mathcal{K} = \mathcal{J}$. The corollary now follows from these identifications by Proposition 45. This completes the proof. □

Remark 47. *In the case of the conductivity equation for a two-component composite, the conditions* $\mathcal{P}_1 \cap \mathcal{U} = \mathcal{P}_2 \cap \mathcal{U} = \{\mathbf{0}\}$ *always hold. This is a direct consequence of the definition of the spaces* $\mathcal{P}_1, \mathcal{P}_2$ *and* $\mathcal{U}$ *(see Remarks 34 and 39) and we deduce that* $\dim \mathcal{V} = \dim \mathcal{U} = d$*. Concerning the conditions* $\mathcal{V} \cap \mathcal{J} = \{\mathbf{0}\}$ *and* $\mathcal{V} \cap \mathcal{E} = \{\mathbf{0}\}$*, they are nearly always satisfied. Indeed, it depends on the geometry shape of the interfaces between the two phases. Suppose that the two phases of the composite are Lipschitz domains. A current density* $\mathbf{J} \in \mathcal{V} \cap \mathcal{J}$ *is (by definition of* $\mathcal{V}$*) a constant* $\mathbf{C}_i$ *in each phase. Moreover, it belongs to* $H_{\mathrm{div}}(\mathcal{D}) = \{\mathbf{U} \in \mathbf{L}^2(\mathcal{D}) \mid \mathrm{div}\, \mathbf{U} \in \mathbf{L}^2(\mathcal{D})\}$*. Thus at each interface point between the two phases, the normal component of* $\mathbf{J}$ *has to be continuous (see chapter I of Monk 2003):*

$$[\mathbf{J} \cdot \mathbf{n}] = (\mathbf{C}_2 - \mathbf{C}_1) \cdot \mathbf{n} = \mathbf{0},$$

where $\mathbf{n}$ *denotes the unit normal vector (oriented for instance from phase 1 to phase 2). Thus, if the set of unit normal of all interface points contains* d *linear independent vectors, then* $\mathbf{J} = \mathbf{C}_2 = \mathbf{C}_1$ *in* $\mathcal{D}$ *and the zero average condition on* $\mathbf{J}$ *then implies* $\mathbf{J} = \mathbf{0}$*. In the same way, if an electrical field* $\mathbf{E} \in \mathcal{V} \cap \mathcal{E}$*, it has to be a constant* $\mathbf{C}_i$ *in each phase and it belongs to* $H_{\mathrm{curl}}(\mathcal{D}) = \{\mathbf{U} \in \mathbf{L}^2(\mathcal{D}) \mid \mathrm{curl}\, \mathbf{U} \in \mathbf{L}^2(\mathcal{D})\}$ *which implies that its tangential component is continuous (see chapter I of Monk 2003). This leads to*

$$[\mathbf{E} \wedge \mathbf{n}] = (\mathbf{C}_2 - \mathbf{C}_1) \wedge \mathbf{n} = \mathbf{0},$$

at each interface point between the two phases. Thus again, if the set of unit normal of all interface points contains d *linear independent vectors, then* $\mathbf{E} = \mathbf{C}_2 = \mathbf{C}_1$ *in* $\mathcal{D}$ *and the zero average condition on* $\mathbf{E}$ *implies that* $\mathbf{E} = \mathbf{0}$*. Nevertheless, one can construct easily for* $d \geq 2$ *counterexamples (with for example linear interfaces) where* $\mathcal{V} \cap \mathcal{J} \neq \{\mathbf{0}\}$ *or* $\mathcal{V} \cap \mathcal{E} \neq \{\mathbf{0}\}$*.*

For instance, consider the example from Milton 2002, pp. 407–408, section 19.4 of a laminate of two phases laminated in direction $\mathbf{n}$*: Let* f_1, f_2 *denote the volume fraction occupied by phase* $i = 1, 2$*, then the piecewise constant average value zero fields*

$$\mathbf{E} = [f_2\chi_1 - f_1\chi_2]\mathbf{n}, \quad \mathbf{J} = [f_2\chi_1 - f_1\chi_2]\mathbf{v}, \text{ with } \mathbf{n} \cdot \mathbf{v} = 0, \tag{10.70}$$

are curl-free and divergence-free, respectively, and therefore lie in $\mathcal{V} \cap \mathcal{E} \neq \{\mathbf{0}\}$ *and* $\mathcal{V} \cap \mathcal{J} \neq \{\mathbf{0}\}$*, respectively.*

Now, that we know under some assumptions that $\dim \mathcal{V} = \dim \mathcal{U}^{(1)}$, the second step is to construct an invertible linear map $\mathbf{K} : \mathcal{U}^{(1)} \to \mathcal{V}$ to show that the operators $\mathbf{Y}_*$ and $\mathbf{L}_*{}^{(1)}$ are $\mathbf{K}$ congruent, in other words that:

$$\mathbf{Y}_*(L_1, L_2) = \mathbf{K}\mathbf{L}_*{}^{(1)}(L_1, L_2)\mathbf{K}^\dagger.$$

This is the purpose of the following theorem.

Theorem 48. *If* $\mathcal{P}_1 \cap \mathcal{U} = \mathcal{P}_2 \cap \mathcal{U} = \{\mathbf{0}\}$ *and* $\mathcal{E} \cap \mathcal{V} = \mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$ *then Assumption 41 holds,* $\dim \mathcal{U} = \dim \mathcal{V} = \dim \mathcal{U}^{(1)}$*, and assuming the operators* $\mathbf{Y}_* = \mathbf{Y}_*(L_1, L_2)$*,* $\mathbf{L}_*{}^{(1)} = \mathbf{L}_*{}^{(1)}(L_1, L_2)$ *operators in (10.37) and (10.60), respectively, are well-defined (which is the case, for instance, if* $\mathbf{L}$ *is coercive) we have*

$$\mathbf{Y}_*(L_1, L_2) = \mathbf{K}\mathbf{L}_*{}^{(1)}(L_1, L_2)\mathbf{K}^\dagger, \tag{10.71}$$

where $\mathbf{K} : \mathcal{U}^{(1)} \to \mathcal{V}$ *is the invertible operator defined by*

$$\mathbf{K} = -(\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_2, \tag{10.72}$$

with the inverse of $\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1$ *taken on* $\mathcal{V}$*.*

Proof. Suppose that $\mathcal{P}_1 \cap \mathcal{U} = \mathcal{P}_2 \cap \mathcal{U} = \{\mathbf{0}\}$, $\mathcal{E} \cap \mathcal{V} = \mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$. It then follows that by Proposition 41 since $\mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$ we know that Assumption 41 holds, by Corollary 46 since $\mathcal{P}_1 \cap \mathcal{U} = \mathcal{P}_2 \cap \mathcal{U} = \{\mathbf{0}\}$ we know that $\dim \mathcal{U} = \dim \mathcal{V}$ and since $\mathcal{E} \cap \mathcal{V} = \mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$ we know that $\dim \mathcal{V} = \dim \mathcal{U}^{(1)}$. Also, since $\mathcal{J} \cap \mathcal{V} = \{\mathbf{0}\}$ and $\mathcal{V}$ is finite-dimensional then $\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1$ is invertible on $\mathcal{V}$ and hence the operator $\mathbf{K} : \mathcal{U}^{(1)} \to \mathcal{V}$ given by (10.72) is well-defined. We will now prove it is invertible. To do this we introduce the operator $\mathbf{K}' : \mathcal{U}^{(1)} \to \mathcal{V}$ defined by

$$\mathbf{K}' = -(\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_2, \tag{10.73}$$

with the inverse of $\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1$ taken on $\mathcal{V}$, which exists since by assumption $\mathcal{E} \cap \mathcal{V} = \{\mathbf{0}\}$ and $\mathcal{V}$ is finite-dimensional. We will now prove that $\mathbf{K}$ is invertible by showing that $\mathbf{K}^\dagger = -(\mathbf{K}')^{-1}$. This follows now from the fact that $\dim \mathcal{V} = \dim \mathcal{U}^{(1)} < \infty$ and on $\mathcal{V}$,

$$\begin{aligned}
\mathbf{K}'\mathbf{K}^\dagger &= (\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_2\mathbf{\Pi}_2\mathbf{\Gamma}_1\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1)^{-1} \\
&= (\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_2(\mathbb{I}_{\mathcal{K}} - \mathbf{\Pi}_1)\mathbf{\Gamma}_1\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1)^{-1} \\
&= -(\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1)^{-1}\mathbf{\Pi}_1\mathbf{\Gamma}_2\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1)^{-1} = -\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1(\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1)^{-1} = -\mathbb{I}_{\mathcal{V}}.
\end{aligned}$$

Now suppose the operators $\mathbf{Y}_* = \mathbf{Y}_*(L_1, L_2)$, $\mathbf{L}_*{}^{(1)} = \mathbf{L}_*{}^{(1)}(L_1, L_2)$ in (10.37) and (10.61), respectively, are well-defined. We will prove the identity (10.71). For any $\mathbf{E}_1 \in \mathcal{V}$ we have $\mathbf{J}_1 := -\mathbf{Y}_*\mathbf{E}_1 \in \mathcal{V}$ and we know that there exists a solution to the Y-problem, i.e., there exists an $\mathbf{E}' \in \mathcal{E}$ and $\mathbf{J}' \in \mathcal{J}$ such that $\mathbf{\Pi}_1\mathbf{E}' = \mathbf{E}_1$, $\mathbf{\Pi}_1\mathbf{J}' = \mathbf{J}_1$ and $\mathbf{J}_2 = \mathbf{L}\mathbf{E}_2$ where $\mathbf{J}_2 := \mathbf{\Pi}_2\mathbf{J}'$ and $\mathbf{E}_2 := \mathbf{\Pi}_2\mathbf{E}'$. Moreover, we remark that

$$\mathbf{E}_2 = \mathbf{\Pi}_2\mathbf{E}' \in \mathbf{\Pi}_2(\mathcal{E}) \subseteq \mathbf{\Pi}_2(\mathcal{E}_1 \oplus \mathcal{U}_1 \oplus \mathcal{V}) = \mathcal{E}_1 \oplus \mathcal{U}_1,$$

and in the same way that $\mathbf{J}_2 \in \mathcal{J}_1 \oplus \mathcal{U}_1$.

But for this $\mathbf{E}_0^{(1)} := \mathbf{\Gamma}_0^{(1)}\mathbf{E}'$ we can construct a solution to the Z-problem on the $Z(2)$ subspace collection $\mathcal{H}^{(1)} = \mathcal{U}^{(1)} \oplus \mathcal{E}^{(1)} \oplus \mathcal{J}^{(1)} = \mathcal{P}_1^{(1)} \oplus \mathcal{P}_2^{(1)}$, namely given $\mathbf{E}_0^{(1)} \in \mathcal{U}^{(1)}$ we see that the fields

$$\mathbf{J}_0^{(1)} := \mathbf{\Gamma}_0^{(1)}\mathbf{J}' \in \mathcal{U}^{(1)}, \quad \mathbf{E} := \mathbf{E}_2 - \mathbf{E}_0^{(1)} \in \mathcal{E}^{(1)}, \quad \mathbf{J} := \mathbf{J}_2 - \mathbf{J}_0^{(1)} \in \mathcal{J}^{(1)}$$

are such that $\mathbf{J}_0^{(1)} + \mathbf{J} = \mathbf{J}_2 = \mathbf{L}(\mathbf{E}_2) = \mathbf{L}(\mathbf{E}_0^{(1)} + \mathbf{E})$. But the Z-operator associated with the Z-problem on this $Z(2)$ subspace collection $\mathcal{H}^{(1)}$ is $\mathbf{L}_*{}^{(1)}$ and hence

$$\mathbf{J}_0^{(1)} = \mathbf{L}_*{}^{(1)}\mathbf{E}_0^{(1)}. \tag{10.74}$$

We will now prove that $\mathbf{J}_1 = \mathbf{K}\mathbf{J}_0^{(1)}$ and $\mathbf{E}_1 = \mathbf{K}'\mathbf{E}_0^{(1)}$. First, since $\mathbf{J}_2 - \mathbf{J}_0^{(1)} = \mathbf{J} \in \mathcal{J}^{(1)} \subseteq \mathcal{J}$, we have

$$\mathbf{0} = \mathbf{\Gamma}_1(\mathbf{J}_2 - \mathbf{J}_0^{(1)}) = \mathbf{\Gamma}_1\mathbf{J}_2 - \mathbf{\Gamma}_1\mathbf{J}_0^{(1)} \tag{10.75}$$

which leads to

$$\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{J}_2 = \mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_2\mathbf{J}_0^{(1)}. \tag{10.76}$$

Then, using the relation $\mathbf{J}_2 = \mathbf{J}' - \mathbf{J}_1$ with $\mathbf{J}' \in \mathcal{J}$ and $\mathbf{J}_1 \in \mathcal{V}$, we get

$$\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_1\mathbf{J}_1 = -\mathbf{\Pi}_1\mathbf{\Gamma}_1\mathbf{\Pi}_2\mathbf{J}_0^{(1)}, \tag{10.77}$$

and it leads to

$$\mathbf{J}_1 = \mathbf{K}\mathbf{J}_0^{(1)}. \tag{10.78}$$

Similarly, since $\mathbf{E}_2 - \mathbf{E}_0^{(1)} = \mathbf{E} \in \mathcal{E}^{(1)} \subseteq \mathcal{E}$ then we have

$$\mathbf{0} = \boldsymbol{\Gamma}_2(\mathbf{E}_2 - \mathbf{E}_0^{(1)}) = \boldsymbol{\Gamma}_2\mathbf{E}_2 - \boldsymbol{\Gamma}_2\mathbf{E}_0^{(1)}, \tag{10.79}$$

and we get that

$$\boldsymbol{\Pi}_1\boldsymbol{\Gamma}_2\boldsymbol{\Pi}_1\mathbf{E}_1 = -\boldsymbol{\Pi}_1\boldsymbol{\Gamma}_2\boldsymbol{\Pi}_2\mathbf{E}_0^{(1)}, \tag{10.80}$$

which leads to

$$\mathbf{E}_1 = \mathbf{K}'\mathbf{E}_0^{(1)}. \tag{10.81}$$

It follows from these facts that

$$-\mathbf{Y}_*\mathbf{E}_1 = \mathbf{J}_1 = \mathbf{K}\mathbf{J}_0^{(1)} = \mathbf{K}\mathbf{L}_*{}^{(1)}\mathbf{E}_0^{(1)} = \mathbf{K}\mathbf{L}_*{}^{(1)}(\mathbf{K}')^{-1}\mathbf{E}^{(1)} = -\mathbf{K}\mathbf{L}_*{}^{(1)}\mathbf{K}^\dagger\mathbf{E}_1.$$

As this is true for every $\mathbf{E}_1 \in \mathcal{V}$ it implies the relation

$$\mathbf{Y}_* = \mathbf{K}\mathbf{L}_*{}^{(1)}\mathbf{K}^\dagger, \tag{10.82}$$

as desired. This completes the proof. □

We want now to give necessary conditions under which the operators $\mathbf{L}_*{}^{(1)}$ is well-defined. This question is not as clear as for $\mathbf{Y}_*$ (see Proposition 44) since the well-posedness of the effective tensors $\mathbf{L}_*{}^{(1)}$ (as for the effective $\mathbf{L}_*$ for the base case) is directly linked to the properties of the operators $\mathbf{L}$ in the subspace decomposition $\mathcal{H}^{(1)} = \mathcal{U}^{(1)} \oplus \mathcal{E}^{(1)} \oplus \mathcal{J}^{(1)}$ and therefore is highly dependent on the physical problem and the structure of the composite.

For instance, the representation formulas (10.60) and (10.61) of $\mathbf{L}_*{}^{(1)}$ are respectively defined under the existence of the inverse $\boldsymbol{\Gamma}_1^{(1)}\mathbf{L}\boldsymbol{\Gamma}_1^{(1)}$ on $\mathcal{E}^{(i)}$ and of the inverse of $(\boldsymbol{\Gamma}_0^{(1)}+\boldsymbol{\Gamma}_2^{(1)})\mathbf{L}^{-1}(\boldsymbol{\Gamma}_0^{(1)}+\boldsymbol{\Gamma}_2^{(1)})$ on $\mathcal{U}^{(1)}\oplus\mathcal{J}^{(1)}$. We saw indeed that the existence of these both inverses is equivalent as soon as $\mathbf{L}$ is invertible on $\mathcal{H}^{(1)}$ (see Proposition 36). By the Remarks 37 and 38, the invertibility of $\mathbf{L}$ on $\mathcal{H}^{(1)}$ follows from the invertibility of $\mathbf{L}$ on $\mathcal{H}$ which is equivalent to $L_1 \neq 0$ and $L_2 \neq 0$. Moreover, if $\mathbf{L}$ is coercive (see Remark 35) which is the case for example under the Herglotz hypothesis (10.49) then $\mathbf{L}_*{}^{(1)}$ is well-defined by both formulas (10.60) and (10.61). But what can we say if we don't suppose this strong coercivity assumption?

As the field equation recursion method is an induction method, we want to establish here some criterion of well-posedness for $\mathbf{L}_*{}^{(1)}$ that it inherits from the well-posedness of $\mathbf{L}_*$ in the base case. For this purpose, the representation formula (10.60) is more suitable since the space $\mathcal{E}^{(1)}$ is constructed as a subspace of $\mathcal{E}$ (which is not the case of the formula (10.61) since $\mathcal{U}^{(1)} \oplus \mathcal{J}^{(1)}$ is not included in $\mathcal{U} \oplus \mathcal{J}$).

Proposition 49. *Let $\mathcal{W}^{(1)}$ be the subspace of $\mathcal{E}$ defined by*

$$\mathcal{W}^{(1)} = \mathcal{E} \ominus \mathcal{E}^{(1)}, \tag{10.83}$$

in other words, the orthogonal complements of $\mathcal{E}^{(1)}$ in $\mathcal{E}$. Then $\mathcal{W}^{(1)}$ is a finite-dimensional space and $\mathbf{L}_{11}$ (the restriction of the operator $\mathbf{\Gamma}_1\mathbf{L}\mathbf{\Gamma}_1$ on $\mathcal{E}$) can be represented as a 2×2 block operator matrix

$$\mathbf{L}_{11}=\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{L}_{11}^{(1)}\end{bmatrix}, \tag{10.84}$$

with respect to the orthogonal decomposition (10.83), where $\mathbf{L}_{11}^{(1)}$ is the restriction of $\mathbf{\Gamma}_1^{(1)}\mathbf{L}\mathbf{\Gamma}_1^{(1)}$ on $\mathcal{E}^{(1)}$. Furthermore, if $\mathbf{L}_{11}$ is invertible then $\mathbf{L}_{11}^{(1)}$ is a index 0 Fredholm operator and the invertibility of $\mathbf{L}_{11}^{(1)}$ is equivalent to the injectivity condition: $\operatorname{Ker}\mathbf{L}_{11}^{(1)}=\{0\}$. Hence, if the representation formula (10.15) holds for $\mathbf{L}_*$ and $\operatorname{Ker}\mathbf{L}_{11}^{(1)}=\{0\}$ then ${\mathbf{L}_*}^{(1)}$ is well-defined by the formula (10.60).

Proof. We have simply to prove here that $\mathcal{W}^{(1)}$ is a finite-dimensional space since all the other conclusions follow immediately from Proposition 32. From the relations (10.25), (10.27) and (10.28), we have that

$$\mathcal{K}=\mathcal{E}\oplus\mathcal{J}=\mathcal{V}\oplus\mathcal{U}^{(1)}\oplus\mathcal{E}^{(1)}\oplus\mathcal{J}^{(1)} \text{ with } \mathcal{E}^{(1)}\subseteq\mathcal{E} \text{ and } \mathcal{J}^{(1)}\subseteq\mathcal{J}.$$

Therefore, $\mathcal{E}=\mathcal{E}^{(1)}\oplus\mathcal{W}^{(1)}\subseteq\mathcal{E}^{(1)}\oplus\mathcal{V}\oplus\mathcal{U}^{(1)}$ and, since $\mathcal{V}$ and $\mathcal{U}^{(1)}$ are finite dimensional spaces by Corollary 46, this implies that $\mathcal{W}^{(1)}$ is finite dimensional. □

10.3.2 The induction step

We now introduce a hierarchy of subspaces

$$\mathcal{K}^{(i-1)}=\mathcal{E}^{(i-1)}\oplus\mathcal{J}^{(i-1)}=\mathcal{V}^{(i-1)}\oplus\mathcal{P}_1^{(i)}\oplus\mathcal{P}_2^{(i)}, \tag{10.85}$$

$$\mathcal{H}^{(i)}=\mathcal{U}^{(i)}\oplus\mathcal{E}^{(i)}\oplus\mathcal{J}^{(i)}=\mathcal{P}_1^{(i)}\oplus\mathcal{P}_2^{(i)}, \tag{10.86}$$

for $i=1,2,3,\ldots$, where

$$\mathcal{P}_1^{(i)}=\mathcal{P}_1\cap\mathcal{K}^{(i-1)},\ \ \mathcal{P}_2^{(i)}=\mathcal{P}_2\cap\mathcal{K}^{(i-1)},\ \ \mathcal{V}^{(i-1)}=\mathcal{K}^{(i-1)}\ominus(\mathcal{P}_1^{(i)}\oplus\mathcal{P}_2^{(i)}), \tag{10.87}$$

$$\mathcal{E}^{(i)}=\mathcal{E}\cap\mathcal{H}^{(i)},\ \ \mathcal{J}^{(i)}=\mathcal{J}\cap\mathcal{H}^{(i)},\ \ \mathcal{U}^{(i)}=\mathcal{H}^{(i)}\ominus(\mathcal{E}^{(i)}\oplus\mathcal{J}^{(i)}). \tag{10.88}$$

For all positive integers i, we denote by $\mathbf{\Pi}_1^{(i)}$ and $\mathbf{\Pi}_2^{(i)}$ the orthogonal projections on $\mathcal{V}^{(i-1)}$ and $\mathcal{H}^{(i)}=\mathcal{P}_1^{(i)}\oplus\mathcal{P}_2^{(i)}$ associated with the orthogonal $Y(2)$ subspace collection (10.85) and by $\mathbf{\Gamma}_0^{(i)}$, $\mathbf{\Gamma}_1^{(i)}$, $\mathbf{\Gamma}_2^{(i)}$, $\mathbf{\Lambda}_1^{(i)}$ and $\mathbf{\Lambda}_2^{(i)}$ the orthogonal projections on $\mathcal{U}^{(i)}$, $\mathcal{E}^{(i)}$, $\mathcal{J}^{(i)}$, $\mathcal{P}_1^{(i)}$ and $\mathcal{P}_2^{(i)}$ associated with the orthogonal $Z(2)$ subspace collection (10.86).

The following theorem defines and gives the properties of the Z- and Y-operators at each level i of the induction argument. We take here as convention that in the indexing of the spaces and operators, the index 0 refers to the base case developed in sections 10.2 and 10.3. Also, we denote by $\mathbf{L}_{11}^{(i)}$ the restriction of the operator $\mathbf{\Gamma}_1^{(i)}\mathbf{L}\mathbf{\Gamma}_1^{(i)}$ on $\mathcal{E}^{(i)}$.

Theorem 50. *Assume that the operators $\mathbf{L}$ on $\mathcal{H}^{(0)}$ and $\mathbf{L}_{11}^{(0)}$ on $\mathcal{E}^{(0)}$ are invertible and let $i\in\mathbb{N}$. If, for each $k=1,....,i$, we have $\operatorname{Ker}\mathbf{L}_{11}^{(k)}=\{0\}$ then for all $k=0,....,i$, the Z-operator ${\mathbf{L}_*}^{(k)}:\mathcal{U}^{(k)}\to\mathcal{U}^{(k)}$ is well-defined by both representation formulas:*

$${\mathbf{L}_*}^{(k)} = \mathbf{\Gamma}_0^{(k)}\mathbf{L}\mathbf{\Gamma}_0^{(k)}-\mathbf{\Gamma}_0^{(k)}\mathbf{L}\mathbf{\Gamma}_1^{(k)}(\mathbf{\Gamma}_1^{(k)}\mathbf{L}\mathbf{\Gamma}_1^{(k)})^{-1}\mathbf{\Gamma}_1^{(k)}\mathbf{L}\mathbf{\Gamma}_0^{(k)}] \tag{10.89}$$

$$= \mathbf{\Gamma}_0^{(k)}[(\mathbf{\Gamma}_0^{(k)}+\mathbf{\Gamma}_2^{(k)})\mathbf{L}^{-1}(\mathbf{\Gamma}_0^{(k)}+\mathbf{\Gamma}_2^{(k)})]^{-1}\mathbf{\Gamma}_0^{(k)}, \tag{10.90}$$

and $\mathbf{L}_*{}^{(k)}$ *satisfies the homogeneity, normalization, and Herglotz properties. Furthermore, if* $\mathcal{V}^{(k)} \cap \mathcal{J}^{(k)} = \{\mathbf{0}\}$ *for each* $k = 0, ..., i$*, then the* Y*-operator* $\mathbf{Y}_*{}^{(k)} : \mathcal{V}^{(k)} \to \mathcal{V}^{(k)}$ *is also well-defined by*

$$\mathbf{Y}_*{}^{(k)} = \mathbf{\Pi}_1^{(k)}\mathbf{\Gamma}_2^{(k)}(\mathbf{\Gamma}_2^{(k)}\mathbf{\Pi}_2^{(k)}\mathbf{L}^{-1}\mathbf{\Pi}_2^{(k)}\mathbf{\Gamma}_2^{(k)})^{-1}\mathbf{\Gamma}_2^{(k)}\mathbf{\Pi}_1^{(k)}, \tag{10.91}$$

and $\mathbf{Y}_*{}^{(k)}$ *satisfies both the homogeneity and Herglotz properties. Moreover, if we also suppose that*

$$\mathcal{V}^{(k)} \cap \mathcal{E}^{(k)} = 0 \text{ and } \mathcal{P}_1^{(k)} \cap \mathcal{U}^{(k)} = \mathcal{P}_2^{(k)} \cap \mathcal{U}^{(k)} = \{\mathbf{0}\}, \text{ for all } k = 0, ..., i-1$$

then for all $k = 0, ..., i-1$ *we have*

$$\dim \mathcal{U}^{(0)} = \dim \mathcal{U}^{(k)} = \dim \mathcal{V}^{(k)} = \dim \mathcal{U}^{(i)}, \tag{10.92}$$

and there exists an invertible operator $\mathbf{K}^{(k)} : \mathcal{U}^{(k+1)} \to \mathcal{V}^{(k)}$ *defined by*

$$\mathbf{K}^{(k)} = -(\mathbf{\Pi}_1^{(k)}\mathbf{\Gamma}_1^{(k)}\mathbf{\Pi}_1^{(k)})^{-1}\mathbf{\Pi}_1^{(k)}\mathbf{\Gamma}_1^{(k)}\mathbf{\Pi}_2^{(k)}, \tag{10.93}$$

such that $\mathbf{Y}_*{}^{(k)}$ *and* $\mathbf{L}_*{}^{(k+1)}$ *are linked by the following relation:*

$$\mathbf{Y}_*{}^{(k)} = \mathbf{K}^{(k)}\mathbf{L}_*{}^{(k+1)}(\mathbf{K}^{(k)})^\dagger. \tag{10.94}$$

Proof. Suppose the operators $\mathbf{L}$ on $\mathcal{H}^{(0)}$ and $\mathbf{L}_{11}^{(0)}$ on $\mathcal{E}^{(0)}$ are invertible and fix an integer $i \in \mathbb{N}$. First, by Proposition 36, in the case $k = 0$ the Z-operator $\mathbf{L}_*{}^{(0)} : \mathcal{U}^{(0)} \to \mathcal{U}^{(0)}$ is well-defined by both formulas (10.89) and (10.90) and hence satisfies the homogeneity, normalization, and Herglotz properties.

Next, suppose that for each $k = 1, ..., i$, we have $\mathrm{Ker}\, \mathbf{L}_{11}^{(k)} = \{0\}$. Hence, as $\mathbf{L}_{11}^{(0)}$ is invertible, we deduce by induction by using the Proposition 49 that all the operators $\mathbf{L}_{11}^{(1)}, \ldots, \mathbf{L}_{11}^{(i)}$ are invertible and that the Z-operator $\mathbf{L}_*{}^{(k)} : \mathcal{U}^{(k)} \to \mathcal{U}^{(k)}$ is well-defined by formula (10.89). Then, as $\mathbf{L}$ is assumed invertible on $\mathcal{H}$, we have $L_1 \neq 0$ and $L_2 \neq 0$ (see Remark 37) and this implies that $\mathbf{L}$ is invertible on $\mathcal{H}^{(k)}$ too (see Remark 38). Hence, it follows from Proposition 36 that the Z-operator $\mathbf{L}_*{}^{(k)}$ is also well-defined by the formula (10.90) and thus satisfies the homogeneity, normalization, and Herglotz properties.

Suppose now that in addition, $\mathcal{V}^{(k)} \cap \mathcal{J}^{(k)} = \{\mathbf{0}\}$ for each $k = 0, ..., i$. Then it follows immediately from Proposition 44 that the Y-operator $\mathbf{Y}_*{}^{(k)} : \mathcal{V}^{(k)} \to \mathcal{V}^{(k)}$ is also well-defined by the formula (10.91) and hence satisfies both the homogeneity and Herglotz properties for all $k = 0, ..., i$.

Finally, suppose that in addition,

$$\mathcal{V}^{(k)} \cap \mathcal{E}^{(k)} = 0 \text{ and } \mathcal{P}_1^{(k)} \cap \mathcal{U}^{(k)} = \mathcal{P}_2^{(k)} \cap \mathcal{U}^{(k)} = \{\mathbf{0}\}, \text{ for all } k = 0, ..., i-1.$$

Then the rest of the proof of this theorem now follows immediately from Theorem 48. This completes the proof. □

Acknowledgements

Graeme Milton is grateful to the National Science Foundation for support through the Research Grant DMS-1211359. Aaron Welters is grateful for the support from the U.S. Air Force Office of Scientific Research (AFOSR) through the Air Force's Young Investigator Research Program (YIP) under the grant FA9550-15-1-0086.

11 Projection Functional Theory for finding excited states

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City UT 84112, USA

Abstract

Density functional theory enormously simplifies the calculation of the electron density in the ground state of a system. Here we extend the ideas of density functional theory to excited states of a system. One starts by integrating the square of the Schrodinger equation, which gives a functional to minimize which is quadratic in the potential. Splitting up the minimization in partial minimizations leads to a "density functional type theory" involving a functional of 3 functions, that each project out information contained in the full-wavefunction: hence the name projectional functional theory. Finding or approximating this functional will still be a challenge, especially since one function is essentially the electron-pair density function, and it is not yet known which such functions are legal in the sense that they derive from a completely antisymmetric wavefunction.

11.1 Introduction

Density functional theory (DFT) enables one to calculate the electron density associated with a ground state of an N electron system, while avoiding the direct calculation of the electron wavefunction ψ which is a complex-valued function of $3N$-spatial dimensions and the spin configuration which takes 2^N values: 3 spatial dimensions for each of the N-electrons whose positions may be correlated, and two values for each electron spin. Due to the large dimensionality involved, direct calculation of the wavefunction quickly becomes hopeless when a large number of electrons are involved, even using current computers. Modern density functional theory began with the pioneering papers of Hohenberg and Kohn (1964) and Kohn and Sham (1965). It got off to a slow start, but interest in it has exploded since 1990, with now over 16,000 publications per year on the topics of "density functional or DFT": see the excellent recent historical reviews of Jones (2015) and Zangwill (2014), and the multitude of references therein. Walter Kohn in 1998 shared the Nobel prize in chemistry "for his development of the density-functional theory", and an article in *Nature* lists it as "easily the most heavily cited concept in the physical sciences", with articles by Becke (1993) and Lee, Yang, and Parr (1988) being listed as two of the top ten most cited articles (with about 60,000 citations each according to Google Scholar in 2015). They quote Becke as saying "But the applications are endless. At a fundamental level, DFT can be used to describe all of chemistry, biochemistry, biology, nanosystems and materials. Everything in out

terrestrial world depends on the motions depends on the motions of electrons-therefore, DFT literally underlies everything." Density Functional Theory has been applied with success to periodic systems (Kresse and Furthmüller 1996; Ismail-Beigi and Arias 2000; Segall, Lindan, Probert, Pickard, Hasnip, Clark, and Payne 2002; Gonze, Beuken, Caracas, Detraux, Fuchs, Rignanese, Sindic, Verstraete, Zerah, Jollet, Torrent, Roy, Mikami, Ghosez, Raty, and Allan 2002) and to non-periodic systems, using finite elements (Suryanarayana, Gavini, Blesgen, Bhattacharya, and Ortiz 2010).

Here we are interested in generalizing density functional theory to the case of finding the excited states (eigenfunctions having associated energies above the ground state energy) of the time-harmonic multielectron Schrödinger equation for the wave-function $\psi(\mathbf{x})$,

$$E\psi(\mathbf{x}) = -\nabla \cdot \mathbf{A}\nabla\psi(\mathbf{x}) + V(\mathbf{x})\psi(\mathbf{x}), \tag{11.1}$$

where the energy E (not to be confused with the field $\mathbf{E}(\mathbf{x})$) must be in the spectrum of the operator for a nonzero solution for $\psi(\mathbf{x})$ to exist, $V(\mathbf{x})$ is the potential, and in the simplest approximation $\mathbf{A} = \hbar^2\mathbf{I}/(2m)$, but which may take other forms to take into account the reduced mass of the electron, or mass polarization terms due to the motion of the atomic nuclei. Here, m is the mass of the electron, $\hbar$ is Planck's constant divided by 2. We have used the notation of Parr and Weitao (1994), where, assuming we are in three spatial dimensions with N electrons, $\mathbf{x}$ represents the multidimensional vector $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ where the $\mathbf{x}_i = (\mathbf{r}_i, s_i)$, and $\mathbf{r}_i$ is a three dimensional vector associated with the position of electron i and s_i is associated with its spin, taking values $+1/2$ or $-1/2$. We use

$$\nabla = (\nabla_1, \nabla_2, \ldots, \nabla_N), \quad \text{where } \nabla_j = \left(\frac{\partial}{\partial r_1^{(j)}}, \frac{\partial}{\partial r_2^{(j)}}, \frac{\partial}{\partial r_3^{(j)}}\right). \tag{11.2}$$

The wavefunction $\psi(\mathbf{x})$ is antisymmetric with respect to interchange of any pair of $\mathbf{x}_j$ and $\mathbf{x}_k$, while $V(\mathbf{x})$ is assumed to be symmetric with respect to such interchanges.

Our analysis applies to isolated systems confined within a body Ω, or to periodic systems with Ω taken as the unit cell of periodicity of the wavefunction. Note that if the material we wish to simulate is periodic, with a unit cell Ω_0 of the potential, then we should take Ω to contain many copies of Ω_0 and not just one: for example if Ω_0 was a cube of side length ℓ, then we could take Ω to be a cube of side length $m\ell$ where m is large. This is to allow the wavevector $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N)$ which enters the Bloch periodicity conditions at the boundary of the cell Ω_0^N in the $3N$ dimensional space to take a reasonably large number of discrete values. In the example where Ω_0 is a cube and m is even the components k_{ip} of $\mathbf{k}_p$ can take the m values $k_{ip} = 2\pi j/(m\ell)$, $j = 1 - m/2, 2 - m/2, \ldots, m/2$. Note, however, that a Bloch solution on Ω_0^N will not in general satisfy the required antisymmetry requirements of the wavefunction, so one will have to antisymmetrize it to obtain an acceptable wavefunction: this will no longer be quasiperiodic on the cell Ω_0^N, but it will be periodic on the cell Ω^N. (I thank Richard James for pointing out the need to do something like this). It may be sufficient to take Ω_0 as the cell of periodicity of the wavefunction if the cell contains a molecule whose associated wavefunction is localized near the center of the cell.

Following the procedure of the **Chapter 1** one can rewrite the equation as

$$\underbrace{\begin{pmatrix} \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{q}(\mathbf{x}) \end{pmatrix}}_{\mathbf{J}(\mathbf{x})} = \underbrace{\begin{pmatrix} -\mathbf{A} & 0 \\ 0 & E - V(\mathbf{x}) \end{pmatrix}}_{\mathbf{L}(\mathbf{x})} \underbrace{\begin{pmatrix} \nabla\psi(\mathbf{x}) \\ \psi(\mathbf{x}) \end{pmatrix}}_{\mathbf{E}(\mathbf{x})}, \tag{11.3}$$

where $\mathbf{q}(\mathbf{x})$ is some current.

Density functional theory only involves a functional of

$$\rho(\mathbf{x}_1) = \frac{1}{|\Omega|^{N-1}} \int_{\Omega^{N-1}} \overline{\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) d\mathbf{x}_2 \, d\mathbf{x}_3 \ldots d\mathbf{x}_N. \tag{11.4}$$

which is the electron density. Here Ω could be the region to which the wavefunction is confined, or it could be a unit cell of periodicity.

By contrast, our theory in the simplest case where the wavefunction and energy are real involves a functional $F_0(S, T, P_2, E)$ of the energy E and three functions:

$$\begin{aligned}
P_2(\mathbf{x}_1, \mathbf{x}_2) &= \frac{1}{|\Omega|^{N-2}} \int_{\Omega^{N-2}} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) d\mathbf{x}_3 \ldots d\mathbf{x}_N, \\
S(\mathbf{x}_1) &= \frac{1}{|\Omega^{N-1}|} \int_{\Omega^{N-1}} \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \nabla \cdot \underbrace{\mathbf{q}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)}_{-\mathbf{A}\nabla\psi} d\mathbf{x}_2 \ldots d\mathbf{x}_N, \\
T(\mathbf{x}_1) &= \frac{1}{|\Omega^{N-1}|} \int_{\Omega^{N-1}} \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) V_{\text{ee}}(\mathbf{x}) \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) d\mathbf{x}_2 \ldots d\mathbf{x}_N,
\end{aligned} \tag{11.5}$$

where $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is the (real-valued) wavefunction and the potential $V_{\text{ee}}(\mathbf{x})$, with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, is the interaction between electrons and other electrons:

$$V_{\text{ee}}(\mathbf{x}) = \sum_{i<j}^{N} U(\mathbf{x}_i, \mathbf{x}_j), \tag{11.6}$$

in which

$$U(\mathbf{x}_i, \mathbf{x}_j) = \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{11.7}$$

is the electron–electron repulsion, and e is the electron charge. Each of these three functions projects out information contained in the three functions

$$\begin{aligned}
&\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N), \\
&\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \nabla \cdot \mathbf{q}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N), \\
&\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) V_{\text{ee}}(\mathbf{x}) \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N),
\end{aligned} \tag{11.8}$$

hence the name "Projection functional theory".

We emphasize that it might require a considerable amount of work to determine appropriate forms for the functional $F_0(S, T, P_2, E)$, given that it involves not just one but three functions including $P_2(\mathbf{x}_1, \mathbf{x}_2)$ which involves a function of two co-ordinates, $\mathbf{x}_1$ and $\mathbf{x}_2$. For nonperiodic systems a lot of attention has been paid to both the electron-pair density

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = \frac{N(N-1)}{2} \int \cdots \int \overline{\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)} \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) d\mathbf{x}_3 \ldots d\mathbf{x}_N, \tag{11.9}$$

(where an integration over $\mathbf{x}_1$ means integrating $\mathbf{r}_i$ over $\mathbb{R}^3$ and summing over s_i) which is obviously closely related to $P_2(\mathbf{x}_1, \mathbf{x}_2)$ when the wavefunction is real, as we have assumed, and also the associated second-order reduced density matrix

$$\gamma_2(\mathbf{x}_1', \mathbf{x}_2'; \mathbf{x}_1, \mathbf{x}_2) = \frac{N(N-1)}{2} \int \cdots \int \overline{\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)} \psi(\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_N) d\mathbf{x}_3 \ldots d\mathbf{x}_N. \tag{11.10}$$

(The electron-pair density can be regarded as the "diagonal component" $\gamma_2(\mathbf{x}_1', \mathbf{x}_2'; \mathbf{x}_1, \mathbf{x}_2)$ of the second-order reduced density matrix.) It is not even known which electron-pair densities, or second-order reduced density matrices, are legitimate in the sense that they can be generated from some antisymmetric wavefunction $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$: this is the famous N-representability problem (Coleman 1981; Erdahl and Smith, Jr. 1987; Parr and Weitao 1994; Davidson 1995; Harrison 2003). A necessary step in the minimization is knowing the set of functions S, T, P_2, and E over which the minimization should be performed. This need not be the set of all legitimate function combinations, but just those legitimate functions for which $F_0(S, T, P_2, E)$ is not too large — but of course identifying these functions may not be easy either.

This chapter is mostly self-contained, and it is only necessary for the reader to have studied Section 1.7.

11.2 A review of density functional theory for the ground state

Let us review the standard density functional theory using the approach of Levy (1979). If one is interested in finding the electronic ground state, with energy E_0, for a multielectron system (with no source terms) one minimizes the Rayleigh quotient

$$E_0 = \min_{\substack{\psi \\ \psi=0 \text{ on } \partial\Omega^N \\ \langle\overline{\psi}\psi\rangle=1}} \langle \nabla\overline{\psi} \cdot \mathbf{A}\nabla\psi + \overline{\psi}V(\mathbf{x})\psi\rangle, \tag{11.11}$$

where the angular brackets denote an average as the spatial coordinates $\mathbf{r}$ range over Ω^N and as the spin configuration $s = (s_1, s_2, \ldots, s_N)$ varies over all 2^N possibilities. Now $V(\mathbf{x})$ has the decomposition

$$V(\mathbf{x}) = V_{\text{ne}}(\mathbf{x}) + V_{\text{ee}}(\mathbf{x}), \tag{11.12}$$

where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_i, \mathbf{x}_j, \ldots, \mathbf{x}_N)$. The potential $V(\mathbf{x})$ is completely symmetric with respect to swapping any pair of co-ordinates $\mathbf{x}_i, \mathbf{x}_j$. The term V_{ne} is the interaction between the M nuclei and N electrons given by

$$V_{\text{ne}}(\mathbf{x}) = \sum_{i=1}^{N}\sum_{j=1}^{M} V_{\text{ne}}^{(j)}(\mathbf{x}_i), \quad \text{[recall } \mathbf{x}_i = (\mathbf{r}_i, s_i) \text{ where } \mathbf{r}_i \in \mathbb{R}^3\text{]}, \tag{11.13}$$

with

$$V_{\text{ne}}^{(j)}(\mathbf{x}_i) = \frac{-Z_j e^2}{|\mathbf{r}_i - \mathbf{R}_j|}, \tag{11.14}$$

where $Z_j e$ is the charge on the nucleus positioned at $\mathbf{R}_j$ and $\mathbf{r}_i$ is the position of electron i having spin s_i. The term V_{ee} given by (11.6) and (11.7) is the interaction between electrons and other electrons. V_{ee} is universal (independent of the position of the nuclei) while V_{ne} is not. Note that $V_{\text{ee}}(\mathbf{x})$ and $V_{\text{ne}}(\mathbf{x})$ are completely symmetric but $V_{\text{ne}}^{(j)}(\mathbf{x}_i)$ is not.

Now from the antisymmetry properties of $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ we can successively swap variables of integration and deduce that

$$\begin{aligned} \langle\overline{\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)}V_{\text{ne}}^{(j)}(\mathbf{x}_i)\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)\rangle &= \langle\overline{\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)}V_{\text{ne}}^{(j)}(\mathbf{x}_1)\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)\rangle \\ &= \frac{1}{|\Omega|}\int_\Omega V_{\text{ne}}^{(j)}(\mathbf{x}_1)\rho(\mathbf{x}_1)d\mathbf{x}_1, \end{aligned} \tag{11.15}$$

where

$$\rho(\mathbf{x}_1) = \frac{1}{|\Omega|^{N-1}} \int_{\Omega^{N-1}} \overline{\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) d\mathbf{x}_2 \, d\mathbf{x}_3 \ldots d\mathbf{x}_N \tag{11.16}$$

is the electron density.

Then the minimization in (11.11) can be split into two minimizations. The first is the calculation of

$$F(\rho_*) = \min_{\substack{\psi \\ \psi=0 \text{ on } \partial\Omega^N \\ \rho(\mathbf{x}_1)=\rho_*(\mathbf{x}_1)}} \langle \nabla\overline{\psi} \cdot \mathbf{A} \nabla\psi + \overline{\psi} V_{\text{ee}}(\mathbf{x}) \psi \rangle, \tag{11.17}$$

which is a *universal functional*, the density functional, that is independent of the positions of the nuclei. If the system is periodic then the constraint that $\psi = 0$ on $\partial\Omega^N$ should be replaced with periodic boundary conditions. The evaluation of this functional is not easy but one may use numerical experiments, approximations, and guesswork to estimate its form. With the density functional in hand one only needs to do the remaining minimization,

$$E_0 = \min_{\substack{\rho_* \\ \langle \rho_* \rangle = 1}} \left[F(\rho_*) + \langle \sum_{i,j}^{N} V_{\text{ne}}^{(j)}(\mathbf{x}_1) \rho_*(\mathbf{x}_1) \rangle \right], \tag{11.18}$$

which greatly simplifies the calculation as the problem is reduced to minimization of a function in a 3-dimensional space, rather than the minimization over the $3N$ dimensional space we started with. In this last expression the angular brackets are now just an average as $\mathbf{r}_1$ ranges over Ω and as s_1 takes the two possible values $+1/2$ and $-1/2$.

Of course the difficulty is knowing what form of the density functional $F(\rho_*)$ to take, and there is an enormous body of work directed to determining a suitable form, that agrees well with numerical results and experimental data.

11.3 Expanding the terms in a variational principle for excited states

Let us define

$$p(\mathbf{x}) = \nabla \cdot \mathbf{A} \nabla\psi(\mathbf{x}) + (E - V(\mathbf{x}))\psi(\mathbf{x}). \tag{11.19}$$

If the potential $V(\mathbf{x})$ is real and the energy E is real and the spectrum is discrete, then at points E on the spectrum, both the real and imaginary parts of the wavefunction satisfy the Schrödinger equation, and so $p(\mathbf{x}) = 0$. Hence, let us assume the wavefunction ψ is real. We have the obvious variational principle that when

$$W(\psi) = \langle [p(\mathbf{x})]^2 \rangle, \quad \text{where } p(\mathbf{x}) = \nabla \cdot \mathbf{A} \nabla\psi(\mathbf{x}) + (E - V(\mathbf{x}))\psi(\mathbf{x}), \tag{11.20}$$

is minimized over ψ, with $\langle \psi^2 \rangle = 1$, where the angular brackets denote an average over $\mathbf{r}$ as it ranges over Ω^N and over s as it ranges over all 2^N configurations. The minimum is zero and occurs at $p(\mathbf{x}) = 0$, provided E is on the discrete spectrum.

The quantity $W(\psi)$ to be minimized, given by (11.20), has the equivalent form

$$W(\psi) \quad = \quad \langle \begin{pmatrix} -\nabla \cdot \mathbf{q} \\ \psi \end{pmatrix} \cdot \begin{pmatrix} 1 & E - V(\mathbf{x}) \\ E - V(\mathbf{x}) & (E - V(\mathbf{x}))^2 \end{pmatrix} \begin{pmatrix} -\nabla \cdot \mathbf{q} \\ \psi \end{pmatrix} \rangle, \tag{11.21}$$

in which $\mathbf{q} = -\mathbf{A}\nabla\psi$. Now we follow the ideas inherent in the standard density functional theory. From the antisymmetry properties of $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ and $\nabla \cdot \mathbf{q}(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ we can successively swap variables of integration and deduce that

$$\begin{aligned}\langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_i) \nabla \cdot \mathbf{q}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle &= \langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_1) \nabla \cdot \mathbf{q}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &= \frac{1}{|\Omega|} \int_\Omega V_{\text{ne}}^{(j)}(\mathbf{x}_1) S(\mathbf{x}_1) d\mathbf{x}_1, \end{aligned} \tag{11.22}$$

where

$$S(\mathbf{x}_1) = \frac{1}{|\Omega^{N-1}|} \int_{\Omega^{N-1}} \psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \nabla \cdot \underbrace{\mathbf{q}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)}_{-\mathbf{A}\nabla\psi} d\mathbf{x}_2 \ldots d\mathbf{x}_N. \tag{11.23}$$

Also the term which is quadratic in the potentials can be expanded:

$$\begin{aligned}(E - V(\mathbf{x}))^2 &= (E - V_{\text{ne}}(\mathbf{x}) - V_{\text{ee}}(\mathbf{x}))^2 = E^2 + (V_{\text{ne}}(\mathbf{x}))^2 + (V_{\text{ee}}(\mathbf{x}))^2 \\ &\quad -2EV_{\text{ne}}(\mathbf{x}) - 2EV_{\text{ee}}(\mathbf{x}) + 2V_{\text{ne}}(\mathbf{x})V_{\text{ee}}(\mathbf{x}). \end{aligned} \tag{11.24}$$

Let's look at the contribution from each of these. We have for $i \neq \ell$:

$$\begin{aligned}&\langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(k)}(\mathbf{x}_i) V_{\text{ne}}^{(j)}(\mathbf{x}_\ell) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(k)}(\mathbf{x}_1) V_{\text{ne}}^{(j)}(\mathbf{x}_2) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \frac{1}{|\Omega|^2} \int_{\Omega^2} V_{\text{ne}}^{(k)}(\mathbf{x}_1) V_{\text{ne}}^{(j)}(\mathbf{x}_2) P_2(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2, \end{aligned} \tag{11.25}$$

where

$$P_2(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{|\Omega|^{N-2}} \int_{\Omega^{N-2}} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) d\mathbf{x}_3 \ldots d\mathbf{x}_N. \tag{11.26}$$

For $\ell = i$ we have

$$\begin{aligned}&\langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(k)}(\mathbf{x}_i) V_{\text{ne}}^{(j)}(\mathbf{x}_i) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(k)}(\mathbf{x}_1) V_{\text{ne}}^{(j)}(\mathbf{x}_1) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \langle V_{\text{ne}}^{(k)}(\mathbf{x}_1) V_{\text{ne}}^{(j)}(\mathbf{x}_1) P_1(\mathbf{x}_1) \rangle, \end{aligned} \tag{11.27}$$

where

$$P_1(\mathbf{x}_1) = \int_\Omega P_2(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2. \tag{11.28}$$

Also we have

$$\begin{aligned}\langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_i) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle &= \langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_1) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &= \frac{1}{|\Omega|} \int_\Omega V_{\text{ne}}^{(j)}(\mathbf{x}_1) P_1(\mathbf{x}_1) d\mathbf{x}_1. \end{aligned} \tag{11.29}$$

Similarly, because $V_{\text{ee}}(\mathbf{x})$, with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_i, \mathbf{x}_j, \ldots, \mathbf{x}_N)$, is symmetric with respect to interchange of any pair $\mathbf{x}_i$ and $\mathbf{x}_k$ we have

$$\begin{aligned}&\langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_i) V_{\text{ee}}(\mathbf{x}) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \langle \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) V_{\text{ne}}^{(j)}(\mathbf{x}_1) V_{\text{ee}}(\mathbf{x}) \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \rangle \\ &\quad = \frac{1}{|\Omega|} \int_\Omega V_{\text{ne}}^{(i)}(\mathbf{x}_1) T(\mathbf{x}_1) d\mathbf{x}_1, \end{aligned} \tag{11.30}$$

where

$$T(\mathbf{x}_1) = \frac{1}{|\Omega^{N-1}|}\int_{\Omega^{N-1}} \psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)V_{\text{ee}}(\mathbf{x})\psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)d\mathbf{x}_2\ldots d\mathbf{x}_N. \tag{11.31}$$

Now we have

$$\begin{aligned}\langle\psi(E-V(\mathbf{x}))\nabla\cdot\mathbf{q}\rangle &= \langle\nabla\cdot\mathbf{q}(E-V(\mathbf{x}))\psi\rangle \\ &= \langle\psi\left(E-\sum_{i,j}^{N}V_{\text{ne}}^{(j)}(\mathbf{x}_i)-V_{\text{ee}}(\mathbf{x})\right)\nabla\cdot\mathbf{q}\rangle \\ &= E\langle\psi\nabla\cdot\mathbf{q}\rangle-\sum_{i}^{N}\langle V_{\text{ne}}^{(j)}(\mathbf{x})S(\mathbf{x})\rangle N \\ &\quad -\langle\psi V_{\text{ee}}(\mathbf{x})\nabla\cdot\mathbf{q}\rangle,\end{aligned} \tag{11.32}$$

and

$$\begin{aligned}\langle\psi[E-V(\mathbf{x})]^2\psi\rangle &= E^2\langle\psi^2\rangle-2E\langle\psi(V_{\text{ne}}(\mathbf{x})+V_{\text{ee}}(\mathbf{x}))\psi\rangle \\ &\quad +\langle\psi(V_{\text{ne}}(\mathbf{x}))^2\psi\rangle+\langle\psi(V_{\text{ee}}(\mathbf{x}))^2\psi\rangle \\ &\quad +2\langle\psi V_{\text{ne}}(\mathbf{x})V_{\text{ee}}(\mathbf{x})\psi\rangle.\end{aligned} \tag{11.33}$$

Here one can make the identifications

$$\begin{aligned}\langle\psi V_{\text{ne}}(\mathbf{x})\psi\rangle &= \langle\psi\sum_{i,j}^{N}V_{\text{ne}}^{(j)}(\mathbf{x}_i)\psi\rangle \\ &= \sum_{j=1}^{N}\langle V_{\text{ne}}^{(j)}(\mathbf{x}_1)P_1(\mathbf{x}_1)\rangle N,\end{aligned} \tag{11.34}$$

and

$$\begin{aligned}\langle\psi(V_{\text{ne}}(\mathbf{x}))^2\psi\rangle &= \langle\psi\sum_{k,i}^{N}V_{\text{ne}}^{(k)}(\mathbf{x}_i)\sum_{\ell,j}^{N}V_{\text{ne}}^{(j)}(\mathbf{x}_\ell)\psi\rangle \\ &= \sum_{k,i}^{N}\sum_{j,\ell}^{N}\langle\psi V_{\text{ne}}^{(k)}(\mathbf{x}_i)V_{\text{ne}}^{(j)}(\mathbf{x}_\ell)\psi\rangle \\ &= \sum_{k,i}^{N}\sum_{j,\ell}^{N}N(N-1)\langle V_{\text{ne}}^{(k)}(\mathbf{x}_1)V_{\text{ne}}^{(j)}(\mathbf{x}_2)P_2(\mathbf{x}_1,\mathbf{x}_2)\rangle \\ &\quad +\sum_{k=1}^{N}\sum_{j=1}^{N}N\langle V_{\text{ne}}^{(k)}(\mathbf{x}_1)V_{\text{ne}}^{(j)}(\mathbf{x}_1)P_1(\mathbf{x}_1)\rangle,\end{aligned} \tag{11.35}$$

and

$$\begin{aligned}\langle\psi V_{\rm ne}(\mathbf{x})V_{\rm ee}(\mathbf{x})\psi\rangle &= \langle\psi\sum_{i,j}^{N}V_{\rm ne}^{(j)}(\mathbf{x}_i)V_{\rm ee}(\mathbf{x})\psi\rangle \\ &= \sum_{j=1}^{N}\langle V_{\rm ne}^{(j)}(\mathbf{x}_1)T(\mathbf{x}_1)\rangle N, \end{aligned} \tag{11.36}$$

and

$$\begin{aligned}\langle\psi(E-V)\psi\rangle &= \langle\psi(E-\sum_{i,j}^{N}V_{\rm ne}^{(j)}(\mathbf{x}_i)-V_{\rm ee}(\mathbf{x}))\psi\rangle \\ &= E\langle\psi^2\rangle-\sum_{j=1}^{N}\langle V_{\rm ne}^{(j)}(\mathbf{x}_1)P_1(\mathbf{x}_1)\rangle N-\langle\psi V_{\rm ee}(\mathbf{x})\psi\rangle. \end{aligned} \tag{11.37}$$

11.4 Using the variational principle to derive Projection Functional Theory

Now noting that P_1 derives from P_2, we can write W as a sum

$$W = W_1(\psi,E)+W_2(S,T,P_2,E) \tag{11.38}$$

of a first term $W_1(\psi,E)$ which is universal, not dependent on the positions of the nuclei, and a remaining term $W_2(S,T,P_2,E)$ which is dependent on the position of the nuclei. If the projection functionals $S(\mathbf{x}_1)$, $T(\mathbf{x}_1)$, and $P_2(\mathbf{x}_1,\mathbf{x}_2)$ are known (and again these do not depend on the nuclei positions) the evaluation of $W_2(S,T,P_2,E,T_0)$ only requires an integration over two variables $\mathbf{x}_1$ and $\mathbf{x}_2$.

Specifically we have

$$\begin{aligned}W_1(\psi,E) &= \langle(\nabla\cdot\mathbf{q})(\nabla\cdot\mathbf{q})\rangle-2E\langle\psi\nabla\cdot\mathbf{q}\rangle \\ &\quad +2\langle\psi V_{\rm ee}(\mathbf{x})\nabla\cdot\mathbf{q}\rangle+E^2\langle\psi^2\rangle \\ &\quad -2E\langle\psi V_{\rm ee}(\mathbf{x})\psi\rangle+\langle\psi(V_{\rm ee}(\mathbf{x}))^2\psi\rangle, \end{aligned} \tag{11.39}$$

and

$$\begin{aligned}W_2(S,T,P_2,E) &= 2\sum_{j=1}^{N}\langle V_{\rm ne}^{(j)}(\mathbf{x}_1)S(\mathbf{x}_1)\rangle-2E\sum_{j=1}^{N}\langle V_{\rm ne}^{(j)}(\mathbf{x}_1)P_1(\mathbf{x})\rangle \\ &\quad +2\sum_{k=1}^{N}\sum_{j=1}^{N}\langle V_{\rm ne}^{(k)}(\mathbf{x}_1)V_{\rm ne}^{(j)}(\mathbf{x}_2)P_2(\mathbf{x}_1,\mathbf{x}_2)\rangle \\ &\quad +2\sum_{k=1}^{N}\sum_{j=1}^{N}N\langle V_{\rm ne}^{(k)}(\mathbf{x}_1)V_{\rm ne}^{(j)}(\mathbf{x}_1)P_1(\mathbf{x}_1)\rangle \\ &\quad +2\sum_{j=1}^{N}\langle V_{\rm ne}^{(j)}(\mathbf{x}_1)T(\mathbf{x}_1)\rangle N, \end{aligned} \tag{11.40}$$

in which $\mathbf{q} = -\mathbf{A}\nabla\psi$. We split the minimization into two parts, the first being

$$\min_{\substack{\psi \\ \langle\psi^2\rangle=1}} W_1(\psi) \equiv F_0(S_*, T_*, P_{2*}, E), \tag{11.41}$$

where the minimum is over all functions ψ with the required antisymmetry, satisfying the constraints

$$\begin{aligned} S(\mathbf{x}_1) &= S_*(\mathbf{x}_1), \\ T(\mathbf{x}_1)) &= T_*(\mathbf{x}_1), \\ P_2(\mathbf{x}_1, \mathbf{x}_2) &= P_{2*}(\mathbf{x}_1, \mathbf{x}_2), \end{aligned} \tag{11.42}$$

where $S_*(\mathbf{x}_1)$, $T_*(\mathbf{x}_1)$ and $P_{2*}(\mathbf{x}_1, \mathbf{x}_2)$ are arbitrary given functionals. It could be that there are no antisymmetric wavefunctions $\psi(\mathbf{x})$ satisfying these constraints, in which case we set $F_0(S_*, T_*, P_{2*}) = \infty$. This is a multidimensional problem that has to be solved (or the solution approximated), but it is independent of the positions of the nuclei. Finally, the positions on the nuclei enter in the final minimization

$$W = \min_{S_*, T_*, P_{2*}} [F_0(S_*, T_*, P_{2*}, E) + W_2(S_*, T_*, P_{2*}, E)], \tag{11.43}$$

which is a minimization over the projective functions S_*, T_*, P_{2*}. It is a nonlinear minimization, but one that only involves functionals of at most six co-ordinates, namely those of $\mathbf{x}_1$ and $\mathbf{x}_2$.

12 The desymmetrization method for solving the Schrödinger equation

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City UT 84112, USA

Abstract

Here we present a new method for solving the multielectron Schrödinger equation in periodic media with source terms. The key point is to write the equation in a form where the matrix entering the constitutive law is desymmetrized, i.e., only depends on the first two, $\mathbf{x}_1$ and $\mathbf{x}_2$, of the N particle coordinates $\mathbf{x}_1$, $\mathbf{x}_2$,..., $\mathbf{x}_N$. Then it is proposed to solve the equation by Fast Fourier Transform methods. Due to the desymmetrization the Fast Fourier Transforms need only be done on the first two coordinates $\mathbf{x}_1$ and $\mathbf{x}_2$. A disadvantage of the method is that one has to store and update the complete wavefunction in Fourier space.

12.1 Introduction

Density functional theory relies on the fact that in the variational principle for the ground state of a N-electron system one can desymmetrize the potential $V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ (where each $\mathbf{x}_i = (\mathbf{r}_i, s_i)$ represents a combination of a spatial coordinate $\mathbf{r}_i$ and a spin coordinate s_i) and replace it by a potential which just depends on $\mathbf{x}_1$ and $\mathbf{x}_2$ due to the antisymmetry property of the wavefunction. For periodic media, Fast Fourier Transforms have become an essential tool in modern density functional theory computations (Kresse and Furthmüller 1996; Ismail-Beigi and Arias 2000; Segall, Lindan, Probert, Pickard, Hasnip, Clark, and Payne 2002; Gonze, Beuken, Caracas, Detraux, Fuchs, Rignanese, Sindic, Verstraete, Zerah, Jollet, Torrent, Roy, Mikami, Ghosez, Raty, and Allan 2002). Here we show that a similar desymmetrization can be made in the constitutive law for the Schrödinger's equation, and thus is applicable to excited states as well as ground states. When desymmetrized, the tensor which enters the constitutive law only depends on 6 coordinates, namely the elements of two vectors $\mathbf{x}_1, \mathbf{x}_2$, rather than the entire $3N$ coordinates that enter the usual three-dimensional Schrödinger equation. With periodic boundary conditions the desymmetrized equation can be solved using the Fast Fourier Transform methods that were developed for composites by Moulinec and Suquet (1994, 1998) and Eyre and Milton (1999) (see also the generalization in Section 14.9 of Milton 2002, in particular equation (14.38), and see also Moulinec and Silva 2014 for a recent review). The desymmetrized nature of the constitutive law allows one to do the Fourier transform steps only on $\mathbf{x}_1$ and $\mathbf{x}_2$, rather than on all co-ordinates $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. At each stage of the iteration one still needs to update the Fourier transform of the full electron wavefunction $\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ and thus it is expected that this method will be applicable only when there are relatively

few electrons. One primarily works in Fourier space. An additional advantage is that the antisymmetrization step is considerably easier as the function to be antisymmetrized is already partially antisymmetrized.

We emphasize that if the material we wish to simulate is periodic, with a unit cell Ω_0 in three-dimensional space, then we should take the period cell Ω of the wavefunction to contain many copies of Ω_0 and not just one: for example if Ω_0 was a cube of side length ℓ, then we could take Ω to be a cube of side length $m\ell$ where m is large. As mentioned in the introduction of the previous chapter this is to allow the wavevector $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N)$ which enters the Bloch periodicity conditions at the boundary of the cell Ω_0^N in the $3N$ dimensional space to take a reasonably large number of discrete values. While the solution we seek will not correspond to a Bloch solution (which generally does not have the required antisymmetry) it should correspond to an antisymmetrization of it and will not be quasiperiodic on the cell Ω_0^N, but will be periodic on the cell Ω^N. Again, it may be sufficient to take Ω_0 as the cell of periodicity of the wavefunction if the cell contains a molecule whose associated wavefunction is localized near the center of the cell.

We are investigating the multielectron Schrödinger equation with a source term, which as follows from (1.92) and Section 1.19 of **Chapter 1**, can be expressed in the form,

$$\begin{pmatrix} \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{v}(\mathbf{x}) + S_0 \end{pmatrix} = \begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & E - V(\mathbf{x}) & h(\mathbf{x}) \\ 0 & \overline{h}(\mathbf{x}) & d(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \nabla\psi \\ \psi \\ \theta_0 \end{pmatrix}. \tag{12.1}$$

where we have added the source term $h(\mathbf{x})$. If we are looking for an eigensolution, then we need to take E almost real and close to an eigenvalue and the precise form of the forcing term $h(\mathbf{x})$ is almost irrelevant except that it is chosen to prevent the null-solution and its amplitude gives the normalization of the eigensolution. (We should avoid those special $h(\mathbf{x})$ which do not excite the relevant eigensolution, such as happens when $h(\mathbf{x})$ and the eigensolution have opposing symmetries that prevent excitation).

In this formulation, the wavefunction $\psi(\mathbf{x})$, the driving source term $h(\mathbf{x})$, and the divergence $\nabla \cdot \mathbf{q}(\mathbf{x})$ are antisymmetric with respect to interchange of any pair of $\mathbf{x}_i$ and $\mathbf{x}_\ell$, with $i \neq \ell$, in $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, while the potential $V(\mathbf{x})$, $\nabla \cdot \mathbf{v}(\mathbf{x})$, and $d(\mathbf{x})$ are assumed to be symmetric with respect to such interchanges. When E'' is positive, the solution to the equation (12.1) will be unique for a specified choice of θ_0, and thus the wavefunction $\psi(\mathbf{x})$ will have the desired symmetries, provided $V(\mathbf{x})$ and $d(\mathbf{x})$ are symmetric under electron interchange and $h(\mathbf{x})$ is antisymmetric. The multidimensional nature of this problem makes it very challenging to solve if there are a large number N of electrons. Here we show how some simplifications can be made to facilitate its solution.

A major assumption we make in our proofs of convergence, and associated estimates of the rate of convergence is that the potential $V(\mathbf{x})$ is bounded. Neither the electron–electron repulsion potential, nor the electron–nuclei attraction potential satisfy this boundedness. It is still conceivable that the iterative solution converges with these unbounded potentials, however it may be best to truncate the potentials to assure convergence.

This chapter is essentially self-contained although the reader is advised to look at **Chapters 1** and **2** beforehand. Readers interested in this chapter may also be interested in **Chapters 8** and **11**.

12.2 The desymmetrized form of the Schrödinger equation

Let us consider the following Hilbert space $\mathcal{H}$ consisting of Ω^N-periodic fields

$$\mathbf{P}(\mathbf{x}) = \begin{pmatrix} \mathbf{P}_1(\mathbf{x}) \\ p_2(\mathbf{x}) \\ p_3(\mathbf{x}) \end{pmatrix}, \tag{12.2}$$

that are such that

$$\int \cdots \int \overline{\mathbf{P}(\mathbf{x})} \cdot \mathbf{P}(\mathbf{x})\, d\mathbf{x}_1 d\mathbf{x}_2 \ldots d\mathbf{x}_N \tag{12.3}$$

is finite where the integral over $d\mathbf{x}_i$, where $\mathbf{x}_i = (\mathbf{r}_i, s_i)$ means an integral of the 3-dimensional vector $\mathbf{r}_i$ over the unit cell Ω and a sum over the spin variable s_i. Here $\mathbf{P}_1(\mathbf{x}) = \mathbf{P}_1(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is a $3N$ dimensional vector where N is the number of particles considered, but it does not have any special symmetries under the interchange of $\mathbf{x}_\ell = (\mathbf{r}_\ell, s_\ell)$, and $\mathbf{x}_j = (\mathbf{r}_j, s_j)$. The functions $p_2(\mathbf{x}), p_3(\mathbf{x})$ are scalars, again with no special symmetries. Define the projection

$$\mathbf{\Lambda} \begin{pmatrix} \mathbf{P}_1(\mathbf{x}) \\ p_2(\mathbf{x}) \\ p_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_A(\mathbf{P}_1(\mathbf{x})) \\ \Lambda_a(p_2(\mathbf{x})) \\ \Lambda_s(p_3(\mathbf{x})) \end{pmatrix}. \tag{12.4}$$

which acts to appropriately symmetrize the fields. Here the operator $\mathbf{\Lambda}_A$ projects any $3N$-dimensional vector field $\mathbf{P}_1(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ onto the space of $3N$-dimensional vector fields with 3-dimensional component fields $\mathbf{B}_m(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, $m =$ 1,2,. . ., N, such that under interchange of j and k:

$$\begin{aligned} \mathbf{B}_j(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_N) &= -\mathbf{B}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N), \\ \mathbf{B}_m(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_N) &= -\mathbf{B}_m(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N), \end{aligned} \tag{12.5}$$

if $m \neq j$ and $m \neq k$, e.g., for a 3 electron system where $\mathbf{P}_1(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ has three-dimensional component vector fields $\mathbf{P}_{11}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, $\mathbf{P}_{12}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, and $\mathbf{P}_{13}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ we have

$$\mathbf{\Lambda}_A \left(\begin{array}{c} \mathbf{P}_{11}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\ \hline \mathbf{P}_{12}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \\ \hline \mathbf{P}_{13}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \end{array} \right) = \left(\begin{array}{c} \mathbf{P}_{11}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) - \mathbf{P}_{11}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_2) - \mathbf{P}_{12}(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) \\ +\mathbf{P}_{12}(\mathbf{x}_3, \mathbf{x}_1, \mathbf{x}_2) - \mathbf{P}_{13}(\mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1) + \mathbf{P}_{13}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_1) \\ \hline \mathbf{P}_{12}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) - \mathbf{P}_{12}(\mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1) - \mathbf{P}_{11}(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) \\ +\mathbf{P}_{11}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_1) - \mathbf{P}_{13}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_2) + \mathbf{P}_{13}(\mathbf{x}_3, \mathbf{x}_1, \mathbf{x}_2) \\ \hline \mathbf{P}_{13}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) - \mathbf{P}_{13}(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) - \mathbf{P}_{11}(\mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1) \\ +\mathbf{P}_{11}(\mathbf{x}_3, \mathbf{x}_1, \mathbf{x}_2) - \mathbf{P}_{12}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_2) + \mathbf{P}_{12}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_1) \end{array} \right). \tag{12.6}$$

This projection has been chosen so the $3N$-dimensional vector field $\mathbf{P}_1(\mathbf{x}) = \nabla\psi(\mathbf{x})$, with component fields $\mathbf{P}_{1j}(\mathbf{x}) = \nabla_j\psi(\mathbf{x})$ where ∇_j is the gradient with respect to $\mathbf{r}_j$, satisfies $\mathbf{\Lambda}_A(\nabla\psi) = \nabla\psi$ when the wavefunction ψ is antisymmetric with respect to interchange of any pair of $\mathbf{x}_i$ and $\mathbf{x}_\ell$, with $i \neq \ell$, in $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$. The operator Λ_s projects onto completely symmetric fields and Λ_a projects onto completely antisymmetric fields, e.g., for a 3 electron system:

$$\begin{aligned} \Lambda_s(p_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)) &= \frac{1}{6}\Big[p_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) + p_3(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3) + p_3(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_1) \\ &\quad + p_3(\mathbf{x}_3, \mathbf{x}_2, \mathbf{x}_1) + p_3(\mathbf{x}_3, \mathbf{x}_1, \mathbf{x}_2) + p_3(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_2)\Big], \end{aligned} \tag{12.7}$$

and

$$\begin{aligned}\Lambda_a(p_2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)) &= \frac{1}{6}\Big[p_2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)-p_2(\mathbf{x}_2,\mathbf{x}_1,\mathbf{x}_3)+p_2(\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_1)\\ &\quad -p_2(\mathbf{x}_3,\mathbf{x}_2,\mathbf{x}_1)+p_2(\mathbf{x}_3,\mathbf{x}_1,\mathbf{x}_2)-p_2(\mathbf{x}_1,\mathbf{x}_3,\mathbf{x}_2)\Big],\end{aligned} \tag{12.8}$$

with obvious generalizations in the case of more electrons. The operator $\mathbf{\Lambda}$ projections onto some space $\mathcal{H}_s$ which has an orthogonal complement $\mathcal{H}_r$, with a projection $\mathbf{\Lambda}_r$ onto it:

$$\mathbf{\Lambda}_r = \mathbf{I}-\mathbf{\Lambda}, \quad \mathbf{\Lambda}_r\mathbf{\Lambda}=0. \tag{12.9}$$

Let $\mathcal{U}$ consist of fields representable in the form

$$\mathbf{U}(\mathbf{x}) = \begin{pmatrix}0\\0\\\theta_0(s)\end{pmatrix}, \tag{12.10}$$

for some choice of function $\theta_0(s)$ only depending on the spin variable $s=(s_1,s_2,\ldots,s_N)$.

Define $\mathcal{E}$ to consist of those Ω^N periodic fields in $\mathcal{H}$ of the form

$$\mathbf{E}(\mathbf{x}) = \begin{pmatrix}\nabla\psi(\mathbf{x})\\\psi(\mathbf{x})\\0\end{pmatrix}, \tag{12.11}$$

as $\psi(\mathbf{x})$ varies over all fields with the usual antisymmetries of wavefunctions, i.e.,

$$\mathbf{\Lambda E} = \mathbf{E}. \tag{12.12}$$

(We emphasize that the field $\mathbf{E}$ should not be confused with the energy E). Define $\mathcal{J}$ to consist of those Ω^N periodic fields $\mathbf{J}(\mathbf{x})$ in $\mathcal{H}$ of the form

$$\mathbf{\Lambda J}(\mathbf{x}) = \begin{pmatrix}\mathbf{q}(\mathbf{x})\\\nabla\cdot\mathbf{q}(\mathbf{x})\\\nabla\cdot\mathbf{v}(\mathbf{x})\end{pmatrix}. \tag{12.13}$$

for some periodic fields $\mathbf{q}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$. The relevance of this analysis is the following. Suppose we are given an "applied field"

$$\mathbf{E}_0 = \begin{pmatrix}0\\0\\1\end{pmatrix} \in \mathcal{U}, \tag{12.14}$$

which clearly satisfies $\mathbf{\Lambda E}_0 = \mathbf{E}_0$. By solving

$$\mathbf{J}_0(s)+\mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})(\mathbf{E}_0(s)+\mathbf{E}(\mathbf{x})), \quad \mathbf{E}_0,\mathbf{J}_0\in\mathcal{U}, \quad \mathbf{J}\in\mathcal{J}, \quad \mathbf{E}\in\mathcal{E}, \tag{12.15}$$

we automatically have a solution to the problem

$$\begin{aligned}\mathbf{\Lambda E} &= \mathbf{E},\\ \mathbf{\Lambda J}_0(s)+\mathbf{\Lambda J} &= (\mathbf{\Lambda L}(\mathbf{x})\mathbf{\Lambda})(\mathbf{E}(\mathbf{x})+\mathbf{E}_0),\end{aligned} \tag{12.16}$$

i.e.,

$$\begin{pmatrix} \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{q}(\mathbf{x}) \\ S_0(s) + \nabla \cdot \mathbf{v}(\mathbf{x}) \end{pmatrix} = \mathbf{\Lambda L}(\mathbf{x})\mathbf{\Lambda} \begin{pmatrix} \nabla\psi(\mathbf{x}) \\ \psi(\mathbf{x}) \end{pmatrix}, \tag{12.17}$$

where $S_0(s)$ is the last component of the field $\mathbf{J}_0(s)$. This is the Schrödinger equation with a source term if $\mathbf{\Lambda L}(\mathbf{x})\mathbf{\Lambda}$ has the form of the matrix entering (12.1). In a sense, this represents an extension of the density functional theory ideas to the Hilbert space setting. For example suppose

$$\mathbf{L}(\mathbf{x}) = \begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & E - a(\mathbf{x}) & g(\mathbf{x}) \\ 0 & \overline{g}(\mathbf{x}) & d(\mathbf{x}) \end{pmatrix}, \tag{12.18}$$

where $g(\mathbf{x})$ and $d(\mathbf{x})$ have no special symmetries. Here we assume

$$a(\mathbf{x}) = \sum_{j=1}^{M} N V_{\text{ne}}^{(j)}(\mathbf{x}_1) + \frac{N(N-1)}{2} U(\mathbf{x}_1, \mathbf{x}_2), \tag{12.19}$$

where

$$V_{\text{ne}}^{(j)}(\mathbf{x}_1) = \frac{-Z_j e^2}{|\mathbf{r}_1 - \mathbf{R}_j|} \tag{12.20}$$

[in which $\mathbf{x}_1 = (\mathbf{r}_1, s_1)$] represents the electrostatic interaction of electron 1 at $\mathbf{r}_1$ having spin s_1 and charge e with the nucleus at $\mathbf{R}_j$, having charge $Z_j e$, M is the number of nuclei and

$$U(\mathbf{x}_1, \mathbf{x}_2) = \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{12.21}$$

[in which $\mathbf{x}_1 = (\mathbf{r}_1, s_1)$ and $\mathbf{x}_1 = (\mathbf{r}_1, s_1)$] represents the electron–electron repulsion of an electron at $\mathbf{r}_1$ with spin s_1 and an electron at $\mathbf{r}_2$ with spin s_2. Then we have

$$\mathbf{\Lambda L \Lambda} = \begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & E - V(\mathbf{x}) & \Lambda_a g(\mathbf{x}) \Lambda_s \\ 0 & \Lambda_s \overline{g}(\mathbf{x}) \Lambda_a & \Lambda_s d(\mathbf{x}) \Lambda_s \end{pmatrix}, \tag{12.22}$$

where

$$V_{\text{ne}}(\mathbf{x}) = \sum_{i=1}^{N} \sum_{j=1}^{M} V_{\text{ne}}^{(j)}(\mathbf{x}_i) = \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{-Z_j e^2}{|\mathbf{r}_1 - \mathbf{R}_j|}, \tag{12.23}$$

and

$$V_{\text{ee}}(\mathbf{x}) = \sum_{i<j}^{N} \frac{e^2}{|\mathbf{x}_i - \mathbf{x}_j|} \tag{12.24}$$

are the total potentials of interaction of all electrons with the nuclei ($V_{\text{ne}}(\mathbf{x})$) and with other electrons ($V_{\text{ee}}(\mathbf{x})$).

Now in Fourier space the projection onto $\mathcal{U}$ is from (12.10)

$$\widehat{\mathbf{\Gamma}}_0(\mathbf{K}) = \begin{cases} 0 & \text{for } \mathbf{k} \neq 0, \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \text{for } \mathbf{k} = 0. \end{cases} \tag{12.25}$$

Here $\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_N)$, and each $\mathbf{K}_i = (\mathbf{k}_i, s_i)$ consists of a three-dimensional Fourier space vector $\mathbf{k}_i$ and a spin s_i. (In quantum mechanics $\mathbf{k}_i$ has the interpretation of the momentum of electron i, aside from a proportionality constant.) We also sometimes write $\mathbf{K} = (\mathbf{k}, s)$ where $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N)$ is the $3N$-dimensional Fourier vector and $s = (s_1, s_2, \ldots, s_N)$ is the spin configuration.

The fields in the space $\mathcal{E}$ in Fourier space look like

$$\widehat{\mathbf{E}}(\mathbf{K}) = \begin{pmatrix} i\mathbf{k}\widehat{\psi}(\mathbf{K}) \\ \widehat{\psi}(\mathbf{K}) \\ 0 \end{pmatrix} \quad \text{for } \mathbf{k} \neq 0, \quad \widehat{\mathbf{E}}(\mathbf{K}) = \begin{pmatrix} 0 \\ \langle \psi((\mathbf{r}, s)) \rangle_{\mathbf{r}} \\ 0 \end{pmatrix} \quad \text{for } \mathbf{k} = 0, \tag{12.26}$$

where $\langle \psi \rangle_{\mathbf{r}}$ denotes a volume average of $\psi(\mathbf{x}) = \psi((\mathbf{r}, s))$ as $\mathbf{r}$ ranges over Ω^N keeping the spin configuration $s = (s_1, s_2, \ldots, s_N)$ fixed. Here we have expressed $\mathbf{x}$ as the pair $(\mathbf{r}, s)$, where $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ denotes the spatial coordinates. In Fourier space the wavefunction $\widehat{\psi}$ is now an antisymmetric function of $\mathbf{K}$. That is $\widehat{\psi}(\mathbf{K})$ changes sign if we swap $\mathbf{K}_\ell = (\mathbf{k}_\ell, s_\ell)$ and $\mathbf{K}_i = (\mathbf{k}_i, s_i)$. Note that the antisymmetry of $\psi(\mathbf{K})$ ensures that the average over spin coordinates of $\langle \psi((\mathbf{r}, s)) \rangle_{\mathbf{r}}$ is zero. The projection operator onto $\mathcal{E}$ is expressed in Fourier space as:

$$\widehat{\mathbf{\Gamma}}_1 \quad = \quad \frac{\mathbf{\Lambda}}{k^2 + 1} \begin{pmatrix} \mathbf{k}\mathbf{k}^T & i\mathbf{k} & 0 \\ -i\mathbf{k}^T & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{\Lambda} \quad \text{for all } \mathbf{k} \text{ including } \mathbf{k} = 0. \tag{12.27}$$

Here $\mathbf{\Lambda}$ is understood to act on the fields to the right including the $\mathbf{k}$'s in the above expression, and we abbreviate $|\mathbf{k}|^2 = k^2$. Next look at

$$\widehat{\mathbf{\Gamma}}_2 = \mathbf{I} - \widehat{\mathbf{\Gamma}}_0 - \widehat{\mathbf{\Gamma}}_1. \tag{12.28}$$

If we define

$$\begin{pmatrix} \widehat{\mathbf{J}}_1(\mathbf{K}) \\ \widehat{J}_2(\mathbf{K}) \\ \widehat{J}_3(\mathbf{K}) \end{pmatrix} = \widehat{\mathbf{\Gamma}}_2 \begin{pmatrix} \widehat{\mathbf{P}}_1(\mathbf{K}) \\ \widehat{p}_2(\mathbf{K}) \\ \widehat{p}_3(\mathbf{K}) \end{pmatrix}, \tag{12.29}$$

where there are no special symmetries on $\widehat{\mathbf{P}}_1(\mathbf{K})$, $\widehat{p}_2(\mathbf{K})$, and $\widehat{p}_3(\mathbf{K})$, then for $\mathbf{k} \neq 0$,

$$\begin{aligned} \mathbf{\Lambda} \begin{pmatrix} \widehat{\mathbf{J}}_1(\mathbf{K}) \\ \widehat{J}_2(\mathbf{K}) \\ \widehat{J}_3(\mathbf{K}) \end{pmatrix} \quad &= \quad (\mathbf{\Lambda} - \mathbf{\Lambda}\widehat{\mathbf{\Gamma}}_1) \begin{pmatrix} \widehat{\mathbf{P}}_1(\mathbf{K}) \\ \widehat{p}_2(\mathbf{K}) \\ \widehat{p}_3(\mathbf{K}) \end{pmatrix} \\ &= \quad \begin{pmatrix} \mathbf{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K})) \\ \Lambda_a(\widehat{p}_2(\mathbf{K})) \\ \Lambda_s(\widehat{p}_3(\mathbf{K})) \end{pmatrix} - \begin{pmatrix} \frac{\mathbf{k}\mathbf{k}^T}{k^2+1}\mathbf{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K})) + \frac{i\mathbf{k}}{k^2+1}\Lambda_a\widehat{p}_2(\mathbf{K}) \\ \frac{-i\mathbf{k}^T}{k^2+1}\mathbf{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K})) + \frac{1}{k^2+1}\Lambda_a\widehat{p}_2(\mathbf{K}) \\ 0 \end{pmatrix} \\ &= \quad \begin{pmatrix} \widehat{\mathbf{q}}(\mathbf{K}) \\ i\mathbf{k} \cdot \widehat{\mathbf{q}}(\mathbf{K}) \\ i\mathbf{K} \cdot \mathbf{v}(\mathbf{K}) \end{pmatrix}, \end{aligned} \tag{12.30}$$

where $\mathbf{v}(\mathbf{K})$ is not unique, but chosen so that

$$i\mathbf{k} \cdot \mathbf{v}(\mathbf{K}) = \Lambda_s(\widehat{p}_3(\mathbf{K})), \tag{12.31}$$

and

$$\begin{aligned}\widehat{\mathbf{q}}(\mathbf{K}) &= \left(\mathbf{I}-\frac{\mathbf{k}\mathbf{k}^T}{k^2+1}\right)\boldsymbol{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K}))-\frac{i\mathbf{k}}{k^2+1}\Lambda_a(\widehat{p}_2(\mathbf{K})),\\ i\mathbf{k}\cdot\widehat{\mathbf{q}}(\mathbf{k}) &= i\left(\mathbf{k}^T-\frac{k^2\mathbf{k}^T}{k^2+1}\right)\boldsymbol{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K}))+\frac{k^2}{k^2+1}\Lambda_a(\widehat{p}_2(\mathbf{K}))\\ &= \frac{i\mathbf{k}^T}{k^2+1}\boldsymbol{\Lambda}_A(\widehat{\mathbf{P}}_1(\mathbf{K}))+\Lambda_a(\widehat{p}_2(\mathbf{K}))-\frac{1}{k^2+1}\Lambda_a(\widehat{p}_2(\mathbf{K})).\end{aligned} \tag{12.32}$$

Also for $\mathbf{k}=0$ we have

$$\begin{pmatrix}\widehat{\mathbf{J}}_1(\mathbf{K})\\ \widehat{J}_2(\mathbf{K})\\ \widehat{J}_3(\mathbf{K})\end{pmatrix}=\begin{pmatrix}\widehat{\mathbf{P}}_1((0,\mathbf{s}))\\ \widehat{p}_2((0,\mathbf{s}))\\ \widehat{p}_3((0,\mathbf{s}))\end{pmatrix}-\begin{pmatrix}0\\ 0\\ \widehat{p}_3((0,\mathbf{s}))\end{pmatrix}-\begin{pmatrix}0\\ \Lambda_a[\widehat{p}_2((0,\mathbf{s}))]\\ 0\end{pmatrix}=\begin{pmatrix}\widehat{\mathbf{P}}_1((0,\mathbf{s}))\\ (1-\Lambda_a)\widehat{p}_2((0,\mathbf{s}))\\ 0\end{pmatrix} \tag{12.33}$$

and this implies

$$\boldsymbol{\Lambda}\begin{pmatrix}\widehat{\mathbf{J}}_1(\mathbf{K})\\ \widehat{J}_2(\mathbf{K})\\ \widehat{J}_3(\mathbf{K})\end{pmatrix}=\begin{pmatrix}\boldsymbol{\Lambda}_A\widehat{\mathbf{P}}_1((0,\mathbf{s}))\\ 0\\ 0\end{pmatrix}, \tag{12.34}$$

when $\mathbf{k}=0$. So it is clear from (12.33) that if $\boldsymbol{\Gamma}_2$ is the projection onto $\mathcal{J}$, then in Fourier space this projection is

$$\widehat{\boldsymbol{\Gamma}}_2(\mathbf{K})=\begin{cases}\mathbf{I}-\widehat{\boldsymbol{\Gamma}}_1(\mathbf{k}) & \text{for } \mathbf{k}\neq 0,\\ \begin{pmatrix}\mathbf{I} & 0 & 0\\ 0 & (1-\Lambda_a) & 0\\ 0 & 0 & 0\end{pmatrix} & \text{for } \mathbf{k}=0.\end{cases} \tag{12.35}$$

Thus $\boldsymbol{\Gamma}_0+\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2=0$ and each projects onto orthogonal subspaces.

12.3 Simplifying the equation using a suitable reference medium

To condense formulas, let's relabel $\mathbf{J}_0(s)+\mathbf{J}(\mathbf{x})$ and $\mathbf{E}_0(s)+\mathbf{E}(\mathbf{x})$ as $\mathbf{J}(\mathbf{x})$ and $\mathbf{E}(\mathbf{x})$ respectively. We want to solve

$$\mathbf{J}(\mathbf{x})=\mathbf{L}(\mathbf{x})\mathbf{E}(\mathbf{x}), \tag{12.36}$$

where $\mathbf{J}\in\mathcal{U}\oplus\mathcal{J}$, $\mathbf{E}\in\mathcal{U}\oplus\mathcal{E}$ and

$$\mathbf{L}(\mathbf{x})=\begin{pmatrix}-\mathbf{A} & 0 & 0\\ 0 & a(\mathbf{x}_1,\mathbf{x}_2) & g(\mathbf{x}_1,\mathbf{x}_2)\\ 0 & \overline{g}(\mathbf{x}_1,\mathbf{x}_2) & d(\mathbf{x}_1,\mathbf{x}_2)\end{pmatrix}. \tag{12.37}$$

We assume $g(\mathbf{x}_1,\mathbf{x}_2)$ is antisymmetric in $\mathbf{x}_1,\mathbf{x}_2$ and $a(\mathbf{x}_1,\mathbf{x}_2), d(\mathbf{x}_1,\mathbf{x}_2)$ are symmetric in $\mathbf{x}_1,\mathbf{x}_2$ with no dependencies on $\mathbf{x}_3,\dots,\mathbf{x}_N$. We further assume a,d are real while g could be complex.

It is common in the theory of composites to aid the analysis to introduce a constant comparison medium, or reference medium;see, for example, Sections 9.6, 12.3, 12.4, 13.5, in Milton (2002). The Fast Fourier

Transform methods for computing fields in composites are based on series expansions about the reference medium. Let us take a reference medium

$$\mathbf{L}_0 = \begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & -t_1 & 0 \\ 0 & 0 & -t_2 \end{pmatrix}, \tag{12.38}$$

where $\mathbf{A}$ is positive definite and t_1 and t_2 are constants. We will take t_1 to be complex.

Introduce the operator $\boldsymbol{\Gamma}$ whose Fourier components for all $\mathbf{k}$, including $\mathbf{k} = 0$, are

$$\widehat{\boldsymbol{\Gamma}}(\mathbf{K}) = \boldsymbol{\Gamma}_1(\mathbf{K})[\boldsymbol{\Gamma}_1(\mathbf{K})\mathbf{L}_0\boldsymbol{\Gamma}_1(\mathbf{K})]^{-1}\boldsymbol{\Gamma}_1(\mathbf{K}). \tag{12.39}$$

Here the inverse is taken on the space $\mathcal{E}$ onto which $\boldsymbol{\Gamma}_1(\mathbf{K})$ projects. Then we have

$$\widehat{\boldsymbol{\Gamma}}(\mathbf{K}) = \frac{-1}{\mathbf{k}^T\mathbf{A}\mathbf{k} + t_1} \begin{pmatrix} \mathbf{k}\mathbf{k}^T & i\mathbf{k} & 0 \\ -i\mathbf{k}^T & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{\Lambda}. \tag{12.40}$$

Now define the polarization field

$$\mathbf{P}(\mathbf{x}) = (\mathbf{L} - \mathbf{L}_0)\mathbf{E}(\mathbf{x}) = \mathbf{J}(\mathbf{x}) - \mathbf{L}_0\mathbf{E}(\mathbf{x}). \tag{12.41}$$

From the definition of $\boldsymbol{\Gamma}$,

$$\boldsymbol{\Gamma}\mathbf{P} = \boldsymbol{\Gamma}\mathbf{J} + \boldsymbol{\Gamma}_0\mathbf{E} - \mathbf{E} = \boldsymbol{\Gamma}_0\mathbf{E} - \mathbf{E}, \tag{12.42}$$

since $\mathbf{J} = (\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_2)\mathbf{J}$ and $\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2 = 0$. This implies that

$$\mathbf{P} = -(\mathbf{L} - \mathbf{L}_0)\boldsymbol{\Gamma}\mathbf{P} + (\mathbf{L} - \mathbf{L}_0)\boldsymbol{\Gamma}_0\mathbf{E}. \tag{12.43}$$

Since

$$\delta\mathbf{L} \equiv \mathbf{L} - \mathbf{L}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a + t_1 & g \\ 0 & \overline{g} & d + t_2 \end{pmatrix}, \tag{12.44}$$

it is clear that $\mathbf{P}$ necessarily takes the form

$$\mathbf{P} = \begin{pmatrix} 0 \\ p_2(\mathbf{x}) \\ p_3(\mathbf{x}) \end{pmatrix}, \tag{12.45}$$

and the Fourier transforms of $\mathbf{P}$ and $(\mathbf{L} - \mathbf{L}_0)\boldsymbol{\Gamma}\mathbf{P}$ must take the same form. So

$$(\mathbf{L} - \mathbf{L}_0)\boldsymbol{\Gamma}\mathbf{P} = \begin{pmatrix} 0 \\ \widetilde{(\mathbf{L} - \mathbf{L}_0)\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{P}}} \end{pmatrix}, \tag{12.46}$$

where in Fourier space the action of $\widetilde{\boldsymbol{\Gamma}}$ on a field can be computed by multiplication of that field by

$$\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K}) = \frac{-1}{\mathbf{k}^T\mathbf{A}\mathbf{k} + t_1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{12.47}$$

while in real space, the action of $\widetilde{\delta\mathbf{L}} \equiv \widetilde{\mathbf{L}-\mathbf{L}_0}$ on a field can be computed by multiplication of that field by

$$\widetilde{\delta\mathbf{L}} \equiv \widetilde{\mathbf{L}-\mathbf{L}_0} = \begin{pmatrix} a(\mathbf{x}_1,\mathbf{x}_2)+t_1 & g(\mathbf{x}_1,\mathbf{x}_2) \\ \overline{g}(\mathbf{x}_1,\mathbf{x}_2) & d(\mathbf{x}_1,\mathbf{x}_2)+t_2 \end{pmatrix}, \tag{12.48}$$

and the action of $\widetilde{\mathbf{\Lambda}}$ in either Fourier or real space is given by

$$\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{P}} = \begin{pmatrix} \Lambda_a p_2 \\ \Lambda_s p_3 \end{pmatrix}. \tag{12.49}$$

Let $p_2(\mathbf{x}) = (t_1 - \nabla\cdot\mathbf{A}\nabla)\phi(\mathbf{x})$, then in Fourier space $\widehat{p}_2(\mathbf{K}) = (t_1 + \mathbf{k}^T\mathbf{A}\mathbf{k})\widehat{\phi}(\mathbf{K})$. So we have

$$(\mathbf{L}-\mathbf{L}_0)\mathbf{\Gamma}\mathbf{P} = -\begin{pmatrix} 0 \\ (a(\mathbf{x}_1,\mathbf{x}_2)+t_1)\Lambda_a\phi \\ \overline{g}(\mathbf{x}_1,\mathbf{x}_2)\Lambda_a\phi \end{pmatrix}, \tag{12.50}$$

and the equations become

$$\begin{pmatrix} 0 \\ (t_1 - \nabla\cdot\mathbf{A}\nabla)\phi(\mathbf{x}) \\ p_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ (t_1 + a(\mathbf{x}_1,\mathbf{x}_2))\Lambda_a(\phi) \\ \overline{g}(\mathbf{x})\Lambda_a\phi \end{pmatrix} + \begin{pmatrix} 0 \\ g(\mathbf{x}) \\ d+t_2 \end{pmatrix}\theta_0, \tag{12.51}$$

assuming $\mathbf{\Gamma}_0\mathbf{E} = (0,0,\theta_0)^T$. Note that for the identity $\Lambda_a\widehat{p}_2 = (t_1+\mathbf{k}^T\mathbf{A}\mathbf{k})\Lambda_a\phi$ to hold, we need to assume that $\mathbf{k}^T\mathbf{A}\mathbf{k}$ is completely symmetric with respect to interchange of $\mathbf{k}_i$ and $\mathbf{k}_j$. Assuming this, since $p_3(\mathbf{x})$ is not subject to any differential constraints, we are left with

$$(t_1 - \nabla\cdot\mathbf{A}\nabla)\phi(\mathbf{x}) = (t_1 + a(\mathbf{x}_1,\mathbf{x}_2))\Lambda_a\phi + g(\mathbf{x})\theta_0, \tag{12.52}$$

where $\phi(\mathbf{x})$ and $g(\mathbf{x})$ are not required to satisfy any symmetry properties.

12.4 Solving for the fields using Fast Fourier Transforms

We look for solutions with $\mathbf{L}(\mathbf{x}) = \mathbf{L}_0 + \epsilon\delta\mathbf{L}$, and expand in powers of ϵ. Following Moulinec and Suquet (1994, 1998), see also **Section 14.11** of Milton (2002), one has the approximants

$$\mathbf{E}^m = \sum_{j=0}^{m}\epsilon^j(-\mathbf{\Gamma}\delta\mathbf{L})^j\mathbf{E}_0, \quad \delta\mathbf{L} = (\mathbf{L}-\mathbf{L}_0) \text{ when } \epsilon = 1, \tag{12.53}$$

for the field $\mathbf{E}$, which satisfies

$$\mathbf{E}^{m+1} = \mathbf{E}_0 - \epsilon\mathbf{\Gamma}\delta\mathbf{L}\mathbf{E}^m. \tag{12.54}$$

So we take

$$\mathbf{E}_0 = \begin{pmatrix} 0 \\ 0 \\ \theta_0 \end{pmatrix}. \tag{12.55}$$

In Fourier space $\widehat{\mathbf{\Gamma}}(\mathbf{K})$ is given by (12.40) and $\delta\mathbf{L}$ is given by (12.44).

It suffices to look at $\widetilde{\mathbf{E}}_m = \mathbf{\Theta}\mathbf{E}_m$ where $\mathbf{\Theta}$ is the projection

$$\mathbf{\Theta} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12.56}$$

Since $\delta\mathbf{L} = \mathbf{\Theta}\delta\mathbf{L}\mathbf{\Theta}$, on multiplying (12.54) on the left by $\mathbf{\Theta}$ we have

$$\widetilde{\mathbf{E}}^{m+1} = \widetilde{\mathbf{E}}^0 - \epsilon\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}\widetilde{\delta\mathbf{L}}\widetilde{\mathbf{E}}^m, \quad \widetilde{\mathbf{E}}^0 = \begin{pmatrix} 0 \\ \theta_0 \end{pmatrix}, \tag{12.57}$$

where in Fourier space the action of $\widetilde{\mathbf{\Lambda}}$ on a field can be easily computed by multiplying it by $\widehat{\widetilde{\mathbf{\Gamma}}}(\mathbf{K})$ given by (12.47). Note that because of the structure of $\widehat{\widetilde{\mathbf{\Gamma}}}(\mathbf{K})$, and the fact that it commutes with $\widetilde{\mathbf{\Lambda}}$, $\widetilde{\mathbf{E}}^m$ necessarily has a Fourier transform of the form

$$\widehat{\widetilde{\mathbf{E}}}^m(\mathbf{K}) = \begin{pmatrix} \widehat{\psi}^m(\mathbf{K}) \\ 0 \end{pmatrix}, \tag{12.58}$$

where $\widehat{\psi}^m(\mathbf{K})$ can be identified with a m-th order approximation to the Fourier transform of the wavefunction $\psi(\mathbf{x})$. The function $\widehat{\psi}^m(\mathbf{K})$ is completely antisymmetric in the sense that it changes sign if we swap $\mathbf{K}_j = (\mathbf{k}_j, s_j)$ with $\mathbf{K}_\ell = (\mathbf{k}_\ell, s_\ell)$ (where $\mathbf{k}_j$ and $\mathbf{k}_\ell$ are the 3-dimensional Fourier vectors, while s_j and s_ℓ are the associated spins).

Now using Fast Fourier Transforms, we compute

$$\mathcal{F}\delta\mathbf{L}\mathcal{F}^{-1}\widehat{\widetilde{\mathbf{E}}}^m(\mathbf{K}) \equiv \begin{pmatrix} \widetilde{p}_2(\mathbf{K}) \\ \widetilde{p}_3(\mathbf{K}) \end{pmatrix}, \tag{12.59}$$

where $\mathcal{F}^{-1}$ is an inverse Fourier transform $\mathcal{F}^{-1}$ on the variables $\mathbf{k}_1$ and $\mathbf{k}_2$ only, leaving $\mathbf{k}_3, \ldots, \mathbf{k}_N$ unchanged, while $\mathcal{F}$ is a Fourier transform on the variables $\mathbf{r}_1$ and $\mathbf{r}_2$ only, leaving $\mathbf{k}_3, \ldots, \mathbf{k}_N$ unchanged. There is no need to compute $\widetilde{p}_3(\mathbf{K})$ since it gets annihilated in the next step, which is the computation of

$$\widehat{\widetilde{\mathbf{\Gamma}}}(\mathbf{K}) \begin{pmatrix} \widetilde{p}_2(\mathbf{K}) \\ \widetilde{p}_3(\mathbf{K}) \end{pmatrix} = \begin{pmatrix} -\widehat{Q}(\mathbf{K}) \\ 0 \end{pmatrix}, \tag{12.60}$$

where $\widehat{Q}(\mathbf{K}) = \widetilde{p}_2(\mathbf{K})/(\mathbf{k}^T\mathbf{A}\mathbf{k} + t_1)$ is antisymmetric with respect to swapping $\mathbf{K}_1, \mathbf{K}_2$ and antisymmetric with respect to swapping $\mathbf{K}_3, \ldots, \mathbf{K}_N$. (Note that $\widehat{\widetilde{\mathbf{\Gamma}}}(\mathbf{K})$ has been defined so its computation, unlike the computation of $\widehat{\mathbf{\Gamma}}(\mathbf{K})$, does not involve any symmetrization.)

Finally because of the symmetry properties of $\widehat{Q}(\mathbf{K})$ we have

$$\begin{aligned}\Lambda_a\widehat{Q}(\mathbf{K}_1,\mathbf{K}_2;\mathbf{K}_3,\dots,\mathbf{K}_N) &= \frac{2}{N(N-1)}\Bigg[\widehat{Q}(\mathbf{K}_1,\mathbf{K}_2;\mathbf{K}_3,\dots,\mathbf{K}_N)\\ &\quad+\sum_{i=3}^{N}(-1)^{i+1}\widehat{Q}(\mathbf{K}_2,\mathbf{K}_i;\mathbf{K}_1,\mathbf{K}_3,\mathbf{K}_4,\dots,\mathbf{K}_{i-1},\mathbf{K}_{i+1},\dots,\mathbf{K}_N)\\ &\quad-\sum_{i=3}^{N}(-1)^{i+1}\widehat{Q}(\mathbf{K}_1,\mathbf{K}_i;\mathbf{K}_2,\mathbf{K}_3,\mathbf{K}_4,\dots,\mathbf{K}_{i-1},\mathbf{K}_{i+1},\dots,\mathbf{K}_N)\\ &\quad+\sum_{i=3}^{N}\sum_{\ell=i+1}^{N}(-1)^{i+\ell+1}\widehat{Q}\Big(\mathbf{K}_i,\mathbf{K}_\ell;\mathbf{K}_1,\mathbf{K}_2,\dots\\ &\qquad\dots,\mathbf{K}_{i-1},\mathbf{K}_{i+1}\dots,\mathbf{K}_{\ell-1},\mathbf{K}_{\ell+1},\dots,\mathbf{K}_N\Big)\Bigg],\end{aligned}\tag{12.61}$$

and so we obtain the next iterate for the field

$$\widehat{\widetilde{\mathbf{E}}}^{m+1}(\mathbf{K})=\begin{pmatrix}\Lambda_a\widehat{Q}(\mathbf{K})\\0\end{pmatrix}\equiv\begin{pmatrix}\widehat{\psi}^{m+1}(\mathbf{K})\\0\end{pmatrix}.\tag{12.62}$$

12.5 Convergence of the series

Let $\boldsymbol{\Xi}$ denote the projection

$$\boldsymbol{\Xi}=\begin{pmatrix}1&0\\0&0\end{pmatrix}.\tag{12.63}$$

Then clearly $\boldsymbol{\Xi}\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})=\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})=\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})\boldsymbol{\Xi}$, and hence $\boldsymbol{\Xi}\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})=\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})=\widehat{\widetilde{\boldsymbol{\Gamma}}}(\mathbf{K})\boldsymbol{\Xi}$. Now we have

$$\begin{aligned}\widetilde{\mathbf{E}}^1 &= \widetilde{\mathbf{E}}^0-\epsilon\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0,\\ \widetilde{\mathbf{E}}^2 &= \widetilde{\mathbf{E}}^0-\epsilon\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0+\epsilon^2\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0,\\ &= \widetilde{\mathbf{E}}^0-\epsilon\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0+\epsilon^2(\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Xi}\widetilde{\delta}\widetilde{\mathbf{L}}\boldsymbol{\Xi})\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0,\\ \widetilde{\mathbf{E}}^3 &= \cdots-\epsilon^3\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0,\\ &= \cdots-\epsilon^3(\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\boldsymbol{\Xi}\widetilde{\delta}\widetilde{\mathbf{L}}\boldsymbol{\Xi})^2\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\delta}\widetilde{\mathbf{L}}\widetilde{\mathbf{E}}^0.\end{aligned}\tag{12.64}$$

To say something about the convergence of this series, we need to introduce a norm on the space of relevant fields. For fields

$$\mathbf{P}(\mathbf{x})=\begin{pmatrix}p_2(\mathbf{x})\\p_3(\mathbf{x})\end{pmatrix},\quad \mathbf{P}'(\mathbf{x})=\begin{pmatrix}p_2'(\mathbf{x})\\p_3'(\mathbf{x})\end{pmatrix},\tag{12.65}$$

where $p_2(\mathbf{x})$, $p_3(\mathbf{x})$, $p_2'(\mathbf{x})$, and $p_3'(\mathbf{x})$ are Ω^N-periodic in each of the variables $\mathbf{r}_j$, $j=1,2,\dots,N$, define the inner product

$$(\mathbf{P},\mathbf{P}')=\langle\overline{\mathbf{P}}\cdot\mathbf{P}'\rangle=\langle\overline{p}_2(\mathbf{x})p_2'(\mathbf{x})+\overline{p}_3(\mathbf{x})p_3'(\mathbf{x})\rangle,\tag{12.66}$$

in which $\langle\cdot\rangle$ denotes the volume and spin average,

$$\langle g(\mathbf{x})\rangle = \frac{1}{2^N|\Omega|^N}\sum_s \int_{\Omega^N} g(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)d\mathbf{r}_1 d\mathbf{r}_2,\ldots d\mathbf{r}_N, \tag{12.67}$$

$|\Omega|$ being the volume of the unit cell Ω, and the sum is over all 2^N spin configurations $s = (s_1, s_2, \ldots, s_N)$ as each s_j takes values $+1/2$ or $-1/2$. Then it is natural to define the norm $|\mathbf{P}|$ of a field $\mathbf{P}$ and the norm $\|\mathbf{A}\|$ of a linear operator $\mathbf{A}$ via

$$|\mathbf{P}| = (\mathbf{P},\mathbf{P})^{1/2}. \tag{12.68}$$

We consider the Hilbert space $\mathcal{H}$ of fields $\mathbf{P}$ of the form (12.65) having finite norm, with inner product (12.66). On $\mathcal{H}$, the norm $\|\mathbf{A}\|$ of a linear operator $\mathbf{A}$ is naturally defined to be

$$\|\mathbf{A}\| = \sup_{\substack{\mathbf{P}\in\mathcal{H}\\ |\mathbf{P}|=1}} |\mathbf{A}\mathbf{P}|. \tag{12.69}$$

From this definition of the norm of an operator it easily follows that for any linear operators $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$ and $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\|+\|\mathbf{B}\|$. From the completeness of a Hilbert space, the series (12.64) will converge if

$$\epsilon\|\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}\mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}\| < 1, \tag{12.70}$$

which guarantees that $\widetilde{\mathbf{E}}^j$, $j = 1, 2, \ldots$, is a Cauchy sequence.

Now let's suppose t_1 is purely imaginary, i.e., $t_1 = -ic$ for some $c \in \mathbb{R}$. We have

$$\|\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}\mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}\| \leq \|\widetilde{\mathbf{\Gamma}}\|\|\widetilde{\mathbf{\Lambda}}\|\|\mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}\|, \tag{12.71}$$

and $\|\widetilde{\mathbf{\Lambda}}\| = 1$ since $\widetilde{\mathbf{\Lambda}}$ is a projection (projecting onto the subspace of fields with the desired symmetries),

$$\begin{aligned} \|\widehat{\widetilde{\mathbf{\Gamma}}}(\mathbf{k})\| &\leq \frac{1}{c}, \\ \|\mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}\| &\leq \max_{\mathbf{x}_1,\mathbf{x}_2} |a(\mathbf{x}_1,\mathbf{x}_2) - ic|. \end{aligned} \tag{12.72}$$

Hence the series (12.64) will converge if

$$r_\epsilon \equiv \epsilon \max_{\mathbf{x}_1,\mathbf{x}_2} \left| \frac{a(\mathbf{x}_1,\mathbf{x}_2)}{c} - i \right| < 1. \tag{12.73}$$

For fixed ϵ it will converge when c is sufficiently large if $a(\mathbf{x}_1,\mathbf{x}_2)$ is bounded, and the value of r_ϵ determines an upper bound on the rate of convergence. If $\epsilon = 1$ it will converge provided $a(\mathbf{x}_1,\mathbf{x}_2) \leq \beta$ and $\mathrm{Im}(a(\mathbf{x}_1,\mathbf{x}_2)) \geq \delta$ for all $\mathbf{x}_1,\mathbf{x}_2$, for some $\beta > 0$ and $\delta > 0$. This is ensured if the energy E has a small positive imaginary part. We should choose c to be the value for which the minimum

$$r_1 = \min_c \max_{\mathbf{x}_1,\mathbf{x}_2} \left| \frac{a(\mathbf{x}_1,\mathbf{x}_2)}{c} - i \right| \tag{12.74}$$

is attained. There are other ways of utilizing the series expansion information too, such as the use of Padé approximants.

12.6 An alternate series expansion with typically faster convergence

Here we use the accelerated scheme of Eyre and Milton (1999) as generalized in Section 14.9 of Milton (2002). From equation (12.43) we have

$$\widetilde{\mathbf{P}} = -\widetilde{\delta\mathbf{L}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{P}} + \widetilde{\delta\mathbf{L}}\widetilde{\mathbf{E}}^0, \tag{12.75}$$

i.e.,

$$\begin{aligned}\widetilde{\mathbf{P}} &= [\mathbf{I} + \widetilde{\delta\mathbf{L}}\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}]^{-1}\widetilde{\delta\mathbf{L}}\widetilde{\mathbf{E}}^0 \\ &= [\mathbf{I} + \widetilde{\delta\mathbf{L}}\mathbf{M} + \widetilde{\delta\mathbf{L}}(\widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}} - \mathbf{M})]^{-1}\widetilde{\delta\mathbf{L}}\widetilde{\mathbf{E}}^0 \\ &= [\mathbf{I} - \mathbf{K}\mathbf{\Upsilon}]^{-1}\mathbf{K}\widetilde{\mathbf{E}}^0_0,\end{aligned} \tag{12.76}$$

where

$$\mathbf{K} = [\mathbf{I} + \widetilde{\delta\mathbf{L}}\mathbf{M}]^{-1}\widetilde{\delta\mathbf{L}}, \quad \mathbf{\Upsilon} = \mathbf{M} - \widetilde{\mathbf{\Gamma}}\widetilde{\mathbf{\Lambda}}. \tag{12.77}$$

Expanding $[\mathbf{I} - \mathbf{K}\mathbf{\Upsilon}]^{-1}$ in powers of $\mathbf{K}\mathbf{\Upsilon}$ gives the expansion

$$\widetilde{\mathbf{P}} = \sum_{j=0}^{\infty} \mathbf{K}(\mathbf{\Upsilon}\mathbf{K})^j\widetilde{\mathbf{E}}^0. \tag{12.78}$$

Let's take

$$\mathbf{M} = \alpha\mathbf{\Xi} = \begin{pmatrix}\alpha & 0\\ 0 & 0\end{pmatrix}. \tag{12.79}$$

From the definition of $\mathbf{K}$:

$$[\mathbf{I} + \widetilde{\delta\mathbf{L}}\mathbf{M}]\mathbf{K} = \widetilde{\delta\mathbf{L}}. \tag{12.80}$$

Multiplying on the left and right by $\mathbf{\Xi}$ we find

$$[\mathbf{\Xi} + \alpha\mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}]\mathbf{\Xi}\mathbf{K}\mathbf{\Xi} = \mathbf{\Xi}\widetilde{\delta\mathbf{L}}\mathbf{\Xi}, \tag{12.81}$$

i.e.,

$$\mathbf{\Xi}\mathbf{K}\mathbf{\Xi} = \frac{a + t_1}{1 + \alpha(a + t_1)}\mathbf{\Xi}. \tag{12.82}$$

Also, in Fourier space we have

$$\mathbf{\Upsilon} = \left[\alpha + \frac{1}{\mathbf{k}^T\mathbf{A}\mathbf{k} + t_1}\right]\Lambda_a\mathbf{\Xi}, \tag{12.83}$$

where Λ_a is understood to act component-wise. Hence we have

$$\mathbf{\Xi}\widetilde{\mathbf{P}} = \sum_{j=0}^{\infty}(\mathbf{\Xi}\mathbf{K}\mathbf{\Xi}\mathbf{\Upsilon})^j\mathbf{\Xi}\mathbf{K}\widetilde{\mathbf{E}}^0 \tag{12.84}$$

and this will converge if

$$\|\mathbf{\Xi}\mathbf{K}\mathbf{\Xi}\mathbf{\Upsilon}\| < 1. \tag{12.85}$$

We have

$$\|\mathbf{\Xi}\mathbf{K}\mathbf{\Xi}\mathbf{\Upsilon}\| \leq \|\mathbf{\Xi}\mathbf{K}\mathbf{\Xi}\|\|\mathbf{\Upsilon}\|. \tag{12.86}$$

We can choose t_1 and α to minimize this product of norms. Suppose t_1 is purely imaginary, i.e., $t_1 = -ic$ for some $c \in \mathbb{R}$, and $\alpha = -\frac{1}{2t_1}$. Then we have

$$\begin{aligned} \|\boldsymbol{\Upsilon}\| &= \frac{1}{2c}, \\ \boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi} &= \frac{a+t_1}{\frac{1}{2}-\frac{a}{2t_1}} = \frac{2(a+t_1)t_1}{t_1-a}, \end{aligned} \tag{12.87}$$

and so

$$\|\boldsymbol{\Upsilon}\|\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi} = \frac{a+t_1}{t_1-a}\frac{t_1}{|t_1|} = \frac{a-ic}{a+ic}i, \tag{12.88}$$

and

$$\|\boldsymbol{\Upsilon}\|\|\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}\| \leq \max_{\mathbf{x}_1,\mathbf{x}_2}\left|\frac{a(\mathbf{x}_1,\mathbf{x}_2)-ic}{a(\mathbf{x}_1,\mathbf{x}_2)+ic}\right|. \tag{12.89}$$

Convergence occurs if

$$|a(\mathbf{x}_1,\mathbf{x}_2)-ic| < \kappa|a(\mathbf{x}_1,\mathbf{x}_2)+ic|, \tag{12.90}$$

for all $\mathbf{x}_1, \mathbf{x}_2$, and for some constant $\kappa \leq 1$. Roughly speaking, convergence occurs when the imaginary part of $a(\mathbf{x}_1, \mathbf{x}_2)$ is positive. We should choose c to be the value which minimizes

$$r_2 = \min_c \max_{\mathbf{x}_1,\mathbf{x}_2}\left|\frac{a(\mathbf{x}_1,\mathbf{x}_2)-ic}{a(\mathbf{x}_1,\mathbf{x}_2)+ic}\right|. \tag{12.91}$$

Note that

$$\left|\frac{a(\mathbf{x}_1,\mathbf{x}_2)-ic}{a(\mathbf{x}_1,\mathbf{x}_2)+ic}\right| = \frac{|a(\mathbf{x}_1,\mathbf{x}_2)/c-i|}{|a(\mathbf{x}_1,\mathbf{x}_2)/c+i|}, \tag{12.92}$$

and if $a(\mathbf{x}_1, \mathbf{x}_2)$ has nonnegative imaginary part

$$|a(\mathbf{x}_1,\mathbf{x}_2)/c+i| > 1. \tag{12.93}$$

Hence $r_2 < r_1$. Thus we expect that generally the convergence using this alternative series expansion would be much faster than the original one, though this is not guaranteed as r_1 and r_2 only give bounds on the rate of convergence, and may not properly reflect the actual convergence rate. This new numerical scheme corresponds to that which Eyre and Milton (1999) developed for conducting composites.

12.7 Further improvements to enhance the rate of convergence

More generally in a periodic material

$$\min_{\mathbf{k}} \mathbf{k}^T\mathbf{A}\mathbf{k} = c_0, \tag{12.94}$$

is a nonzero positive constant. So the bound on the norm of $\boldsymbol{\Upsilon}$ is

$$f_\Upsilon(\alpha, t_1) = \max_{\substack{c \text{ real} \\ c \geq c_0}}\left|\alpha + \frac{1}{c+t_1}\right|, \tag{12.95}$$

where α and t_1 could be complex. We then have

$$\|\boldsymbol{\Upsilon}\| \leq f_{\Upsilon}(\alpha, t_1). \tag{12.96}$$

The bound on the norm of $\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}$, as implied by (12.82), is

$$f_K(\alpha, t_1) = \max_{\mathbf{x}_1, \mathbf{x}_2} \left| \frac{a(\mathbf{x}_1, \mathbf{x}_2) + t_1}{1 + \alpha(a(\mathbf{x}_1, \mathbf{x}_2) + t_1)} \right|, \tag{12.97}$$

and so

$$\|\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}\| \leq f_K(\alpha, t_1). \tag{12.98}$$

Finally a good choice of the complex numbers α and t_1 should be those that attain the infimum (or close to it) in

$$\inf_{\alpha} \inf_{t_1} f_{\Upsilon}(\alpha, t_1) f_K(\alpha, t_1). \tag{12.99}$$

Finally having converged to some $\boldsymbol{\Xi}\widetilde{\mathbf{P}} = \begin{pmatrix} \widetilde{p}_2 \\ 0 \end{pmatrix}$ we have

$$\widetilde{p}_2 = (a(\mathbf{x}_1, \mathbf{x}_2) + t_1)\psi + g(\mathbf{x}_1, \mathbf{x}_2), \quad \text{i.e., } \psi = \frac{\widetilde{p}_2 - g(\mathbf{x}_1, \mathbf{x}_2)}{a(\mathbf{x}_1, \mathbf{x}_2) + t_1}. \tag{12.100}$$

The expression above does not make the antisymmetry of ψ apparent, so a better approach may be to note that from (12.42), $\boldsymbol{\Gamma}\mathbf{P} = \boldsymbol{\Gamma}_0\mathbf{E} - \mathbf{E}$ and the fact that $\mathbf{P}(\mathbf{x})$ takes the form (12.45), we get in Fourier space

$$\frac{-\Lambda_a \widetilde{p}_2(\mathbf{K})}{\mathbf{k}^T\mathbf{A}\mathbf{k} + t_1} = -\psi(\mathbf{K}), \tag{12.101}$$

and then we can transform $\psi(\mathbf{K})$ back to real space.

Finally if we define the approximants

$$\widetilde{\mathbf{P}}^m = \sum_{j=0}^{m} \mathbf{K}(\boldsymbol{\Upsilon}\mathbf{K})^j \widetilde{\mathbf{E}}_0, \tag{12.102}$$

we have

$$\widetilde{\mathbf{P}}^{m+1} = \mathbf{K}\widetilde{\mathbf{E}}_0 + \mathbf{K}\boldsymbol{\Upsilon}\widetilde{\mathbf{P}}^m, \quad \widetilde{\mathbf{P}}^0 = \mathbf{K}\widetilde{\mathbf{E}}_0. \tag{12.103}$$

Hence for $m \geq 0$,

$$\boldsymbol{\Xi}\widetilde{\mathbf{P}}^{m+1} = \boldsymbol{\Xi}\mathbf{K}\widetilde{\mathbf{E}}_0 + \boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}\boldsymbol{\Upsilon}\boldsymbol{\Xi}\widetilde{\mathbf{P}}^m, \tag{12.104}$$

where $\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}$ is given by (12.81) and only depends on $\mathbf{x}_1, \mathbf{x}_2$. We work primarily in Fourier space, so to evaluate $\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}\boldsymbol{\Upsilon}\boldsymbol{\Xi}\widetilde{\mathbf{P}}^m$ we take

$$\mathcal{F}\boldsymbol{\Xi}\mathbf{K}\boldsymbol{\Xi}\mathcal{F}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\Xi}\widetilde{\mathbf{P}}^m, \tag{12.105}$$

where $\mathcal{F}^{-1}$ is the Fourier transform with respect to the variables $\mathbf{k}_1$ and $\mathbf{k}_2$ only while $\mathcal{F}$ is the Fourier transform with respect to the variables $\mathbf{r}_1, \mathbf{r}_2$ only. Also note we need only keep track of $\boldsymbol{\Xi}\widetilde{\mathbf{P}}^m$, rather than the full field $\widetilde{\mathbf{P}}^m$.

We also remark that our choice of starting field need not be $\mathbf{K}\widetilde{\mathbf{E}}_0$ but instead it is probably better to choose an approximation which one believes will be closer to the actual field. Additionally, it may be the case that the acceleration technique developed in **Chapter 8** is applicable and leads to an improvement in the rate of convergence.

12.8 Proof of the antisymmetrizing action of the projection Λ_a

Here we give the proof that Λ_a defined by (12.61) defines a projection onto fully antisymmetric functions given that $\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N)$ is antisymmetric in $\mathbf{K}_1, \mathbf{K}_2$, and in $\mathbf{K}_3, \ldots, \mathbf{K}_N$.

The first observation is that it is antisymmetric with respect to interchange of $\mathbf{K}_1, \mathbf{K}_2$. To check for antisymmetry with respect to $\mathbf{K}_3, \mathbf{K}_4$, note that the only terms where these enter the first pair are

$$\begin{aligned}
&\widehat{Q}(\mathbf{K}_2, \mathbf{K}_3; \mathbf{K}_1, \mathbf{K}_4, \mathbf{K}_5, \ldots) - \widehat{Q}(\mathbf{K}_2, \mathbf{K}_4; \mathbf{K}_1, \mathbf{K}_3, \mathbf{K}_5, \ldots) \\
&-\widehat{Q}(\mathbf{K}_1, \mathbf{K}_3; \mathbf{K}_2, \mathbf{K}_4, \mathbf{K}_5, \ldots) + \widehat{Q}(\mathbf{K}_1, \mathbf{K}_4; \mathbf{K}_2, \mathbf{K}_3, \mathbf{K}_5, \ldots) \\
&+\widehat{Q}(\mathbf{K}_3, \mathbf{K}_4; \mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_5, \ldots, \mathbf{K}_N) \\
&+\sum_{\ell=5}^{N} (-1)^{\ell} \widehat{Q}(\mathbf{K}_3, \mathbf{K}_\ell; \mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_4, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N) \\
&-\sum_{\ell=5}^{N} (-1)^{\ell} \widehat{Q}(\mathbf{K}_4, \mathbf{K}_\ell; \mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_3, \mathbf{K}_5, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N),
\end{aligned} \tag{12.106}$$

and this is clearly antisymmetric with respect to interchange of $\mathbf{K}_3$ and $\mathbf{K}_4$. Also note because of the antisymmetry of $\widehat{Q}$ with respect to $\mathbf{K}_3, \ldots, \mathbf{K}_N$, this immediately implies the function on the right of (12.61) is antisymmetric with respect to any $\mathbf{K}_j$ and $\mathbf{K}_\ell$ for $j \neq \ell, j, \ell \geq 4$.

To check for antisymmetry with respect to $\mathbf{K}_1, \mathbf{K}_3$, note that the only terms where $\mathbf{K}_1$ and $\mathbf{K}_3$ enter the first pair are

$$\begin{aligned}
&\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_N) + \widehat{Q}(\mathbf{K}_2, \mathbf{K}_3; \mathbf{K}_1, \mathbf{K}_4, \ldots, \mathbf{K}_N) \\
&-\widehat{Q}(\mathbf{K}_1, \mathbf{K}_3; \mathbf{K}_2, \mathbf{K}_4, \ldots, \mathbf{K}_N) \\
&-\sum_{i=4}^{N} (-1)^{i+1} \widehat{Q}(\mathbf{K}_1, \mathbf{K}_i; \mathbf{K}_2, \mathbf{K}_3, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N) \\
&+\sum_{\ell=4}^{N} (-1)^{\ell} \widehat{Q}(\mathbf{K}_3, \mathbf{K}_\ell; \mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N),
\end{aligned} \tag{12.107}$$

and this is clearly antisymmetric with respect to interchange of $\mathbf{K}_1$ and $\mathbf{K}_3$. Hence the expression is invariant with respect to interchange of either $\mathbf{K}_1$ or $\mathbf{K}_2$ with any of $\mathbf{K}_3, \ldots, \mathbf{K}_4$.

Finally suppose $\widehat{Q}$ was already fully antisymmetric. To go from

$$\widehat{Q}(\mathbf{K}_2, \mathbf{K}_i; \mathbf{K}_1, \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N) \tag{12.108}$$

to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \mathbf{K}_4, \mathbf{K}_4, \ldots, \mathbf{K}_N), \tag{12.109}$$

we first go to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_i, \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N), \tag{12.110}$$

which requires 2 swaps, then a further $i - 3$ swaps to get to the desired ordering. So

$$\widehat{Q}(\mathbf{K}_2, \mathbf{K}_i; \mathbf{K}_1, \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N) = (-1)^{i-1} \widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N). \tag{12.111}$$

To go from

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_i; \mathbf{K}_2, \mathbf{K}_3, \ldots, \mathbf{K}_N) \tag{12.112}$$

to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \mathbf{K}_3, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_i, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N), \tag{12.113}$$

we first go to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_i, \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_N), \tag{12.114}$$

with one swap. Then it requires a further $i-3$ swaps to reach the desired ordering. So

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_i; \mathbf{K}_2, \mathbf{K}_3, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_N) = (-1)^{i-2}\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N). \tag{12.115}$$

To go from

$$\widehat{Q}(\mathbf{K}_i, \mathbf{K}_\ell; \mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N) \tag{12.116}$$

to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N), \tag{12.117}$$

we first go to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_i, \mathbf{K}_\ell, \mathbf{K}_3, \mathbf{K}_4, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N), \tag{12.118}$$

which takes 2 swaps, then to

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_i, \mathbf{K}_3, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_\ell, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N), \tag{12.119}$$

which takes $\ell - 4$ swaps, and finally another $i-3$ swaps to get the correct ordering. So

$$\widehat{Q}(\mathbf{K}_i, \mathbf{K}_\ell; \mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_{i-1}, \mathbf{K}_{i+1}, \ldots, \mathbf{K}_{\ell-1}, \mathbf{K}_{\ell+1}, \ldots, \mathbf{K}_N) = (-1)^{i+\ell-5}\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N). \tag{12.120}$$

The upshot is that every term in the sum equals $\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N)$ as the minus signs cancel out. The number of terms in the sum is

$$1 + (N-2) + (N-2) + \frac{(N-2)(N-3)}{2} = \frac{N^2}{2} + 2N - \frac{5N}{2} = \frac{N(N-1)}{2}. \tag{12.121}$$

So the right hand side of (12.61) is simply

$$\widehat{Q}(\mathbf{K}_1, \mathbf{K}_2; \mathbf{K}_3, \ldots, \mathbf{K}_N). \tag{12.122}$$

This establishes that Λ_a projects onto completely antisymmetric functions.

13 Variational Principles for the Schrödinger equation when the energy is complex, and associated Q_C^*-convex functions.

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

Minimization variational principles for the time-harmonic acoustic equation in a body containing lossy materials have been obtained by Milton, Seppecher, and Bouchitté (2009) and Milton and Willis (2010), generalizing the quasistatic minimization variational principles of Cherkaev and Gibiansky (1994). As the time-harmonic acoustic equation is fundamentally the same as the multielectron time-harmonic Schrödinger equation, these variational principles naturally extend to the multielectron Schrödinger equation. We allow for sources, and find that due to the form of the Schrödinger equation the minimization variational principles simplify. They involve minimizing a functional of the real part of the wavefunction, subject to appropriate boundary conditions. We also find a variety of quadratic functions of the wavefunction, and of the associated current fields, that are Q_C^*-convex. These may prove to be important in deriving bounds, and for accelerating Fast Fourier Transform methods for calculating the wavefunction.

13.1 Introduction

In this chapter we obtain variational principles for the time-harmonic multielectron Schrödinger equation with a source term $h(\mathbf{x})\theta_0$,

$$E\psi(\mathbf{x}) = -\nabla \cdot \mathbf{A}\nabla\psi(\mathbf{x}) + V(\mathbf{x})\psi(\mathbf{x}) - h(\mathbf{x})\theta_0, \tag{13.1}$$

when the energy E is complex, where θ_0 scales the magnitude of the source term. Physically, an energy with positive imaginary part E'', and real part E' corresponds to a wavefunction which increases exponentially with time, since the wavefunction has a time-dependence $e^{-iEt/\hbar} = e^{E''t/\hbar}e^{-iE't/\hbar}$. Here $V(\mathbf{x})$ is the potential, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ where each $\mathbf{x}_i = (\mathbf{r}_i, s_i)$ represents a combination of a spatial coordinate $\mathbf{r}_i$ and a spin

coordinate s_i, and in the simplest approximation $\mathbf{A} = \hbar^2\mathbf{I}/(2m)$, where m is the mass of the electron, but which may take other forms to take into account the reduced mass of the electron, or mass polarization terms due to the motion of the atomic nuclei. We find it helpful to introduce the real-valued source term $h(\mathbf{x})\theta_0 \in \mathbb{R}$ which may be used to excite the electrons. Here we use

$$\nabla = (\nabla_1, \nabla_2, \dots, \nabla_N), \quad \text{where} \quad \nabla_j = \left(\frac{\partial}{\partial r_1^{(j)}}, \frac{\partial}{\partial r_2^{(j)}}, \frac{\partial}{\partial r_3^{(j)}} \right), \tag{13.2}$$

where N is the number of electrons. The source term $h(\mathbf{x}) \in \mathbb{R}$, like the wavefunction $\psi(\mathbf{x})$, is antisymmetric with respect to interchange of any pair of $\mathbf{x}_i$ and $\mathbf{x}_k$, while $V(\mathbf{x})$ is assumed to be symmetric with respect to such interchanges.

Throughout this chapter we let a' and a'' denote the real and imaginary parts of a quantity $a = a' + ia''$. We assume the energy E has a positive imaginary part denoted E''. The source term $h(\mathbf{x})$ excites the system, while E'' damps the response. When E'' is small and one is near a resonance (eigenvalue) associated with E' then the response ψ, should grow in proportion to the eigenfunction as $E'' \to 0$.

Minimization variational principles have been obtained for the time-harmonic acoustic equation (Milton, Seppecher, and Bouchitté 2009; Milton and Willis 2010) based on extensions of variational principles Cherkaev and Gibiansky (1994) obtained for quasistatics. The key to this generalization was the recognition that the acoustic equations could be written in the canonical form (3.7). The acoustic minimization variational principles have been used by Richins and Dobson (2012) to obtain numerical minimization schemes based on simple iterative descent methods with preconditioning. Since the time-harmonic Schrödinger equation is virtually identical to the time-harmonic acoustic equation with a constant density those variational principles immediately also apply to the Schrödinger equation. However some simplification can be made due to the structure of the matrix entering the constitutive law of the Schrödinger equation.

Our analysis applies to isolated systems confined within a body Ω, or to periodic systems with Ω taken as the unit cell of periodicity. Note that if the material we wish to simulate is periodic, with a unit cell Ω_0, then we should take Ω to contain many copies of Ω_0 and not just one: for example if Ω_0 was a cube of side length ℓ, then we could take Ω to be a cube of side length $m\ell$ where m is large. Again, as emphasized in the previous two chapters, this is to allow the wavevector $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_N)$ which enters the Bloch periodicity conditions at the boundary of the cell Ω_0^N in the $3N$ dimensional space to take a reasonably large number of discrete values. Again, while the solution we seek will not correspond to a Bloch solution (which generally does not have the required antisymmetry if $\mathbf{k}$ itself does not have this symmetry) it should correspond to an antisymmetrization of it and will not be quasiperiodic on the cell Ω_0^N, but will be periodic on the cell $C = \Omega^N$.

We also find Q_C^*-convex translations that could be useful in accelerating numerical methods for solving the Schrödinger equation in periodic systems (see Section 8.9). A Q_C^*-convex function is one which is convex on the space of C-periodic fields $\mathcal{E}$ (or $\mathcal{J}$) appropriate to the problem at hand, which is slightly different to the meaning of Q_C^*-convexity given by Milton (2013b)). Q_C^*-convexity is a natural generalization of quasiconvexity to wave equations, or more generally to equations where the differential constraints on the fields involve derivatives of different orders.

Quasiconvexity was introduced by Morrey (1952) for establishing a necessary condition for a functional to be lower semicontinuous in the calculus of variations (briefly a functional $I(\mathbf{u})$ is lower semicontinuous if for every sequence of functions $\mathbf{u}_\epsilon(\mathbf{x})$ converging as $\epsilon \to 0$ in an appropriate topology to a function $\mathbf{u}_0(\mathbf{x})$, the limit infimum of $I(\mathbf{u}_\epsilon)$ is greater than or equal to $I(\mathbf{u}_0)$)-i.e, fluctuations are penalized. As shown by Ball (1977) quasiconvexity of the energy $W(\mathbf{x}, \nabla\mathbf{u}(\mathbf{x}))$ rather than convexity seems to be the natural framework for establishing a theory of nonlinear elasticity, in which one can show the existence of minimizers and the

same time have the desired invariance properties of the energy under rotation of coordinates. Quasiconvexity is also important in the theory of shape memory materials, where a breakdown of quasiconvexity is associated with the appearance of microstructure such as twinning (Ball and James 1987, 1992). For a good introduction to quasiconvexity see the book of Dacorogna (2007). The problem is that it is very difficult to determine if a function is quasiconvex unless it is quadratic or polyconvex (Ball 1977).

Following the pivotal contributions of Murat and Tartar (Tartar 1979b, particularly theorem 8; Murat and Tartar 1985; Tartar 1985) and Lurie and Cherkaev (1982, 1984), quadratic quasiconvex functions have played an important role, through the translation method (also called the method of compensated compactness), in bounding effective tensors of composite materials and for giving clues for finding microstructures which attain them: see the books of Allaire (2002), Cherkaev (2000), Milton (2002), Tartar (2009), and Torquato (2002) and references therein. They have also played an important role in bounding the response of bodies, or inversely in bounding the volume occupied by an inclusion phase (Kang, Kim, and Milton 2012; Kang, Milton, and Wang 2014; Milton and Nguyen 2012; Kang, Milton, and Wang 2014; Kang and Milton 2013; Kang, Kim, Lee, Li, and Milton 2014, Thaler and Milton 2015). It is expected that the Q_C^*-convex functions here may also have a role in bounding the volume occupied by an inclusion phase for the acoustic equation (which, in the case the density is constant, is mathematically identical to the single electron Schrödinger equations in the time-harmonic case).

This chapter is mostly self-contained, although the reader is advised to begin by studying **Chapters 1** and **2**. Those readers interested in this chapter may also like to look at **Chapters 3**, **5**, **8**, **11**, and **12**.

13.2 The basic variational theorem: a direct proof

Following the procedure of **Chapter 1**, the Schrödinger equation with a source term can be rewritten as

$$\underbrace{\begin{pmatrix} \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{q}(\mathbf{x}) \\ \nabla \cdot \mathbf{r}(\mathbf{x}) \end{pmatrix}}_{\mathbf{J}(\mathbf{x})} = \underbrace{\begin{pmatrix} -\mathbf{A} & 0 & 0 \\ 0 & E - V(\mathbf{x}) & h(\mathbf{x}) \\ 0 & h(\mathbf{x}) & d(\mathbf{x}) \end{pmatrix}}_{\mathbf{L}(\mathbf{x})} \underbrace{\begin{pmatrix} \nabla\psi \\ \psi \\ \theta_0 \end{pmatrix}}_{\mathbf{E}(\mathbf{x})}, \tag{13.3}$$

where $\mathbf{q}(\mathbf{x})$ is some current, and $d(\mathbf{x})$ like $V(\mathbf{x})$ is assumed to be symmetric with respect to electron interchanges (swapping $\mathbf{x}_i$ and $\mathbf{x}_\ell$). Again, the energy E is not to be confused with the field $\mathbf{E}(\mathbf{x})$. We assume $\mathbf{A}$ is a real, constant, and positive-definite matrix.

Let $W(\psi', p)$ be the following functional:

$$W(\psi', p) = \sum_s \int_{\Omega^N} \underbrace{[p(\mathbf{x})]^2 + (E'')^2[\psi'(\mathbf{x})]^2 + 2\theta_0 p(\mathbf{x})h(\mathbf{x})}_{I(p,\psi')} \, d\mathbf{r} \tag{13.4}$$

where the sum is over all 2^N spin configurations $s = (s_1, s_2, \ldots, s_N)$ as each s_j takes values $+1/2$ or $-1/2$. Further define $W(\psi')$ to be $W(\psi', p)$ with the constraint that

$$p(\mathbf{x}) = p(\mathbf{x}, \psi') = \nabla \cdot \mathbf{A}\nabla\psi' + (E' - V(\mathbf{x}))\psi', \tag{13.5}$$

The claim is the variational principle that in any body Ω with appropriate boundary conditions on ψ' on $\partial\Omega^N$, that when $W(\psi')$ is minimized over $\psi'(\mathbf{x})$, the minimum occurs when ψ' is the real part of the wave function

ψ which satisfies the Schrödinger equation

$$0 = \nabla \cdot \mathbf{A}\nabla\psi + (E - V(\mathbf{x}))\psi + h(\mathbf{x})\theta_0, \tag{13.6}$$

where the source term $h(\mathbf{x})$, the potential $V(\mathbf{x})$, and θ_0 are real with E and ψ complex. We emphasize that we are minimizing over all functions $\psi'(\mathbf{x})$ satisfying the appropriate boundary conditions including functions $\psi'(\mathbf{x})$ that are *not* antisymmetric under interchange of any pair $\mathbf{x}_i$ and $\mathbf{x}_k$. Of course the boundary conditions must be chosen so the minimizer is antisymmetric under interchange of any pair $\mathbf{x}_i$ and $\mathbf{x}_k$.

From the real and imaginary parts of the Schrödinger equation we obtain

$$\begin{aligned} 0 &= \nabla \cdot \mathbf{A}\nabla\psi' + (E' - V(\mathbf{x}))\psi' + h(\mathbf{x})\theta_0 - E''\psi'', \\ 0 &= \nabla \cdot \mathbf{A}\nabla\psi'' + (E' - V)\psi'' + E''\psi'. \end{aligned} \tag{13.7}$$

First note that $I(p, \psi')$ in (13.4) is a positive definite quadratic form in $p(\mathbf{x})$ and $\psi'(\mathbf{x})$, and so a minimum should exist when we further constrain $p(\mathbf{x})$ and $\psi'(\mathbf{x})$ to satisfy the linear relation (13.5).

Now, taking variations in (13.5) gives

$$\delta p = \nabla \cdot \mathbf{A}\nabla\delta\psi' + (E' - V(\mathbf{x}))\delta\psi'. \tag{13.8}$$

Then, using this, the variation of (13.4) equates to

$$\begin{aligned} \delta W &= \sum_s \int_{\Omega^N} 2p(\mathbf{x})\delta p + 2(E'')^2(\psi')\delta\psi' + 2\theta_0 h(\mathbf{x})\delta p \\ &= \sum_s \int_{\Omega^N} 2\underbrace{(p(\mathbf{x}) + h(\mathbf{x})\theta_0)}_{t(\mathbf{x})}\delta p + 2(E'')^2(\psi')\delta\psi' \\ &= \sum_s \int_{\Omega^N} 2t(\mathbf{x})[\nabla \cdot \mathbf{A}\nabla\delta\psi' + (E' - V(\mathbf{x}))\delta\psi'] + 2(E'')^2\psi'\delta\psi'. \end{aligned} \tag{13.9}$$

Looking at the first term we get

$$\begin{aligned} \sum_s \int_{\Omega^N} 2t(\mathbf{x})[\nabla \cdot \mathbf{A}\nabla\delta\psi'] &= \sum_s \int_{\Omega^N} -2\underbrace{(\nabla t(\mathbf{x})) \cdot \mathbf{A}\nabla\delta\psi'}_{[\mathbf{A}\nabla t(\mathbf{x})]\cdot\nabla\delta\psi'} + \sum_s \int_{\partial\Omega^N} 2\mathbf{n} \cdot [t(\mathbf{x})\mathbf{A}\nabla\delta\psi'] \\ &= \sum_s \int_{\Omega^N} 2[\nabla \cdot \mathbf{A}\nabla t(\mathbf{x})] \cdot \delta\psi' + \sum_s \int_{\partial\Omega^N} 2\mathbf{n} \cdot [t(\mathbf{x})\mathbf{A}\nabla\delta\psi'] \\ &\quad - \sum_s \int_{\partial\Omega^N} 2\mathbf{n} \cdot [\mathbf{A}\nabla t(\mathbf{x})]\delta\psi'. \end{aligned} \tag{13.10}$$

Here $\mathbf{n}$ is the $3N$-dimensional vector that is the outwards normal to Ω^N. We will see that the variational principle will hold if the boundary conditions are such that the sum of the last two boundary integrals in (13.10) is zero. In particular, the boundary terms will be zero if $\delta\psi' = 0$ on $\partial\Omega^N$ and $\mathbf{n} \cdot \mathbf{A}\nabla\delta\psi' = 0$ on $\partial\Omega^N$ where $\mathbf{n}$ is normal to the boundary.

Excluding these boundary terms, the variation in W is

$$\delta W = \sum_s \int_{\Omega^N} 2[\nabla \cdot \mathbf{A}\nabla t(\mathbf{x}) + (E' - V(\mathbf{x}))t(\mathbf{x}) + (E'')^2\psi']\delta\psi', \tag{13.11}$$

so at the minimum

$$0 = \nabla \cdot \mathbf{A}\nabla t(\mathbf{x}) + (E' - V(\mathbf{x}))t(\mathbf{x}) + (E'')^2\psi', \tag{13.12}$$

where

$$t(\mathbf{x}) = h(\mathbf{x})\theta_0 + \nabla \cdot \mathbf{A}\nabla\psi' + (E' - V(\mathbf{x}))\psi'. \tag{13.13}$$

Let's check that this equation is satisfied when the Schrödinger equation is satisfied. Substituting (13.7) in the expression for $t(\mathbf{x})$ we see that

$$t(\mathbf{x}) = E''\psi''(\mathbf{x}). \tag{13.14}$$

Substituting this into (13.12) we get

$$0 = E''\nabla \cdot \mathbf{A}\nabla\psi'' + (E' - V(\mathbf{x}))E''\psi'' + (E'')^2\psi', \tag{13.15}$$

that is

$$0 = \nabla \cdot \mathbf{A}\nabla\psi'' + (E' - V(\mathbf{x}))\psi'' + E''\psi', \tag{13.16}$$

which is exactly the second equation in (13.7).

To have a variational principle we also need to choose boundary conditions so that the surface terms in (13.10) vanish. This is the case if the conditions $\delta\psi' = 0$ on $\partial\Omega^N$ and $\mathbf{n} \cdot \nabla\delta\psi' = 0$ are satisfied, as ensured if we fix

$$\begin{aligned} \psi' &= \psi'_0, \\ \mathbf{n} \cdot \mathbf{q}' &= -\mathbf{n} \cdot (\mathbf{A}\nabla\psi') = f'_0 \quad \text{on } \partial\Omega^N. \end{aligned} \tag{13.17}$$

Naturally $\psi'_0(\mathbf{x})$ and $f'_0(\mathbf{x})$ must have appropriate symmetries if the minimizing $\psi'(\mathbf{x})$ is to be antisymmetric under interchange of $\mathbf{x}_j$ and $\mathbf{x}_\ell$ for all $\mathbf{x}_j \neq \mathbf{x}_\ell$. Thus $\psi'_0(\mathbf{x})$ and $f'_0(\mathbf{x})$ should be chosen so they too are antisymmetric under interchange of $\mathbf{x}_j$ and $\mathbf{x}_\ell$ for all $j \neq \ell$. Suppose that with such a choice of boundary conditions the minimizer is not antisymmetric under interchange of $\mathbf{x}_j$ and $\mathbf{x}_\ell$ for some $j \neq \ell$. Then if we make this interchange we obtain another function which is also a minimizer. But the convexity of $I(p, \psi')$ in both p and ψ' implies this minimizer is unique. So we arrive at a contradiction and we conclude that the minimizer does satisfy the desired antisymmetries.

Another choice which ensures that the boundary terms in (13.10) vanish, for example, would be to take

$$\psi' = 0, \quad t = h\theta_0 + \nabla \cdot \mathbf{A}\nabla\psi' + (E' - V)\psi' = 0 \quad \text{on } \partial\Omega^N, \tag{13.18}$$

which, due to (13.14), corresponds at the minimum to $\psi = \psi' + i\psi'' = 0$ as would be appropriate if the boundary $\partial\Omega$ was impermeable to electrons (corresponding to an infinite potential there). Alternatively, we could assume our fields are periodic with period cell Ω^N and use periodic boundary conditions on $\partial\Omega^N$.

13.3 Variational principles with other boundary conditions

Other boundary conditions, such as specifying nonzero values of $\psi = \psi' + i\psi''$ on $\partial\Omega^N$ can be used by following the treatment by Milton and Willis (2010). To the functional $W(\psi')$ we add the functional

$$W_0(\psi') = \sum_s \int_{\Omega^N} -2c_0 \begin{pmatrix} \nabla\psi''_0 \\ \psi''_0 \\ \mathbf{q}''_0 \\ \nabla \cdot \mathbf{q}''_0 \end{pmatrix} \cdot \begin{pmatrix} -\mathbf{q}' \\ -\nabla \cdot \mathbf{q}' \\ \nabla\psi' \\ \psi' \end{pmatrix}, \tag{13.19}$$

where c_0 is a real constant, and we minimize over ψ' the total functional $W(\psi')+W_0(\psi')$ with $\mathbf{q}' = -\mathbf{A}\nabla\psi'$: here ψ_0'' and $\mathbf{q}_0''$ are arbitrary fields which we are free to choose. Due to the differential constraints on the fields, the key identity holds and $W_0(\psi')$ can be expressed entirely in terms of surface terms:

$$W_0(\psi') = \sum_s \int_{\partial\Omega^N} 2c_0\mathbf{n}\cdot(\mathbf{q}'\psi_0'' - \mathbf{q}_0''\psi'). \tag{13.20}$$

Hence if ψ' is a solution to the Schrödinger equation, satisfying (13.7), and we vary ψ' by $\delta\psi'$ then to first order in the variations the only variation in $W + W_0$ comes from the variation in (13.20) and from the surface terms in (13.10):

$$\delta(W+W_0) = 2\sum_s \int_{\partial\Omega^N} \mathbf{n}\cdot\{[t(\mathbf{x})\mathbf{A}\nabla\delta\psi'] - [\mathbf{A}\nabla t(\mathbf{x})]\delta\psi' + c_0[\psi_0''\mathbf{A}\nabla\delta\psi'] + c_0\mathbf{q}_0''\delta\psi'\}. \tag{13.21}$$

When doing the minimization we can choose any boundary conditions that ensure this surface contribution is zero. For instance we could choose

$$\delta\psi' = 0, \quad c_0 = -E'', \quad t(\mathbf{x}) = E''\psi_0'', \tag{13.22}$$

which corresponds to fixing both the real and imaginary parts of ψ at the boundary $\partial\Omega^N$. Another natural choice would be to take

$$c_0 = -E'', \quad \mathbf{n}\cdot\mathbf{A}\nabla t(\mathbf{x}) = -E''\mathbf{n}\cdot\mathbf{q}_0'', \quad \mathbf{n}\cdot\mathbf{A}\nabla\delta\psi' = 0, \tag{13.23}$$

which corresponds to fixing the real and imaginary parts of $\mathbf{n}\cdot\mathbf{q}(\mathbf{x})$ at $\partial\Omega^N$.

13.4 Original derivation of the variational principle

The variational principle that (13.4) is minimized, when $p(\mathbf{x})$ is given by (13.5), seems to pop out of nowhere. For completeness we feel it is instructive to provide the derivation by which (13.4) was first obtained. The derivation is rather circuitous, and so this section can easily be skipped by the reader.

The starting point is the variational principles that have been obtained for the time-harmonic acoustic equation (Milton, Seppecher, and Bouchitté 2009) based on extensions of variational principles that Cherkaev and Gibiansky (1994) obtained for quasistatics. In this derivation of the new variational principles, we assume the fields are all periodic with period cell Ω^N, and we let the angular brackets denote an average of $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ over Ω^N and $s = (s_1, s_2, \ldots, s_N)$ over all 2^N spin configurations as each s_i takes values $+1/2$ or $-1/2$. In this section, unlike what was done in the previous chapter and unlike what will be done in the subsequent sections of this chapter, we restrict our space $\mathcal{U}$ of "applied fields" to fields expressible in the form

$$\mathbf{U}(\mathbf{x}) = \begin{pmatrix} 0 \\ 0 \\ \theta_0 \end{pmatrix}, \tag{13.24}$$

where θ_0 is a constant independent of the spin configuration s. [Fields of the form (13.60) where the average of $\theta_0(s)$ over spin configurations s is zero, are added to the space $\mathcal{J}$ so as not to affect the solution, as was done similarly in 3.6.] This has the advantage that when the dual problem is considered, $\mathbf{J}_0(\mathbf{x}) \in \mathcal{U}$ is also

independent of $\mathbf{x}$, and in particular independent of s. In the analysis below this corresponds to taking S_0' independent of s. It has the disadvantage that the projection operators onto $\mathcal{U}$ and $\mathcal{J}$ are then no longer local in Fourier space but require spin averages when $\mathbf{k} = 0$. This is the reason we define $\mathcal{U}$ and $\mathcal{J}$ differently in the previous chapter, and in subsequent sections of this chapter. In this section no Fourier analysis is required, so it is not a problem.

We assume $d(\mathbf{x})$ has a positive imaginary part $d''(\mathbf{x})$. Note that we can rewrite the equation (13.3) as

$$\widetilde{\mathbf{J}}(\mathbf{x}) = \widetilde{\mathbf{L}}(\mathbf{x})\mathbf{E}(\mathbf{x}), \tag{13.25}$$

with

$$\widetilde{\mathbf{J}}(\mathbf{x}) = e^{-i\theta}\mathbf{J}(\mathbf{x}), \quad \widetilde{\mathbf{L}}(\mathbf{x}) = e^{-i\theta}\mathbf{L}(\mathbf{x}), \tag{13.26}$$

and by choosing θ appropriately we can ensure $\widetilde{\mathbf{L}}(\mathbf{x})$ has a strictly positive definite imaginary part. So without loss of generality let us drop the tildes and assume $\mathbf{L}(\mathbf{x})$ has strictly positive definite imaginary part:

$$\mathbf{L}(\mathbf{x}) = \mathbf{L}'(\mathbf{x}) + i\mathbf{L}''(\mathbf{x}), \tag{13.27}$$

where $\mathbf{L}''(\mathbf{x})$ is self adjoint and positive semidefinite and $\mathbf{L}'(\mathbf{x})$ is self adjoint. Similarly let

$$\mathbf{J}(\mathbf{x}) = \mathbf{J}'(\mathbf{x}) + i\mathbf{J}''(\mathbf{x}), \quad \mathbf{E}(\mathbf{x}) = \mathbf{E}'(\mathbf{x}) + i\mathbf{E}''(\mathbf{x}). \tag{13.28}$$

Following Cherkaev and Gibiansky (1994) and Milton, Seppecher, and Bouchitté (2009) we can rewrite the constitutive equation as

$$\begin{pmatrix} \operatorname{Im}\mathbf{E} \\ \operatorname{Im}\mathbf{J} \end{pmatrix} = \mathcal{L}\begin{pmatrix} -\operatorname{Re}\mathbf{J} \\ \operatorname{Re}\mathbf{E} \end{pmatrix}, \tag{13.29}$$

with

$$\mathcal{L}(\mathbf{x}) = \begin{pmatrix} (\mathbf{L}'')^{-1} & (\mathbf{L}'')^{-1}\mathbf{L}' \\ \mathbf{L}'(\mathbf{L}'')^{-1} & \mathbf{L}'' + \mathbf{L}'(\mathbf{L}'')^{-1}\mathbf{L}' \end{pmatrix}. \tag{13.30}$$

See, e.g., **Section 11.6** of Milton (2002).

Now $\mathbf{L}(\mathbf{x})$ separates into a $3N \times 3N$ block, involving the constant matrix $-\mathbf{A}$, and a 2×2 block $\mathbf{B}$:

$$\mathbf{L}(\mathbf{x}) = e^{-i\theta}\begin{pmatrix} -\mathbf{A} & 0 \\ 0 & \mathbf{B}(x) \end{pmatrix}, \tag{13.31}$$

with

$$\mathbf{B}(\mathbf{x}) = \begin{pmatrix} a(\mathbf{x}) & h(\mathbf{x}) \\ \overline{h}(\mathbf{x}) & d(\mathbf{x}) \end{pmatrix}. \tag{13.32}$$

Since we assume $h(\mathbf{x}) \in \mathbb{R}$ we have $\overline{h}(\mathbf{x}) = h(\mathbf{x})$. So we obtain

$$\mathcal{L}(\mathbf{x}) = \left(\begin{array}{cc|cc} (\mathbf{A}'')^{-1} & 0 & (-\mathbf{A}'')^{-1}\mathbf{A}' & 0 \\ 0 & (\mathbf{B}'')^{-1} & 0 & (\mathbf{B}'')^{-1}\mathbf{B}' \\ \hline -\mathbf{A}'(\mathbf{A}'')^{-1} & 0 & \mathbf{A}'' + \mathbf{A}'(\mathbf{A}'')^{-1}\mathbf{A}' & 0 \\ 0 & \mathbf{B}'(\mathbf{B}'')^{-1} & 0 & \mathbf{B}'' + \mathbf{B}'(\mathbf{B}')^{-1}\mathbf{B}' \end{array}\right), \tag{13.33}$$

where

$$\begin{aligned} -e^{-i\theta}\mathbf{A} &= -\mathbf{A}' + i\mathbf{A}'', \quad \mathbf{A}'' > 0, \\ e^{-i\theta}\mathbf{B} &= \mathbf{B}' + i\mathbf{B}'', \end{aligned} \tag{13.34}$$

with $\mathbf{A}, \mathbf{A}', \mathbf{A}'', \mathbf{B}'$ and $\mathbf{B}''$ being all Hermitian. However, note $\mathbf{B}$ is not Hermitian because E has a positive imaginary part (i.e., $E = E' + iE''$ with $E'' > 0$) and $d(\mathbf{x}) = d'(\mathbf{x}) + id''(\mathbf{x})$ has an imaginary part $d''(\mathbf{x}) > 0$.

We have

$$\begin{aligned} \mathbf{A}' &= \cos(\theta)\mathbf{A}, \quad \mathbf{A}'' = \sin(\theta)\mathbf{A}, \\ \mathbf{B}' &= \cos(\theta)\begin{pmatrix} E' - V(\mathbf{x}) & h(\mathbf{x}) \\ h(\mathbf{x}) & d'(\mathbf{x}) \end{pmatrix} + \sin(\theta)\begin{pmatrix} E'' & 0 \\ 0 & d''(\mathbf{x}) \end{pmatrix}, \\ \mathbf{B}'' &= \sin(\theta)\begin{pmatrix} E' - V(\mathbf{x}) & h(\mathbf{x}) \\ h(\mathbf{x}) & d'(\mathbf{x}) \end{pmatrix} + \cos(\theta)\begin{pmatrix} E'' & 0 \\ 0 & d''(\mathbf{x}) \end{pmatrix}. \end{aligned} \tag{13.35}$$

This gives

$$\mathcal{L}(\mathbf{x}) = \left(\begin{array}{cc|cc} \mathbf{A}^{-1}\csc(\theta) & 0 & -\mathbf{I}\cot(\theta) & 0 \\ 0 & (\mathbf{B}'')^{-1} & 0 & (\mathbf{B}'')^{-1}\mathbf{B}' \\ \hline -\mathbf{I}\cot(\theta) & 0 & \mathbf{A}\csc(\theta) & 0 \\ 0 & \mathbf{B}'(\mathbf{B}'')^{-1} & 0 & \mathbf{B}'' + \mathbf{B}'(\mathbf{B}'')^{-1}\mathbf{B}' \end{array}\right), \tag{13.36}$$

and so the variational principle reads

$$\min_{\mathbf{q}', S', \psi'} \left\langle \begin{pmatrix} -\mathbf{q}'(\mathbf{x}) \\ -\nabla\cdot\mathbf{q}'(\mathbf{x}) \\ -S'(\mathbf{x}) \\ \nabla\psi' \\ \psi' \\ \theta_0' \end{pmatrix} \cdot \mathcal{L}(\mathbf{x}) \begin{pmatrix} -\mathbf{q}'(\mathbf{x}) \\ -\nabla\cdot\mathbf{q}'(\mathbf{x}) \\ S'(\mathbf{x}) \\ \nabla\psi' \\ \psi' \\ \theta_0' \end{pmatrix} \right\rangle, \tag{13.37}$$

where the angular brackets denote an average of $\mathbf{r}$ over the unit cell of periodicity Ω^N and over all 2^N spin configurations s, and the minimum is over all potentials $\mathbf{q}'(\mathbf{x})$ and $\psi'(\mathbf{x})$, and scalar fields $S'(\mathbf{x})$ such that $\langle S'(\mathbf{x})\rangle = S_0'$. Note the vectors on the left and right of $\mathcal{L}(\mathbf{x})$ are real.

Alternatively the variational principle can be written as

$$\begin{aligned} \min_{\mathbf{q}', S', \psi'} \Big\langle & \mathbf{q}'(\mathbf{x})\cdot\mathbf{A}^{-1}\mathbf{q}'(\mathbf{x})\csc(\theta) + \mathbf{q}'(\mathbf{x})\cdot\nabla\psi'\cot(\theta) \\ & + (\nabla\psi')\cdot\mathbf{q}'(\mathbf{x})\cot(\theta) + (\nabla\psi')\cdot\mathbf{A}(\nabla\psi')\csc(\theta) \\ & + \begin{pmatrix} \nabla\cdot\mathbf{q}' \\ S'(\mathbf{x}) \end{pmatrix}\cdot(\mathbf{B}'')^{-1}\begin{pmatrix} \nabla\cdot\mathbf{q}' \\ S'(\mathbf{x}) \end{pmatrix} + \begin{pmatrix} \nabla\cdot\mathbf{q}' \\ S'(\mathbf{x}) \end{pmatrix}\cdot(\mathbf{B}'')^{-1}\mathbf{B}'\begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix} \\ & + \begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix}\cdot\mathbf{B}'(\mathbf{B}'')^{-1}\begin{pmatrix} \nabla\mathbf{q}' \\ S'(\mathbf{x}) \end{pmatrix} + \begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix}\cdot(\mathbf{B}'' + \mathbf{B}'(\mathbf{B}'')^{-1}\mathbf{B}')\begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix} \Big\rangle. \end{aligned} \tag{13.38}$$

In order that the term involving $\mathbf{A}$ does not blow up as $\theta \to 0$ we should choose

$$\mathbf{q}'(\mathbf{x}) = -\mathbf{A}\nabla\psi'. \tag{13.39}$$

Then we see that

$$\begin{aligned}
&\mathbf{q}'(\mathbf{x}) \cdot \mathbf{A}^{-1}\mathbf{q}'(\mathbf{x}) \csc(\theta) + \mathbf{q}'(\mathbf{x}) \cdot \nabla\psi' \cot(\theta) \\
&+(\nabla\psi') \cdot \mathbf{q}'(\mathbf{x}) \cot(\theta) + (\nabla\psi') \cdot \mathbf{A}(\nabla\psi') \csc(\theta) \\
&= 2\left(\frac{1-\cos(\theta)}{\sin(\theta)}\right) \nabla\psi' \cdot \mathbf{A}\nabla\psi',
\end{aligned} \tag{13.40}$$

which goes to zero as $\theta \to 0$. After choosing $\mathbf{q}'(\mathbf{x}) = -\mathbf{A}\nabla\psi'$, let's now take the limit $\theta \to 0$ in the variational principle. It becomes

$$\min_{\substack{\psi' \\ S' \\ \langle S' \rangle = S_0' \\ \theta_0' \text{ fixed}}} \left\langle \begin{pmatrix} -\nabla \cdot \mathbf{q}'(\mathbf{x}) \\ -S'(\mathbf{x}) \\ \psi' \\ \theta_0' \end{pmatrix} \cdot \mathcal{L}^B(\mathbf{x}) \begin{pmatrix} -\nabla \cdot \mathbf{q}'(\mathbf{x}) \\ -S'(\mathbf{x}) \\ \psi' \\ \theta_0' \end{pmatrix} \right\rangle, \tag{13.41}$$

where $\mathbf{q}'(\mathbf{x}) = -\mathbf{A}\nabla\psi'$, θ_0' is fixed, and

$$\mathcal{L}^B(\mathbf{x}) = \begin{pmatrix} (\mathbf{B}'')^{-1} & (\mathbf{B}'')^{-1}\mathbf{B}' \\ \mathbf{B}'(\mathbf{B}'')^{-1} & \mathbf{B}'' + \mathbf{B}'(\mathbf{B}'')^{-1}\mathbf{B}' \end{pmatrix}, \tag{13.42}$$

where now (as we have taken the limit $\theta \to 0$)

$$\begin{aligned}
\mathbf{B}' &= \begin{pmatrix} E' - V(\mathbf{x}) & h(\mathbf{x}) \\ h(\mathbf{x}) & d'(\mathbf{x}) \end{pmatrix}, \\
\mathbf{B}'' &= \begin{pmatrix} E'' & 0 \\ 0 & d''(\mathbf{x}) \end{pmatrix}.
\end{aligned} \tag{13.43}$$

Now we have

$$\begin{aligned}
(\mathbf{B}'')^{-1} &= \begin{pmatrix} 1/E'' & 0 \\ 0 & 1/d'' \end{pmatrix}, \\
(\mathbf{B}'')^{-1}\mathbf{B}' &= \begin{pmatrix} [E' - V(\mathbf{x})]/E'' & h(\mathbf{x})/E'' \\ h(\mathbf{x})/d'' & d'(\mathbf{x})/d'' \end{pmatrix}, \\
\mathbf{B}'(\mathbf{B}'')^{-1} &= \begin{pmatrix} [E' - V(\mathbf{x})]/E'' & h(\mathbf{x})/d'' \\ h(\mathbf{x})/E'' & d'(\mathbf{x})/d'' \end{pmatrix}, \\
\mathbf{B}'' + \mathbf{B}'(\mathbf{B}'')^{-1}\mathbf{B}' &= \begin{pmatrix} E'' + \frac{(E'-V)^2}{E''} + \frac{hh}{d''} & \frac{(E'-V)h}{E''} + \frac{h(\mathbf{x})d'(\mathbf{x})}{d''} \\ \frac{h(E'-V)}{E''} + \frac{d'(\mathbf{x})h(\mathbf{x})}{d''} & d'' + \frac{hh}{E''} + \frac{(d'(\mathbf{x}))^2}{d''} \end{pmatrix}, \\
\begin{pmatrix} \nabla \cdot \mathbf{q}'(\mathbf{x}) \\ S'(\mathbf{x}) \end{pmatrix} \cdot (\mathbf{B}'')^{-1} \begin{pmatrix} \nabla \cdot \mathbf{q}'(\mathbf{x}) \\ S'(\mathbf{x}) \end{pmatrix} &= \frac{(\nabla \cdot \mathbf{q}'(\mathbf{x}))^2}{E''} + \frac{(S'(\mathbf{x}))^2}{d''}, \\
\begin{pmatrix} \nabla \cdot \mathbf{q}'(\mathbf{x}) \\ S'(\mathbf{x}) \end{pmatrix} \cdot (\mathbf{B}'')^{-1}\mathbf{B}' \begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix} &= (\nabla \cdot \mathbf{q}'(\mathbf{x})) \left(\frac{(E' - V(\mathbf{x}))\psi' + h(\mathbf{x})\theta_0'}{E''} \right) \\
&\quad + \frac{S'(\mathbf{x})h(\mathbf{x})\psi' + S'(\mathbf{x})d'(\mathbf{x})\theta_0'}{d''}, \\
\begin{pmatrix} \psi' \\ \theta_0' \end{pmatrix} \cdot \mathbf{B}'(\mathbf{B}'')^{-1} \begin{pmatrix} \nabla \cdot \mathbf{q}(\mathbf{x}) \\ S' \end{pmatrix} &= (\nabla \cdot \mathbf{q}'(\mathbf{x})) \left(\frac{(E' - V(\mathbf{x}))\psi + h(\mathbf{x})\theta_0'}{E''} \right) \\
&\quad + \frac{S'(\mathbf{x})h(\mathbf{x})\psi' + S'(\mathbf{x})d'(\mathbf{x})\theta_0'}{d''}. \qquad (13.44)
\end{aligned}$$

Collecting the terms in $S'(\mathbf{x})$ and minimizing over $S'(\mathbf{x})$, we need to minimize

$$\langle \frac{S'(\mathbf{x})^2 - S'(\mathbf{x})w(\mathbf{x})}{d''} \rangle, \qquad (13.45)$$

subject to $\langle S' \rangle = S_0'$ and where

$$w(\mathbf{x}) = 2h(\mathbf{x})\psi'(\mathbf{x}) + 2d'(\mathbf{x})\theta_0'. \qquad (13.46)$$

We solve this by introducing a Lagrange multiplier λ and minimizing

$$\langle S'(\mathbf{x})^2 - S'(\mathbf{x})(w(\mathbf{x}) - \lambda) \rangle \qquad (13.47)$$

over all Ω^N periodic $S'(\mathbf{x})$ without the constraint that $\langle S' \rangle = S_0'$. The solution satisfies

$$S'(\mathbf{x}) = \frac{w(\mathbf{x}) - \lambda}{2}, \qquad (13.48)$$

and we choose $\lambda = \langle w \rangle - 2S_0'$ to ensure $\langle S' \rangle = S_0'$. So the minimum is at

$$\begin{aligned}
S'(\mathbf{x}) &= S_0' - \frac{\langle w \rangle - w(\mathbf{x})}{2}, \\
S'(\mathbf{x})^2 &= \left(S_0' - \frac{\langle w \rangle}{2} \right)^2 + \frac{w(\mathbf{x})^2}{4} + w(\mathbf{x}) \left(S_0' - \frac{\langle w \rangle}{2} \right)^2, \\
-S'(\mathbf{x})w(\mathbf{x}) &= -\left(S_0' - \frac{\langle w \rangle}{2} \right) w(\mathbf{x}) - \frac{w(\mathbf{x})^2}{2}. \qquad (13.49)
\end{aligned}$$

The minimum value of $\langle (S')^2 - S'w\rangle/d''$ is

$$\left[\left(S'_0 - \frac{\langle w\rangle}{2}\right)^2 - \frac{\langle w^2\rangle}{4}\right]\frac{1}{d''}, \tag{13.50}$$

with $w(\mathbf{x}) = 2(h(\mathbf{x})\psi'(\mathbf{x}) + d'(\mathbf{x})\theta'_0)$.

So the contribution from the S' terms is

$$\begin{aligned} -\frac{1}{4d''}\left[\langle(2h(\mathbf{x})\psi' + 2d'\theta'_0)^2\rangle - \langle 2h(\mathbf{x})\psi' + 2d'\theta'_0 - 2S'_0\rangle^2\right] \\ = \frac{-1}{d''}\left[\langle(h\psi' + d'\theta'_0)^2\rangle - \langle h(\mathbf{x})\psi' + d'\theta'_0 - S'_0\rangle^2\right]. \end{aligned} \tag{13.51}$$

Finally we have

$$\begin{aligned} \begin{pmatrix}\psi'\\ \theta'_0\end{pmatrix}\cdot(\mathbf{B}'' + \mathbf{B}'(\mathbf{B}'')^{-1}\mathbf{B}')\begin{pmatrix}\psi'\\ \theta'_0\end{pmatrix} &= (\psi')^2\left(E'' + \frac{(E'-V)^2}{E''} + \frac{hh}{d''}\right) \\ &\quad +\psi'\theta'_0\left(\frac{E'-V}{E''} + \frac{d'(\mathbf{x})}{d''}\right)2h \\ &\quad +(\theta'_0)^2\left(d'' + \frac{hh}{E''} + \frac{(d'(\mathbf{x}))^2}{d''}\right), \end{aligned} \tag{13.52}$$

where the last term is a constant we can ignore. So the total quantity to be minimized is

$$\begin{aligned} &\langle\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{E''} & \frac{E'-V(\mathbf{x})}{E''}\\ \frac{E'-V(\mathbf{x})}{E''} & E'' + \frac{(E'-V)^2}{E''}\end{pmatrix}\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\rangle \\ &+\theta'_0\langle\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\cdot\begin{pmatrix}2h/E''\\ 2h(E'-V)/E''\end{pmatrix}\rangle \\ &+\frac{1}{d''}\langle h\psi' + d'\theta'_0 - S'_0\rangle^2, \end{aligned} \tag{13.53}$$

over ψ', with $\nabla\cdot\mathbf{q}' = -\nabla\cdot\mathbf{A}\nabla\psi$ and where we have neglected the terms in $(\theta'_0)^2$ which can be thrown away (since they are constant).

Multiplying by E'' the quantity to be minimized is

$$\begin{aligned} W &= \langle\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\cdot\begin{pmatrix}1 & E'-V(\mathbf{x})\\ E'-V(\mathbf{x}) & (E'')^2 + (E'-V(\mathbf{x}))^2\end{pmatrix}\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\rangle \\ &\quad +2\theta'_0\langle\begin{pmatrix}-\nabla\cdot\mathbf{q}'\\ \psi'\end{pmatrix}\cdot\begin{pmatrix}h\\ (E'-V)h\end{pmatrix}\rangle \\ &\quad +\frac{E''}{d''}\langle h\psi' + d'\theta'_0 - S'_0\rangle^2, \end{aligned} \tag{13.54}$$

with $\nabla\cdot\mathbf{q}' = -\nabla\cdot\mathbf{A}\nabla\psi'$. We can choose d'' extremely large so this last term vanishes.

Introduce

$$p(\mathbf{x}) = -\nabla\cdot\mathbf{q}' + (E' - V(\mathbf{x}))\psi', \tag{13.55}$$

then we have

$$[p(\mathbf{x})]^2 = (\nabla \cdot \mathbf{q}')^2 + (E' - V(\mathbf{x}))^2(\psi')^2 - 2(E' - V(\mathbf{x}))\psi'\nabla \cdot \mathbf{q}', \tag{13.56}$$

and

$$p(\mathbf{x})h = -(\nabla \cdot \mathbf{q}')h + (E' - V)\psi' h. \tag{13.57}$$

So the quantity to be minimized is

$$\langle [p(\mathbf{x})]^2 + (E'')^2(\psi')^2 + 2\theta_0' p(\mathbf{x})h(\mathbf{x})\rangle, \tag{13.58}$$

over ψ' where

$$p(\mathbf{x}) = \nabla \cdot \mathbf{A}\nabla\psi' + (E' - V(\mathbf{x}))\psi'. \tag{13.59}$$

This is the variational principle stated in (13.4) and (13.5) for the case where we have periodic boundary conditions on the fields in Ω^N.

13.5 The basic subspaces associated with the Schrödinger equation with sources in a periodic medium

Here we consider the Schrödinger equation (13.3) with a source term $h(\mathbf{x})$ in a periodic medium. We let $\mathcal{U}$ be the space of all fields representable in the form

$$\mathbf{U}(\mathbf{x}) = \begin{pmatrix} 0 \\ 0 \\ \theta_0(s) \end{pmatrix}, \tag{13.60}$$

for some choice of the function $\theta_0(s)$. Here $\theta_0(s)$ is independent of $\mathbf{r}$ but can depend on the spin configuration $s = (s_1, s_2, \ldots, s_N)$.

By the uniqueness of the solutions the wavefunction $\psi(\mathbf{x})$ will necessarily satisfy the required antisymmetry properties if $\theta_0(s)$ and hence the "applied field" $\mathbf{U}(\mathbf{x})$ is symmetric under interchange of any pair of spins s_j and s_ℓ. Then since θ_0 appears in the Schrödinger equation multiplied by $h(\mathbf{x})$, which we are free to multiply by a symmetric function of s, it suffices to take an "applied field" $\mathbf{U}$ with $\theta_0 = 1$. We do not require that $\theta_0(s)$ in (13.60) satisfy these symmetry properties.

We let $\mathcal{E}$ be the space of Ω^N-periodic fields, that are square integrable in the unit cell Ω^N, of the form

$$\mathbf{E} = \begin{pmatrix} \nabla\psi \\ \psi \\ 0 \end{pmatrix}, \tag{13.61}$$

where ψ is Ω^N-periodic. It is not required that $\psi(\mathbf{x})$ satisfy the antisymmetry properties appropriate to an electron wavefunction. We let $\mathcal{J}$ be the space of Ω^N-periodic fields, that are square integrable in the unit cell Ω^N, of the form

$$\mathbf{J} = \begin{pmatrix} \mathbf{q} \\ \nabla \cdot \mathbf{q} \\ S(\mathbf{x}) \end{pmatrix}, \tag{13.62}$$

where $\mathbf{q}$ is Ω^N-periodic and $S(\mathbf{x})$ is Ω^N-periodic, with zero average value $\langle S(\mathbf{x})\rangle_{\mathbf{r}} = 0$ where the average is over the spatial coordinates $\mathbf{r}$ and not over the spin coordinates s. Again, it is not required that $\mathbf{q}(\mathbf{x})$ satisfy any symmetry properties.

We introduce the inner product

$$(\mathbf{P},\mathbf{Q}) = \sum_s \int_{\Omega^N} \overline{\begin{pmatrix}\mathbf{p}_1(\mathbf{x})\\ p_2(\mathbf{x})\\ p_3(\mathbf{x})\end{pmatrix}} \cdot \begin{pmatrix}\mathbf{q}_1(\mathbf{x})\\ q_2(\mathbf{x})\\ q_3(\mathbf{x})\end{pmatrix} = \sum_s \int_{\Omega^N} \overline{\mathbf{p}}_1(\mathbf{x})\cdot\mathbf{q}_1(\mathbf{x}) + \overline{p}_2(\mathbf{x})q_2(\mathbf{x}) + \overline{p}_3(\mathbf{x})q_3(\mathbf{x}), \quad (13.63)$$

where $\mathbf{p}_1, \mathbf{q}_1$ are $3N$-dimensional vectors and p_2, p_3, q_2, q_3 are scalars, while the overline denotes complex conjugation. With respect to this inner product the spaces $\mathcal{E}$ and $\mathcal{J}$ are both clearly orthogonal to $\mathcal{U}$. Also the spaces $\mathcal{E}$ and $\mathcal{J}$ are orthogonal:

$$(\mathbf{J},\mathbf{E}) = \sum_s \int_{\Omega^N} \overline{\mathbf{q}}\cdot\nabla\psi(\mathbf{x}) + \psi(\mathbf{x})\nabla\cdot\overline{\mathbf{q}} = \sum_s \int_{\Omega^N} \nabla\cdot(\overline{\mathbf{q}}\psi(\mathbf{x})) = 0, \quad (13.64)$$

since $\overline{\mathbf{q}}\psi(\mathbf{x})$ is Ω^N-periodic.

Now in Fourier space the fields in $\mathcal{E}$ have the form

$$\widehat{\mathbf{E}}(\mathbf{K}) = \begin{pmatrix} i\mathbf{k}\widehat{\psi}\\ \widehat{\psi}\\ 0\end{pmatrix} \text{ for } \mathbf{k}\neq 0, \quad \widehat{\mathbf{E}}((0,s)) = \begin{pmatrix}0\\ \langle\psi(\mathbf{x})\rangle_{\mathbf{r}}\\ 0\end{pmatrix}, \quad (13.65)$$

where $\mathbf{K} = (\mathbf{k}, s)$, in which $\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_N)$ are the Fourier variables and $s = (s_1, s_2, \ldots, s_N)$ is the spin configuration, and $\langle\cdot\rangle_{\mathbf{r}}$ denotes a volume average of $\mathbf{r}$ over Ω^N keeping s fixed. The projection onto these fields is given in Fourier space as

$$\mathbf{\Gamma}_1((\mathbf{k},s)) = \frac{1}{k^2+1}\begin{pmatrix}\mathbf{k}\mathbf{k}^T & i\mathbf{k} & 0\\ -i\mathbf{k}^T & 1 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad (13.66)$$

for all $\mathbf{k}$ including $\mathbf{k} = 0$. Here we abbreviate $|\mathbf{k}|^2 = k^2$. The fields in the space $\mathcal{J}$ have, from (13.62), Fourier components of the form

$$\widehat{\mathbf{J}}(\mathbf{K}) = \begin{pmatrix}\widehat{\mathbf{q}}\\ i\mathbf{k}\cdot\widehat{\mathbf{q}}\\ \widehat{S}\end{pmatrix} \text{ for} \mathbf{k}\neq 0, \quad \widehat{\mathbf{J}}((0,s)) = \begin{pmatrix}\langle\mathbf{q}\rangle_{\mathbf{r}}\\ 0\\ 0\end{pmatrix}. \quad (13.67)$$

The projection onto $\mathcal{J}$ in Fourier space is

$$\begin{aligned}\mathbf{\Gamma}_2((\mathbf{k},s)) &= \frac{1}{k^2+1}\begin{pmatrix}(k^2+1)\mathbf{I}-\mathbf{k}\mathbf{k}^T & -i\mathbf{k} & 0\\ i\mathbf{k}^T & k^2 & 0\\ 0 & 0 & k^2+1\end{pmatrix} \text{ for } \mathbf{k}\neq 0\\ \mathbf{\Gamma}_2((0,s)) &= \begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}. \end{aligned} \quad (13.68)$$

The projection onto $\mathcal{U}$ is

$$\mathbf{\Gamma}_0((\mathbf{k},s)) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ for } \mathbf{k} \neq 0, \quad \mathbf{\Gamma}_0((0,s)) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{13.69}$$

As expected one has $\mathbf{\Gamma}_0 + \mathbf{\Gamma}_1 + \mathbf{\Gamma}_2 = \mathbf{I}$. Our Hilbert space consists of fields of the form

$$\mathbf{P} = \begin{pmatrix} \mathbf{p}_1(\mathbf{x}) \\ p_2(\mathbf{x}) \\ p_3(\mathbf{x}) \end{pmatrix}, \tag{13.70}$$

where $\mathbf{p}_1(\mathbf{x})$ is a $3N$-dimensional vector field taking complex values, while $p_2(\mathbf{x})$ and $p_3(\mathbf{x})$ are scalar fields.

13.6 Q_C^*-convex functions associated with the subspace $\mathcal{E}$

A translation on $\mathcal{E}$ is a self adjoint matrix $\mathbf{T}$ (not necessarily positive definite) such that

$$\sum_s \int_{\Omega^N} \overline{\mathbf{E}}(\mathbf{x}) \cdot \mathbf{T}\mathbf{E}(\mathbf{x}) \geq 0, \tag{13.71}$$

for all C-periodic fields $\mathbf{E} \in \mathcal{E}$, with $C = \Omega^N$. In Fourier space this becomes

$$\overline{\widehat{\mathbf{E}}((\mathbf{k},s))} \cdot \mathbf{T}\widehat{\mathbf{E}}((\mathbf{k},s)) \geq 0, \tag{13.72}$$

for all spin configurations $s = (s_1, s_2, \ldots, s_N)$ and all vectors $\mathbf{k}$ in the reciprocal lattice (including $\mathbf{k} = 0$) and $\widehat{\mathbf{E}}((\mathbf{k},s))$ of the form

$$\widehat{\mathbf{E}}((\mathbf{k},s)) = \begin{pmatrix} i\mathbf{k}\widehat{\psi}((\mathbf{k},s)) \\ \widehat{\psi}((\mathbf{k},s)) \\ 0 \end{pmatrix} \quad \text{for all } \mathbf{k} \neq 0, \quad \widehat{\mathbf{E}}((0,s)) = \begin{pmatrix} 0 \\ \psi_0(s) \\ 0 \end{pmatrix}. \tag{13.73}$$

The translation $\mathbf{T}$ has the form

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{t}_{12} & \mathbf{t}_{13} \\ \overline{\mathbf{t}}_{12}^T & t_{22} & t_{23} \\ \overline{\mathbf{t}}_{13}^T & \overline{t}_{23} & t_{33} \end{pmatrix}, \tag{13.74}$$

where $\mathbf{T}_{11}$ is a $3N \times 3N$ self adjoint matrix, $\mathbf{t}_{12}, \mathbf{t}_{13}$ are $3N$-dimensional vectors and t_{22}, t_{23}, t_{33} are scalars. The Q_C^*-convexity condition reads as

$$\begin{pmatrix} -i\mathbf{k}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{T}_{11} & \mathbf{t}_{12} \\ \overline{\mathbf{t}}_{12}^T & t_{22} \end{pmatrix} \begin{pmatrix} i\mathbf{k} \\ 1 \end{pmatrix} \geq 0, \tag{13.75}$$

or

$$\mathbf{k} \cdot \mathbf{T}_{11}\mathbf{k} + 2(\operatorname{Im} \mathbf{t}_{12}) \cdot \mathbf{k} + t_{22} \geq 0, \tag{13.76}$$

for all real $\mathbf{k} \neq 0$ vectors in the reciprocal lattice, with the additional constraint that

$$t_{22} \geq 0, \tag{13.77}$$

to ensure that (13.72) is satisfied when $\mathbf{k} = 0$. Note there are no restrictions on $\mathrm{Re}(\mathbf{t}_{12}), \mathbf{t}_{13}, t_{23}$, and t_{33} so they can be anything.

Since the vectors in the reciprocal lattice with $\mathbf{k} \neq 0$ have a minimum value of $|\mathbf{k}|$ the matrix entering (13.75) need not be positive definite. As an example, suppose for simplicity the vectors in the reciprocal lattice consisted of vectors $\mathbf{k} = (k_1, k_2, \ldots, k_{3N})$ where all the k_i are integers. Suppose too, for simplicity, we were looking for $\mathbf{T}$'s with $\mathbf{T}_{11} = \mathbf{I}$ and $t_{22} = 0$. Then (13.76) is satisfied if we choose $|\mathrm{Im}\, \mathbf{t}_{12}| < 1/2$.

13.7 Q_C^*-convex functions associated with the subspace $\mathcal{J}$

A translation on $\mathcal{J}$ is a self adjoint matrix (not necessarily positive definite) such that

$$\sum_s \int_{\Omega^N} \overline{\mathbf{J}}(\mathbf{x}) \cdot \mathbf{T}\mathbf{J}(\mathbf{x}) \geq 0, \tag{13.78}$$

for all C-periodic $\mathbf{J} \in \mathcal{J}$, with $C = \Omega^N$. In Fourier space this becomes

$$\overline{\widehat{\mathbf{J}}(\mathbf{k})} \cdot \mathbf{T}\widehat{\mathbf{J}}(\mathbf{k}) \geq 0, \tag{13.79}$$

for all $\mathbf{k}$ in the reciprocal lattice with $\widehat{\mathbf{J}}(\mathbf{k})$ of the form

$$\widehat{\mathbf{J}}(\mathbf{k}) = \begin{pmatrix} \widehat{\mathbf{q}} \\ i\mathbf{k} \cdot \mathbf{q} \\ \widehat{S} \end{pmatrix}, \quad \text{for all } \mathbf{k} \neq 0, \quad \widehat{\mathbf{J}}(0) = \begin{pmatrix} \mathbf{q}_0 \\ 0 \\ 0 \end{pmatrix}. \tag{13.80}$$

Assume $\mathbf{T}$ has the form

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{t}_{12} & \mathbf{t}_{13} \\ \overline{\mathbf{t}}_{12}^T & t_{22} & t_{23} \\ \overline{\mathbf{t}}_{13}^T & \overline{t}_{23} & t_{33} \end{pmatrix}, \tag{13.81}$$

where as before $\mathbf{T}_{11}$ is a $3N \times 3N$ self adjoint matrix, $\mathbf{t}_{12}, \mathbf{t}_{13}$ are $3N$-dimensional vectors and t_{22}, t_{23}, t_{33} are scalars. The condition for Q_C^*-convexity becomes first that

$$\mathbf{T}_{11} \geq 0, \tag{13.82}$$

to ensure (13.79) holds when $\mathbf{k} = 0$, and second that

$$\begin{aligned} \overline{\hat{\mathbf{q}}} \cdot \mathbf{T}_{11}\hat{\mathbf{q}} + i(\overline{\hat{\mathbf{q}}} \cdot \mathbf{t}_{12})(\mathbf{k} \cdot \mathbf{q}) + (\overline{\hat{\mathbf{q}}} \cdot \mathbf{t}_{13})\hat{S} - i(\mathbf{k} \cdot \overline{\hat{\mathbf{q}}})(\overline{\mathbf{t}}_{12}^T\hat{\mathbf{q}}) + t_{22}(\mathbf{k} \cdot \overline{\hat{\mathbf{q}}})(\mathbf{k} \cdot \mathbf{q}) \\ -i(\mathbf{k} \cdot \overline{\hat{\mathbf{q}}})t_{23}\hat{S} + \overline{\hat{S}}\overline{\mathbf{t}}_{13}^T\hat{\mathbf{q}} + \overline{\hat{S}}\overline{t}_{23}i\mathbf{k} \cdot \hat{\mathbf{q}} + \overline{\hat{S}}t_{33}\hat{S} \geq 0, \end{aligned} \tag{13.83}$$

or

$$\begin{pmatrix} \overline{\hat{\mathbf{q}}} & \overline{\hat{S}} \end{pmatrix} \mathbf{M}_k \begin{pmatrix} \hat{\mathbf{q}} \\ \hat{S} \end{pmatrix} \geq 0, \tag{13.84}$$

holds for all $\hat{\mathbf{q}}, \hat{S}$ and for all $\mathbf{k} \neq 0$ in the reciprocal lattice, where

$$\mathbf{M}_k = \begin{pmatrix} \mathbf{T}_{11} - i\mathbf{k}\bar{\mathbf{t}}_{12}^T + i\mathbf{t}_{12}\mathbf{k}^T + t_{22}\mathbf{k}\mathbf{k}^T & \mathbf{t}_{13} - it_{23}\mathbf{k} \\ \bar{\mathbf{t}}_{13}^T + i\bar{t}_{23}\mathbf{k}^T & t_{33} \end{pmatrix}. \tag{13.85}$$

Thus we require that $\mathbf{M}_k$ be positive definite for all $\mathbf{k}$ in the reciprocal lattice, $\mathbf{k} \neq 0$. In the case where (for simplicity) $\mathbf{T}$ is diagonal and $\mathbf{T}_{11} = t_{11}\mathbf{I}$, then the Q_C^*-convexity conditions read

$$t_{11} \geq 0, \quad t_{33} \geq 0, \quad t_{11}\mathbf{I} + t_{22}\mathbf{k}\mathbf{k}^T \geq 0, \tag{13.86}$$

for all $\mathbf{k} \neq 0$ in the reciprocal lattice. This implies

$$t_{11} \geq 0, \quad t_{22} \geq 0, \quad \text{and } t_{33} \geq 0. \tag{13.87}$$

So $\mathbf{T}$ is positive semi-definite which is not so interesting.

Let's assume $\mathbf{T}_{11} = t_{11}\mathbf{I}$ and note that $\mathbf{M}_k$ is positive semidefinite if and only if for any invertible matrix $\mathbf{Q}$ the matrix

$$\overline{\mathbf{Q}}^T \mathbf{M}_k \mathbf{Q} = \mathbf{M}'_k \tag{13.88}$$

is positive semidefinite. In particular we may choose

$$\mathbf{Q} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & 1 \end{pmatrix}, \tag{13.89}$$

where $\mathbf{R}$ is a $3N \times 3N$ real rotation matrix with $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ and

$$\mathbf{R} \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{k}, \tag{13.90}$$

where $\alpha = |\mathbf{k}|$ is a real scalar. Letting

$$\mathbf{R}^T(\mathbf{t}_{12}) = \mathbf{t}'_{12} \quad \text{and } \mathbf{R}^T(\mathbf{t}_{13}) = \mathbf{t}'_{13}, \tag{13.91}$$

we then have

$$\mathbf{R}(\bar{\mathbf{t}}'_{12}) = \bar{\mathbf{t}}_{12}, \quad \text{and } \mathbf{R}(\bar{\mathbf{t}}'_{13}) = \bar{\mathbf{t}}_{13}. \tag{13.92}$$

This rotation preserves the length of the vectors $\mathbf{t}_{12}, \mathbf{t}_{13}$. Then the matrix $\mathbf{M}'_k$ takes the form

$$\mathbf{M}'_k = \begin{pmatrix} \mathbf{M}'_{k11} & \mathbf{M}'_{k12} \\ (\overline{\mathbf{M}}'_{k12})^T & M'_{k22} \end{pmatrix}, \tag{13.93}$$

where

$$\begin{aligned} \mathbf{M}'_{k11} &= \begin{pmatrix} t_{11} - i\alpha\{\bar{\mathbf{t}}'_{12}\}_1 + i\alpha\{\mathbf{t}'_{12}\}_1 + t_{22}\alpha^2 & -i\alpha\{\bar{\mathbf{t}}'_{12}\}_2^T \\ i\alpha\{\mathbf{t}'_{12}\}_2 & t_{11}\mathbf{I} \end{pmatrix}, \\ \mathbf{M}'_{k12} &= \begin{pmatrix} \{\mathbf{t}'_{13}\}_1 - i\alpha t_{23} \\ \{\mathbf{t}'_{13}\}_2 \end{pmatrix}, \\ M'_{k22} &= t_{33}. \end{aligned} \tag{13.94}$$

Here $\{\mathbf{t}'_{12}\}_2$ and $\{\mathbf{t}'_{13}\}_2$ are $(3N-1)$-dimensional vectors which could be complex, $\mathbf{I}$ is the $(3N-1)\times(3N-1)$ identity matrix and all other quantities are scalar. Note that $\{\mathbf{t}'_{12}\}_1$ denotes the first component of the vector $\mathbf{t}'_{12}$ while $\{\mathbf{t}'_{12}\}_2$ denotes the remaining $3N-1$ components of $\mathbf{t}'_{12}$ (respectively for $\{\mathbf{t}'_{13}\}_1, \{\mathbf{t}'_{13}\}_2$).

We want the matrix $\mathbf{M}'_k$ to be positive semidefinite for all rotations $\mathbf{R}$, where

$$\mathbf{t}'_{12} = \mathbf{R}^T\mathbf{t}_{12}, \quad \mathbf{t}'_{13} = \mathbf{R}^T\mathbf{t}_{13}, \tag{13.95}$$

and for all α such that $\alpha = |\mathbf{k}|$ for some $\mathbf{k} \neq 0$ in the reciprocal lattice. In particular, suppose $\mathbf{t}_{12}$ (and hence $\mathbf{t}'_{12}$) is real, $\mathbf{t}_{13} = \mathbf{t}'_{13} = 0$. Then the form of the matrix $\mathbf{M}'_k$ reduces to

$$\mathbf{M}'_k = \begin{pmatrix} t_{11} + t_{22}\alpha^2 & -i\alpha\{\bar{\mathbf{t}}'_{12}\}_2^T & -i\alpha t_{23} \\ i\alpha\{\mathbf{t}'_{12}\}_2 & t_{11}\mathbf{I} & 0 \\ i\alpha\bar{t}_{23} & 0 & t_{33} \end{pmatrix}. \tag{13.96}$$

This must be positive semidefinite in particular for large α, hence

$$\begin{pmatrix} \alpha^{-1} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{M}'_k \begin{pmatrix} \alpha^{-1} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & 1 \end{pmatrix} \geq 0 \tag{13.97}$$

must be positive semidefinite in the limit of large α:

$$\mathbf{M}_R^\infty = \begin{pmatrix} t_{22} & -i\{\mathbf{t}'_{12}\}_2^T & -it_{23} \\ i\{\mathbf{t}'_{12}\}_2 & t_{11}\mathbf{I} & 0 \\ i\bar{t}_{23} & 0 & t_{33} \end{pmatrix} \geq 0. \tag{13.98}$$

Conversely if this is satisfied, then $t_{11} \geq 0$ and hence

$$\mathbf{M}'_k = \begin{pmatrix} t_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{M}_R^\infty \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & 1 \end{pmatrix} \geq 0. \tag{13.99}$$

Finally we can choose a basis where

$$\{\mathbf{t}'_{12}\}_2 = \begin{pmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{13.100}$$

in which $\beta = |\{\mathbf{t}'_{12}\}_2| \leq |\mathbf{t}_{12}|$. Then to ensure Q_C^*-convexity we may choose the parameters t_{ij} such that $\mathbf{t}_{12}$ is real, $\mathbf{t}_{13} = \mathbf{t}'_{13} = 0$ and

$$\begin{pmatrix} t_{22} & -i|\mathbf{t}_{12}| & -it_{23} \\ i|\mathbf{t}_{12}| & t_{11} & 0 \\ it_{23} & 0 & t_{33} \end{pmatrix} \geq 0. \tag{13.101}$$

14 Green's functions for self-adjoint and non-self-adjoint operators

Graeme W. Milton

Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA

Abstract

Starting with a solution to the equation $\mathcal{L}\mathbf{u} = \mathbf{f}$, and a solution to the equation $\mathcal{L}^\dagger \mathbf{u}' = \mathbf{f}'$, we add and subtract them to obtain an equation $\mathbf{MU} = \mathbf{F}$, where $\mathbf{F}$ involves $\mathbf{f}$ and $\mathbf{f}'$, $\mathbf{U}$ involves $\mathbf{u}$ and $\mathbf{u}'$, and $\mathbf{M}$ is self adjoint and involves $\mathcal{L}$ and $\mathcal{L}^\dagger$. From the Green's function of this new problem, with a self-adjoint operator, we can recover the Green's function for the original equation. Of course the boundary conditions should be such that $\mathbf{M}$ is not just formally self-adjoint, but a proper self-adjoint operator. Technicalities arise if $\mathcal{L}$ and $\mathcal{L}^\dagger$ don't have the same domain. We also briefly review the theory of Green's functions in homogeneous infinite media.

14.1 Introduction

Consider the equation $\mathcal{L}\mathbf{u} = \mathbf{f}$. When $\mathcal{L}$ is self-adjoint, this is usually solved with eigenvalues and eigenfunctions and these can be obtained from the resolvent $(\mathcal{L} - \lambda \mathbf{I})^{-1}$ which is an analytic function of λ and only has singularities when λ is real. When $\mathcal{L}$ is not self-adjoint one can of course look for eigenvalues and eigenfunctions of $\mathcal{L}$, but these are generally complex, and this becomes problematical when $\mathcal{L}$ is a non-self-adjoint operator and not a finite-dimensional matrix.

For non-self-adjoint operators if one is interested in the Green's function one can combine the equation $\mathcal{L}\mathbf{u} = \mathbf{f}$ with the adjoint equation $\mathcal{L}^\dagger \mathbf{u}' = \mathbf{f}'$ where $\dagger$ denotes adjoint, and then to rewrite the equation as an equation involving a self-adjoint operator. I learned of the essence of this trick when I visited Andrej Cherkaev and Leonid Gibiansky in Russia in 1987. They were concerned with equations where the tensor entering the constitutive equation was symmetric but complex, and their goal was not finding Green's functions, but rather to derive minimization variational principles for such problems. It was not until 1994 when their work was published (Cherkaev and Gibiansky 1994). In the meantime I realized their approach could be extended to any equation where the tensor entering the constitutive law is not self-adjoint (see Milton 1990), and later Fannjiang and Papanicolaou (1994) and Norris (1997) developed the idea further but still in the context of

variational principles, not Green's functions. The transformation has been successfully used to bound effective complex dielectric tensors of composites (Milton 1990), the effective complex bulk and shear moduli of composites (Gibiansky and Milton 1993; Gibiansky and Lakes 1993; Milton and Berryman 1997; Gibiansky, Milton, and Berryman 1999), and the effective moduli for magnetotransport in strong magnetic fields (Briane and Milton 2011). It has also led to minimization principles for acoustics, elastodynamics, and electromagnetism in inhomogeneous bodies containing lossy media (Milton, Seppecher, and Bouchitté 2009; Milton and Willis 2010) used by Richins and Dobson (2012).

As was pointed out to me independently by Jim Keener, of the University of Utah, and Mihai Putinar, of the University of Santa Barbara, after this chapter was written, a related idea is simply to write the equation and adjoint equation in block form,

$$\begin{pmatrix} 0 & \boldsymbol{\mathcal{L}}^\dagger \\ \boldsymbol{\mathcal{L}} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{u}' \end{pmatrix} = \begin{pmatrix} \mathbf{f}' \\ \mathbf{f} \end{pmatrix}, \tag{14.1}$$

where the operator on the left is clearly self-adjoint, and so in this form one can directly try to use the spectral theory of self-adjoint operators to obtain the Green's function. According to Mihai Putinar this is common knowledge among experts in spectral theory, but I was unable to find an exact reference in the literature. The procedure I use here is more complicated, and it is not clear if there is any advantage to using it rather than (14.1) but the success of the transformations of Cherkaev and Gibiansky suggest that there could be some advantage.

It is to be emphasized that we are not developing spectral theory for non-self-adjoint operators. Spectral theory allows one to easily take powers of an operator, or indeed take any polynomial expression of it. For dissipative operators, which (modulo multiplication by a complex number) have a positive semi-definite imaginary part, Gohberg and Kreĭn (1969) show that one can embed the Hilbert space on which $\boldsymbol{\mathcal{L}}$ acts in a larger Hilbert space and find a Hermitian operator $\mathbf{H}$ such that $f(\mathbf{H})\mathbf{x} = f(\boldsymbol{\mathcal{L}})\mathbf{x}$ for all $\mathbf{x}$ in the Hilbert space where $\boldsymbol{\mathcal{L}}$ acts. The spectral theory for $\mathbf{H}$ then allows one to compute $f(\mathbf{H})\mathbf{x}$. The Cayley transform $\boldsymbol{\mathcal{L}}' = (\mathbf{I} - \boldsymbol{\mathcal{L}})(\mathbf{I} + \boldsymbol{\mathcal{L}})^{-1}$ maps dissipative operators to contractive ones, with operator norm less than 1, and there is an analogous result for contractive operators: the Bela Sz.-Nagy dilation theorem (Sz.-Nagy and Foiaş 1970) embeds the Hilbert space on which $\boldsymbol{\mathcal{L}}'$ acts in a larger Hilbert space and finds a unitary operator $\mathbf{U}$ (satisfying $\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$) such that $f(\mathbf{U})\mathbf{x} = f(\boldsymbol{\mathcal{L}}')\mathbf{x}$ for all $\mathbf{x}$ in the Hilbert space where $\boldsymbol{\mathcal{L}}'$ acts. However, as pointed out to me by Kirill Cherednichenko of Bath University, there is no easy way of passing between the results of Gohberg and Kreĭn (1969) and Sz.-Nagy and Foiaş (1970).

This chapter is mostly self-contained, although the reader is advised to begin by studying **Chapters 1** and **2**, and Section 5.2.

14.2 Green's functions in homogeneous media

The Green's function $V(\mathbf{x})$ for the Poisson's equation satisfies

$$\nabla^2 V(\mathbf{x}) = \delta(\mathbf{x}), \tag{14.2}$$

where $\delta(\mathbf{x})$ is the delta function and in dimension three the solution is well known to be $V(\mathbf{x}) = -1/(4\pi|\mathbf{x}|)$. Given any solution $V(\mathbf{x})$ to $\nabla^2 V(\mathbf{x}) = f(\mathbf{x})$, i.e.,

$$\nabla \cdot \mathbf{j}(\mathbf{x}) = f(\mathbf{x}), \quad \mathbf{j}(\mathbf{x}) = \nabla V(\mathbf{x}), \tag{14.3}$$

we can make an affine change of coordinates from $\mathbf{x}$ to $\mathbf{x}' = \mathbf{A}\mathbf{x}$, and using the chain rule of differentiation we have

$$\begin{aligned} \frac{\partial V}{\partial x_i} &= \frac{\partial x'_k}{\partial x_i}\frac{\partial V}{\partial x'_k} = A_{ki}\frac{\partial V}{\partial x'_k}, \\ \frac{\partial j_i}{\partial x_i} &= \frac{\partial x'_k}{\partial x_i}\frac{\partial j_i}{\partial x'_k} = A_{ki}\frac{\partial j_i}{\partial x'_k}. \end{aligned} \tag{14.4}$$

This implies

$$\nabla V(\mathbf{x}) = \mathbf{A}^T\nabla' V(\mathbf{A}^{-1}\mathbf{x}'), \quad f(\mathbf{A}^{-1}\mathbf{x}') = f(\mathbf{x}) = \nabla\cdot\mathbf{j} = \nabla'\cdot(\mathbf{A}\mathbf{j}), \tag{14.5}$$

where

$$\nabla' = \left(\frac{\partial}{\partial x'_1}, \frac{\partial}{\partial x'_2}, \frac{\partial}{\partial x'_3}\right). \tag{14.6}$$

Letting $V'(\mathbf{x}') = V(\mathbf{A}^{-1}\mathbf{x}') = V(\mathbf{x})$, where $V(\mathbf{x}) = -1/(4\pi|\mathbf{x}|)$ solves (14.2) we get

$$\nabla'\cdot\mathbf{A}\mathbf{A}^T\nabla' V'(\mathbf{x}') = \delta(\mathbf{A}^{-1}\mathbf{x}') = \delta(\mathbf{x}')\det(\mathbf{A}), \tag{14.7}$$

where the last identity follows from the scaling property of the delta function. Dropping the primes, and defining $\boldsymbol{\sigma} = \mathbf{A}\mathbf{A}^T$ we see that Green's function $V_0(\mathbf{x})$ for the conductivity equation in a constant anisotropic medium with conductivity $\boldsymbol{\sigma}$, which satisfies

$$\nabla\cdot\boldsymbol{\sigma}\nabla V_0(\mathbf{x}) = \delta(\mathbf{x}), \tag{14.8}$$

is given by

$$V_0(\mathbf{x}) = \frac{-1}{4\pi\sqrt{(\det\boldsymbol{\sigma})\mathbf{x}\cdot\boldsymbol{\sigma}^{-1}\mathbf{x}}}. \tag{14.9}$$

For anisotropic elasticity, or more generally for equations of the form

$$\frac{\partial}{\partial x_i}L_{i\alpha j\beta}\frac{\partial u_\beta(\mathbf{x})}{\partial x_j} + f_\alpha(\mathbf{x}) = 0, \tag{14.10}$$

where $\mathbf{L}$ is a constant tensor with elements $L_{i\alpha j\beta}$ there is no simple formula for the Green's function in an infinite three-dimensional medium, unless $\mathbf{L}$ has some special symmetries. For elasticity Fredholm (1900) transformed the three second order differential equations to one sixth order equation which he solved. An alternative approach (Lifšic and Rozencveĭg 1947; Synge 2012) is to take $\mathbf{f}(\mathbf{x}) = \mathbf{v}\delta(\mathbf{x})$, where $\delta(\mathbf{x})$ is the delta function, and $\mathbf{v}$ a constant vector, and Fourier transform both sides of (14.10) to get

$$Q_{\alpha\beta}\widehat{u}_\beta(\mathbf{k}) = v_\alpha, \tag{14.11}$$

where

$$Q_{\alpha\beta}(\mathbf{k}) = L_{i\alpha j\beta}k_i k_j \tag{14.12}$$

is called the acoustic tensor in elasticity, and

$$\widehat{\mathbf{u}}(\mathbf{k}) = \int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dx_2\int_{-\infty}^{\infty}dx_3 e^{-i\mathbf{k}\cdot\mathbf{x}}\mathbf{u}(\mathbf{x}) \tag{14.13}$$

is the Fourier transform of $\mathbf{u}(\mathbf{x})$ (the Fourier transform of the delta function being 1). Solving (14.11) for $\widehat{\mathbf{u}}(\mathbf{k})$ and applying the inverse Fourier transform gives

$$\mathbf{u}(\mathbf{x}) = \frac{1}{8\pi^3}\int_{-\infty}^{\infty} dk_1 \int_{-\infty}^{\infty} dk_2 \int_{-\infty}^{\infty} dk_3 \mathbf{Q}^{-1}(\mathbf{k})\mathbf{v}e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{14.14}$$

Using the fact that $\mathbf{Q}^{-1}(\mathbf{k})$ is homogeneous in degree -2 in $\mathbf{k}$, this reduces (Synge 2012) to

$$\mathbf{u}(\mathbf{x}) = \frac{1}{8\pi^2|x|}\int \mathbf{Q}^{-1}(\mathbf{n})\mathbf{v}\, d\theta(\mathbf{n}), \tag{14.15}$$

where the integral is around the unit circle that is the intersection of the unit sphere $\mathbf{n} = 1$ with the plane $\mathbf{x}\cdot\mathbf{n} = 0$. For elasticity Lifšic and Rozencveĭg (1947) expressed the integral in (14.15) in terms of the poles of $\mathbf{Q}^{-1}(\mathbf{n})$. For materials with transverse isotropy Kröner (1953) gave an explicit expression for the Green's function given by Fredholm (1900), which Willis (1965) showed could also be obtained from the formula of Synge (2012). For more general elastic materials, an expression for the elastic Green's function which is explicit when the poles are known (even in the degenerate case where some of the poles coincide) was obtained by Ting and Lee (1997).

Another method, which is also applicable to wave equations in infinite homogeneous media, where one wishes to find the Green's function $\mathbf{u}(\mathbf{x})$ in an infinite medium associated with a constant differential operator $\boldsymbol{\mathcal{L}}$, i.e.,

$$\boldsymbol{\mathcal{L}}\mathbf{u}(\mathbf{x}) = \delta(\mathbf{x})\mathbf{v}, \tag{14.16}$$

was introduced by Burridge (1967), Willis and Acton (1976) and Willis (1980, 1981a, 1982): see also Norris (1994) and Milton and Willis (2010). The idea is to use the plane-wave decomposition of the three-dimensional delta function (see page 680 of Gel'fand and Shilov 1964),

$$\delta(\mathbf{x}) = -\frac{1}{8\pi^2}\int_{|\boldsymbol{\xi}|=1} \delta''(\boldsymbol{\xi}\cdot\mathbf{x})\, dS. \tag{14.17}$$

This plane-wave decomposition can be obtained by starting from the easily checked identity

$$2\pi|\mathbf{x}| = \int_{|\boldsymbol{\xi}|=1} |\boldsymbol{\xi}\cdot\mathbf{x}|\, dS, \tag{14.18}$$

applying ∇^2 twice to both sides, and using the fact that $\nabla^2(1/|\mathbf{x}|) = -4\pi\delta(\mathbf{x})$.

Then instead of using a source term $\mathbf{f}(\mathbf{x}) = \delta(\mathbf{x})\mathbf{v}$, one starts with a source term $\mathbf{f}(\mathbf{x}) = \delta(\xi\cdot\mathbf{x})\mathbf{v}$, and looks for a plane wave solution $\mathbf{u}(\mathbf{x}) = \mathbf{U}(\xi\cdot\mathbf{x})$, which when plugged in the equation $\boldsymbol{\mathcal{L}}\mathbf{U} = \mathbf{f}$ gives an ordinary differential equation to solve for $\mathbf{U}$ (or a one-dimensional integral equation if there are nonlocal terms in $\boldsymbol{\mathcal{L}}$). Finally by twice differentiating the solution with respect to $y = \xi\cdot\mathbf{x}$ one gets the response to a source $\mathbf{f}(\mathbf{x}) = \delta''(\xi\cdot\mathbf{x})\mathbf{v}$, and by integrating the solution over the sphere $|\xi| = 1$ and using (14.17) one arrives at the Green's function. An essentially equivalent approach is to use Radon transforms (Wang and Achenbach 1993,1995; Buroni and Sáez 2010).

If we are interested in the Green's function in a homogeneous body, rather than in an infinite medium, we may of course take the solution for the Green's function in an infinite medium and find its boundary values. Then we should add to this a solution (usually numerical) of the equations with no sources inside the body, and appropriate boundary conditions so the total field meets the desired (homogeneous) boundary conditions for the finite body Green's function.

14.3 Symmetrizing the equation

We consider the equation $\boldsymbol{\mathcal{L}}\mathbf{u} = \mathbf{f}$ together with its adjoint equation $\boldsymbol{\mathcal{L}}^\dagger \mathbf{u}' = \mathbf{f}'$ where $\dagger$ denotes the adjoint. We may write

$$\begin{aligned} \boldsymbol{\mathcal{L}} &= (\boldsymbol{\mathcal{L}}_s + \boldsymbol{\mathcal{L}}_a), & \boldsymbol{\mathcal{L}}^\dagger &= (\boldsymbol{\mathcal{L}}_s - \boldsymbol{\mathcal{L}}_a), \\ \mathbf{u} &= (\mathbf{u}_s + \mathbf{u}_a), & \mathbf{u}' &= (\mathbf{u}_s - \mathbf{u}_a), \\ \mathbf{f} &= (\mathbf{f}_s + \mathbf{f}_a), & \mathbf{f}' &= (\mathbf{f}_s - \mathbf{f}_a), \end{aligned} \tag{14.19}$$

where

$$\begin{aligned} \boldsymbol{\mathcal{L}}_s &= (\boldsymbol{\mathcal{L}} + \boldsymbol{\mathcal{L}}^\dagger)/2, & \boldsymbol{\mathcal{L}}_a &= (\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{L}}^\dagger)/2, \\ \mathbf{u}_s &= (\mathbf{u} + \mathbf{u}')/2, & \mathbf{u}_a &= (\mathbf{u} - \mathbf{u}')/2, \\ \mathbf{f}_s &= (\mathbf{f} + \mathbf{f}')/2, & \mathbf{f}_a &= (\mathbf{f} - \mathbf{f}')/2. \end{aligned} \tag{14.20}$$

in which we have assumed the operators $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{L}}^\dagger$ have the same domain. So the equations become

$$\begin{aligned} \underbrace{(\boldsymbol{\mathcal{L}}_s + \boldsymbol{\mathcal{L}}_a)(\mathbf{u}_s + \mathbf{u}_a)} &= \mathbf{f}_s + \mathbf{f}_a, \\ \boldsymbol{\mathcal{L}}_s\mathbf{u}_s + \boldsymbol{\mathcal{L}}_s\mathbf{u}_a + \boldsymbol{\mathcal{L}}_a\mathbf{u}_s + \boldsymbol{\mathcal{L}}_a\mathbf{u}_a & \\ \underbrace{(\boldsymbol{\mathcal{L}}_s - \boldsymbol{\mathcal{L}}_a)(\mathbf{u}_s - \mathbf{u}_a)} &= \mathbf{f}_s - \mathbf{f}_a. \\ \boldsymbol{\mathcal{L}}_s\mathbf{u}_s - \boldsymbol{\mathcal{L}}_a\mathbf{u}_s + \boldsymbol{\mathcal{L}}_s\mathbf{u}_a + \boldsymbol{\mathcal{L}}_a\mathbf{u}_a & \end{aligned} \tag{14.21}$$

Adding these equations we get:

$$\boldsymbol{\mathcal{L}}_s\mathbf{u}_s + \boldsymbol{\mathcal{L}}_a\mathbf{u}_a = \mathbf{f}_s. \tag{14.22}$$

Subtracting these equations we get:

$$\boldsymbol{\mathcal{L}}_s\mathbf{u}_a + \boldsymbol{\mathcal{L}}_a\mathbf{u}_s = \mathbf{f}_a. \tag{14.23}$$

Then the equations 14.21 can be written as

$$\underbrace{\begin{pmatrix} \boldsymbol{\mathcal{L}}_s & \boldsymbol{\mathcal{L}}_a \\ -\boldsymbol{\mathcal{L}}_a & -\boldsymbol{\mathcal{L}}_s \end{pmatrix}}_{\mathbf{M}} \underbrace{\begin{pmatrix} \mathbf{u}_s \\ \mathbf{u}_a \end{pmatrix}}_{\mathbf{U}} = \underbrace{\begin{pmatrix} \mathbf{f}_s \\ -\mathbf{f}_a \end{pmatrix}}_{\mathbf{F}}. \tag{14.24}$$

The operator $\mathbf{M}$ is now self-adjoint and the regular spectral theory applies, although there is no simple relation between the spectrum of $\mathbf{M}$ and that of the original operator $\boldsymbol{\mathcal{L}}$. (However see the last section in this chapter). If we are only interested in solutions for the original problem, then we could take $\mathbf{f}' = 0$, in which case $\mathbf{f}_a = \mathbf{f}_s = \mathbf{f}/2$ so that

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} -\mathbf{f}(\mathbf{x})/2 \\ \mathbf{f}(\mathbf{x})/2 \end{pmatrix}. \tag{14.25}$$

14.4 Eigenfunctions, eigenvalues, and symmetry of the measure

Let's look at the eigenfunctions. Suppose the spectrum of the operator $\mathbf{M}$ is discrete and we have found an eigenfunction at $\lambda = \lambda_0$:

$$\begin{pmatrix} \boldsymbol{\mathcal{L}}_s & \boldsymbol{\mathcal{L}}_a \\ -\boldsymbol{\mathcal{L}}_a & -\boldsymbol{\mathcal{L}}_s \end{pmatrix} \begin{pmatrix} \mathbf{u}_s \\ \mathbf{u}_a \end{pmatrix} = \lambda_0 \begin{pmatrix} \mathbf{u}_s \\ \mathbf{u}_a \end{pmatrix}, \tag{14.26}$$

i.e.,

$$\begin{aligned} \boldsymbol{\mathcal{L}}_s \mathbf{u}_s + \boldsymbol{\mathcal{L}}_a \mathbf{u}_a &= \lambda_0 \mathbf{u}_s, \\ -\boldsymbol{\mathcal{L}}_a \mathbf{u}_s - \boldsymbol{\mathcal{L}}_s \mathbf{u}_a &= \lambda_0 \mathbf{u}_a. \end{aligned} \tag{14.27}$$

Switching these and changing signs gives:

$$\begin{aligned} \boldsymbol{\mathcal{L}}_s \mathbf{u}_a + \boldsymbol{\mathcal{L}}_a \mathbf{u}_s &= -\lambda_0 \mathbf{u}_a, \\ -\boldsymbol{\mathcal{L}}_a \mathbf{u}_a - \boldsymbol{\mathcal{L}}_s \mathbf{u}_s &= -\lambda_0 \mathbf{u}_s, \end{aligned} \tag{14.28}$$

i.e.,

$$\begin{pmatrix} \boldsymbol{\mathcal{L}}_s & \boldsymbol{\mathcal{L}}_a \\ -\boldsymbol{\mathcal{L}}_a & -\boldsymbol{\mathcal{L}}_s \end{pmatrix} \begin{pmatrix} \mathbf{u}_a \\ \mathbf{u}_s \end{pmatrix} = -\lambda_0 \begin{pmatrix} \mathbf{u}_a \\ \mathbf{u}_s \end{pmatrix}. \tag{14.29}$$

So we have also found an associated eigenvector at $\lambda = -\lambda_0$.

Now the resolvent has an integral representation formula:

$$(\mathbf{M} - \lambda \mathbf{I})^{-1} = \int_{-\infty}^{\infty} \frac{d\boldsymbol{\mu}(\lambda')}{\lambda' - \lambda} d\lambda', \tag{14.30}$$

where $\boldsymbol{\mu}(\lambda')$ is a nonnegative operator-valued measure. If the spectrum is discrete at λ_0 and there is no degeneracy, then near λ_0,

$$d\boldsymbol{\mu}(\lambda') = \begin{pmatrix} \mathbf{u}_s \mathbf{u}_s^\dagger & \mathbf{u}_s \mathbf{u}_a^\dagger \\ \mathbf{u}_a \mathbf{u}_s^\dagger & \mathbf{u}_a \mathbf{u}_a^\dagger \end{pmatrix} \delta(\lambda' - \lambda_0), \tag{14.31}$$

where we have not assumed the eigenfunction has been normalized, and near $-\lambda_0$

$$d\boldsymbol{\mu}(\lambda') = \begin{pmatrix} \mathbf{u}_a \mathbf{u}_a^\dagger & \mathbf{u}_a \mathbf{u}_s^\dagger \\ \mathbf{u}_s \mathbf{u}_a^\dagger & \mathbf{u}_s \mathbf{u}_s^\dagger \end{pmatrix} \delta(\lambda' + \lambda_0). \tag{14.32}$$

More generally, let's for simplicity assume that the spectrum is absolutely continuous,

$$d\boldsymbol{\mu}(\lambda') = \begin{pmatrix} \mathbf{A}(\lambda) & \mathbf{B}(\lambda) \\ \mathbf{B}^\dagger(\lambda) & \mathbf{C}(\lambda) \end{pmatrix} d\lambda', \tag{14.33}$$

where $\mathbf{A}(\lambda)$ and $\mathbf{C}(\lambda)$ are Hermitian and the above matrix is a positive semi-definite operator. Then the spectral measure will have the symmetry property:

$$d\boldsymbol{\mu}(-\lambda') = \begin{pmatrix} \mathbf{C}(\lambda) & \mathbf{B}^\dagger(\lambda) \\ \mathbf{B}(\lambda) & \mathbf{A}(\lambda) \end{pmatrix}. \tag{14.34}$$

To say something about the eigenvalue λ_0 lets assume the operator $\boldsymbol{\mathcal{L}}_s$ is coercive in the sense that there exists a real constant $\alpha > 0$ such that

$$(\mathbf{p}, \boldsymbol{\mathcal{L}}_s \mathbf{p}) \geq \alpha(\mathbf{p}, \mathbf{p}), \tag{14.35}$$

for all $\mathbf{p}$ in the Hilbert space. Then from (14.27) it is clear that

$$\begin{aligned} (\mathbf{u}_s, \boldsymbol{\mathcal{L}}_s \mathbf{u}_s) + (\mathbf{u}_s, \boldsymbol{\mathcal{L}}_a \mathbf{u}_a) &= \lambda_0(\mathbf{u}_s, \mathbf{u}_s), \\ (\mathbf{u}_a, \boldsymbol{\mathcal{L}}_a \mathbf{u}_s) + (\mathbf{u}_a, \boldsymbol{\mathcal{L}}_s \mathbf{u}_a) &= -\lambda_0(\mathbf{u}_a, \mathbf{u}_a). \end{aligned} \tag{14.36}$$

From the first equation it is clear that $(\mathbf{u}_s, \boldsymbol{\mathcal{L}}_a \mathbf{u}_a)$ is real since all other terms in the expression are real. This, and the fact that $\boldsymbol{\mathcal{L}}_a^\dagger = -\boldsymbol{\mathcal{L}}_a$ implies

$$(\mathbf{u}_s, \boldsymbol{\mathcal{L}}_a \mathbf{u}_a) = (\boldsymbol{\mathcal{L}}_a \mathbf{u}_a, \mathbf{u}_s) = (\mathbf{u}_a, \boldsymbol{\mathcal{L}}_a^\dagger \mathbf{u}_s) = -(\mathbf{u}_a, \boldsymbol{\mathcal{L}}_a \mathbf{u}_s). \tag{14.37}$$

So adding the two expressions in (14.36), and using the coercivity gives

$$\lambda_0[(\mathbf{u}_s, \mathbf{u}_s) - (\mathbf{u}_a, \mathbf{u}_a)] \geq \alpha[(\mathbf{u}_s, \mathbf{u}_s) + (\mathbf{u}_a, \mathbf{u}_a)], \tag{14.38}$$

or equivalently

$$(\lambda_0 - \alpha)(\mathbf{u}_s, \mathbf{u}_s) \geq (\alpha + \lambda_0)(\mathbf{u}_a, \mathbf{u}_a). \tag{14.39}$$

If $\lambda_0 \geq 0$ we see that the right side is positive, and so from the left we conclude that $\lambda_0 > \alpha$. Alternatively, if $\lambda_0 \leq 0$ we see that the left side is negative, and so from the right we conclude that $\lambda_0 < -\alpha$. In summary we have that

$$|\lambda_0| > \alpha, \tag{14.40}$$

and so the spectrum of the operator $\mathbf{M}$ is bounded away from zero, and the resolvent $(\mathbf{M} - \lambda \mathbf{I})^{-1}$ is nonsingular at $\lambda = 0$, which when applied to the Dirac delta function source term $\delta(\mathbf{x} - \mathbf{x}_0)\mathbf{v}$ gives us the Green's function.

14.5 Boundary conditions

If we are solving equations within a body Ω rather than looking for say periodic solutions, then associated with the operator is its boundary conditions. If we want to ensure the operator is self-adjoint (which we need for spectral theory), and not just formally self-adjoint, then we need to consider the boundary conditions of the operator equation.

To fix ideas lets look at the equation

$$\mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})\mathbf{E}(\mathbf{x}) + \mathbf{F}(\mathbf{x}), \tag{14.41}$$

with differential constraints on the fields $\mathbf{J}(\mathbf{x})$ and $\mathbf{E}(\mathbf{x})$ which imply

$$\mathbf{J} \in \mathcal{U} \oplus \mathcal{J}, \quad \mathbf{E} \in \mathcal{U} \oplus \mathcal{E}, \tag{14.42}$$

which we wish to solve subject to homogeneous boundary conditions which are expressed as constraints on the fields $\boldsymbol{\Gamma}_0 \mathbf{E} = \mathbf{E}_0$ and/or $\boldsymbol{\Gamma}_0 \mathbf{J} = \mathbf{J}_0$, where $\boldsymbol{\Gamma}_0$ is the projection onto $\mathcal{U}$ (these are our "applied fields"). Consider also the adjoint problem:

$$\mathbf{J}'(\mathbf{x}) = \mathbf{L}^\dagger(\mathbf{x})\mathbf{E}'(\mathbf{x}) + \mathbf{F}'(\mathbf{x}), \tag{14.43}$$

with homogeneous boundary conditions involving $\mathbf{\Gamma}_0\mathbf{E}' = \mathbf{E}'_0$ and/or $\mathbf{\Gamma}_0\mathbf{J} = \mathbf{J}_0$. Let

$$\begin{aligned}
\mathbf{L} &= \mathbf{L}_s + \mathbf{L}_a, & \mathbf{L}^\dagger &= \mathbf{L}_s - \mathbf{L}_a, \\
\mathbf{E} &= \mathbf{E}_s + \mathbf{E}_a, & \mathbf{E}' &= \mathbf{E}_s - \mathbf{E}_a, \\
\mathbf{J} &= \mathbf{J}_s + \mathbf{J}_a, & \mathbf{J}' &= \mathbf{J}_s - \mathbf{J}_a, \\
\mathbf{F} &= \mathbf{F}_s + \mathbf{F}_a, & \mathbf{F}' &= \mathbf{F}_s - \mathbf{F}_a, \\
\mathbf{E}_0 &= \mathbf{E}_{0s} + \mathbf{E}_{0a}, & \mathbf{E}'_0 &= \mathbf{E}_{0s} - \mathbf{E}_{0a}, \\
\mathbf{J}_0 &= \mathbf{J}_{0s} + \mathbf{J}_{0a}, & \mathbf{J}'_0 &= \mathbf{J}_{0s} - \mathbf{J}_{0a}.
\end{aligned} \tag{14.44}$$

Then we have

$$\begin{aligned}
\mathbf{J}_s + \mathbf{J}_a &= (\mathbf{L}_s + \mathbf{L}_a)(\mathbf{E}_s + \mathbf{E}_a) + \mathbf{F}_s + \mathbf{F}_a, \\
\mathbf{J}_s - \mathbf{J}_a &= (\mathbf{L}_s - \mathbf{L}_a)(\mathbf{E}_s - \mathbf{E}_a) + \mathbf{F}_s - \mathbf{F}_a,
\end{aligned} \tag{14.45}$$

and adding and subtracting gives

$$\begin{aligned}
2\mathbf{J}_s &= 2\mathbf{L}_s\mathbf{E}_s + 2\mathbf{L}_a\mathbf{E}_a + 2\mathbf{F}_s, \\
-2\mathbf{J}_a &= -2\mathbf{L}_s\mathbf{E}_a - 2\mathbf{L}_a\mathbf{E}_s - 2\mathbf{F}_a.
\end{aligned} \tag{14.46}$$

We write this as the formally self-adjoint linear system

$$\underbrace{\begin{pmatrix} \mathbf{J}_s \\ -\mathbf{J}_a \end{pmatrix}}_{\mathbb{J}(\mathbf{x})} = \underbrace{\begin{pmatrix} \mathbf{L}_s & \mathbf{L}_a \\ -\mathbf{L}_a & -\mathbf{L}_s \end{pmatrix}}_{\mathbb{L}(\mathbf{x})} \underbrace{\begin{pmatrix} \mathbf{E}_s \\ \mathbf{E}_a \end{pmatrix}}_{\mathbb{E}(\mathbf{x})} + \begin{pmatrix} \mathbf{F}_s \\ -\mathbf{F}_a \end{pmatrix}, \tag{14.47}$$

where the fields satisfy the differential constraints

$$\mathbf{J}_s,\ \mathbf{J}_a \in \mathcal{J}, \quad \mathbf{E}_s,\ \mathbf{E}_a \in \mathcal{E}, \tag{14.48}$$

and are subject to homogeneous boundary conditions involving

$$\mathbf{\Gamma}_0\mathbf{E}_s = \mathbf{E}_{0s}, \quad \mathbf{\Gamma}_0\mathbf{E}_a = \mathbf{E}_{0a}, \quad \mathbf{\Gamma}_0\mathbf{J}_s = \mathbf{J}_{0s}, \quad \mathbf{\Gamma}_0\mathbf{J}_a = \mathbf{J}_{0a}. \tag{14.49}$$

Now $\mathbb{L}(\mathbf{x})$ is formally self-adjoint. To see whether it is self-adjoint with boundary conditions, look at

$$\Big(\widetilde{\mathbb{E}}(\mathbf{x}), \mathbb{L}(\mathbf{x})\mathbb{E}(\mathbf{x})\Big) = \Big(\widetilde{\mathbb{E}}(\mathbf{x}), \mathbb{J}(\mathbf{x})\Big) = \Big(\mathbf{\Gamma}_0\widetilde{\mathbb{E}}, \mathbf{\Gamma}_0\mathbb{J}\Big) = \Big(\widetilde{\mathbb{E}}_0, \mathbb{J}_0\Big) = \mathbf{J}_{0s}\cdot\widetilde{\mathbf{E}}_{0s} - \mathbf{J}_{0a}\cdot\widetilde{\mathbf{E}}_{0a}, \tag{14.50}$$

where

$$\begin{aligned}
\mathbf{J}_{0s} &= \mathbf{\Gamma}_0\mathbf{J}_s = \mathbf{\Gamma}_0[(\mathbf{J} + \mathbf{J}')/2] = (\mathbf{J}_0 + \mathbf{J}'_0)/2, \\
\mathbf{J}_{0a} &= \mathbf{\Gamma}_0\mathbf{J}_a = \mathbf{\Gamma}_0[(\mathbf{J} - \mathbf{J}')/2] = (\mathbf{J}_0 - \mathbf{J}'_0)/2, \\
\widetilde{\mathbf{E}}_{0s} &= \mathbf{\Gamma}_0\widetilde{\mathbf{E}}_s = \mathbf{\Gamma}_0[(\widetilde{\mathbf{E}} + \widetilde{\mathbf{E}}')/2] = (\widetilde{\mathbf{E}}_0 + \widetilde{\mathbf{E}}'_0)/2, \\
\widetilde{\mathbf{E}}_{0a} &= \mathbf{\Gamma}_0\widetilde{\mathbf{E}}_a = \mathbf{\Gamma}_0[(\widetilde{\mathbf{E}} - \widetilde{\mathbf{E}}')/2] = (\widetilde{\mathbf{E}}_0 - \widetilde{\mathbf{E}}'_0)/2.
\end{aligned} \tag{14.51}$$

And so we have

$$\begin{aligned}\left(\widetilde{\mathbb{E}}_0, \mathbb{J}_0\right) &= \frac{1}{4}[(\widetilde{\mathbf{E}}_0 + \widetilde{\mathbf{E}}_0') \cdot (\mathbf{J}_0 + \mathbf{J}_0') - (\mathbf{J}_0 - \mathbf{J}_0') \cdot (\widetilde{\mathbf{E}}_0 - \widetilde{\mathbf{E}}_0')] \\ &= \frac{1}{4}[\mathbf{J}_0 \cdot (\widetilde{\mathbf{E}}_0 + \widetilde{\mathbf{E}}_0' - \widetilde{\mathbf{E}}_0 + \widetilde{\mathbf{E}}_0') + \mathbf{J}_0' \cdot (\widetilde{\mathbf{E}}_0' + \widetilde{\mathbf{E}}_0 + \widetilde{\mathbf{E}}_0 - \widetilde{\mathbf{E}}_0')] \\ &= \frac{1}{2}[\mathbf{J}_0 \cdot \widetilde{\mathbf{E}}_0' + \mathbf{J}_0' \cdot \widetilde{\mathbf{E}}_0]. \end{aligned} \tag{14.52}$$

Similarly it follows that

$$\left(\mathbb{L}(\mathbf{x})\widetilde{\mathbb{E}}(\mathbf{x}), \mathbb{E}(\mathbf{x})\right) = \left(\widetilde{\mathbb{J}}_0, \mathbb{E}_0\right) = \frac{1}{2}[\widetilde{\mathbf{J}}_0 \cdot \mathbf{E}_0' + \widetilde{\mathbf{J}}_0' \cdot \mathbf{E}_0]. \tag{14.53}$$

So the operator will be self-adjoint if

$$\frac{1}{2}(\mathbf{J}_0 \cdot \widetilde{\mathbf{E}}_0' + \mathbf{J}_0' \cdot \widetilde{\mathbf{E}}_0) = \frac{1}{2}(\widetilde{\mathbf{J}}_0 \cdot \mathbf{E}_0' + \widetilde{\mathbf{J}}_0' \cdot \mathbf{E}_0). \tag{14.54}$$

This will be satisfied, for example, if we impose the constraints that $\mathbf{J}_0 = \mathbf{J}_0' = \widetilde{\mathbf{J}}_0 = \widetilde{\mathbf{J}}_0' = 0$, or alternatively the constraint $\mathbf{E}_0 = \mathbf{E}_0' = \widetilde{\mathbf{E}}_0 = \widetilde{\mathbf{E}}_0' = 0$, but as we will see in our next section, other boundary conditions are possible too.

14.6 A class of examples

As a more concrete example, suppose we have the equation

$$\mathbf{J}(\mathbf{x}) = \mathbf{L}(\mathbf{x})\mathbf{E}(\mathbf{x}) + \mathbf{F}(\mathbf{x}), \tag{14.55}$$

with $\mathbf{E}(\mathbf{x})$ and $\mathbf{J}(\mathbf{x})$ subject to the differential constraints

$$\mathbf{E}(\mathbf{x}) = \begin{pmatrix} \nabla \mathbf{u} \\ \mathbf{u} \end{pmatrix}, \quad \mathbf{J}(\mathbf{x}) = \begin{pmatrix} \mathbf{G}(\mathbf{x}) \\ \nabla \cdot \mathbf{G}(\mathbf{x}) \end{pmatrix}, \tag{14.56}$$

where $\mathbf{u}$ is an $m-$component vector, $\mathbf{G}$ is a $d \times m$ matrix valued field. These fields satisfy the *key identity*

$$\int_\Omega \mathbf{E}(\mathbf{x}) \cdot \mathbf{J}(\mathbf{x}) = \int_\Omega \underbrace{(\nabla \mathbf{u}) : \mathbf{G}(\mathbf{x}) + \mathbf{u} \cdot \nabla \cdot \mathbf{G}(\mathbf{x})}_{\nabla \cdot \mathbf{Q} \text{ with } \mathbf{Q} = \mathbf{G}(\mathbf{x})\mathbf{u}(\mathbf{x})} = \int_{\partial\Omega} \mathbf{n} \cdot \underbrace{\mathbf{G}(\mathbf{x})\mathbf{u}(\mathbf{x})}_{\mathbf{Q}(\mathbf{x})}. \tag{14.57}$$

For simplicity we choose a comparison or reference medium having tensor $\mathbf{Z}_0 = \mathbf{I}$. Then the space $\mathcal{U}_0$ consists of fields such that

$$\mathbf{E}(\mathbf{x}) = \mathbf{J}(\mathbf{x}) \text{ in } \Omega, \tag{14.58}$$

that is

$$\begin{aligned} \mathbf{G}(\mathbf{x}) &= \nabla \mathbf{u}, \quad \nabla \cdot \mathbf{G}(\mathbf{x}) = \mathbf{u} \\ \implies \nabla \cdot \mathbf{G}(\mathbf{x}) &= \nabla^2 \mathbf{u} = \mathbf{u}. \end{aligned} \tag{14.59}$$

Suppose some boundary condition $\mathbf{u} = \mathbf{u}_0$ is prescribed. Consider

$$\min_{\substack{\mathbf{u} \\ \mathbf{u}=\mathbf{u}_0 \text{ on } \partial\Omega}} \int_\Omega \underbrace{\mathbf{E}(\mathbf{x}) \cdot \mathbf{E}(\mathbf{x})}_{(\nabla\mathbf{u})^2+\mathbf{u}^2}. \tag{14.60}$$

Provided $\mathbf{u}_0$ is sufficiently regular the minimum exists, is unique and at the minimum, from the Euler–Lagrange equation, the minimizer satisfies $\nabla^2\mathbf{u} = \mathbf{u}$. So for any sufficiently regular boundary condition $\mathbf{u} = \mathbf{u}_0$ there exists a unique field $\mathbf{E}_0(\mathbf{x}) \in \mathcal{U}$ which corresponds to it. Similarly if we consider

$$\min_{\substack{\mathbf{G}(\mathbf{x}) \\ \mathbf{n}\cdot\mathbf{G}(\mathbf{x})=\mathbf{t}(\mathbf{x}) \text{ on } \partial\Omega}} \int_\Omega \underbrace{\mathbf{J}(\mathbf{x}) \cdot \mathbf{J}(\mathbf{x}),}_{\mathbf{G}(\mathbf{x}):\mathbf{G}(\mathbf{x})+[\nabla\cdot\mathbf{G}(\mathbf{x})]\cdot[\nabla\cdot\mathbf{G}(\mathbf{x})]} , \tag{14.61}$$

then if $\mathbf{t}(\mathbf{x})$ is sufficiently regular the minimum exists, is unique and from the Euler–Lagrange equations the minimizer satisfies

$$\mathbf{G}(\mathbf{x}) = \nabla(\nabla \cdot \mathbf{G}(\mathbf{x})). \tag{14.62}$$

That is, if we define $\mathbf{u}(\mathbf{x})$ to be $\nabla \cdot \mathbf{G}(\mathbf{x})$, then from 14.62 we have $\mathbf{G}(\mathbf{x}) = \nabla\mathbf{u}(\mathbf{x})$, and from the definition of $\mathbf{u}(\mathbf{x})$

$$\mathbf{u}(\mathbf{x}) = \nabla \cdot \mathbf{G}(\mathbf{x}) = \nabla^2\mathbf{u}(\mathbf{x}). \tag{14.63}$$

So for any boundary condition $\mathbf{n} \cdot \mathbf{G}(\mathbf{x}) = \mathbf{t}(\mathbf{x})$ there exists a unique field $\mathbf{J}_0(\mathbf{x}) \in \mathcal{U}$ which corresponds to it. The condition 14.54 for the self-adjoint operator $\mathbb{L}(\mathbf{x})$ to be self-adjoint in this class of examples reduces to

$$\int_{\partial\Omega} \mathbf{t}_0 \cdot \widetilde{\mathbf{u}}_0' + \mathbf{t}_0' \cdot \widetilde{\mathbf{u}}_0 = \int_{\partial\Omega} \widetilde{\mathbf{t}}_0 \cdot \mathbf{u}_0' + \widetilde{\mathbf{t}}_0' \cdot \mathbf{u}_0, \tag{14.64}$$

where we have used the key identity to transform volume integrals into surface integrals, and we have defined $\mathbf{t} = \mathbf{n} \cdot \mathbf{G}$ with $\mathbf{n}$ being the outward normal to the surface. The operator will be self-adjoint if, for example, $\mathbf{t}_0 = \widetilde{\mathbf{t}}_0 = \mathbf{t}_0' = \widetilde{\mathbf{t}}_0' = 0$ which corresponds to Neumann boundary conditions, or if $\mathbf{u}_0 = \widetilde{\mathbf{u}}_0 = \mathbf{u}_0' = \widetilde{\mathbf{u}}_0' = 0$ which corresponds to Dirichlet boundary conditions, or if $\mathbf{t}_0 = \beta\mathbf{u}_0, \widetilde{\mathbf{t}}_0 = \beta\widetilde{\mathbf{u}}_0, \mathbf{t}_0' = -\beta\mathbf{u}_0', \widetilde{\mathbf{t}}_0' = -\beta\widetilde{\mathbf{u}}_0'$ which corresponds to Robin boundary conditions.

14.7 A specific example

On July 30th, 2015, after emailing my notes to Jim Keener at the University of Utah, he asked about how one would apply it to a one-dimensional problem with the operator

$$\boldsymbol{\mathcal{L}}u = a\frac{du}{dx} + \frac{d^2u}{dx^2}, \tag{14.65}$$

with boundary constraints

$$au + \frac{du}{dx} = 0 \quad \text{at } x = 0 \text{ and } x = 1. \tag{14.66}$$

where the given function $a = a(x)$ depends on x. He mentioned on August 25th, 2015 that these boundary conditions correspond to a flux, or current $au + \frac{du}{dx}$ which is zero at the boundary. So it makes sense to introduce the current

$$j(x) = au + \frac{du}{dx}, \tag{14.67}$$

and to rewrite the equation $\boldsymbol{\mathcal{L}}u = f$ as

$$\underbrace{\begin{pmatrix} j \\ \frac{dj}{dx} \end{pmatrix}}_{\mathbf{J}(x)} = \underbrace{\begin{pmatrix} 1 & a \\ 0 & \frac{da}{dx} \end{pmatrix}}_{\mathbf{L}(x)} \underbrace{\begin{pmatrix} \frac{du}{dx} \\ u \end{pmatrix}}_{\mathbf{E}(x)} + \underbrace{\begin{pmatrix} 0 \\ f \end{pmatrix}}_{\mathbf{F}(x)}. \tag{14.68}$$

Taking a reference medium with tensor $\mathbf{L}_0 = \mathbf{I}$, we define $\mathcal{U}$ as the space of fields on $[0,1]$ which solve the homogeneous problem

$$\begin{pmatrix} j_0 \\ \frac{dj_0}{dx} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{du_0}{dx} \\ u_0 \end{pmatrix}, \tag{14.69}$$

i.e.,

$$\begin{aligned} \frac{d^2u_0}{dx^2} = u_0 \implies \qquad & u_0 = k_1 e^x + k_2 e^{-x} = \frac{dj_0}{dx}, \\ & \frac{du_0}{dx} = k_1 e^x - k_2 e^{-x} = j_0, \end{aligned} \tag{14.70}$$

i.e., $\mathcal{U}$ is spanned by the fields

$$e^x \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and } e^{-x} \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{14.71}$$

Let $\mathcal{E}$ be the space of fields spanned by

$$\mathbf{E}(x) = \begin{pmatrix} \frac{du}{dx} \\ u \end{pmatrix} \quad \text{with } u(0) = u(1) = 0, \tag{14.72}$$

as u varies. Let $\mathcal{J}$ be the space of fields spanned by

$$\mathbf{J}(x) = \begin{pmatrix} j \\ \frac{dj}{dx} \end{pmatrix} \quad \text{with } j(0) = j(1) = 0, \tag{14.73}$$

as j varies. Note that $\mathcal{E}, \mathcal{J}$ and $\mathcal{U}$ are all orthogonal spaces with respect to the inner product

$$\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \int_0^1 p_1(x)q_1(x) + p_2(x)q_2(x)dx, \tag{14.74}$$

as can be seen from the identities

$$\begin{aligned} \int_0^1 \mathbf{E}(x) \cdot \mathbf{J}(x) &= \int_0^1 \frac{d}{dx}(uj) = 0, \\ \int_0^1 \mathbf{E}(x) \cdot \begin{pmatrix} j_0 \\ \frac{dj_0}{dx} \end{pmatrix} &= \int_0^1 \frac{d}{dx}(uj_0) = 0, \\ \int_0^1 \begin{pmatrix} \frac{du_0}{dx} \\ u_0 \end{pmatrix} \cdot \mathbf{J}(x)dx &= \int_0^1 \frac{d}{dx}(u_0 j) = 0. \end{aligned} \tag{14.75}$$

We solve the equation 14.68 subject to the boundary condition $j(x) = 0$ at $x = 0$ and $x = 1$. The equation (14.47), to which we can apply spectral analysis, takes the form

$$\underbrace{\begin{pmatrix} j_s \\ \frac{dj_s}{dx} \\ -j_a \\ -\frac{dj_a}{dx} \end{pmatrix}}_{\mathbb{J}(\mathbf{x})} = \underbrace{\begin{pmatrix} 1 & a/2 & 0 & a/2 \\ a/2 & \frac{da}{dx} & -a/2 & 0 \\ 0 & -a/2 & -1 & -a/2 \\ a/2 & 0 & -a/2 & -\frac{da}{dx} \end{pmatrix}}_{\mathbb{L}(\mathbf{x})} \underbrace{\begin{pmatrix} \frac{du_s}{dx} \\ u_s \\ \frac{du_a}{dx} \\ u_a \end{pmatrix}}_{\mathbb{E}(\mathbf{x})} + \begin{pmatrix} 0 \\ f/2 \\ 0 \\ -f/2 \end{pmatrix}. \tag{14.76}$$

14.8 Recovering the eigenvalues and eigenvectors of the original equation

Here we follow Section 18 of Milton (1990) to obtain an equation for the eigenvalues and eigenvectors of the original operator $\boldsymbol{\mathcal{L}}$ and its adjoint $\boldsymbol{\mathcal{L}}^\dagger$. As is evident from the characteristic equation, recall that if λ is an eigenvalue of $\boldsymbol{\mathcal{L}}$ then its complex conjugate $\overline{\lambda}$ is an eigenvalue of $\boldsymbol{\mathcal{L}}^\dagger$. Let $\mathbf{u}$ and $\mathbf{u}'$ be corresponding eigenvectors so that

$$(\boldsymbol{\mathcal{L}}_s + \boldsymbol{\mathcal{L}}_a)\mathbf{u} = \lambda\mathbf{u}, \quad (\boldsymbol{\mathcal{L}}_s - \boldsymbol{\mathcal{L}}_a)\mathbf{u}' = \overline{\lambda}\mathbf{u}'. \tag{14.77}$$

Replacing $\mathbf{f}$ by $\lambda\mathbf{u}$ and $\mathbf{f}'$ by $\overline{\lambda}\mathbf{u}'$ we see that (14.24) implies

$$\begin{pmatrix} \boldsymbol{\mathcal{L}}_s & \boldsymbol{\mathcal{L}}_a \\ -\boldsymbol{\mathcal{L}}_a & -\boldsymbol{\mathcal{L}}_s \end{pmatrix} \begin{pmatrix} \mathbf{u} + \mathbf{u}' \\ \mathbf{u} - \mathbf{u}' \end{pmatrix} = \begin{pmatrix} \lambda\mathbf{u} + \overline{\lambda}\mathbf{u}' \\ -\lambda\mathbf{u} + \overline{\lambda}\mathbf{u}' \end{pmatrix}. \tag{14.78}$$

Assuming the operator $\boldsymbol{\mathcal{L}}$ is bounded and that $\boldsymbol{\mathcal{L}}_s$ is coercive, this can be rewritten, using the transformation of Cherkaev and Gibiansky (1994) and Milton (1990) as

$$\boldsymbol{\mathcal{K}} \begin{pmatrix} \mathbf{u} + \mathbf{u}' \\ -\lambda\mathbf{u} + \overline{\lambda}\mathbf{u}' \end{pmatrix} = \begin{pmatrix} \lambda\mathbf{u} + \overline{\lambda}\mathbf{u}' \\ \mathbf{u}' - \mathbf{u} \end{pmatrix}, \tag{14.79}$$

where $\boldsymbol{\mathcal{K}}$ is the operator

$$\boldsymbol{\mathcal{K}} = \begin{pmatrix} \boldsymbol{\mathcal{L}}_s - \boldsymbol{\mathcal{L}}_a\boldsymbol{\mathcal{L}}_s^{-1}\boldsymbol{\mathcal{L}}_a & \boldsymbol{\mathcal{L}}_a\boldsymbol{\mathcal{L}}_s^{-1} \\ -\boldsymbol{\mathcal{L}}_s^{-1}\boldsymbol{\mathcal{L}}_a & \boldsymbol{\mathcal{L}}_s^{-1} \end{pmatrix}, \tag{14.80}$$

which is positive definite, and for this reason the transformation to $\boldsymbol{\mathcal{K}}$ has proved important for deriving bounds, via variational principles Note that the rule for the fields in going from (14.78) to (14.79) is to leave the "top" component of the fields, i.e., $\mathbf{u} + \mathbf{u}'$ and $\lambda\mathbf{u} + \overline{\lambda}\mathbf{u}'$ in the same positions, while switching positions of the bottom components $\mathbf{u} - \mathbf{u}'$ and $-\lambda\mathbf{u} + \overline{\lambda}\mathbf{u}$, and changing the sign of the component on the bottom right of the second constitutive law from $\mathbf{u} - \mathbf{u}'$ to $\mathbf{u}' - \mathbf{u}$.

At this point it is interesting to investigate the eigenvalues and eigenvectors of $\boldsymbol{\mathcal{K}}$. These satisfy

$$\boldsymbol{\mathcal{K}} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = \eta \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix}, \tag{14.81}$$

where η is the eigenvalue. Transforming back gives

$$\begin{pmatrix} \boldsymbol{\mathcal{L}}_s & \boldsymbol{\mathcal{L}}_a \\ -\boldsymbol{\mathcal{L}}_a & -\boldsymbol{\mathcal{L}}_s \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ -\eta\mathbf{w} \end{pmatrix} = \begin{pmatrix} \eta\mathbf{v} \\ \mathbf{w} \end{pmatrix}, \tag{14.82}$$

or equivalently

$$\mathcal{L}_s\mathbf{v} - \eta\mathcal{L}_a\mathbf{w} = \eta\mathbf{v}, \quad \eta\mathcal{L}_s\mathbf{w} - \mathcal{L}_a\mathbf{v} = \mathbf{w}. \tag{14.83}$$

Defining

$$\mathbf{v}' = \eta\mathbf{w}, \quad \mathbf{w}' = \eta\mathbf{v}, \quad \eta' = 1/\eta, \tag{14.84}$$

these equations become

$$\eta'\mathcal{L}_s\mathbf{w}' - \mathcal{L}_a\mathbf{v}' = \mathbf{w}', \quad \mathcal{L}_s\mathbf{v}' - \eta'\mathcal{L}_a\mathbf{w}' = \eta'\mathbf{v}'. \tag{14.85}$$

These equations are exactly of the same form as (14.83) but with primes on the fields. Hence we directly have that (14.81) is satisfied with $\mathbf{v}$, $\mathbf{w}$ and η replaced by $\mathbf{v}'$, $\mathbf{w}'$, and η', giving

$$\mathcal{K}\begin{pmatrix}\mathbf{w}\\ \mathbf{v}\end{pmatrix} = \frac{1}{\eta}\begin{pmatrix}\mathbf{w}\\ \mathbf{v}\end{pmatrix}. \tag{14.86}$$

So the eigenvalues of $\mathcal{K}$ come in reciprocal pairs.

Now following Section 18 of Milton (1990), let us introduce the family of operators $\mathbf{A}(\alpha)$

$$\mathbf{A}(\alpha) = \begin{pmatrix}\mathbf{I} & \mathbf{I}\\ -\alpha\mathbf{I} & \overline{\alpha}\mathbf{I}\end{pmatrix}, \tag{14.87}$$

which have inverses

$$[\mathbf{A}(\alpha)]^{-1} = \frac{1}{2\,\mathrm{Re}\,\alpha}\begin{pmatrix}\overline{\alpha}\mathbf{I} & -\mathbf{I}\\ \alpha\mathbf{I} & \mathbf{I}\end{pmatrix} = \frac{1}{2\,\mathrm{Re}\,\alpha}\begin{pmatrix}\alpha\mathbf{I} & \overline{\alpha}\mathbf{I}\\ -\mathbf{I} & \mathbf{I}\end{pmatrix}^{\dagger}. \tag{14.88}$$

Using these relations (14.79) becomes

$$2\,\mathrm{Re}\,\lambda[\mathbf{A}(\lambda)^{\dagger}]^{-1}\begin{pmatrix}\mathbf{u}\\ \mathbf{u}'\end{pmatrix} = \mathcal{K}\mathbf{A}(\lambda)\begin{pmatrix}\mathbf{u}\\ \mathbf{u}'\end{pmatrix}. \tag{14.89}$$

So if we let $\mathbf{B}(\alpha)$ denote the Hermitian operator

$$\mathbf{B}(\alpha) = \frac{1}{2\,\mathrm{Re}\,\alpha}\mathbf{A}(\alpha)^{\dagger}\mathcal{K}\mathbf{A}(\alpha), \tag{14.90}$$

then (14.89) implies

$$\mathbf{B}(\lambda)\begin{pmatrix}\mathbf{u}\\ \mathbf{u}'\end{pmatrix} = \begin{pmatrix}\mathbf{u}\\ \mathbf{u}'\end{pmatrix}. \tag{14.91}$$

Note that $\mathbf{B}(\alpha)$ is positive or negative definite according to whether $\mathrm{Re}\,\alpha > 0$ or $\mathrm{Re}\,\alpha < 0$ and hence can only have an eigenvalue of 1 if $\mathrm{Re}\,\alpha > 0$. As discussed in Section 18 of Milton (1990) one can vary α in the right half plane until it has eigenvalue 1. Then α and the eigenvector of $\mathbf{B}(\alpha)$ give the eigenvalue λ and eigenvectors $\mathbf{u}$ and $\mathbf{u}'$ of $\mathcal{L}$ and its adjoint. Some numerical examples of this are shown in **Figure 14.1**.

Note the converse is true too. By reversing the steps in the argument one sees that if for some α with $\mathrm{Re}\,\alpha > 0$, the operator $\mathbf{B}(\alpha)$ has eigenvalue 1 then α is an eigenvalue of $\mathcal{L}$ and the corresponding eigenvector associated with $\mathbf{B}(\alpha)$ gives the eigenvectors of $\mathcal{L}$ and its adjoint.

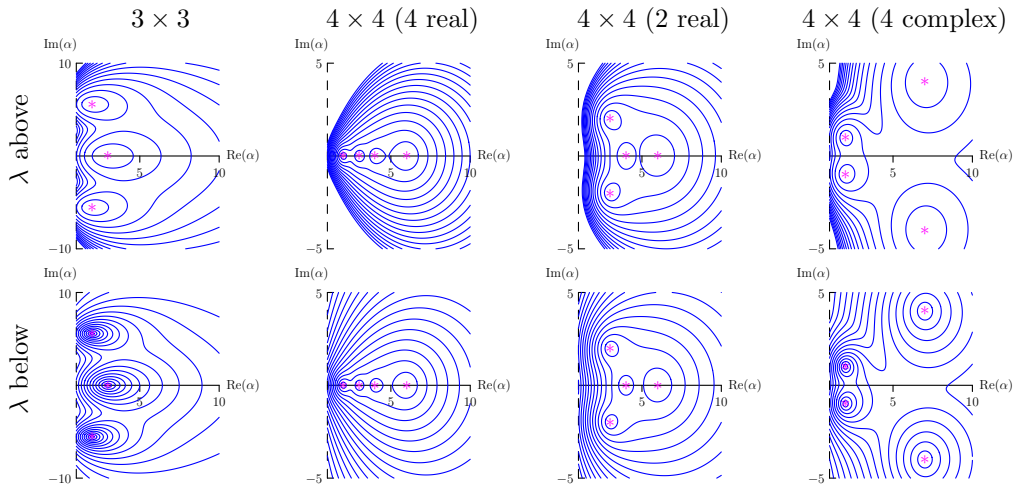


Figure 14.1: Contours, obtained by Patrick Bardsley, showing the eigenvalues just below 1 and just above 1 of the Hermitian matrix $\mathbf{B}(\alpha)$ as α is varied in the right hand side of the complex α plane. The contours encircle the eigenvalues of the matrix $\boldsymbol{\mathcal{L}}$ which are marked by asterisks, and at these points $\mathbf{B}(\alpha)$ has eigenvalue 1. From left to right $\boldsymbol{\mathcal{L}}$ is taken to be a 3×3 matrix, and then a 4×4 matrix with all real, 2 complex, and 4 complex eigenvalues.

Acknowledgments

G.W. Milton wishes to especially thank Jim Keener for pointing out the need to assume in the development of the theory for general operators $\boldsymbol{\mathcal{L}}$, that $\boldsymbol{\mathcal{L}}^\dagger$ and $\boldsymbol{\mathcal{L}}$ need to have the same domain. However, this constraint does not apply to the class of boundary value problems considered in Section 14.5, provided the condition (14.54) holds. He is grateful to Mihai Putinar for references to many results in the spectral theory of non-self-adjoint operators, which allow one to take powers of an operator. The help of Patrick Bardsley is gratefully acknowledged in producing **Figure 14.1**.

Bibliography

Each entry is followed by a braced list of page numbers where it is cited. Personal names are indexed in the separate author/editor index on page 402.
International Standard Book Numbers (ISBNs) are unique 10- and 13-digit identifiers that are convenient for searching bookseller and library-catalog Web sites. Many publishers use them in document Web addresses.
International Standard Serial Numbers (ISSNs) are unique 8-digit identifiers of periodicals, and differ for print and electronic editions. The *linking ISSN*, ISSN-L, serves as a common identifier for both forms, and in most cases, is identical to the print edition ISSN. Like ISBNs, ISSNs are useful for catalog and database searches.
Convert Digital Object Identifier (DOI) values to Web addresses by replacing the prefix `doi:` by `http://dx.doi.org/`. Uniform Resource Locator (URL) values are document Web addresses. They are omitted here if a DOI is shown, but remain available in the original bibliographic databases.
The BIBTEX files used to create this bibliography are freely available at the book's Web site.

Agler, J. and J. E. McCarthy 2002. *Pick Interpolation and Hilbert Function Spaces*, Volume 44 of *Graduate Studies in Mathematics*. Providence, RI, USA: American Mathematical Society. xix + 308 pp. ISSN-L 1065-7339. ISBN 0-8218-2898-3. ISBN-13 978-0-8218-2898-4. {*68, 151, 258, 289, 369*}

Agler, J., J. E. McCarthy, and N. J. Young 2012, November. Operator monotone functions and Löwner functions of several variables. *Annals of Mathematics (2)* **176**(**3**):1783–1826. ISSN-L 0003-486X. doi:10.4007/annals.2012.176.3.7 {*98, 369*}

Ajaib, M. A. 2015, December. A fundamental form of the Schrödinger equation. *Foundations of Physics* **45**(**12**):1586–1598. ISSN-L 0015-9018. doi:10.1007/s10701-015-9944-z {*41, 369*}

Albanese, R. and P. B. Monk 2006, May. The inverse source problem for Maxwell's equations. *Inverse Problems* **22**(**3**):1023. ISSN-L 0266-5611. doi:10.1088/0266-5611/22/3/018 {*96, 112, 369*}

Alessandrini, G. and S. Vessella 2005, August. Lipschitz stability for the inverse conductivity problem. *Advances in Applied Mathematics* **35**(**2**):207–241. ISSN-L 0196-8858. doi:10.1016/j.aam.2004.12.002 {*87, 369*}

Allaire, G. 1992. Homogenization and two-scale convergence. *SIAM Journal on Mathematical Analysis* **23**(**6**):1482–1518. ISSN-L 0036-1410. doi:10.1137/0523084 {*49, 369*}

Allaire, G. 1997. Mathematical approaches and methods. In U. Hornung (ed.), *Homogenization and Porous Media*, Volume 6 of *Interdisciplinary Applied Mathematics*, pp. 225–257 (of xvi + 279). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-94786-8. ISBN-13 978-0-387-94786-0. {*49, 369*}

Allaire, G. 2002. *Shape Optimization by the Homogenization Method*, Volume 146 of *Applied Mathematical Sciences*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xv + 456 pp. ISBN 0-387-95298-5. ISBN-13 978-0-387-95298-7. doi:10.1007/978-1-4684-9286-6 {*2, 50, 131, 150, 257, 258, 289, 339, 369*}

Allaire, G., E. Bonnetier, G. Francfort, and F. Jouve 1997, March. Shape optimization by the homogenization method. *Numerische Mathematik* **76**(**1**):27–68. ISSN-L 0029-599X. doi:10.1007/s002110050253 {*2, 369*}

Allaire, G. and M. Briane 1996. Multiscale convergence and reiterated homogenisation. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **126**(**2**):297–342. ISSN-L 0308-2105. doi:10.1017/S0308210500022757 {*49, 370*}

Allaire, G. and R. V. Kohn 1994, June. Optimal lower bounds on the elastic energy of a composite made from two non well-ordered isotropic materials. *Quarterly of Applied Mathematics* **52**(**2**):311–333. ISSN-L 0033-569X. URL http://dl.acm.org/citation.cfm?id=186414 {*18, 370*}

Alpay, D. 2001. *The Schur Algorithm, Reproducing Kernel Spaces and System Theory*, Volume 5 of *SMF/AMS Texts and Monographs*. Providence, RI, USA: American Mathematical Society. viii + 150 pp. Translated from the French by Stephen S. Wilson. ISSN-L 1525-2302. ISBN 0-8218-2155-5. ISBN-13 978-0-8218-2155-8. {*258, 370*}

Ammari, H., J. Garnier, W. Jing, H. Kang, M. Lim, K. Sølna, and H. Wang 2013. *Mathematical and Statistical Methods for Multistatic Imaging*, Volume 2098 of *Lecture Notes in Mathematics*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xvii + 361 pp. ISBN 3-319-02584-8. ISBN-13 978-3-319-02584-1. doi:10.1007/978-3-319-02585-8 {*124, 370*}

Armand, J.-L., K. A. Lurie, and A. V. Cherkaev 1984. Optimal control theory and structural design. In E. Atrek, R. H. Gallagher, K. M. Ragsdell, and O. C. Zienkiewicz (eds.), *New Directions in Optimum Structural Design*, pp. 211–229. New York / London / Sydney, Australia: John Wiley and Sons. {*2, 370*}

Astala, K. and L. Päivärinta 2006. Calderón's inverse conductivity problem in the plane. *Annals of Mathematics* **163**(**1**): 265–299. ISSN-L 0003-486X. doi:10.4007/annals.2006.163.265 {*142, 370*}

Atkinson, C. and R. V. Craster 1992, May 15. The application of invariant integrals in diffusive elastic solids. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* **339**(**1653**):231–263. ISSN-L 1364-503X. doi:10.1098/rsta.1992.0033 {*14, 370*}

Attouch, H. 1984. *Variational Convergence for Functions and Operators*. Applicable Mathematics Series. London: Pitman Publishing Ltd. xiv + 423 pp. ISSN-L 0743-0353. ISBN 0-273-08583-2. ISBN-13 978-0-273-08583-6. {*49, 370*}

Auld, B. A. 1973. *Acoustic Waves and Fields in Solids*, Volume 1. New York / London / Sydney, Australia: John Wiley and Sons. vii + 423 pp. ISBN 0-471-03700-1 (vol. 1), 0-471-03701-X (vol. 2), 0-471-03702-8 (set). ISBN-13 978-0-471-03700-2 (vol. 1), 978-0-471-03701-9 (vol. 2), 978-0-471-03702-6 (set). {*29, 370*}

Avellaneda, M. and O. P. Bruno 1990, August. Effective conductivity and average polarizability of random polycrystals. *Journal of Mathematical Physics* **31**(**8**):2047–2056. ISSN-L 0022-2488. doi:10.1063/1.528656 {*118, 370*}

Baker, Jr., G. A. 1969, May. Best error bounds for Padé approximants to convergent series of Stieltjes. *Journal of Mathematical Physics* **10**(**5**):814–820. ISSN-L 0022-2488. doi:10.1063/1.1664911 {*184, 258, 289, 370*}

Baker, Jr., G. A. and P. R. Graves-Morris 1981. *Padé Approximants: Basic Theory. Part I. Extensions and Applications. Part II*, Volume 13 & 14 of *Encyclopedia of Mathematics and its Applications*. Reading, Massachusetts: Addison-Wesley. xviii + 215 pp. With a foreword by Peter A. Carruthers. ISBN 0-201-13512-4 (part 1), 0-521-30233-1 (part 1), 0-201-13513-2 (part 2), 0-521-30234-X (part 2). ISBN-13 978-0-201-13512-1 (part 1), 978-0-521-30233-3 (part 1), 978-0-201-13513-8 (part 2), 978-0-521-30234-0 (part 2). {*184, 246, 289, 370*}

Bakhvalov, N. S. and G. P. Panasenko 1989. *Homogenisation: Averaging Processes in Periodic Media. Mathematical Problems in the Mechanics of Composite Materials*, Volume 36 of *Mathematics and its Applications (Soviet Series)*. Norwell, Massachusetts / Dordrecht, The Netherlands: Kluwer Academic Publishers Group. xxxv + 366 pp. Translated from the Russian by D. Leĭtes. ISBN 0-7923-0049-1. ISBN-13 978-0-7923-0049-6. doi:10.1007/978-94-009-2247-1 {*11, 48, 49, 370*}

Ball, J. A. and S. ter Horst 2010. *Robust control, multidimensional systems and multivariable Nevanlinna–Pick interpolation*, Volume 203 of *Operator Theory: Advances and Applications*, pp. 13–88. Basel, Switzerland: Birkhäuser Verlag. ISBN 3-0346-0160-3. ISBN-13 978-3-0346-0160-3. doi:10.1007/978-3-0346-0161-0_2 {*68, 151, 370*}

Ball, J. A. and T. T. Trent 1998, August 1. Unitary colligations, reproducing kernel Hilbert spaces, and Nevanlinna–Pick interpolation in several variables. *Journal of Functional Analysis* **157**(**1**):1–61. ISSN-L 0022-1236. doi:10.1006/jfan.1998.3278 {*68, 151, 371*}

Ball, J. M. 1977. Convexity conditions and existence theorems in non-linear elasticity. *Archive for Rational Mechanics and Analysis* **63**(**4**):337–403. ISSN-L 0003-9527. doi:10.1007/BF00279992 {*2, 14, 17, 338, 339, 371*}

Ball, J. M. 1996. Some recent developments in nonlinear elasticity and its applications to materials science. In P. J. Aston (ed.), *Nonlinear Mathematics and its Applications: A Spring School in Applied Nonlinear Mathematics for Postgraduate Students Held at the University of Surrey on 3–7 April 1995*, pp. 93–119 (of vii + 256). Cambridge, UK: Cambridge University Press. ISBN 0-521-57190-1 (hardback), 0-521-57676-8 (paperback). ISBN-13 978-0-521-57190-6 (hardback), 978-0-521-57676-5 (paperback). {*2, 371*}

Ball, J. M., J. C. Currie, and P. J. Olver 1981, April. Null Lagrangians, weak continuity, and variational problems of arbitrary order. *Journal of Functional Analysis* **41**(**2**):135–174. ISSN-L 0022-1236. doi:10.1016/0022-1236(81)90085-9 {*14, 371*}

Ball, J. M. and R. D. James 1987. Fine phase mixtures as minimizers of energy. *Archive for Rational Mechanics and Analysis* **100**(**1**):13–52. ISSN-L 0003-9527. doi:10.1007/BF00281246 {*2, 17, 339, 371*}

Ball, J. M. and R. D. James 1992, February. Proposed experimental tests of a theory of fine microstructure and the two-well problem. *Philosophical Transactions of the Royal Society of London Series A* **338**(**1650**):389–450. ISSN-L 1364-503X. doi:10.1098/rsta.1992.0013 {*2, 339, 371*}

Ball, J. M. and F. Murat 1984, October. $W^{1,p}$-quasiconvexity and variational problems for multiple integrals. *Journal of Functional Analysis* **58**(**3**):225–253. ISSN-L 0022-1236. doi:10.1016/0022-1236(84)90041-7 {*2, 371*}

Banerjee, B. 2011. *An Introduction to Metamaterials and Waves in Composites*. Boca Raton, Florida: CRC Press. xiii + 361 pp. ISBN 1-4398-4157-8 (hardcover). ISBN-13 978-1-4398-4157-0 (hardcover). {*52, 371*}

Barabash, S. and D. Stroud 1999, December 20. Spectral representation for the effective macroscopic response of a polycrystal: application to third-order non-linear susceptibility. *Journal of Physics: Condensed Matter* **11**(**50**):10323. See also the erratum in arXiv:cond-mat/9910246v2. ISSN-L 0953-8984. URL http://stacks.iop.org/0953-8984/11/i=50/a=324 {*118, 371*}

Barretta, R., R. Luciano, and J. R. Willis 2015, November 15. On torsion of random composite beams. *Composite Structures* **132**:915–922. ISSN-L 0263-8223. doi:10.1016/j.compstruct.2015.06.069 {*154, 371*}

Batchelor, G. K. 1974. Transport properties of two-phase materials with random structure. *Annual Review of Fluid Mechanics* **6**:227–255. ISSN-L 0066-4189. doi:10.1146/annurev.fl.06.010174.001303 {*5, 371*}

Baumgärtel, H. 1985. *Analytic Perturbation Theory for Matrices and Operators*, Volume 15 of *Operator Theory*. Basel, Switzerland: Birkhäuser Verlag. 427 pp. ISBN 3-7643-1664-0. ISBN-13 978-3-7643-1664-8. {*96, 371*}

Becke, A. D. 1993, April. Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics* **98**(**7**):5648–5652. ISSN-L 0021-9606. doi:10.1063/1.464913 {*309, 371*}

Bendsøe, M. P. and O. Sigmund 2004. *Topology Optimization: Theory, Methods, and Applications* (Second ed.). Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xiv + 370 pp. ISBN 3-540-42992-1 (hardcover), 3-642-07698-X (softcover 2011). ISBN-13 978-3-540-42992-0 (hardcover), 978-3-642-07698-5 (softcover 2011). doi:10.1007/978-3-662-05086-6 {*2, 371*}

Bensoussan, A., J.-L. Lions, and G. C. Papanicolaou 1978. *Asymptotic Analysis for Periodic Structures*, Volume 5 of *Studies in Mathematics and its Applications*. Amsterdam: North-Holland Publishing Co. xxiv + 700 pp. ISBN 0-444-85172-0. ISBN-13 978-0-444-85172-7. {*11, 48, 49, 264, 371*}

Benveniste, Y. and T. Chen 2003, December. The Saint-Venant torsion of a circular bar consisting of a composite cylinder assemblage with cylindrically orthotropic constituents. *International Journal of Solids and Structures* **40**(**25**):7093–7107. ISSN-L 0020-7683. doi:10.1016/S0020-7683(03)00356-1 {*154, 371*}

Beran, M. J. 1965. Use of the variational approach to determine bounds for the effective permittivity in random media. *Il Nuovo Cimento (1955–1965)* **38**(**2**):771–782. ISSN-L 1827-6121. doi:10.1007/BF02748596 {*54, 55, 67, 257, 279, 372*}

Beran, M. J. 1968. *Statistical Continuum Theories*, Volume 9 of *Monographs in Statistical Physics and Thermodynamics*. New York: Interscience Publishers. xv + 424 pp. ISBN 0-470-06861-2. ISBN-13 978-0-470-06861-8. {*48, 49, 257, 288, 372*}

Beran, M. J. and J. J. McCoy 1970, August. Mean field variations in a statistical sample of heterogeneous linear elastic solids. *International Journal of Solids and Structures* **6**(**8**):1035–1054. ISSN-L 0020-7683. doi:10.1016/0020-7683(70)90046-6 {*257, 288, 372*}

Beran, M. J. and J. Molyneux 1963, December. Statistical properties of the electric field in a medium with small random variations in permittivity. *Il Nuovo Cimento (10)* **30**(**6**):1406–1422. ISSN-L 0029-6341. doi:10.1007/BF02749819 {*257, 288, 372*}

Beran, M. J. and J. Molyneux 1966. Use of classical variational principles to determine bounds for the effective bulk modulus in heterogeneous media. *Quarterly of Applied Mathematics* **24**:107–118. ISSN-L 0033-569X. {*54, 55, 279, 372*}

Beretta, E., M. V. de Hoop, and L. Qiu 2013. Lipschitz stability of an inverse boundary value problem for a Schrödinger-type equation. *SIAM Journal on Mathematical Analysis* **45**(**2**):679–699. ISSN-L 0036-1410. doi:10.1137/120869201 {*87, 372*}

Beretta, E. and E. Francini 2011. Lipschitz stability for the electrical impedance tomography problem: The complex case. *Communications in Partial Differential Equations* **36**(**10**):1723–1749. ISSN-L 0360-5302. doi:10.1080/03605302.2011.552930 {*87, 372*}

Beretta, E., E. Francini, and S. Vessella 2014, August. Uniqueness and Lipschitz stability for the identification of Lamé parameters from boundary measurements. *Inverse Problems and Imaging* **8**(**3**):611–644. ISSN-L 1930-8337. doi:10.3934/ipi.2014.8.611 {*87, 372*}

Berg, C. 2008. Stieltjes–Pick–Bernstein–Schoenberg and their connection to complete monotonicity. In J. Mateu and E. Porcu (eds.), *Positive Definite Functions: From Schoenberg to Space–Time Challenges*, pp. 15–45 (of xiv + 198). Castellón de la Plana, Spain: Editorial Universitat Jaume I, Department of Mathematics. ISBN 84-612-8282-5. ISBN-13 978-84-612-8282-1. URL http://www.math.ku.dk/~berg/manus/castellon.pdf {*39, 98, 372*}

Bergman, D. J. 1976, August 15. Variational bounds on some bulk properties of a two-phase composite material. *Physical Review B: Solid State* **14**(**4**):1531–1542. ISSN-L 0556-2805. doi:10.1103/PhysRevB.14.1531 {*258, 372*}

Bergman, D. J. 1978, July. The dielectric constant of a composite material — A problem in classical physics. *Physics Reports* **43**(**9**):377–407. ISSN-L 0370-1573. doi:10.1016/0370-1573(78)90009-1 {*62, 68, 96, 116, 136, 140, 149, 150, 153, 156, 167–169, 178, 183, 236, 258, 288, 372*}

Bergman, D. J. 1980, May 12. Exactly solvable microscopic geometries and rigorous bounds for the complex dielectric constant of a two-component composite material. *Physical Review Letters* **44**(**19**):1285–1287. ISSN-L 0031-9007. doi:10.1103/PhysRevLett.44.1285 {*68, 96, 149, 151, 152, 178, 372*}

Bergman, D. J. 1982. Rigorous bounds for the complex dielectric constant of a two-component composite. *Annals of Physics* **138**(**1**):78–114. ISSN-L 0003-4916. doi:10.1016/0003-4916(82)90176-2 {*68, 96, 151, 372*}

Bergman, D. J. 1986. The effective dielectric coefficient of a composite medium: Rigorous bounds from analytic properties. In J. L. Ericksen, D. Kinderlehrer, R. V. Kohn, and J.-L. Lions (eds.), *Homogenization and Effective Moduli of Materials and Media*, Volume 1 of *The IMA Volumes in Mathematics and its Applications*, pp. 27–51 (of x + 263). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-96306-5. ISBN-13 978-0-387-96306-8. doi:10.1007/978-1-4613-8646-9 {*96, 184, 372*}

Bergman, D. J. 1993. Hierarchies of Stieltjes functions and their application to the calculation of bounds for the dielectric constant of a two-component composite medium. *SIAM Journal on Applied Mathematics* **53**(**4**):915–930. ISSN-L 0036-1399. doi:10.1137/0153045 {*68, 136, 288, 372*}

Beris, A. N. and B. J. Edwards 1993, September. On the admissibility criteria for linear viscoelasticity kernels. *Rheologica Acta* **32**(**5**):505–510. ISSN-L 0035-4511. doi:10.1007/BF00396182 {*152, 373*}

Bernland, A., A. Luger, and M. Gustafsson 2011, April 8. Sum rules and constraints on passive systems. *Journal of Physics A: Mathematical and Theoretical* **44**(**14**):145205. ISSN-L 1751-8113. doi:10.1088/1751-8113/44/14/145205 {*96, 110, 116, 373*}

Berreman, D. W. 1972. Optics in stratified and anisotropic media: 4×4-matrix formulation. *Journal of the Optical Society of America* **62**(**4**):502–510. ISSN-L 0030-3941. {*101, 373*}

Berryman, J. G. 1982. Effective medium theory for elastic composites. In V. K. Varadan and V. V. Varadan (eds.), *Elastic Wave Scattering and Propagation: Based on Presentations made at a Special Session of the Midwestern Mechanics Conference Held at the University of Michigan, Ann Arbor, May 7–9, 1981*, pp. 111–129 (of v + 210). Ann Arbor, Michigan: Ann Arbor Science. ISBN 0-250-40534-2. ISBN-13 978-0-250-40534-3. {*55, 67, 373*}

Berryman, J. G. 2009, April. Frequency dependent thermal expansion in binary viscoelastic composites. *Mechanics of Materials: an International Journal* **41**(**4**):463–480. Special Issue in Honor of Graeme W. Milton, 2007 Winner of the William Prager Medal of the Society of Engineering Science. ISSN-L 0167-6636. doi:10.1016/j.mechmat.2009.01.013 {*10, 373*}

Berryman, J. G. and R. V. Kohn 1990, July 16. Variational constraints for electrical-impedance tomography. *Physical Review Letters* **65**(**3**):325–328. ISSN-L 0031-9007. doi:10.1103/PhysRevLett.65.325 {*129, 373*}

Berryman, J. G. and G. W. Milton 1988. Microgeometry of random composites and porous media. *Journal of Physics D: Applied Physics* **21**(**1**):87–94. ISSN-L 0022-3727. doi:10.1088/0022-3727/21/1/013 {*55, 373*}

Bhattacharya, K. 1993, September. Comparison of the geometrically nonlinear and linear theories of martensitic transformation. *Continuum Mechanics and Thermodynamics* **5**(**3**):205–242. ISSN-L 0935-1175. doi:10.1007/BF01126525 {*2, 373*}

Bhattacharya, K., N. B. Firoozye, R. D. James, and R. V. Kohn 1994. Restrictions on microstructure. *Proceedings of the Royal Society of Edinburgh* **124A**(**5**):843–878. ISSN-L 0080-4541. doi:10.1017/S0308210500022381 {*2, 373*}

Bhattacharya, K. and R. V. Kohn 1997, October. Elastic energy minimization and the recoverable strains of polycrystalline shape-memory materials. *Archive for Rational Mechanics and Analysis* **139**(**2**):99–180. ISSN-L 0003-9527. doi:10.1007/s002050050049 {*2, 19, 373*}

Biot, M. A. 1962, April 1. Mechanics of deformation and acoustic propogation in porous media. *Journal of Applied Physics* **33**(**4**):1482–1498. ISSN-L 0021-8979. doi:10.1063/1.1728759 {*9, 33, 373*}

Boley, B. A. and J. H. Weiner 1997. *Theory of Thermal Stresses*. New York: Dover Publications, Inc. xvi + 586 pp. ISBN 0-486-69579-4 (paperback). ISBN-13 978-0-486-69579-2 (paperback). {*8, 373*}

Bonifasi-Lista, C. and E. Cherkaev 2008, December. Analytical relations between effective material properties and microporosity: Application to bone mechanics. *International Journal of Engineering Science* **46**(**12**):1239–1252. ISSN-L 0020-7225. doi:10.1016/j.ijengsci.2008.06.011 {*153, 373*}

Bonnet-Ben Dhia, A.-S., L. Chesnel, and X. Claeys 2013, August. Radiation condition for a non-smooth interface between a dielectric and a metamaterial. *Mathematical Models and Methods in Applied Sciences* **23**(**9**):1629–1662. ISSN-L 0218-2025. doi:10.1142/S0218202513500188 {*248, 373*}

Borcea, L., G. A. Gray, and Y. Zhang 2003, September 15. Variationally constrained numerical solution of electrical impedance tomography. *Inverse Problems* **19**(**5**):1159–1184. ISSN-L 0266-5611. doi:10.1088/0266-5611/19/5/309 {*129, 373*}

Bouchitté, G. and D. Felbacq 2004. Homogenization near resonances and artificial magnetism from dielectrics. *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique* **339**(**5**):377–382. ISSN-L 0249-6291. doi:10.1016/j.crma.2004.06.018 {*21, 373*}

Boulanger, P. and M. A. Hayes 1993. *Bivectors and Waves in Mechanics and Optics*, Volume 4 of *Applied Mathematics and Mathematical Computation*. Boca Raton, Florida: CRC Press. xiii + 281 pp. ISBN 0-412-46460-8. ISBN-13 978-0-412-46460-7. URL https://www.crcpress.com/Bivectors-and-Waves-in-Mechanics-and-Optics/Boulanger-Hayes/9780412464607 {*11, 374*}

Boverman, G., D. Isaacson, T.-J. Kao, G. J. Saulnier, and J. C. Newell 2008. Methods for direct image reconstruction for EIT in two and three dimensions. In R. H. Bayford and A. Hartov (eds.), *9th International Conference on Electrical Impedance Tomography (Dartmouth College, New Hampshire, USA, 16–18 June 2008)*, p. 4 (of 224). Philadelphia, PA: IOP Publishers. URL https://engineering.dartmouth.edu/eit2008/EIT_Conference_2008.pdf {*142, 374*}

Briane, M. and G. W. Milton 2009, September. Homogenization of the three-dimensional Hall effect and change of sign of the Hall coefficient. *Archive for Rational Mechanics and Analysis* **193**(**3**):715–736. ISSN-L 0003-9527. doi:10.1007/s00205-008-0200-y {*50, 374*}

Briane, M. and G. W. Milton 2011, October. Bounds on strong field magneto-transport in three-dimensional composites. *Journal of Mathematical Physics* **52**(**10**):103705 (18 pp.). ISSN-L 0022-2488. doi:10.1063/1.3644377 {*153, 356, 374*}

Brown, Jr., W. F. 1955, August. Solid mixture permittivities. *Journal of Chemical Physics* **23**(**8**):1514–1517. ISSN-L 0021-9606. doi:10.1063/1.1742339 {*168, 257, 288, 374*}

Bruggeman, D. A. G. 1935. Berechnung verschiedener physikalischer Konstanten von heterogen Substanzen. I Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen. (German) [Calculation of various physical constants in heterogeneous substances. I. Dielectric constants and conductivity of composites from isotropic substances]. *Annals of Physics* **24**:636–679. ISSN-L 0003-4916. {*67, 374*}

Brun, M., S. Guenneau, and A. B. Movchan 2009. Achieving control of in-plane elastic waves. *Applied Physics Letters* **94**(**6**):061903. ISSN-L 0003-6951. doi:10.1063/1.3068491 {*29, 374*}

Bruno, O. P. 1991a. The effective conductivity of strongly heterogeneous composites. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **433**(**1888**):353–381. ISSN-L 0080-4630. doi:10.1098/rspa.1991.0053 {*96, 236, 247, 374*}

Bruno, O. P. 1991b. Taylor expansions and bounds for the effective conductivity and the effective elastic moduli of multicomponent composites and polycrystals. *Asymptotic Analysis* **4**(**4**):339–365. ISSN-L 0921-7134. doi:10.3233/ASY-1991-4404 {*59, 86, 257, 288, 374*}

Bruno, O. P. and P. H. Leo 1992, December. On the stiffness of materials containing a disordered array of microscopic holes or hard inclusions. *Archive for Rational Mechanics and Analysis* **121**(**4**):303–338. ISSN-L 0003-9527. doi:10.1007/BF00375624 {*59, 86, 374*}

Bückmann, T., M. Kadic, R. Schittny, and M. Wegener 2015, July. Mechanical metamaterials with anisotropic and negative effective mass-density tensor made from one constituent material. *Physica Status Solidi. B, Basic Solid State Physics* **252**(**7**):1671–1674. ISSN-L 0370-1972. doi:10.1002/pssb.201451698 {*51, 374*}

Bückmann, T., R. Schittny, M. Thiel, M. Kadic, G. W. Milton, and M. Wegener 2014, March. On three-dimensional dilational elastic metamaterials. *New Journal of Physics* **16**(**3**):033032. ISSN-L 1367-2630. doi:10.1088/1367-2630/16/3/033032 {*50, 52, 374*}

Bückmann, T., N. Stenger, M. Kadic, J. Kaschke, A. Frölich, T. Kennerknecht, C. Eberl, M. Thiel, and M. Wegener 2012, May 22. Tailored 3D mechanical metamaterials made by dip-in direct-laser-writing optical lithography. *Advanced Materials* **24**(**20**):2710–2714. ISSN-L 0935-9648. doi:10.1002/adma.201200584 {*52, 374*}

Buroni, F. C. and A. Sáez 2010, February 8. Three-dimensional Green's function and its derivative for materials with general anisotropic magneto-electro-elastic coupling. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **466**(**2114**):515–537. ISSN-L 1364-5021. doi:10.1098/rspa.2009.0389 {*358, 374*}

Burridge, R. 1967. The singularity on the plane lids of the wave surface of elastic media with cubic symmetry. *Quarterly Journal of Mechanics and Applied Mathematics* **20**(**1**):41–56. ISSN-L 0033-5614. doi:10.1093/qjmam/20.1.41 {*358, 375*}

Buttazzo, G. 1989. *Semicontinuity, Relaxation and Integral Representations in the Calculus of Variations*, Volume 207 of *Pitman Research Notes in Mathematics Series*. Harlow, Essex, UK: Longman Scientific and Technical. iv + 222 pp. ISBN 0-582-01859-5. ISBN-13 978-0-582-01859-4. {*49, 375*}

Cai, W. and V. Shalaev 2010. *Optical Metamaterials: Fundamentals and Applications*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xii + 200 pp. ISBN 1-4419-1150-2 (hardcover). ISBN-13 978-1-4419-1150-6 (hardcover). doi:10.1007/978-1-4419-1151-3 {*52, 375*}

Calderón, A.-P. 1980. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics: 24 a 28 de Março 1980*, Volume 12 of *Coleção Atas*, pp. 65–73 (of ii + 216). Rio de Janeiro, Brazil: Sociedade Brasiliera de Mathemática. {*142, 375*}

Callen, H. B. 1960a. *Thermodynamics: An Introduction to the Physical Theories of Equilibrium Thermostatics and Irreversible Thermodynamics*. New York / London / Sydney, Australia: John Wiley and Sons. xv + 376 pp. ISBN 0-471-13036-2. ISBN-13 978-0-471-13036-9. {*4, 126, 375*}

Callen, H. B. 1960b. *Thermodynamics: An Introduction to the Physical Theories of Equilibrium Thermostatics and Irreversible Thermodynamics*, pp. 293–302 (of xv + 376). New York / London / Sydney, Australia: John Wiley and Sons. {*15, 375*}

Capdeboscq, Y. and M. S. Vogelius 2003, March/April. Optimal asymptotic estimates for the volume of internal inhomogeneities in terms of multiple boundary measurements. *Mathematical Modelling and Numerical Analysis = Modelisation mathématique et analyse numérique:* M^2AN **37**(**2**):227–240. ISSN-L 0764-583X. doi:10.1051/m2an:2003024 {*78, 375*}

Capdeboscq, Y. and M. S. Vogelius 2004. A review of some recent work on impedance imaging for inhomogeneities of low volume fraction. In C. Conca, R. Manásevich, G. Uhlmann, and M. S. Vogelius (eds.), *Partial Differential Equations and Inverse Problems*, Volume 362 of *Contemporary Mathematics*, pp. 69–87. Providence, RI, USA: American Mathematical Society. ISBN 0-8218-3448-7 (paperback). ISBN-13 978-0-8218-3448-0 (paperback). doi:10.1090/conm/362/06606 {*78, 375*}

Carini, A. and O. Mattei 2015, November/December. Variational formulations for the linear viscoelastic problem in the time domain. *European Journal of Mechanics, A, Solids* **54**:146–159. ISSN-L 0997-7538. doi:10.1016/j.euromechsol.2015.05.007 {*126, 153, 375*}

Carsey, F. D. (ed.) 1992. *Microwave Remote Sensing of Sea Ice*, Volume 68 of *Geophysical monograph*. Washington, DC: American Geophysical Union. 462 pp. ISSN-L 0065-8448. ISBN 0-87590-033-X. ISBN-13 978-0-87590-033-9. {*118, 375*}

Cessenat, M. 1996. *Mathematical Methods in Electromagnetism: Linear Theory and Applications*, Volume 41 of *Series on advances in mathematics for applied sciences*. Singapore / Philadelphia / River Edge, New Jersey: World Scientific Publishing Co. xii + 376 pp. ISBN 981-02-2467-2. ISBN-13 978-981-02-2467-7. doi:10.1142/2938 {*96, 112, 113, 375*}

Chandrasekharaiah, D. S. 1986, March. Thermoelasticity with second sound: a review. *Applied Mechanics Reviews* **39** (**3**):355–376. ISSN-L 0003-6900. doi:10.1115/1.3143705 {*9, 35, 375*}

Charina, M., M. Putinar, C. Scheiderer, and J. Stöckler 2015, September. An algebraic perspective on multivariate tight wavelet frames. II. *Applied and Computational Harmonic Analysis. Time-Frequency and Time-Scale Analysis, Wavelets, Numerical Algorithms, and Applications* **39**(**2**):185–213. ISSN-L 1063-5203. doi:10.1016/j.acha.2014.09.003 {*68, 151, 375*}

Chaulet, N. 2014. The electromagnetic scattering problem with generalized impedance boundary condition. *Mathematical Modelling and Numerical Analysis = Modelisation mathématique et analyse numérique:* M^2AN. To appear. See also arXiv:1312.1089 [math.AP]. ISSN-L 0764-583X. doi:10.1051/m2an/2014064 {*112, 375*}

Chen, G. and Ç. K. Koç 1994, May/June. Computing matrix-valued Nevanlinna–Pick interpolation. *Linear Algebra and its Applications* **203/204**:253–263. ISSN-L 0024-3795. doi:10.1016/0024-3795(94)90205-4 {*68, 136, 147, 151, 258, 376*}

Chen, G. and Ç. K. Koç 1995. A fast algorithm for matrix-valued Nevanlinna–Pick interpolation. In C. K. Chui and L. L. Schumaker (eds.), *Approximation theory VIII: College Station, Texas, USA, 8–12 January 1995*, Volume 1 of *Ser. Approx. Decompos.*, pp. 129–136 (of xxxi + 570 (vol. 1), xxix + 422 (vol. 2)). Singapore / Philadelphia / River Edge, New Jersey: World Scientific Publishing Co. ISBN 981-02-2318-8 (Set), 981-02-2971-2 (Vol. 1), 981-02-2972-0 (Vol. 2). ISBN-13 978-981-02-2318-2 (Set), 978-981-02-2971-9 (Vol. 1), 978-981-02-2972-6 (Vol. 2). {*68, 136, 147, 151, 258, 376*}

Chen, T. 1995, December 15. Exact moduli of phase interchange invariant composites with coupled fields. *Journal of Applied Physics* **78**(**12**):7413–7415. ISSN-L 0021-8979. doi:10.1063/1.360396 {*16, 258, 376*}

Chen, T. 1997, March. Exact moduli and bounds of two-phase composites with coupled multifield linear responses. *Journal of the Mechanics and Physics of Solids* **45**(**3**):385–398. ISSN-L 0022-5096. doi:10.1016/S0022-5096(96)00092-0 {*16, 258, 376*}

Cherepanov, G. P. 1967. The propagation of cracks in a continuous medium. *Prikladnaia matematika i mekhanika (Leningrad, R.S.F.S.R.)* **31**(**3**):476–488. Translated as Crack propagation in continuous media in Journal of Applied Mathematics and Mechanics, **31**(**3**), pp. 503–512. ISSN-L 0032-8235. doi:10.1016/0021-8928(67)90034-2 {*14, 376*}

Cherkaev, A. V. 2000. *Variational Methods for Structural Optimization*, Volume 140 of *Applied Mathematical Sciences*. Berlin / Heidelberg / London / etc.: Springer-Verlag. xxvi + 545 pp. ISBN 0-387-98462-3. ISBN-13 978-0-387-98462-9. doi:10.1007/978-1-4612-1188-4 {*2, 50, 126, 131, 150, 257, 258, 289, 339, 376*}

Cherkaev, A. V. and L. V. Gibiansky 1992. The exact coupled bounds for effective tensors of electrical and magnetic properties of two-component two-dimensional composites. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **122**(**1–2**):93–125. ISSN-L 0308-2105. doi:10.1017/S0308210500020990 {*5, 55, 71, 133, 135, 258, 284, 285, 376*}

Cherkaev, A. V. and L. V. Gibiansky 1993. Coupled estimates for the bulk and shear moduli of a two-dimensional isotropic elastic composite. *Journal of the Mechanics and Physics of Solids* **41**(**5**):937–980. ISSN-L 0022-5096. doi:10.1016/0022-5096(93)90006-2 {*5, 133, 376*}

Cherkaev, A. V. and L. V. Gibiansky 1994, January. Variational principles for complex conductivity, viscoelasticity, and similar problems in media with complex moduli. *Journal of Mathematical Physics* **35**(**1**):127–145. ISSN-L 0022-2488. doi:10.1063/1.530782 {*viii, 11, 12, 68, 78, 125–127, 129, 152, 153, 337, 338, 342, 343, 355, 366, 376*}

Cherkaev, E. 2001, August. Inverse homogenization for evaluation of effective properties of a mixture. *Inverse Problems* **17**(**4**):1203–1218. ISSN-L 0266-5611. doi:10.1088/0266-5611/17/4/341 {*125, 258, 376*}

Cherkaev, E. and C. Bonifasi-Lista 2011, January. Characterization of structure and properties of bone by spectral measure method. *Journal of Biomechanics* **44**(**2**):345–351. ISSN-L 0021-9290. doi:10.1016/j.jbiomech.2010.10.031 {*125, 143, 258, 376*}

Cherkaev, E. and K. M. Golden 1998. Inverse bounds for microstructural parameters of composite media derived from complex permittivity measurements. *Waves in Random Media* **8**(**4**):437–450. ISSN-L 0959-7174. doi:10.1088/0959-7174/8/4/004 {*129, 376*}

Chluba, C., W. Ge, R. Lima de Miranda, J. Strobel, L. Kienle, E. Quandt, and M. Wuttig 2015, May 29. Ultralow-fatigue shape memory alloy films. *Science* **348**(**6238**):1004–1007. ISSN-L 0036-8075. doi:10.1126/science.1261164 {*4, 376*}

Christensen, R. M. 1971. *Theory of Viscoelasticity; an Introduction*. New York, USA: Academic Press. xi + 245 pp. ISBN 0-12-174250-4. ISBN-13 978-0-12-174250-8. {*150, 376*}

Christensen, R. M. 2003. *Theory of Viscoelasticity* (Second ed.). New York: Dover Publications, Inc. xii + 364 pp. ISBN 0-486-42880-X (paperback). ISBN-13 978-0-486-42880-2 (paperback). {*21, 377*}

Clark, K. E. 1997, September. A continued fraction representation for the effective conductivity of a two-dimensional polycrystal. *Journal of Mathematical Physics* **38**(**9**):4528–4541. ISSN-L 0022-2488. doi:10.1063/1.532141 {*234, 288, 290, 377*}

Clark, K. E. and G. W. Milton 1994. Modeling the effective conductivity function of an arbitrary two-dimensional polycrystal using sequential laminates. *Proceedings of the Royal Society of Edinburgh* **124A**(**4**):757–783. ISSN-L 0080-4541. doi:10.1017/S030821050002864X {*152, 183, 234, 278, 288, 290, 377*}

Clark, K. E. and G. W. Milton 1995, February 8. Optimal bounds correlating electric, magnetic and thermal properties of two-phase, two-dimensional composites. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **448**(**1933**):161–190. ISSN-L 0080-4630. doi:10.1098/rspa.1995.0011 {*68, 152, 258, 284, 290, 377*}

Coleman, A. J. 1981. Calculation of first- and second-order reduced density matrices. In B. M. Deb (ed.), *The Force Concept in Chemistry*, Chapter 8, pp. 418–448 (of xvii + 502). New York: Van Nostrand Reinhold. ISBN 0-442-26106-3. ISBN-13 978-0-442-26106-1. {*312, 377*}

Coleman, B. D. and W. Noll 1959, January. On the thermostatics of continuous media. *Archive for Rational Mechanics and Analysis* **4**(**1**):97–128. ISSN-L 0003-9527. doi:10.1007/BF00281381 {*2, 3, 377*}

COMSOL 2013, May. *Acoustics Module User's Guide, Version 4.3b*. Stockholm, Sweden: COMSOL. May 2013. xii + 418 + xii pp. Part number CM020201. See also http://www.comsol.com/blogs/theory-thermoacoustics-acoustics-thermal-viscous-losses/. URL http://hpc.mtech.edu/comsol/pdf/Acoustics_Module/AcousticsModuleUsersGuide.pdf {*24, 377*}

Craster, R. V. and S. Guenneau (eds.) 2013. *Acoustic Metamaterials: Negative Refraction, Imaging, Lensing and Cloaking*, Volume 166 of *Springer Series in Materials Science*, Berlin, Germany / Heidelberg, Germany / London, UK / etc. Springer-Verlag. xvi + 323 pp. ISBN 94-007-4813-2. ISBN-13 978-94-007-4813-2. {*52, 377*}

Craster, R. V. and Y. V. Obnosov 2001. Four-phase checkerboard composites. *SIAM Journal on Applied Mathematics* **61** (**6**):1839–1856. ISSN-L 0036-1399. doi:10.1137/S0036139900371825 {*247, 377*}

Dacorogna, B. 2007. *Direct Methods in the Calculus of Variations* (Second ed.), Volume 78 of *Applied Mathematical Sciences, .* Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xii + 621 pp. Series editors: S. S. Antman, J. E. Marsden, and L. Sirovich. ISBN 0-387-35779-3. ISBN-13 978-0-387-35779-9. doi:10.1007/978-0-387-55249-1 {*17, 339, 377*}

Dal Maso, G. 1993. *An Introduction to* Γ*-Convergence*, Volume 8 of *Progress in Nonlinear Differential Equations and Their Applications*. Basel, Switzerland: Birkhäuser Verlag. xiv + 340 pp. ISBN 0-8176-3679-X. ISBN-13 978-0-8176-3679-1. {*49, 377*}

Dantzig, G. B. 1998. *Linear Programming and Extensions*. Princeton, New Jersey: Princeton University Press. xvi + 627 pp. ISBN 0-691-08000-3. ISBN-13 978-0-691-08000-0. {*169, 377*}

Davidson, E. R. 1995, November 24. N-representability of the electron pair density. *Chemical Physics Letters* **246**(**3**): 209–213. ISSN-L 0009-2614. doi:10.1016/0009-2614(95)01102-F {*312, 377*}

Day, A. R., A. R. Grant, A. J. Sievers, and M. F. Thorpe 2000, February 28. Spectral function of composites from reflectivity measurements. *Physical Review Letters* **84**(**9**):1978–1981. ISSN-L 0031-9007. doi:10.1103/PhysRevLett.84.1978 {*125, 258, 377*}

Day, A. R. and M. F. Thorpe 1999, March 29. The spectral function of composites: The inverse problem. *Journal of Physics: Condensed Matter* **11**(**12**):2551–2568. ISSN-L 0953-8984. doi:10.1088/0953-8984/11/12/010 {*125, 258, 377*}

De Giorgi, E. 1984. G-operators and Γ-convergence. In *Proceedings of the International Congress of Mathematicians (Warsaw, 1983)*, Volume 1, 2, pp. 1175–1191. Warsaw, Poland: PWN. {*49, 377*}

Dederichs, P. H. and R. Zeller 1973, April. Variational treatment of the elastic constants of disordered materials. *Zeitschrift für Physik* **259**(**2**):103–116. ISSN-L 0044-3328. doi:10.1007/BF01392841 {*257, 288, 378*}

Delbary, F., P. C. Hansen, and K. Knudsen 2011. A direct numerical reconstruction algorithm for the 3D Calderón problem. In Y.-C. Hon and L. Ling (eds.), *International Conference on Inverse Problems 2010, 13–17 December 2010 Hong Kong*, Volume 290 of *Journal of Physics: Conference Series*, p. 012003. Bristol, PA: IOP. {*142, 378*}

Dell'Antonio, G. F., R. Figari, and E. Orlandi 1986. An approach through orthogonal projections to the study of inhomogeneous or random media with linear response. *Annales de l'institut Henri Poincaré (A) Physique théorique* **44**(**1**): 1–28. ISSN-L 0020-2339. URL http://eudml.org/doc/76310 {*54, 96, 183, 378*}

Delsarte, P., Y. Genin, and Y. Kamp 1979, February. The Nevanlinna–Pick problem for matrix-valued functions. *SIAM Journal on Applied Mathematics* **36**(**1**):47–61. ISSN-L 0036-1399. doi:10.1137/0136005 {*68, 136, 146, 151, 258, 378*}

Delsarte, P., Y. Genin, and Y. Kamp 1981, April. On the role of the Nevanlinna–Pick problem in circuit and system theory. *International Journal of Circuit Theory and Applications* **9**(**2**):177–187. ISSN-L 0098-9886. doi:10.1002/cta.4490090204 {*136, 258, 378*}

Dolin, L. S. 1961. To the possibility of comparison of three-dimensional electromagnetic systems with nonuniform anisotropic filling. *Izvestiya Vysshikh Uchebnykh Zavedeniĭ. Radiofizika* **4**(**5**):964–967. ISSN-L 0021-3462. {*29, 378*}

Doyle, P. G. and L. Snell 1984. *Random Walks and Electrical Networks*, Volume 2 of *Carus Mathematical Monographs*. Washington, DC, USA: Mathematical Association of America. xiii + 159 pp. ISBN 0-88385-024-9. ISBN-13 978-0-88385-024-4. {*73, 378*}

Dukhin, A. S. and P. J. Goetz 2009, March 28. Bulk viscosity and compressibility measurement using acoustic spectroscopy. *Journal of Chemical Physics* **130**(**12**):124519. ISSN-L 0021-9606. doi:10.1063/1.3095471 {*22, 25, 378*}

Dykhne, A. M. 1970, July. Conductivity of a two-dimensional two-phase system. *Zhurnal eksperimental'noi i teoreticheskoi fiziki / Akademiia Nauk SSSR* **59**:110–115. English translation in Soviet Physics JETP **32**(**1**):63–65 (January 1971). ISSN-L 0044-4510. URL http://jetp.ac.ru/cgi-bin/dn/e_032_01_0063.pdf {*16, 267, 378*}

Erdahl, R. and V. H. Smith, Jr. (eds.) 1987. *Density Matrices and Density Functionals. Proceedings of the A. John [Albert John] Coleman Symposium*, Dordrecht, The Netherlands; Boston, MA, USA; Lancaster, UK; Tokyo, Japan. D. Reidel. xvi + 722 pp. ISBN 90-277-2477-6. ISBN-13 978-90-277-2477-9. {*312, 378*}

Eshelby, J. D. 1951, November 6. The force on an elastic singularity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* **244**(**877**):87–112. ISSN-L 1364-503X. doi:10.1098/rsta.1951.0016 {*14, 378*}

Eshelby, J. D. 1970. Energy relations and the energy-momentum tensor in continuum mechanics. In M. F. Kanninen (ed.), *Inelastic Behavior of Solids*, McGraw-Hill series in materials science and engineering, pp. 77–115. New York, NY, USA: McGraw-Hill. Reprinted in Fundamental Contributions to the Continuum Theory of Evolving Phase Interfaces in Solids, ed. by J. M. Ball et al, Springer-Verlag, 1999, ISBN-13 978-3-642-64188-6. ISBN 0-07-033274-6. ISBN-13 978-0-07-033274-4. doi:10.1007/978-3-642-59938-5_5 {*14, 378*}

Estakhri, N. M. and A. Alù 2013, May 15. Physics of unbounded, broadband absorption/gain efficiency in plasmonic nanoparticles. *Physical Review B: Condensed Matter and Materials Physics* **87**(**20**):205418. ISSN-L 1098-0121. doi:10.1103/PhysRevB.87.205418 {*248, 378*}

Eyre, D. J. and G. W. Milton 1999, April. A fast numerical scheme for computing the response of composites using grid refinement. *European Physical Journal. Applied Physics* **6**(**1**):41–47. ISSN-L 1286-0042. doi:10.1051/epjap:1999150 {*45, 58, 78, 235–239, 245, 246, 249, 288, 319, 331, 332, 378*}

Eyre, D. J., G. W. Milton, and R. S. Lakes 2002, August. Bounds for interpolating complex effective moduli of viscoelastic materials from measured data. *Rheologica acta* **41**(**5**):461–470. ISSN-L 0035-4511. doi:10.1007/s003970100172 {*151, 378*}

Fannjiang, A. and G. C. Papanicolaou 1994. Convection enhanced diffusion for periodic flows. *SIAM Journal on Applied Mathematics* **54**(**2**):333–408. ISSN-L 0036-1399. doi:10.1137/S0036139992236785 {*53, 126, 355, 379*}

Fedčina, I. P. 1972. A criteron for the solvability of the Nevanlinna–Pick tangent problem. *Matematicheskie Issledovaniya* **7**(**4 (26)**):213–227, 258. ISSN-L 0542-9994. {*136, 258, 379*}

Field, D. A. 1976, May. Series of Stieltjes, Padé approximants and continued fractions. *Journal of Mathematical Physics* **17**(**5**):843–844. ISSN-L 0022-2488. doi:10.1063/1.522983 {*289, 379*}

Foiaş, C. and A. E. Frazho 1990. *The Commutant Lifting Approach to Interpolation Problems*, Volume 44 of *Operator Theory: Advances and Applications*. Basel, Switzerland: Birkhäuser Verlag. xxiii + 632 pp. ISBN 3-7643-2461-9, 0-8176-2461-9. ISBN-13 978-3-7643-2461-2, 978-0-8176-2461-3. doi:10.1007/978-3-0348-7712-1 {*136, 258, 379*}

Foiaş, C., M. S. Jolly, and W.-S. Li 2002, November. Nevanlinna–Pick interpolation of attractors. *Nonlinearity (Bristol)* **15**(**6**):1881–1903. ISSN-L 0951-7715. doi:10.1088/0951-7715/15/6/308 {*289, 379*}

Fokin, A. G. 1982, May 1. Iteration method in the theory of nonhomogeneous dielectrics. *Physica Status Solidi. B, Basic Research* **111**(**1**):281–288. ISSN-L 0370-1972. doi:10.1002/pssb.2221110131 {*54, 183, 379*}

Fokin, A. G. and T. D. Shermergor 1969, January. Calculation of effective elastic moduli of composite materials with multiphase interactions taken into account. *Journal of Applied Mechanics and Technical Physics* **10**(**1**):48–54. ISSN-L 0021-8944. doi:10.1007/BF00916252 {*257, 288, 379*}

Francfort, G. A. and G. W. Milton 1994. Sets of conductivity and elasticity tensors stable under lamination. *Communications on Pure and Applied Mathematics (New York)* **47**(**3**):257–279. ISSN-L 0010-3640. doi:10.1002/cpa.3160470302 {*4, 133, 379*}

Francfort, G. A. and P. M. Suquet 1986. Homogenization and mechanical dissipation in thermoviscoelasticity. *Archive for Rational Mechanics and Analysis* **96**(**3**):265–293. ISSN-L 0003-9527. doi:10.1007/BF00251909 {*10, 379*}

Fredholm, I. 1900, December. Sur les équations de l'équilbre d'un corps solide elastique. (French) [On the equations of the equilibrium of an elastic solid body]. *Acta Mathematica* **23**(**1**):1–42. ISSN-L 0001-5962. doi:10.1007/BF02418668 {*357, 358, 379*}

Garcia, S. R., E. Prodan, and M. Putinar 2014. Mathematical and physical aspects of complex symmetric operators. *Journal of Physics A: Mathematical and Theoretical* **47**(**35**):353001 (54 pp.). ISSN-L 1751-8113. doi:10.1088/1751-8113/47/35/353001 {*11, 379*}

Gel'fand, I. M. and G. E. Shilov 1964. *Generalized Functions. Volume 1: Properties and Operations*. New York, USA: Academic Press. xviii + 423 pp. Translated from the Russian by Eugene Saletan. ISBN 0-12-279501-6. ISBN-13 978-0-12-279501-5. {*358, 379*}

Gesztesy, F. and E. Tsekanovskii 2000. On matrix-valued Herglotz functions. *Mathematische Nachrichten* **218**(**1**):61–138. ISSN-L 0025-584X. DOI http://doi.org/bpcx3m {*96, 98, 138, 379*}

Gibiansky, L. V. and R. S. Lakes 1993, September. Bounds on the complex bulk modulus of a two-phase viscoelastic composite with arbitrary volume fractions of the components. *Mechanics of Materials: An International Journal* **16**(**3**):317–331. ISSN-L 0167-6636. doi:10.1016/0167-6636(93)90060-5 {*152, 356, 379*}

Gibiansky, L. V. and R. S. Lakes 1997, March. Bounds on the complex bulk and shear moduli of a two-dimensional two-phase viscoelastic composite. *Mechanics of Materials: an International Journal* **25**(**2**):79–95. ISSN-L 0167-6636. doi:10.1016/S0167-6636(96)00046-4 {*152, 379*}

Gibiansky, L. V. and G. W. Milton 1993, January 8. On the effective viscoelastic moduli of two-phase media. I. Rigorous bounds on the complex bulk modulus. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **440**(**1908**):163–188. ISSN-L 0080-4630. doi:10.1098/rspa.1993.0010 {*56, 67, 152, 356, 379*}

Gibiansky, L. V., G. W. Milton, and J. G. Berryman 1999, June 8. On the effective viscoelastic moduli of two-phase media: III. Rigorous bounds on the complex shear modulus in two dimensions. *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences* **455**(**1986**):2117–2149. ISSN-L 1364-5021. doi:10.1098/rspa.1999.0395 {*152, 356, 380*}

Gibiansky, L. V. and S. Torquato 1995, October. Geometrical-parameter bounds on the effective moduli of composites. *Journal of the Mechanics and Physics of Solids* **43**(**10**):1587–1613. ISSN-L 0022-5096. doi:10.1016/0022-5096(95)00049-O {*55, 380*}

Gilbarg, D. and N. S. Trudinger 1983. *Elliptic Partial Differential Equations of Second Order*, Volume 224 of *Applied Mathematical Sciences*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xiii + 513 pp. ISBN 3-540-13025-X. ISBN-13 978-3-540-13025-3. doi:10.1007/978-3-642-61798-0 {*280, 380*}

Gillman, A., G. Amadio, K. Matouš, and T. L. Jackson 2015, May 8. Third-order thermo-mechanical properties for packs of Platonic solids using statistical micromechanics. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **471**(**2177**):20150060. ISSN-L 0080-4630. doi:10.1098/rspa.2015.0060 {*55, 380*}

Gohberg, I. C., S. Goldberg, and M. A. Kaashoek 1990. *Classes of linear operators, volume 1*, Volume 49 of *Operator theory: Advances and Applications*. Basel, Switzerland: Birkhäuser Verlag. xiii + 468 pp. ISBN 3-0348-7511-8. ISBN-13 978-3-0348-7511-0. doi:10.1007/978-3-0348-7509-7 {*295, 380*}

Gohberg, I. C., S. Goldberg, and M. A. Kaashoek 2003. *Basic Classes of Linear Operators*. Basel, Switzerland: Birkhäuser Verlag. xvii + 423 pp. ISBN 3-0348-7980-6. ISBN-13 978-3-0348-7980-4. doi:10.1007/978-3-0348-7980-4 {*291, 299, 380*}

Gohberg, I. C. and M. G. Kreĭn 1969. *Introduction to the Theory of Linear Nonselfadjoint Operators*, Volume 18 of *Translations of Mathematical Monographs*. Providence, RI, USA: American Mathematical Society. xv + 378 pp. Translated from the Russian by A. Feinstein. ISBN 0-8218-1568-7. ISBN-13 978-0-8218-1568-7. {*356, 380*}

Golden, K. M. 1995, July 15. Bounds on the complex permittivity of sea ice. *Journal of Geophysical Research (Oceans)* **100**(**C7**):13699–13711. ISSN-L 2169-9275. doi:10.1029/94JC03007 {*118, 380*}

Golden, K. M. 1998. The interaction of microwaves with sea ice. In G. C. Papanicolaou (ed.), *Wave Propagation in Complex Media*, Volume 96 of *IMA Volumes in Mathematics and its Applications*, pp. 75–94 (of x + 287). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-98309-0. ISBN-13 978-0-387-98309-7. {*236, 380*}

Golden, K. M. 2009. Climate change and the mathematics of transport in sea ice. *Notices of the American Mathematical Society* **56**(**5**):562–584. ISSN-L 0002-9920. URL http://www.ams.org/notices/200905/tx090500562p.pdf {*118, 380*}

Golden, K. M. and G. C. Papanicolaou 1983. Bounds for effective parameters of heterogeneous media by analytic continuation. *Communications in Mathematical Physics* **90**(**4**):473–491. ISSN-L 0010-3616. doi:10.1007/BF01216179 {*11, 49, 54, 62, 68, 96, 151, 153, 167, 183, 184, 236, 264, 380*}

Golden, K. M. and G. C. Papanicolaou 1985, September. Bounds for effective parameters of multicomponent media by analytic continuation. *Journal of Statistical Physics* **40**(**5–6**):655–667. ISSN-L 0022-4715. doi:10.1007/BF01009895 {*96, 119, 151, 167, 236, 380*}

Gonze, X., J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D. C. Allan 2002, November. First-principles computation of material properties: the ABINIT software project. *Computational Materials Science* **25**(**3**):478–492. ISSN-L 0927-0256. doi:10.1016/S0927-0256(02)00325-7 {*310, 319, 380*}

Grabovsky, Y. 1998. Exact relations for effective tensors of polycrystals. I: Necessary conditions. *Archive for Rational Mechanics and Analysis* **143**(**4**):309–329. ISSN-L 0003-9527. doi:10.1007/s002050050107 {*17, 78, 96, 183, 380*}

Grabovsky, Y. 2004. Algebra, geometry and computations of exact relations for effective moduli of composites. In G. Capriz and P. M. Mariano (eds.), *Advances in Multifield Theories of Continua with Substructure*, Modelling and Simulation in Science, Engineering and Technology, pp. 167–197 (of ix + 263). Boston, MA: Birkhäuser Verlag. ISBN 0-8176-4324-9. ISBN-13 978-0-8176-4324-9. doi:10.1007/978-0-8176-8158-6 {*17, 78, 96, 183, 381*}

Grabovsky, Y. and G. W. Milton 1998. Exact relations for composites: Towards a complete solution. *Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung* **Extra Volume ICM III**:623–632. ISSN-L 1431-0635. URL http://www.emis.ams.org/journals/DMJDMV/xvol-icm/16/Milton.MAN.html {*17, 183, 381*}

Grabovsky, Y., G. W. Milton, and D. S. Sage 2000, March. Exact relations for effective tensors of composites: Necessary conditions and sufficient conditions. *Communications on Pure and Applied Mathematics (New York)* **53**(**3**):300–353. ISSN-L 0010-3640. DOI http://doi.org/d8k4vw {*17, 58, 78, 96, 183, 381*}

Grabovsky, Y. and D. S. Sage 1998. Exact relations for effective tensors of polycrystals. II: Applications to elasticity and piezoelectricity. *Archive for Rational Mechanics and Analysis* **143**(**4**):331–356. ISSN-L 0003-9527. doi:10.1007/s002050050108 {*17, 78, 96, 183, 381*}

Gragg, W. B. 1968, June. Truncation error bounds for g-fractions. *Numerische Mathematik* **11**(**5**):370–379. ISSN-L 0029-599X. doi:10.1007/BF02161885 {*288, 289, 381*}

Greaves, G. N., A. L. Greer, R. S. Lakes, and T. Rouxel 2011, November. Poisson's ratio and modern materials. *Nature Materials* **10**(**11**):823–837. ISSN-L 1476-1122. doi:10.1038/nmat3134 {*50, 381*}

Greenleaf, A., M. Lassas, and G. Uhlmann 2003a, April. Anisotropic conductivities that cannot be detected by EIT. *Physiological Measurement* **24**(**2**):413–419. ISSN-L 0967-3334. URL http://iopscience.iop.org/0967-3334/24/2/353 {*29, 381*}

Greenleaf, A., M. Lassas, and G. Uhlmann 2003b. On non-uniqueness for Calderón's inverse problem. *Mathematical Research Letters* **10**(**5**):685–693. ISSN-L 1073-2780. doi:10.4310/MRL.2003.v10.n5.a11 {*29, 381*}

Grima, J. N. and K. E. Evans 2000. Auxetic behaviour from rotating squares. *Journal of Materials Science Letters* **19**: 1563–1565. ISSN-L 0261-8028. {*50, 381*}

Gubernatis, J. E. and J. A. Krumhansl 1975, May 1. Macroscopic engineering properties of polycrystalline materials: Elastic properties. *Journal of Applied Physics* **46**(**5**):1875–1883. ISSN-L 0021-8979. doi:10.1063/1.321884 {*257, 288, 381*}

Guevara Vasquez, F., G. W. Milton, and D. Onofrei 2011, January. Complete characterization and synthesis of the response function of elastodynamic networks. *Journal of Elasticity* **102**(**1**):31–54. ISSN-L 0374-3535. doi:10.1007/s10659-010-9260-y {*96, 381*}

Guevara Vasquez, F., G. W. Milton, D. Onofrei, and P. Seppecher 2013. Transformation elastodynamics and active exterior acoustic cloaking. In R. V. Craster and S. Guenneau (eds.), *Acoustic Metamaterials: Negative Refraction, Imaging, Lensing and Cloaking*, Volume 166 of *Springer Series in Materials Science*, pp. 289–318. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. ISBN 94-007-4813-2. ISBN-13 978-94-007-4813-2. doi:10.1007/978-94-007-4813-2 {*29, 381*}

Gully, A., J. Lin, E. Cherkaev, and K. M. Golden 2015. Bounds on the complex permittivity of polycrystalline materials by analytic continuation. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **471** (**2174**):20140702. ISSN-L 1364-5021. doi:10.1098/rspa.2014.0702 {*118, 381*}

Gustafsson, M. and D. Sjöberg 2010, April. Sum rules and physical bounds on passive metamaterials. *New Journal of Physics* **12**(**4**):043046. ISSN-L 1367-2630. doi:10.1088/1367-2630/12/4/043046 {*96, 110, 116, 381*}

Harrison, N. M. 2003. An introduction to density functional theory. In C. R. A. Catlow and E. A. Kotomin (eds.), *Computational Materials Science*, Volume 187 of *NATO Science Series: Computer & Systems Sciences*, pp. 45–70 (of vii + 419). Amsterdam, The Netherlands: IOS Press. ISBN 1-58603-335-2. ISBN-13 978-1-58603-335-4. {*312, 381*}

Harutyunyan, D. and G. W. Milton 2015a, October. Explicit examples of extremal quasiconvex quadratic forms that are not polyconvex. *Calculus of Variations and Partial Differential Equations* **54**(**2**):1575–1589. See also arXiv:1403.3718 [math.AP]. ISSN-L 0944-2669. doi:10.1007/s00526-015-0836-z {*19, 71, 382*}

Harutyunyan, D. and G. W. Milton 2015b, November 16. On the relation between extremal elasticity tensors with orthotropic symmetry and extremal polynomials. *arxiv.org*:1–17 (17 pp.). URL http://arxiv.org/abs/1411.4216 {*x, 1, 18, 19, 71, 382*}

Harutyunyan, D. and G. W. Milton 2016, January 4. Towards characterization of all 3×3 extremal quasiconvex quadratic forms. *arxiv.org*:1–25. URL http://arxiv.org/abs/1512.04174 {*19, 382*}

Hashin, Z. 1965a. On elastic behavior of fiber reinforced materials of arbitrary transverse phase geometry. *Journal of the Mechanics and Physics of Solids* **13**(**3**):119–134. ISSN-L 0022-5096. doi:10.1016/0022-5096(65)90015-3 {*150, 382*}

Hashin, Z. 1965b, September. Viscoelastic behavior of heterogeneous media. *Journal of Applied Mechanics* **32**(**3**):630–636. ISSN-L 0021-8936. doi:10.1115/1.3627270 {*150, 382*}

Hashin, Z. 1983, September. Analysis of composite materials — A survey. *Journal of Applied Mechanics* **50**(**3**):481–505. ISSN-L 0021-8936. doi:10.1115/1.3167081 {*48, 150, 257, 258, 382*}

Hashin, Z. and S. Shtrikman 1962, October 1. A variational approach to the theory of the effective magnetic permeability of multiphase materials. *Journal of Applied Physics* **33**(**10**):3125–3131. ISSN-L 0021-8979. doi:10.1063/1.1728579 {*55, 67, 150, 257, 279, 382*}

Hashin, Z. and S. Shtrikman 1963. A variational approach to the theory of the elastic behavior of multiphase materials. *Journal of the Mechanics and Physics of Solids* **11**(**2**):127–140. ISSN-L 0022-5096. doi:10.1016/0022-5096(63)90060-7 {*54, 55, 150, 382*}

Haus, H. A. and J. R. Melcher 1989. *Electromagnetic Fields and Energy*. Upper Saddle River, New Jersey: Prentice-Hall. xxiii + 742 pp. See also http://web.mit.edu/6.013˙book/www/. ISBN 0-13-249020-X. ISBN-13 978-0-13-249020-7. {*125, 382*}

Hehl, F. W., P. von der Heyde, G. D. Kerlick, and J. M. Nester 1976, July. General Relativity with spin and torsion: Foundations and prospects. *Reviews of Modern Physics* **48**(**3**):393–416. ISSN-L 0034-6861. doi:10.1103/RevModPhys.48.393 {*29, 382*}

Helsing, J., H. Kang, and M. Lim 2016. Classification of spectrum of the Neumann–Poincaré operator on planar domains with corners by resonance: a numerical study. Preprint. {*248, 382*}

Helsing, J., R. C. McPhedran, and G. W. Milton 2011. Spectral super-resolution in metamaterial composites. *New Journal of Physics* **13**(**11**):115005. ISSN-L 1367-2630. doi:10.1088/1367-2630/13/11/115005 {*247, 382*}

Herring, C. 1960, November 1. Effect of random inhomogeneities on electrical and galvanomagnetic measurements. *Journal of Applied Physics* **31**(**11**):1939–1953. ISSN-L 0021-8979. doi:10.1063/1.1735477 {*257, 288, 382*}

Hetherington, J. H. and M. F. Thorpe 1992, September 8. The conductivity of a sheet containing inclusions with sharp corners. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **438**(**1904**):591–604. ISSN-L 0080-4630. doi:10.1098/rspa.1992.0128. {*247, 382*}

Hill, R. 1952. The elastic behavior of a crystalline aggregate. *Proceedings of the Physical Society, London, Section A* **65**(**5**):349–354. ISSN-L 0370-1298. doi:10.1088/0370-1298/65/5/307 {*64, 150, 382*}

Hill, R. 1957. On uniqueness and stability in the theory of finite elastic strain. *Journal of the Mechanics and Physics of Solids* **5**(**4**):229–241. ISSN-L 0022-5096. doi:10.1016/0022-5096(57)90016-9 {*3, 382*}

Hill, R. 1963, September. Elastic properties of reinforced solids: Some theoretical principles. *Journal of the Mechanics and Physics of Solids* **11**(**5**):357–372. ISSN-L 0022-5096. doi:10.1016/0022-5096(63)90036-X {*54, 55, 143, 382*}

Hlushkou, D., H. Liasneuski, U. Tallarek, and S. Torquato 2015, September 28. Effective diffusion coefficients in random packings of polydisperse hard spheres from two-point and three-point correlation functions. *Journal of Applied Physics* **118**(**12**):124901. ISSN-L 0021-8979. doi:10.1063/1.4931153 {*55, 383*}

Hohenberg, P. and W. Kohn 1964, November. Inhomogeneous electron gas. *Physical Review (2)* **136**(**3B**):B864–B871. ISSN-L 0031-899X. doi:10.1103/PhysRev.136.B864 {*309, 383*}

Hori, M. 1973, April. Statistical theory of effective electrical, thermal, and magnetic properties of random heterogeneous materials. I. Perturbation expansions for the effective permittivity of cell materials. *Journal of Mathematical Physics* **14**(**4**):514–523. ISSN-L 0022-2488. doi:10.1063/1.1666347 {*257, 288, 383*}

Hori, M. and S. Nemat-Nasser 1995, October 1. Universal bounds for overall properties of linear and nonlinear heterogeneous solids. *ASME Journal of Engineering Materials and Technology* **117**(**4**):412–432. ISSN-L 0094-4289. doi:10.1115/1.2804735 {*78, 383*}

Hori, M. and S. Nemat-Nasser 1998, September. Universal bounds for effective piezoelectric moduli. *Mechanics of Materials: an International Journal* **30**(**1**):1–19. ISSN-L 0167-6636. doi:10.1016/S0167-6636(98)00029-5 {*78, 383*}

Hörmander, L. 1990. *An Introduction to Complex Analysis in Several Variables* (Third revised ed.), Volume 7 of *North-Holland mathematical library*. Amsterdam: Elsevier. xii + 254 pp. ISBN 0-444-88446-7. ISBN-13 978-0-444-88446-6. {*98, 383*}

Huet, C. 1990. Application of variational concepts to size effects in elastic heterogeneous bodies. *Journal of the Mechanics and Physics of Solids* **38**(**6**):813–841. ISSN-L 0022-5096. doi:10.1016/0022-5096(90)90041-2 {*78, 383*}

Huet, C. 1995. Bounds for the overall properties of viscoelastic heterogeneous and composite materials. *Archives of Mechanics = Archiwum Mechaniki Stosowanej* **47**(**6**):1125–1155. ISSN-L 0373-2029. {*153, 383*}

Isakov, V. 1990. *Inverse Source Problems*, Volume 34 of *Mathematical Surveys and Monographs*. Providence, RI, USA: American Mathematical Society. xiv + 193 pp. ISBN 0-8218-1532-6. ISBN-13 978-0-8218-1532-8. doi:10.1090/surv/034 {*144, 383*}

Isakov, V. 2006. *Inverse Problems for Partial Differential Equations* (Second ed.), Volume 127 of *Applied Mathematical Sciences*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xiii + 344 pp. ISBN 0-387-25364-5. ISBN-13 978-0-387-25364-0. doi:10.1007/0-387-32183-7 {*124, 144, 383*}

Ismail-Beigi, S. and T. A. Arias 2000, June 9. New algebraic formulation of density functional calculation. *Computer Physics Communications* **128**(**1–2**):1–45. ISSN-L 0010-4655. doi:10.1016/S0010-4655(00)00072-2 {*310, 319, 383*}

Jackson, J. D. 1999. *Classical Electrodynamics* (Third ed.). New York, NY: John Wiley and Sons. ISBN 0-471-30932-X. ISBN-13 978-0-471-30932-1. {*110, 383*}

James, R. D. and D. Kinderlehrer 1989. Theory of diffusionless phase transitions. In M. Rascle, D. Serre, and M. Slemrod (eds.), *PDE's and Continuum Models of Phase Transitions: Proceedings of an NSF-CNRS Joint Seminar Held in Nice, France, January 18–22, 1988*, Volume 344 of *Lecture Notes in Physics*, pp. 51–84 (of v + 229). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-51617-4. ISBN-13 978-0-387-51617-2. {*2, 383*}

Jog, C. S. 2002. *Foundations and Applications of Mechanics, Volume 1: Continuum Mechanics*. New Delhi, India: Narosa. xv + 254 pp. ISBN 81-7319-448-3. ISBN-13 978-81-7319-448-1. {*6, 383*}

Jones, R. O. 2015, March. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics* **87**(**3**):897–923. ISSN-L 0034-6861. doi:10.1103/RevModPhys.87.897 {*x, 309, 383*}

Kabanikhin, S. I. 2012. *Inverse and Ill-Posed Problems: Theory and Applications*, Volume 55 of *Inverse and Ill-Posed Problems Series*. New York, NY, USA and Berlin, Germany: Walter de Gruyter. xv + 459 pp. ISBN 3-11-022400-3. ISBN-13 978-3-11-022400-9. {*124, 383*}

Kadic, M., T. Bückmann, N. Stenger, M. Thiel, and M. Wegener 2012, May 7. On the practicability of pentamode mechanical metamaterials. *Applied Physics Letters* **100**(**19**):191901. ISSN-L 0003-6951. doi:10.1063/1.4709436 {*52, 384*}

Kadic, M., R. Schittny, T. Bückmann, C. Kern, and M. Wegener 2015, April/June. Hall-effect sign inversion in a realizable 3D metamaterial. *Physical Review X* **5**(**2**):021030. ISSN-L 2160-3308. doi:10.1103/PhysRevX.5.021030 {*50, 384*}

Kang, H., E. Kim, and G. W. Milton 2012, November. Sharp bounds on the volume fractions of two materials in a two-dimensional body from electrical boundary measurements: the translation method. *Calculus of Variations and Partial Differential Equations* **45**(**3–4**):367–401. ISSN-L 0944-2669. doi:10.1007/s00526-011-0462-3 {*18, 78, 131, 339, 384*}

Kang, H., K. Kim, H. Lee, X. Li, and G. W. Milton 2014. Bounds on the size of an inclusion using the translation method for two-dimensional complex conductivity. *SIAM Journal on Applied Mathematics* **74**(**4**):939–958. ISSN-L 0036-1399. doi:10.1137/130940426 {*18, 78, 129, 131, 339, 384*}

Kang, H. and H. Lee 2004, December. Identification of simple poles via boundary measurements and an application of EIT. *Inverse Problems* **20**(**6**):1853–1863. ISSN-L 0266-5611. doi:10.1088/0266-5611/20/6/010 {*144, 384*}

Kang, H. and G. W. Milton 2013. Bounds on the volume fractions of two materials in a three-dimensional body from boundary measurements by the translation method. *SIAM Journal on Applied Mathematics* **73**(**1**):475–492. ISSN-L 0036-1399. doi:10.1137/120879713 {*18, 78, 131, 150, 339, 384*}

Kang, H., G. W. Milton, and J.-N. Wang 2014, January. Bounds on the volume fraction of the two-phase shallow shell using one measurement. *Journal of Elasticity* **114**(**1**):41–53. ISSN-L 0374-3535. doi:10.1007/s10659-012-9425-y {*18, 78, 131, 339, 384*}

Kantor, Y. and D. J. Bergman 1984. Improved rigorous bounds on the effective elastic moduli of a composite material. *Journal of the Mechanics and Physics of Solids* **32**:41–62. ISSN-L 0022-5096. {*152, 154, 183, 384*}

Kaplan, C. and F. D. Murnaghan 1930, April. On the fundamental constitutive equations in electromagnetic theory. *Physical Review* **35**(**7**):763–777. ISSN-L 0031-899X. doi:10.1103/PhysRev.35.763 {*40, 384*}

Kato, T. 1995. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xxi + 619 pp. ISBN 3-540-58661-X. ISBN-13 978-3-540-58661-6. URL http://www.springerlink.com/content/978-3-540-58661-6 {*96, 98, 384*}

Keller, J. B. 1964, April. A theorem on the conductivity of a composite medium. *Journal of Mathematical Physics* **5**(**4**):548–549. ISSN-L 0022-2488. doi:10.1063/1.1704146 {*16, 169, 171, 267, 384*}

Khachaturyan, A. G. 1966. Some questions concerning the theory of phase transitions in solids. *Fizika Tverdogo Tela.* **8**:2709–2717. English translation in Soviet Phys. Solid State, **8**:2163–2168 (1967). {*2, 384*}

Khachaturyan, A. G. 1983. *Theory of Structural Transformations in Solids*. New York / London / Sydney, Australia: John Wiley and Sons. xiii + 574 pp. ISBN 0-471-07873-5. ISBN-13 978-0-471-07873-9. {*2, 384*}

Kirsch, A. 2011. *An Introduction to the Mathematical Theory of Inverse Problems* (Second ed.), Volume 120 of *Applied Mathematical Sciences*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. 310 pp. ISBN 1-4419-8473-9. ISBN-13 978-1-4419-8473-9. doi:10.1007/978-1-4419-8474-6 {*124, 384*}

Kirsch, A. and F. Hettlich 2015. *The Mathematical Theory of Time-Harmonic Maxwell's Equations: Expansion-, Integral-, and Variational Methods*, Volume 190 of *Applied Mathematical Sciences*. Cham, Switzerland: Springer International Publishing. 337 pp. ISBN 3-319-11085-3 (print), 3-319-11086-1 (e-book). ISBN-13 978-3-319-11085-1 (print), 978-3-319-11086-8 (e-book). doi:10.1007/978-3-319-11086-8 {*96, 110, 112, 113, 115, 384*}

Kochmann, D. M. and G. W. Milton 2014, November. Rigorous bounds on the effective moduli of composites and inhomogeneous bodies with negative-stiffness phases. *Journal of the Mechanics and Physics of Solids* **71**:46–63. ISSN-L 0022-5096. doi:10.1016/j.jmps.2014.06.010 {*150, 385*}

Kohler, W. and G. C. Papanicolaou 1982. Bounds for the effective conductivity of random media. In R. Burridge, S. Childress, and G. C. Papanicolaou (eds.), *Macroscopic Properties of Disordered Media: Proceedings of a Conference Held at the Courant Institute, June 1–3, 1981*, Volume 154 of *Lecture Notes in Physics*, pp. 111–130 (of vi + 307). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-8224-8461-7, 0-273-08461-5. ISBN-13 978-0-8224-8461-5, 978-0-273-08461-7. {*54, 183, 385*}

Kohn, R. V. 1991, September. The relaxation of a double-well energy. *Continuum Mechanics and Thermodynamics* **3** (**3**):193–236. ISSN-L 0935-1175. doi:10.1007/BF01135336 {*2, 4, 5, 385*}

Kohn, R. V. 1992. Numerical structural optimization via a relaxed formulation. In M. C. Delfour and G. Sabidussi (eds.), *Shape Optimization and Free Boundaries: Proceedings of the NATO Advanced Study Institute and Séminaire de Mathématiques Supérieures on Shape Optimization and Free Boundaries, Montréal, Canada, June 25–July 13, 1990*, Volume 380 of *NATO ASI Series. Series C, Mathematical and Physical Sciences*, pp. 173–210 (of xviii + 462). Norwell, Massachusetts / Dordrecht, The Netherlands: Kluwer Academic Publishers Group. ISBN 0-7923-1944-3. ISBN-13 978-0-7923-1944-3. {*2, 385*}

Kohn, R. V. and T. D. Little 1998. Some model problems of polycrystal plasticity with deficient basic crystals. *SIAM Journal on Applied Mathematics* **59**(**1**):172–197. ISSN-L 0036-1399. doi:10.1137/S0036139997320019 {*19, 385*}

Kohn, R. V. and S. Müller 1992. Branching of twins near an austenite–twinned-martensite interface. *Philosophical Magazine A: Physics of Condensed Matter: Structure, Defects and Mechanical Properties* **66**(**5**):697–715. ISSN-L 1364-2804. doi:10.1080/01418619208201585 {*4, 385*}

Kohn, R. V. and G. Strang 1986. Optimum Design and Relaxation of Variational Problems, I, II, and III. *Communications on Pure and Applied Mathematics (New York)* **39**:113–377. ISSN-L 0010-3640. {*2, 385*}

Kohn, W. and L. J. Sham 1965, November. Self-consistent equations including exchange and correlation effects. *Physical Review (2)* **140**(**4A**):A1133–A1138. ISSN-L 0031-899X. doi:10.1103/PhysRev.140.A1133 {*309, 385*}

Kolokolnikov, T. and A. E. Lindsay 2015. Recovering multiple small inclusions in a three-dimensional domain using a single measurement. *Inverse Problems in Science and Engineering* **23**(**3**):377–388. ISSN-L 1741-5977. doi:10.1080/17415977.2014.906414 {*144, 385*}

Korányi, A. and L. Pukánszky 1963, September. Holmorphic functions with positive real part on polycylinders. *Transactions of the American Mathematical Society* **108**(**3**):449–456. ISSN-L 0002-9947. doi:10.2307/1993592 {*119, 385*}

Kozlov, S. M. 1978. Averaging of random structures. *Doklady Akademii Nauk SSSR* **241**(**5**):1016–1019. English translation in Soviet Math. Dokl. **19**(**4**):950–954 (1978). ISSN-L 0002-3264. {*11, 48, 49, 264, 385*}

Kreĭn, M. G. and A. A. Nudel'man 1974. *The Markov Moment Problem and Extremal Problems*, Volume 50 of *Translations of Mathematical Monographs*. Providence, RI, USA: American Mathematical Society. 182 pp. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development. Translated from the Russian by D. Louvish. ISBN 0-8218-4500-4. ISBN-13 978-0-8218-4500-4. {*68, 136, 151, 258, 385*}

Kresse, G. and J. Furthmüller 1996, October. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Physical Review B: Condensed Matter and Materials Physics* **54**(**16**):11169–11186. ISSN-L 1098-0121. doi:10.1103/PhysRevB.54.11169 {*310, 319, 385*}

Kröner, E. 1953. Das Fundamentalintegral der anisotropen elastischen Differentialgleichungen. (German) [The fundamental integral of isotropic elastic differential equations]. *Zeitschrift für Physik* **136**(**4**):402–410. ISSN-L 0044-3328. doi:10.1007/BF01343450 {*358, 385*}

Kröner, E. 1977, April. Bounds for the effective elastic moduli of disordered materials. *Journal of the Mechanics and Physics of Solids* **25**(**2**):137–155. ISSN-L 0022-5096. doi:10.1016/0022-5096(77)90009-6 {*193, 257, 288, 386*}

Kummert, A. 1989. Synthesis of two-dimensional lossless m-ports with prescribed scattering matrix. *Circuits, Systems, and Signal Processing: CSSP* **8**(**1**):97–119. ISSN-L 0278-081X. {*68, 151, 386*}

Lakes, R. S. 1987, February 27. Foam structures with a negative Poisson's ratio. *Science* **235**(**4792**):1038–1040. ISSN-L 0036-8075. doi:10.1126/science.235.4792.1038 {*50, 386*}

Lakes, R. S. 1996, January. Cellular solid structures with unbounded thermal expansion. *Journal of Materials Science Letters* **15**(**6**):475–477. ISSN-L 0261-8028. doi:10.1007/BF00275406 {*50, 386*}

Lakes, R. S. and W. J. Drugan 2002. Dramatically stiffer elastic composite materials due to a negative stiffness phase? *Journal of the Mechanics and Physics of Solids* **50**(**5**):979–1009. ISSN-L 0022-5096. doi:10.1016/S0022-5096(01)00116-8 {*150, 386*}

Lakes, R. S. and J. Quackenbush 1996. Viscoelastic behaviour in indium tin alloys over a wide range of frequency and time. *Philosophical Magazine Letters* **74**(**4**):227–232. ISSN-L 0950-0839. doi:10.1080/095008396180155 {*156, 386*}

Landau, L. D., E. M. Lifshitz, and L. P. Pitaevskiĭ 1984. *Electrodynamics of Continuous Media* (Second ed.), Volume 8 of *Landau and Lifshitz Course of Theoretical Physics*. Oxford, UK: Elsevier Butterworth-Heinemann. xiii + 460 pp. ISBN 0-7506-2634-8. ISBN-13 978-0-7506-2634-7. {*53, 118, 386*}

Landauer, R. 1978. Electrical conductivity in inhomogeneous media. In J. C. Garland and D. B. Tanner (eds.), *Electrical Transport and Optical Properties of Inhomogeneous Media*, Volume 40 of *AIP Conference Proceedings*, pp. 2–43 (of 416). Woodbury, New York: American Institute of Physics. ISBN 0-88318-139-8. ISBN-13 978-0-88318-139-3. {*47, 386*}

Larsen, A. A., B. Laksafoss, J. S. Jensen, and O. Sigmund 2009, February. Topological material layout in plates for vibration suppression and wave propagation control. *Structural and Multidisciplinary Optimization* **37**(**6**):585–594. ISSN-L 1615-147X. doi:10.1007/s00158-008-0257-0 {*31, 386*}

Lautrup, B. 2005. *Physics of Continuous Matter: Exotic and Everyday Phenonema in the Macroscopic World*, pp. 118–119 (of xiv + 608). Bristol, UK and Philadelphia, PA, USA: Institute of Physics. ISBN 0-7503-0752-8. ISBN-13 978-0-7503-0752-9. {*29, 386*}

Lawden, D. F. 1982. *An Introduction to Tensor Calculus, Relativity and Cosmology* (Third ed.). New York / London / Sydney, Australia: John Wiley and Sons. xiii + 205 pp. ISBN 0-471-10096-X. ISBN-13 978-0-471-10096-6. {*6, 386*}

Lebensohn, R. A. 2001, August 16. N-site modeling of a 3D viscoplastic polycrystal using Fast Fourier Transform. *Acta Materialia* **49**(**14**):2723–2737. ISSN-L 1359-6454. doi:10.1016/S1359-6454(01)00172-0 {*236, 386*}

Lee, C., W. Yang, and R. G. Parr 1988, January 15. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B: Condensed Matter and Materials Physics* **37**(**2**):785–789. ISSN-L 1098-0121. doi:10.1103/PhysRevB.37.785 {*309, 386*}

Lee, S.-B., R. A. Lebensohn, and A. D. Rollett 2011, May. Modeling the viscoplastic micromechanical response of two-phase materials using Fast Fourier Transforms. *International Journal of Plasticity* **27**(**5**):707–727. ISSN-L 0749-6419. doi:10.1016/j.ijplas.2010.09.002 {*236, 386*}

Levy, M. 1979, December 15. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the ν-representability problem. *Proceedings of the National Academy of Sciences of the United States of America* **76**(**12**):6062–6065. ISSN-L 0027-8424. doi:10.1073/pnas.76.12.6062 {*312, 386*}

Lifšic, I. M. and L. N. Rozencveĭg 1947. On the construction of the Green's tensor for the basic equation of the theory of elasticity of an anisotropic infinite medium. *Zhurnal eksperimental'noi i teoreticheskoi fiziki / Akademiia Nauk SSSR* **17**(**9**):783–791. ISSN-L 0044-4510. {*357, 358, 387*}

Lipton, R. 2000, October 1. Optimal bounds on electric field fluctuations for random composites. *Journal of Applied Physics* **88**(**7**):4287–4293. ISSN-L 0021-8979. doi:10.1063/1.1290734 {*96, 387*}

Lipton, R. 2001. Optimal inequalities for gradients of solutions of elliptic equations occurring in two-phase heat conductors. *SIAM Journal on Mathematical Analysis* **32**(**5**):1081–1093. ISSN-L 0036-1410. doi:10.1137/S0036141000366625 {*96, 387*}

Liu, Y., S. Guenneau, and B. Gralak 2013, October 15. Causality and passivity properties of effective parameters of electromagnetic multilayered structures. *Physical Review B: Condensed Matter and Materials Physics* **88**(**16**): 165104. ISSN-L 1098-0121. doi:10.1103/PhysRevB.88.165104 {*96, 387*}

Liu, Z., C. T. Chan, and P. Sheng 2005, January 1. Analytic model of phononic crystals with local resonances. *Physical Review B: Condensed Matter and Materials Physics* **71**(**1**):014103. ISSN-L 1098-0121. doi:10.1103/PhysRevB.71.014103 {*21, 50, 52, 387*}

Love, A. E. H. 1906. *A Treatise on the Mathematical Theory of Elasticity* (Second ed.), pp. 312–313. Cambridge, UK: Cambridge University Press. Article 226. URL https://goo.gl/E8pZlP {*153, 387*}

Lurie, K. A. and A. V. Cherkaev 1982. Accurate estimates of the conductivity of mixtures formed of two materials in a given proportion (two-dimensional problem). *Doklady Akademii Nauk SSSR* **264**:1128–1130. English translation in Soviet Phys. Dokl. **27**:461–462 (1982). ISSN-L 0002-3264. {*19, 130, 284, 289, 339, 387*}

Lurie, K. A. and A. V. Cherkaev 1984. Exact estimates of conductivity of composites formed by two isotropically conducting media taken in prescribed proportion. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **99**(**1–2**):71–87. ISSN-L 0308-2105. doi:10.1017/S030821050002597X {*19, 130, 284, 289, 339, 387*}

Lurie, K. A. and A. V. Cherkaev 1986. Exact estimates of the conductivity of a binary mixture of isotropic materials. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **104**(**1–2**):21–38. ISSN-L 0308-2105. doi:10.1017/S0308210500019041 {*55, 387*}

Luskin, M. 1996. On the computation of crystalline microstructure. *Acta Numerica* **5**:191–257. ISSN-L 0962-4929. ISBN 0-521-57234-7. ISBN-13 978-0-521-57234-7. doi:10.1017/S0962492900002658 {*2, 387*}

Martin, P. C., O. Parodi, and P. S. Pershan 1972, December. Unified hydrodynamic theory for crystals, liquid crystals, and normal fluids. *Physical Review A (Atomic, Molecular, and Optical Physics)* **6**(**6**):2401–2420. ISSN-L 1050-2947. doi:10.1103/PhysRevA.6.2401 {*29, 387*}

Mattei, O. 2016. On bounding the response of linear viscoelastic composites in the time domain: The variational approach and the analytic method. Ph.D. Thesis, University of Brescia, Brescia, Italy. xiv + 178 pp. {*v, vii, 150, 387*}

Mattner, L. 2001. Complex differentiation under the integral. *Nieuw Archief voor Wiskunde (Groningen)* **5/2**(**2**):32–35. ISSN-L 0028-9825. {*98, 387*}

McCoy, J. J. 1970. On the displacement field in an elastic medium with random variations of material properties. In A. C. Eringen (ed.), *Proceedings of the Sixth Annual Meeting of the Society of Engineering Science, Held at Princeton University, Princeton, New Jersey, November 11–13, 1968*, Volume 5 of *Recent Advances in Engineering Sciences*, pp. 235–254. Gordon and Breach: Gordon and Breach. ISBN 0-677-13780-X. ISBN-13 978-0-677-13780-3. {*55, 387*}

McPhedran, R. C., D. R. McKenzie, and G. W. Milton 1982, September. Extraction of structural information from measured transport properties of composites. *Applied Physics A: Materials Science & Processing* **29**(**1**):19–27. ISSN-L 0721-7250. doi:10.1007/BF00618111 {*125, 129, 258, 387*}

McPhedran, R. C. and G. W. Milton 1981, December. Bounds and exact theories for the transport properties of inhomogeneous media. *Applied Physics A: Materials Science & Processing* **26**(**4**):207–220. ISSN-L 0721-7250. doi:10.1007/BF00617840 {*246, 257, 258, 288, 388*}

McPhedran, R. C. and G. W. Milton 1990. Inverse transport problems for composite media. *Materials Research Society Symposium Proceedings* **195**:257–274. ISSN-L 0272-9172. doi:10.1557/PROC-195-257 {*129, 388*}

Mendelson, K. S. 1975, November 1. A theorem on the effective conductivity of a two-dimensional heterogeneous medium. *Journal of Applied Physics* **46**(**11**):4740–4741. ISSN-L 0021-8979. doi:10.1063/1.321549 {*16, 267, 388*}

Meyers, N. G. 1963. An L^p-estimate for the gradient of solutions of second order elliptic divergence form equations. *Annali della Scuola Normale Superiore di Pisa. Serie III* **17**(**3**):189–206. URL http://archive.numdam.org/item?id=ASNSP_1963_3_17_3_189_0 {*270, 388*}

Meza, L. R., S. Das, and J. R. Greer 2014, September 12. Strong, lightweight, and recoverable three-dimensional ceramic nanolattices. *Science* **345**(**6202**):1322–1326. ISSN-L 0036-8075. doi:10.1126/science.1255908 {*52, 388*}

Milgrom, M. 1990, June 15. Linear response of general composite systems to many coupled fields. *Physical Review B: Condensed Matter and Materials Physics* **41**(**18**):12484–12494. ISSN-L 1098-0121. doi:10.1103/PhysRevB.41.12484 {*16, 78, 256, 258, 259, 388*}

Milgrom, M. 1997, March. Some more exact results concerning multifield moduli of two-phase composites. *Journal of the Mechanics and Physics of Solids* **45**(**3**):399–404. ISSN-L 0022-5096. doi:10.1016/S0022-5096(96)00082-8 {*16, 258, 388*}

Milgrom, M. 2002, February 15. Forces in nonlinear media. *Journal of Physics A: Mathematical and General* **35**(**6**):1437–1453. ISSN-L 0305-4470. doi:10.1088/0305-4470/35/6/308 {*53, 388*}

Milgrom, M. and S. Shtrikman 1989a, September 15. Linear response of polycrystals to coupled fields: Exact relations among the coefficients. *Physical Review B: Condensed Matter and Materials Physics* **40**(**9**):5991–5994. ISSN-L 1098-0121. doi:10.1103/PhysRevB.40.5991 {*16, 258, 388*}

Milgrom, M. and S. Shtrikman 1989b, August. Linear response of two-phase composites with cross moduli: Exact universal relations. *Physical Review A (Atomic, Molecular, and Optical Physics)* **40**(**3**):1568–1575. ISSN-L 1050-2947. doi:10.1103/PhysRevA.40.1568 {*16, 258, 388*}

Miller, O. D., C. W. Hsu, M. T. H. Reid, W. Qiu, B. G. DeLacy, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson 2014, March 28. Fundamental limits to extinction by metallic nanoparticles. *Physical Review Letters* **112**(**12**):123903. ISSN-L 0031-9007. doi:10.1103/PhysRevLett.112.123903 {*151, 388*}

Miller, O. D., A. G. Polimeridis, M. T. H. Reid, C. W. Hsu, B. G. DeLacy, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson 2016, February 22. Fundamental limits to optical response in absorptive systems. *Optics Express* **24**(**4**):3329–3364. ISSN-L 1094-4087. doi:10.1364/OE.24.003329 {*151, 388*}

Milton, G. W. 1979. Theoretical studies of the transport properties of inhomogeneous media. Unpublished report **TP/79/1**, University of Sydney, Sydney, Australia. 1–65 pp. doi:10.13140/RG.2.1.2184.8482 {*68, 150, 183, 236, 388*}

Milton, G. W. 1980, August 1. Bounds on the complex dielectric constant of a composite material. *Applied Physics Letters* **37**(**3**):300–302. ISSN-L 0003-6951. doi:10.1063/1.91895 {*68, 96, 149, 151, 152, 178, 388*}

Milton, G. W. 1981a, August 1. Bounds on the complex permittivity of a two-component composite material. *Journal of Applied Physics* **52**(**8**):5286–5293. ISSN-L 0021-8979. doi:10.1063/1.329385 {*62, 68, 96, 151–153, 155, 167, 168, 178, 183, 184, 236, 388*}

Milton, G. W. 1981b, February 23. Bounds on the electromagnetic, elastic, and other properties of two-component composites. *Physical Review Letters* **46**(**8**):542–545. ISSN-L 0031-9007. doi:10.1103/PhysRevLett.46.542 {*54, 55, 67, 68, 388*}

Milton, G. W. 1981c, August 1. Bounds on the transport and optical properties of a two-component composite material. *Journal of Applied Physics* **52**(**8**):5294–5304. ISSN-L 0021-8979. doi:10.1063/1.329386 {*62, 96, 118, 136, 151, 152, 169, 178, 184, 246, 257, 258, 288, 389*}

Milton, G. W. 1982. Bounds on the elastic and transport properties of two-component composites. *Journal of the Mechanics and Physics of Solids* **30**(**3**):177–191. ISSN-L 0022-5096. doi:10.1016/0022-5096(82)90022-9 {*54, 55, 67, 389*}

Milton, G. W. 1984. Correlation of the electromagnetic and elastic properties of composites and microgeometries corresponding with effective medium theory. In D. L. Johnson and P. N. Sen (eds.), *Physics and Chemistry of Porous Media: Papers from a Symposium Held at Schlumberger–Doll Research, Oct. 24–25, 1983*, Volume 107 of *AIP Conference Proceedings*, pp. 66–77 (of 223). Woodbury, New York: American Institute of Physics. ISBN 0-88318-306-4. ISBN-13 978-0-88318-306-9. {*67, 389*}

Milton, G. W. 1985. The coherent potential approximation is a realizable effective medium scheme. *Communications in Mathematical Physics* **99**(**4**):463–500. ISSN-L 0010-3616. doi:10.1007/BF01215906 {*67, 389*}

Milton, G. W. 1987a. Multicomponent composites, electrical networks and new types of continued fraction. I. *Communications in Mathematical Physics* **111**(**2**):281–327. ISSN-L 0010-3616. doi:10.1007/BF01217763 {*54, 55, 67, 96, 118, 152, 181, 183, 184, 224, 227, 234, 257, 269–271, 279, 280, 283, 288, 389*}

Milton, G. W. 1987b. Multicomponent composites, electrical networks and new types of continued fraction. II. *Communications in Mathematical Physics* **111**(**3**):329–372. ISSN-L 0010-3616. doi:10.1007/BF01238903 {*55, 68, 96, 118, 151, 152, 181, 183–185, 224, 227, 234, 257, 269, 279, 282, 283, 288–290, 389*}

Milton, G. W. 1990. On characterizing the set of possible effective tensors of composites: The variational method and the translation method. *Communications on Pure and Applied Mathematics (New York)* **43**(**1**):63–125. ISSN-L 0010-3640. doi:10.1002/cpa.3160430104 {*viii, 18, 54, 68, 71, 118, 126, 130, 152, 153, 183, 355, 356, 366, 367, 389*}

Milton, G. W. 1991. The field equation recursion method. In G. Dal Maso and G. F. Dell'Antonio (eds.), *Composite Media and Homogenization Theory: Proceedings of the Workshop on Composite Media and Homogenization Theory Held in Trieste, Italy, from January 15 to 26, 1990*, Volume 5 of *Progress in Nonlinear Differential Equations and Their Applications*, pp. 223–245 (of xiv + 258). Basel, Switzerland: Birkhäuser Verlag. ISBN 0-8176-3511-4, 3-7643-3511-4. ISBN-13 978-0-8176-3511-4, 978-3-7643-3511-3. doi:10.1007/978-1-4684-6787-1_13 book-doi:10.1007/978-1-4684-6787-1 {*55, 56, 71, 181, 183, 185, 234, 285, 288, 290, 389*}

Milton, G. W. 1992, July. Composite materials with Poisson's ratios close to -1. *Journal of the Mechanics and Physics of Solids* **40**(**5**):1105–1137. ISSN-L 0022-5096. doi:10.1016/0022-5096(92)90063-8 {*50, 389*}

Milton, G. W. 1994. A link between sets of tensors stable under lamination and quasiconvexity. *Communications on Pure and Applied Mathematics (New York)* **47**(**7**):959–1003. ISSN-L 0010-3640. doi:10.1002/cpa.3160470704 {*4, 133, 389*}

Milton, G. W. 2001, October. Proof of a conjecture on the conductivity of checkerboards. *Journal of Mathematical Physics* **42**(**10**):4873–4882. ISSN-L 0022-2488. doi:10.1063/1.1385564 {*247, 389*}

Milton, G. W. 2002. *The Theory of Composites*, Volume 6 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge, UK: Cambridge University Press. xxviii + 719 pp. Series editors: P. G. Ciarlet, A. Iserles, Robert V. Kohn, and M. H. Wright. ISBN 0-521-78125-6. ISBN-13 978-0-521-78125-1. doi:10.1017/CBO9780511613357 {*v, ix, 2, 3, 5, 15–17, 19, 26, 45, 50, 55, 56, 58, 59, 62, 63, 65, 70, 71, 73, 77, 78, 81, 82, 86, 87, 90, 96, 97, 110, 116–119, 127, 130, 131, 133, 136, 150, 152, 154, 169, 170, 180, 181, 183–186, 192, 193, 203, 207, 210, 214, 216, 221, 224, 227, 229, 232, 234, 236, 237, 240, 245–247, 249, 250, 257, 258, 288–290, 293, 294, 296, 298, 300, 303, 319, 325, 327, 331, 339, 343, 389*}

Milton, G. W. 2007. New metamaterials with macroscopic behavior outside that of continuum elastodynamics. *New Journal of Physics* **9**(**10**):359. ISSN-L 1367-2630. doi:10.1088/1367-2630/9/10/359 {*29, 389*}

Milton, G. W. 2012. Universal bounds on the electrical and elastic response of two-phase bodies and their application to bounding the volume fraction from boundary measurements. *Journal of the Mechanics and Physics of Solids* **60**(**1**): 139–155. ISSN-L 0022-5096. doi:10.1016/j.jmps.2011.09.007 {*78, 390*}

Milton, G. W. 2013a, July. Complete characterization of the macroscopic deformations of periodic unimode metamaterials of rigid bars and pivots. *Journal of the Mechanics and Physics of Solids* **61**(**7**):1543–1560. ISSN-L 0022-5096. doi:10.1016/j.jmps.2012.08.011 {*50, 390*}

Milton, G. W. 2013b. Sharp inequalities that generalize the divergence theorem: an extension of the notion of quasi-convexity. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **469**(**2157**): 20130075 (18 pp.). See addendum (Milton 2015a). ISSN-L 1364-5021. doi:10.1098/rspa.2013.0075 {*17, 19, 71, 131, 252, 253, 338, 390*}

Milton, G. W. 2015a, March 4. Addendum to "Sharp inequalities that generalize the divergence theorem: an extension of the notion of quasi-convexity". *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **471**(**2176**). See (Milton 2013b). ISSN-L 1364-5021. doi:10.1098/rspa.2014.0886 {*390*}

Milton, G. W. 2015b, July. New examples of three-dimensional dilational materials. *Physica Status Solidi. B, Basic Solid State Physics* **252**(**7**):1426–1430. ISSN-L 0370-1972. doi:10.1002/pssb.201552297 {*50, 390*}

Milton, G. W. 1986a. Modeling the properties of composites by laminates. In J. L. Ericksen, D. Kinderlehrer, R. V. Kohn, and J.-L. Lions (eds.), *Homogenization and Effective Moduli of Materials and Media*, Volume 1 of *The IMA Volumes in Mathematics and its Applications*, pp. 150–174 (of x + 263). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-96306-5. ISBN-13 978-0-387-96306-8. doi:10.1007/978-1-4613-8646-9 {*55, 68, 151, 184, 284, 390*}

Milton, G. W. 1986b. A proof that laminates generate all possible effective conductivity functions of two-dimensional, two-phase media. In G. C. Papanicolaou (ed.), *Advances in Multiphase Flow and Related Problems: Proceedings of the Workshop on Cross Disciplinary Research in Multiphase Flow, Leesburg, Virginia, June 2–4, 1986*, pp. 136–146 (of x + 295). Philadelphia: SIAM Press. ISBN 0-89871-212-2. ISBN-13 978-0-89871-212-4. {*152, 278, 390*}

Milton, G. W. and J. G. Berryman 1997, September 8. On the effective viscoelastic moduli of two-phase media. II. Rigorous bounds on the complex shear modulus in three dimensions. *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences* **453**(**1964**):1849–1880. ISSN-L 1364-5021. doi:10.1098/rspa.1997.0100 {*152, 356, 390*}

Milton, G. W., M. Briane, and J. R. Willis 2006. On cloaking for elasticity and physical equations with a transformation invariant form. *New Journal of Physics* **8**(**10**):248. ISSN-L 1367-2630. doi:10.1088/1367-2630/8/10/248 {*21, 24, 27, 29, 40, 50, 53, 390*}

Milton, G. W. and A. V. Cherkaev 1995, October 1. Which elasticity tensors are realizable? *ASME Journal of Engineering Materials and Technology* **117**(**4**):483–493. ISSN-L 0094-4289. doi:10.1115/1.2804743 {*4, 133, 390*}

Milton, G. W. and K. M. Golden 1985. Thermal conduction in composites. In T. Ashworth and D. R. Smith (eds.), *Thermal Conductivity*, Volume 18, pp. 571–582 (of xix + 764). New York / London: Plenum Press. ISBN 0-306-41918-1. ISBN-13 978-0-306-41918-8. {*55, 68, 181, 234, 257, 288, 290, 390*}

Milton, G. W. and K. M. Golden 1990. Representations for the conductivity functions of multicomponent composites. *Communications on Pure and Applied Mathematics (New York)* **43**(**5**):647–671. ISSN-L 0010-3640. doi:10.1002/cpa.3160430504 {*96, 119, 390*}

Milton, G. W. and R. C. McPhedran 1982. A comparison of two methods for deriving bounds on the effective conductivity of composites. In R. Burridge, S. Childress, and G. C. Papanicolaou (eds.), *Macroscopic Properties of Disordered Media: Proceedings of a Conference Held at the Courant Institute, June 1–3, 1981*, Volume 154 of *Lecture Notes in Physics*, pp. 183–193 (of vi + 307). Berlin / Heidelberg / London / etc.: Springer-Verlag. ISBN 0-387-11202-2. ISBN-13 978-0-387-11202-2. {*68, 246, 257, 288, 390*}

Milton, G. W., R. C. McPhedran, and D. R. McKenzie 1981. Transport properties of arrays of intersecting cylinders. *Applied Physics* **25**(**1**):23–30. ISSN-L 0947-8396. doi:10.1007/BF00935387 {*151, 391*}

Milton, G. W. and L. H. Nguyen 2012. Bounds on the volume fraction of 2-phase, 2-dimensional elastic bodies and on (stress, strain) pairs in composites. *Comptes Rendus M'ecanique* **340**(**4–5**):193–204. ISSN-L 1631-0721. doi:10.1016/j.crme.2012.02.002 {*18, 78, 131, 339, 391*}

Milton, G. W. and N. Phan-Thien 1982. New bounds on effective elastic moduli of two-component materials. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **380**(**1779**):305–331. ISSN-L 0080-4630. doi:10.1098/rspa.1982.0044 {*55, 257, 288, 391*}

Milton, G. W., P. Seppecher, and G. Bouchitté 2009, February 8. Minimization variational principles for acoustics, elastodynamics and electromagnetism in lossy inhomogeneous bodies at fixed frequency. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **465**(**2102**):367–396. ISSN-L 1364-5021. doi:10.1098/rspa.2008.0195 {*12, 21, 24, 78, 79, 127–129, 137, 153, 337, 338, 342, 343, 356, 391*}

Milton, G. W. and J. R. Willis 2007. On modifications of Newton's second law and linear continuum elastodynamics. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **463**(**2079**):855–880. ISSN-L 1364-5021. doi:10.1098/rspa.2006.1795 {*21, 391*}

Milton, G. W. and J. R. Willis 2010. Minimum variational principles for time-harmonic waves in a dissipative medium and associated variational principles of Hashin–Shtrikman type. *Proceedings of the Royal Society A: Mathematical, Physical, & Engineering Sciences* **466**(**2122**):3013–3032. ISSN-L 1364-5021. doi:10.1098/rspa.2010.0006 {*12, 78, 153, 337, 338, 341, 356, 358, 391*}

Mindlin, R. D. 1951. Influence of rotary inertia and shear on flexural motions of isotropic, elastic plates. *Journal of Applied Mechanics* **18**(**1**):31–38. ISSN-L 0021-8936. URL http://en.journals.sid.ir/ViewPaper.aspx?ID=294064 {*31, 391*}

Misner, C. W., K. S. Thorne, and J. A. Wheeler 1973. *Gravitation* (Third ed.). San Francisco, CA: Freeman. xxvi + 1279 pp. ISBN 0-7167-0344-0. ISBN-13 978-0-7167-0344-0. {*7, 391*}

Monk, P. B. 2003. *Finite Element Methods for Maxwell's Equations*. Numerical mathematics and scientific computation. New York: Clarendon Press. xiv + 450 pp. ISBN 0-19-850888-3. ISBN-13 978-0-19-850888-5. {*96, 113, 303, 391*}

Morrey, C. B. 1952. Quasiconvexity and the lower semicontinuity of multiple integrals. *Pacific Journal of Mathematics* **2**(**1**):25–53. ISSN-L 0030-8730. doi:10.2140/pjm.1952.2.25 {*17, 338, 391*}

Morrey, C. B. 1966. *Multiple Integrals in the Calculus of Variations*, Volume 130 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berucksichtigung der Anwendungsgebiete*, p. 122 (of ix + 506). Berlin / Heidelberg / London / etc.: Springer-Verlag. {*17, 391*}

Mortola, S. and S. Steffé 1985. A two-dimensional homogenization problem. (Italian). *Atti della Accademia Nazionale dei Lincei. Rendiconti. Classe di Scienze Fisiche, Matematiche e Naturali. Serie VIII* **78**(**3**):77–82. ISSN-L 0392-7881. {*247, 391*}

Moulinec, H. and F. Silva 2014, March 30. Comparison of three accelerated FFT-based schemes for computing the mechanical response of composite materials. *International Journal for Numerical Methods in Engineering* **97**(**13**):960–985. ISSN-L 0029-5981. doi:10.1002/nme.4614 {*45, 78, 236, 319, 391*}

Moulinec, H. and P. M. Suquet 1994. A fast numerical method for computing the linear and non-linear properties of composites. *Comptes Rendus des Séances de l'Académie des Sciences. Série II* **318**:1417–1423. {*45, 58, 78, 235–237, 245, 246, 248, 288, 319, 327, 391*}

Moulinec, H. and P. M. Suquet 1998. A numerical method for computing the overall response of non-linear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering* **157**(**1–2**):69–94. ISSN-L 0045-7825. doi:10.1016/S0045-7825(97)00218-1 {*45, 58, 78, 235–237, 245, 246, 248, 288, 319, 327, 391*}

Movchan, A. B. and S. Guenneau 2004, September 15. Split-ring resonators and localized modes. *Physical Review B: Condensed Matter and Materials Physics* **70**(**12**):125116. ISSN-L 1098-0121. doi:10.1103/PhysRevB.70.125116 {*21, 51, 392*}

Mueller, J. L. and S. Siltanen 2012. *Linear and Nonlinear Inverse Problems with Practical Applications*. Computational Science & Engineering. Philadelphia: SIAM Press. xiii + 351 pp. ISBN 1-61197-233-7. ISBN-13 978-1-61197-233-7. doi:10.1137/1.9781611972344 {*124, 142, 392*}

Mujica, J. 1986. *Complex Analysis in Banach Spaces: Holomorphic Functions and Domains of Holomorphy in Finite and Infinite Dimensions*, Volume 120 of *North-Holland Mathematical Studies*. Amsterdam, The Netherlands: Elsevier Science Publishers B.V. xi + 434 pp. ISBN 0-444-87886-6. ISBN-13 978-0-444-87886-1. {*98, 290, 392*}

Müller, S. 1987, September. Homogenization of nonconvex integral functionals and cellular elastic materials. *Archive for Rational Mechanics and Analysis* **99**(**3**):189–212. ISSN-L 0003-9527. doi:10.1007/BF00284506 {*2, 392*}

Müller, S. 1998. *Variational Models for Microstructure and Phase Transitions*, Volume 2 of *Lecture Notes*. Leipzig, Germany: Max Planck Institute for Mathematics in the Sciences. 127 pp. doi:10.1007/BFb0092670 {*2, 392*}

Murat, F. 1978. Compacité par compensation. (French) [Compactness through compensation]. *Annali della Scuola normale superiore di Pisa, Classe di scienze. Serie IV* **5**(**3**):489–507. ISSN-L 0391-173X. URL http://www.numdam.org/item?id=ASNSP_1978_4_5_3_489_0 {*11, 14, 17, 392*}

Murat, F. 1981. Compacité par compensation: Condition nécessaire et suffisante de continuité faible sous une hypothèse de rang constant. (French) [Compensated compactness: Necessary and sufficient conditions for weak continuity under a constant-rank hypothesis]. *Annali della Scuola normale superiore di Pisa, Classe di scienze. Serie IV* **8**(**1**): 69–102. ISSN-L 0391-173X. URL http://www.numdam.org/item?id=ASNSP_1981_4_8_1_69_0 {*11, 14, 17, 392*}

Murat, F. 1987. A survey on compensated compactness. In L. Cesari (ed.), *Contributions to Modern Calculus of Variations*, Volume 148 of *Pitman Research Notes in Mathematics Series*, pp. 145–183 (of 231). Harlow, Essex, UK: Longman Scientific and Technical. Papers from the symposium marking the centenary of the birth of Leonida Tonelli held in Bologna, May 13–14, 1985. ISBN 0-470-20378-1 (Wiley), 0-582-99458-6. ISBN-13 978-0-470-20378-1 (Wiley), 978-0-582-99458-4. {*11, 14, 17, 392*}

Murat, F. and L. Tartar 1985. Calcul des variations et homogénísation. (French) [Calculus of variation and homogenization]. In *Les méthodes de l'homogénéisation: théorie et applications en physique*, Volume 57 of *Collection de la Direction des études et recherches d'Électricité de France*, pp. 319–369 (of 405). Paris: Eyrolles. English translation in Topics in the Mathematical Modelling of Composite Materials, pp. 139–173, ed. by A. Cherkaev and R. Kohn, ISBN 0-8176-3662-5. ISSN-L 0399-4198. {*11, 19, 55, 130, 131, 253, 284, 289, 339, 392*}

Murat, F. and L. Tartar 1997. H-convergence. In A. V. Cherkaev and R. V. Kohn (eds.), *Topics in the Mathematical Modelling of Composite Materials*, Volume 31 of *Progress in Nonlinear Differential Equations and Their Applications*, pp. 21–43 (of xiv + 317). Basel, Switzerland: Birkhäuser Verlag. ISBN 0-8176-3662-5 (Boston), 3-7643-3662-5 (Basel). ISBN-13 978-0-8176-3662-3 (Boston), 978-3-7643-3662-2 (Basel). {*49, 392*}

Naboko, S. N. 1996, July. Zygmund's theorem and the boundary behavior of operator R-functions. *Functional Analysis and Its Applications* **30**(**3**):211–213. ISSN-L 0016-2663. doi:10.1007/BF02509510 {*96, 392*}

Nedelec, J.-C. 2001. *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*, Volume 144 of *Applied Mathematical Sciences*. New York, NY: Springer Science & Business Media. ix + 316 pp. ISBN 0-387-95155-5. ISBN-13 978-0-387-95155-3. doi:10.1007/978-1-4757-4393-7 {*96, 113, 392*}

Nemat-Nasser, S. and M. Hori 1993. *Micromechanics: Overall Properties of Heterogeneous Materials*, Volume 37 of *North-Holland Series in Applied Mathematics and Mechanics*. Amsterdam: North-Holland Publishing Co. xx + 687 pp. ISBN 0-444-89881-6. ISBN-13 978-0-444-89881-4. {*78, 392*}

Nemat-Nasser, S. and M. Hori 1999. *Micromechanics: Overall Properties of Heterogeneous Materials* (Second ed.). Amsterdam: Elsevier. xxiv + 786 pp. ISBN 0-444-50084-7. ISBN-13 978-0-444-50084-7. {*2, 48, 150, 392*}

Nesi, V. and E. Rogora 2007, January/March. A complete characterization of invariant jointly rank-r convex quadratic forms and applications to composite materials. *ESAIM: Control, Optimisation and Calculus of Variations* **13**(**1**): 1–34. ISSN-L 1262-3377. doi:10.1051/cocv:2007002 {*19, 393*}

Nevanlinna, R. 1919. Über beschränkte Funktionen, die in gegebenen Punkten vorgeschrieben Werte annehmen. (German) [On limited functions that take prescribed values at given points]. *Annales Academiae Scientiarum Fennicae, Series A* **13**(**1**):71. {*68, 136, 151, 258, 393*}

Nevanlinna, R. 1929. Über beschränkte analytische Funktionen. (German) [On limited analytic functions]. *Annales Academiae Scientiarum Fennicae, Series A* **32**(**7**):75. {*68, 136, 151, 258, 393*}

Nguetseng, G. 1989. A general convergence result for a functional related to the theory of homogenization. *SIAM Journal on Mathematical Analysis* **20**(**3**):608–623. ISSN-L 0036-1410. doi:10.1137/0520043 {*49, 393*}

Nicorovici, N. A., R. C. McPhedran, and G. W. Milton 1993, September 8. Transport properties of a three-phase composite material: The square array of coated cylinders. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **442**(**1916**):599–620. ISSN-L 0080-4630. doi:10.1098/rspa.1993.0124 {*224, 393*}

Noether, E. 1918. Invariante Variationsprobleme. (German) [Invariant variational problems]. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-Physikalische Klasse*:235–257. ISSN-L 0369-6650. {*14, 393*}

Norris, A. N. 1992, February 1. On the correspondence between poroelasticity and thermoelasticity. *Journal of Applied Physics* **71**(**3**):1138–1141. ISSN-L 0021-8979. doi:10.1063/1.351278 {*9, 393*}

Norris, A. N. 1994. Dynamic Green's functions in anisotropic piezoelectric, thermoelastic and poroelastic solids. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **447**(**1929**):175–188. ISSN-L 0080-4630. doi:10.1098/rspa.1994.0134 {*29, 33, 35, 358, 393*}

Norris, A. N. and A. L. Shuvalov 2011, September. Elastic cloaking theory. *Wave Motion* **48**(**6**):525–538. ISSN-L 0165-2125. doi:10.1016/j.wavemoti.2011.03.002 {*29, 393*}

Norris, J. R. 1997, November. Long time behavior of heat flow: Global estimates and exact asymptotics. *Archive for Rational Mechanics and Analysis* **140**(**2**):161–195. ISSN-L 0003-9527. doi:10.1007/s002050050063 {*126, 355, 393*}

Nowacki, W. 1986. *Thermoelasticity* (Second ed.). New York: Pergamon. xi + 566 pp. Translated from the Polish by Henryk Zorski. ISBN 0-08-024767-9. ISBN-13 978-0-08-024767-0. {*8, 393*}

Obnosov, Y. V. 1999. Periodic heterogeneous structures: New explicit solutions and effective characteristics of refraction of an imposed field. *SIAM Journal on Applied Mathematics* **59**(**4**):1267–1287. ISSN-L 0036-1399. doi:10.1137/S0036139997314770 {*247, 393*}

Ola, P., L. Päivärinta, and E. Somersalo 2012. An inverse boundary value problem in electrodynamics. *Duke Mathematical Journal* **70**(**3**):617–653. ISSN-L 0012-7094. doi:10.1215/S0012-7094-93-07014-7 {*96, 112, 393*}

Orum, C., E. Cherkaev, and K. M. Golden 2012, January. Recovery of inclusion separations in strongly heterogeneous composites from effective property measurements. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **468**(**2139**):784–809. ISSN-L 1364-5021. doi:10.1098/rspa.2011.0527 {*236, 393*}

Ou, M. Y. 2012. Two-parameter integral representation formula for the effective elastic moduli of two-phase composites. *Complex Variables and Elliptic Equations* **57**(**2–4**):411–424. ISSN-L 1747-6933. doi:10.1080/17476933.2011.611942 {*152, 154, 393*}

Papanicolaou, G. C. and S. R. S. Varadhan 1982. Boundary value problems with rapidly oscillating random coefficients. *Colloquia Mathematica Societatis János Bolyai* **27**:835–873. ISSN-L 0139-3383. URL http://math.stanford.edu/~papanico/pubftp/pubs_old/pap_vara_79.pdf {*11, 49, 54, 183, 393*}

Parr, R. G. and Y. Weitao 1994. *Density-Functional Theory of Atoms and Molecules*, Volume 16 of *International Series of Monographs on Chemistry*. Oxford, UK: Oxford University Press. ix + 333 pp. ISBN 0-19-504279-4, 0-19-509276-7 (paperback), 1-4294-0606-2 (e-book). ISBN-13 978-0-19-504279-5, 978-0-19-509276-9 (paperback), 978-1-4294-0606-2 (e-book). URL http://catdir.loc.gov/catdir/enhancements/fy0639/88025157-d.html; http://catdir.loc.gov/catdir/enhancements/fy0639/88025157-t.html {*23, 310, 312, 394*}

Pedregal, P. 1989. Weak continuity and weak lower semicontinuity for some compensation operators. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **113**(**3–4**):267–279. ISSN-L 0308-2105. doi:10.1017/S0308210500024136 {*14, 394*}

Pendry, J. B., A. J. Holden, D. J. Robbins, and W. J. Stewart 1999. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE transactions on microwave theory and techniques* **47**(**11**):2075–2084. ISSN-L 0018-9480. doi:10.1109/22.798002 {*50, 394*}

Pendry, J. B., D. Schurig, and D. R. Smith 2006, June 23. Controlling electromagnetic fields. *Science* **312**(**5781**): 1780–1782. ISSN-L 0036-8075. doi:10.1126/science.1125907 {*29, 394*}

Pendry, J. B. and D. R. Smith 2004, June. Reversing light with negative refraction. *Physics Today* **57**(**6**):37–43. ISSN-L 0031-9228. doi:10.1063/1.1784272 {*52, 394*}

Perfekt, K.-M. and M. Putinar 2016, January 13. The essential spectrum of the Neumann–Poincaré operator on a domain with corners. *arxiv.org*:1–15. URL http://arxiv.org/abs/1601.03183 {*248, 394*}

Persson, L.-E., L. Persson, N. Svanstedt, and J. Wyller 1993. *The Homogenization Method: An Introduction*. Lund, Sweden: Studentlitteratur. viii + 80 pp. ISBN 91-44-37661-8. ISBN-13 978-91-44-37661-5. {*49, 394*}

Phan-Thien, N. and G. W. Milton 1982. New bounds on the effective thermal conductivity of N-phase materials. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **380**(**1779**):333–348. ISSN-L 0080-4630. doi:10.1098/rspa.1982.0045 {*257, 288, 394*}

Phan-Thien, N. and G. W. Milton 1983. New third-order bounds on the effective moduli of N-phase composites. *Quarterly of Applied Mathematics* **41**(**1**):59–74. ISSN-L 0033-569X. {*257, 394*}

Pick, G. 1915. Über die Beschränkte analytische Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden. (German) [On limited analytical functions that pass through given function values]. *Mathematische Annalen* **77**(**1**): 7–23. ISSN-L 0025-5831. doi:10.1007/BF01456817 {*68, 136, 151, 258, 394*}

Pierce, A. D. 1981. *Acoustics: An Introduction to its Physical Principles and Applications*, pp. 515–517 (of xxiii + 642). McGraw-Hill Series in Mechanical Engineering. New York, NY, USA: McGraw-Hill. ISBN 0-07-049961-6. ISBN-13 978-0-07-049961-4. {*24, 394*}

Pitkonen, M. 2010. A closed-form solution for the polarizability of a dielectric double half-cylinder. *Journal of Electromagnetic Waves and Applications* **24**(**8–9**):1267–1628. ISSN-L 0920-5071. doi:10.1163/156939310791586106 {*248, 394*}

Plona, T. J. 1980, February 15. Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies. *Applied Physics Letters* **36**(**4**):259–261. ISSN-L 0003-6951. doi:10.1063/1.91445 {*33, 394*}

Post, E. J. 1962. *Formal Structure of Electromagnetics: General Covariance and Electromagnetics*. Amsterdam: North-Holland Publishing Co. xiv + 204 pp. Dover paperback edition in 1997. ISBN 0-486-65427-3. ISBN-13 978-0-486-65427-0. {*37, 40, 394*}

Prager, S. 1960, July 1. Diffusion in inhomogeneous media. *Journal of Chemical Physics* **33**(**1**):122–127. ISSN-L 0021-9606. doi:10.1063/1.1731066 {*257, 288, 394*}

Prager, S. 1969, May 15. Improved variational bounds on some bulk properties of a two-phase random medium. *Journal of Chemical Physics* **50**(**10**):4305–4312. ISSN-L 0021-9606. doi:10.1063/1.1670895 {*136, 150, 258, 394*}

Qiu, C.-W. and B. Luk'yanchuk 2008. Peculiarities in light scattering by spherical particles with radial anisotropy. *Journal of the Optical Society of America. A, Optics, image science, and vision* **25**(**7**):1623–1628. ISSN-L 1084-7529. doi:10.1364/JOSAA.25.001623 {*248, 394*}

Raitums, U. 1997. Lecture notes on G-convergence, convexification and optimal control problems for elliptic equations. Lecture Notes **39**, Department of Mathematics, University of Jyväskylä, Jyväskylä, Finland. iii + 86 pp. {*49, 395*}

Reuss, A. 1929. Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle. (German) [Calculation of the yield point of mixed crystals from plasticity conditions for single crystals]. *Zeitschrift für Angewandte Mathematik und Mechanik* **9**(**1**):49–58. ISSN-L 0044-2267. doi:10.1002/zamm.19290090104 {*64, 395*}

Rice, J. R. 1968, June. A path independent integral and the approximate analysis of strain concentration by notches and cracks. *Journal of Applied Mechanics* **35**(**2**):379–386. ISSN-L 0021-8936. doi:10.1115/1.3601206 {*14, 395*}

Richins, R. B. and D. C. Dobson 2012, January/February. A numerical minimization scheme for the complex Helmholtz equation. *Mathematical Modelling and Numerical Analysis = Modelisation mathématique et analyse numérique:* M^2AN **46**(**1**):39–57. ISSN-L 0764-583X. doi:10.1051/m2an/2011017 {*338, 356, 395*}

Riley, K. F., M. P. Hobson, and S. J. Bence 2002. *Mathematical Methods for Physics and Engineering* (Second ed.). Cambridge, UK: Cambridge University Press. xxiii + 1232 pp. ISBN 0-521-81372-7 (hardcover), 0-521-89067-5 (paperback). ISBN-13 978-0-521-81372-3 (hardcover), 978-0-521-89067-0 (paperback). {*6, 395*}

Ring, W. 1995, June. Identification of a core from boundary data. *SIAM Journal on Applied Mathematics* **55**(**3**):677–706. ISSN-L 0036-1399. doi:10.1137/S0036139993256308 {*144, 395*}

Roitburd, A. L. 1978. Martensitic transformation as a typical phase transformation in solids. *Solid State Physics (New York. 1955)* **33**:317–390. ISSN-L 0081-1947. doi:10.1016/S0081-1947(08)60471-3 {*2, 395*}

Roytburd, A. L. 1967. Orientational and habit relationships between crystalline phases in solid state transformations. *Kristallografiya* **12**:567–574. English translation in Sov. Phys. Crystallography **12**:499–504 (1968). {*2, 395*}

Roytburd, A. L. 1968. Domain structure of crystals formed in a solid phase. *Fizika Tverdogo Tela* **10**:2870–2876. English translation in Sov. Phys. Solid State **10**:2870–2876 (1968). {*2, 395*}

Roytburd, A. L. 1993. Elastic domains and polydomain phases in solids. *Phase Transitions* **45**(**1**):1–33. ISSN-L 0141-1594. doi:10.1080/01411599308203516 {*2, 395*}

Rudin, W. 1969. *Function Theory in Polydiscs*. New York: W. A. Benjamin, Inc. vii + 188 pp. {*119, 395*}

Rudin, W. 1987. *Real and Complex Analysis*. New York: McGraw-Hill. xiv + 416 pp. ISBN 0-07-054234-1. ISBN-13 978-0-07-054234-1. {*86, 395*}

Sanchez-Palencia, E. 1980. *Non-Homogeneous Media and Vibration Theory*, Volume 127 of *Lecture Notes in Physics*. Berlin / Heidelberg / London / etc.: Springer-Verlag. ix + 398 pp. ISBN 0-540-10000-8. ISBN-13 978-0-540-10000-2. {*49, 395*}

Sawicz, R. and K. M. Golden 1995, December 15. Bounds on the complex permittivity of matrix-particle composites. *Journal of Applied Physics* **78**(**12**):7240–7246. ISSN-L 0021-8979. doi:10.1063/1.360436 {*236, 395*}

Schapery, R. A. 1974. Viscoelastic behavior and analysis of composite materials. In G. P. Sendeckyj (ed.), *Composite Materials, Volume 2: Mechanics of Composite Materials*, pp. 85–168 (of xv + 503). New York, USA: Academic Press. ISBN 0-12-136502-6. ISBN-13 978-0-12-136502-8. {*153, 395*}

Schelkunoff, S. A. and H. T. Friis 1952. *Antennas: the theory and practice*, pp. 584–585 (of xxii + 639). New York / London / Sydney, Australia: John Wiley and Sons. {*50, 395*}

Schoenberg, M. and P. N. Sen 1983, January. Properties of a periodically stratified acoustic half-space and its relation to a Biot fluid. *Journal of the Acoustical Society of America* **73**(**1**):61–67. ISSN-L 0001-4966. doi:10.1121/1.388724 {*27, 50, 395*}

Schulgasser, K. 1989, September. Environmentally-induced expansion of heterogeneous media. *Journal of Applied Mechanics* **56**(**3**):546–549. ISSN-L 0021-8936. doi:10.1115/1.3176125 {*9, 395*}

Schulgasser, K. and Z. Hashin 1976, February 1. Bounds for effective permittivities of lossy dielectric composites. *Journal of Applied Physics* **47**(**2**):424–427. ISSN-L 0021-8979. doi:10.1063/1.322664 {*150, 396*}

Schur, J. 1917a, January. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind I. (German) [On power series that are limited in the inside of the unit circle I]. *Journal für die reine und angewandte Mathematik* **147**: 205–232. ISSN-L 0075-4102. doi:10.1515/crll.1917.147.205 {*289, 396*}

Schur, J. 1917b. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind II. (German) [On power series that are limited in the inside of the unit circle II]. *Journal für die reine und angewandte Mathematik* **148**:122–145. ISSN-L 0075-4102. doi:10.1515/crll.1918.148.122 {*289, 396*}

Segall, M. D., P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne 2002, March 25. First-principles simulation: ideas, illustrations and the CASTEP code. *Journal of Physics: Condensed Matter* **14** (**11**):2717–2744. ISSN-L 0953-8984. URL http://stacks.iop.org/0953-8984/14/i=11/a=301 {*310, 319, 396*}

Sen, A. K. and S. Torquato 1989, March 1. Effective conductivity of anisotropic two-phase composite media. *Physical Review B: Condensed Matter and Materials Physics* **39**(**7**):4504–4515. ISSN-L 1098-0121. doi:10.1103/PhysRevB.39.4504 {*257, 288, 396*}

Serdikukov, A., I. Semchenko, S. Tretkyakov, and A. Sihvola 2001. *Electromagnetics of Bi-anisotropic Materials, Theory and Applications*. Amsterdam, The Netherlands: Gordon and Breach. xiv + 337 pp. ISBN 90-5699-327-5. ISBN-13 978-90-5699-327-6. {*27, 40, 396*}

Sheng, P., X. X. Zhang, Z. Liu, and C. T. Chan 2003, October. Locally resonant sonic materials. *Physica. B, Condensed Matter* **338**(**1–4**):201–205. ISSN-L 0921-4526. doi:10.1016/S0921-4526(03)00487-3 {*21, 50, 52, 396*}

Shipman, S. P. and A. T. Welters 2013, October. Resonant electromagnetic scattering in anisotropic layered media. *Journal of Mathematical Physics* **54**(**10**):103511. ISSN-L 0022-2488. doi:10.1063/1.4824686 {*101, 102, 396*}

Sigmund, O. 1995, June. Tailoring materials with prescribed elastic properties. *Mechanics of Materials: An International Journal* **20**(**4**):351–368. ISSN-L 0167-6636. doi:10.1016/0167-6636(94)00069-7 {*50, 396*}

Sigmund, O. and S. Torquato 1996, November 18. Composites with extremal thermal expansion coefficients. *Applied Physics Letters* **69**(**21**):3203–3205. ISSN-L 0003-6951. doi:10.1063/1.117961 {*50, 396*}

Sigmund, O. and S. Torquato 1997. Design of materials with extreme thermal expansion using a three-phase topology optimization method. *Journal of the Mechanics and Physics of Solids* **45**(**6**):1037–1067. ISSN-L 0022-5096. doi:10.1016/S0022-5096(96)00114-7 {*50, 396*}

Smyshlyaev, V. P. 2009, April. Propagation and localization of elastic waves in highly anisotropic periodic composites via two-scale homogenization. *Mechanics of Materials: an International Journal* **41**(**4**):434–447. Special Issue in Honor of Graeme W. Milton, 2007 Winner of the William Prager Medal of the Society of Engineering Science. ISSN-L 0167-6636. doi:10.1016/j.mechmat.2009.01.009 {*21, 396*}

Song, Y., X. Chen, V. Dabade, T. W. Shield, and R. D. James 2013, October 3. Enhanced reversibility and unusual microstructure of a phase-transforming material. *Nature* **502**(**7469**):85–88. ISSN-L 0028-0836. doi:10.1038/nature12532 {*4, 396*}

Stefanov, P., A. Vasy, and M. Zworski (eds.) 2014. *Inverse Problems and Applications: Conference in Honor of Gunther Uhlmann on Inverse Problems, June 18–22, 2012, University of California, Irvine, CA: International Conference in Honor of Gunther Uhlmann's 60th Birthday on Inverse Problems and Applications, September 17–21, 2012, Yuquan Campus, Zhejiang University, Hangzhou, China*, Volume 615 of *Contemporary Mathematics*, Providence, RI, USA. American Mathematical Society. vii + 309 pp. ISSN-L 0271-4132. ISBN 1-4704-1079-6. ISBN-13 978-1-4704-1079-7. doi:10.1090/conm/615 {*124, 396*}

Stogryn, A. 1987, March. An analysis of the tensor dielectric constant of sea ice at microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing* **GE-25**(**2**):147–158. ISSN-L 0196-2892. doi:10.1109/TGRS.1987.289814 {*118, 396*}

Straley, J. P. 1981, November 14. Thermoelectric properties of inhomogeneous materials. *Journal of Physics D: Applied Physics* **14**(**11**):2101–2105. ISSN-L 0022-3727. URL http://iopscience.iop.org/0022-3727/14/11/017 {*16, 258, 397*}

Strang, G. 1986. *Introduction to Applied Mathematics*. Wellesley, Massachusetts: Wellesley–Cambridge Press. ix + 758 pp. ISBN 0-9614088-0-4. ISBN-13 978-0-9614088-0-0. {*vi, 2, 38, 73, 275, 397*}

Strang, G. 1988, June. A framework for equilibrium equations. *SIAM Review* **30**(**2**):283–297. ISSN-L 0036-1445. doi:10.1137/1030048 {*vi, 2, 397*}

Strang, G. 2007. *Computational Science and Engineering*. Wellesley, Massachusetts: Wellesley–Cambridge Press. xi + 716 pp. ISBN 0-9614088-1-2 (hardcover). ISBN-13 978-0-9614088-1-7 (hardcover). URL http://math.mit.edu/~gs/cse/ {*vi, 2, 397*}

Suryanarayana, P., V. Gavini, T. Blesgen, K. Bhattacharya, and M. Ortiz 2010, February. Non-periodic finite-element formulation of Kohn–Sham density functional theory. *Journal of the Mechanics and Physics of Solids* **58**(**2**):256–280. ISSN-L 0022-5096. doi:10.1016/j.jmps.2009.10.002 {*310, 397*}

Šverák, V. 1992. Rank-one convexity does not imply quasiconvexity. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **120**(**1–2**):185–189. ISSN-L 0308-2105. doi:10.1017/S0308210500015080 {*3, 397*}

Sylvester, J. 1993. Linearizations of anisotropic inverse problems. In L. Päivärinta and E. Somersalo (eds.), *Inverse Problems in Mathematical Physics: Proceedings of The Lapland Conference on Inverse Problems Held at Saariselkä, Finland, 14–20 June 1992*, Volume 422 of *Lecture Notes in Physics*, pp. 231–241 (of xviii + 256). Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. ISBN 3-540-57195-7. ISBN-13 978-3-540-57195-7. doi:10.1007/3-540-57195-7_27 book-doi:10.1007/3-540-57195-7 {*142, 397*}

Sylvester, J. and G. Uhlmann 1987. A global uniqueness theorem for an inverse boundary value problem. *Annals of Mathematics* **125**(**1**):153–169. ISSN-L 0003-486X. doi:10.2307/1971291 {*142, 397*}

Synge, J. L. 2012. *The Hypercircle in Mathematical Physics: a Method for the Approximate Solution of Boundary Value Problems* (Second ed.). Cambridge, UK: Cambridge University Press. 440 pp. First edition appeared in 1957. ISBN 1-107-66655-4 (paperback). ISBN-13 978-1-107-66655-9 (paperback). {*357, 358, 397*}

Sz.-Nagy, B. and C. Foiaş 1970. *Harmonic Analysis of Operators on Hilbert Space*. Amsterdam: North-Holland Publishing Co. xiii + 387 pp. ISBN 0-7204-2035-0, 0-444-10046-6. ISBN-13 978-0-7204-2035-7, 978-0-444-10046-7. {*356, 397*}

Tartar, L. 1975. Problèmes de contrôle des coefficients dans des équations aux dérivées partielles. (French) [Coefficient-control problems in partial differential equations]. In A. Bensoussan and J.-L. Lions (eds.), *Control theory, numerical methods, and computer systems modelling: international symposium, Rocquencourt, June 17–21, 1974*, Volume 107 of *Lecture Notes in Economics and Mathematical Systems*, pp. 420–426 (of viii + 757). Berlin / Heidelberg / London / etc.: Springer-Verlag. English translation in Topics in the Mathematical Modelling of Composite Materials, pp. 1–8, ed. by A. Cherkaev and R. Kohn. ISBN 0-8176-3662-5. ISBN 0-387-07020-6. ISBN-13 978-0-387-07020-9. {*2, 11, 397*}

Tartar, L. 1979a. Compensated compactness and applications to partial differential equations. In R. J. Knops (ed.), *Nonlinear Analysis and Mechanics, Heriot–Watt Symposium, Volume IV*, Volume 39 of *Research Notes in Mathematics*, pp. 136–212 (of 212). London: Pitman Publishing Ltd. ISBN 0-8224-8461-7, 0-273-08461-5. ISBN-13 978-0-8224-8461-5, 978-0-273-08461-7. {*11, 17, 253, 397*}

Tartar, L. 1979b. Estimation de coefficients homogénéisés. (French) [Estimation of homogenization coefficients]. In R. Glowinski and J.-L. Lions (eds.), *Computing Methods in Applied Sciences and Engineering: Third International Symposium, Versailles, France, December 5–9, 1977*, Volume 704 of *Lecture Notes in Mathematics*, pp. 364–373 (of vi + 391). Berlin / Heidelberg / London / etc.: Springer-Verlag. English translation in Topics in the Mathematical Modelling of Composite Materials, pp. 9–20, ed. by A. Cherkaev and R. Kohn. ISBN 0-8176-3662-5. ISBN 0-387-09123-8. ISBN-13 978-0-387-09123-5. {*11, 19, 130, 131, 284, 289, 339, 397*}

Tartar, L. 1985. Estimations fines des coefficients homogénéisés. (French) [Fine estimations of homogenized coefficients]. In P. Krée (ed.), *Ennio de Giorgi Colloquium: Papers Presented at a Colloquium Held at the H. Poincaré Institute in November 1983*, Volume 125 of *Pitman Research Notes in Mathematics*, pp. 168–187 (of 196). London: Pitman Publishing Ltd. ISSN-L 0743-0337. ISBN 0-273-08680-4. ISBN-13 978-0-273-08680-2. {*19, 130, 131, 253, 284, 289, 339, 398*}

Tartar, L. 1987. The appearance of oscillations in optimization problems. In R. J. Knops and A. A. Lacey (eds.), *Non-Classical Continuum Mechanics: Proceedings of the London Mathematical Society Symposium, Durham, July 1986*, Volume 122 of *London Mathematical Society Lecture Note Series*, pp. 129–150 (of xii + 335). Cambridge, UK: Cambridge University Press. ISBN 0-521-34935-4. ISBN-13 978-0-521-34935-2. {*2, 398*}

Tartar, L. 1989. H-measures and small amplitude homogenization. In R. V. Kohn and G. W. Milton (eds.), *Proceedings of the SIAM Workshop on Random Media and Composites, Leesburg, Virginia, December 7–10, 1988*, pp. 89–99 (of 214). Philadelphia: SIAM Press. ISBN 0-89871-246-7. ISBN-13 978-0-89871-246-9. {*257, 288, 398*}

Tartar, L. 1990. H-measures, a new approach for studying homogenization, oscillations and concentration effects in partial differential equations. *Proceedings of the Royal Society of Edinburgh. Section A, Mathematical and Physical Sciences* **115**(**3–4**):193–230. ISSN-L 0308-2105. doi:10.1017/S0308210500020606 {*257, 288, 398*}

Tartar, L. 2009. *The General Theory of Homogenization: a Personalized Introduction*, Volume 7 of *Lecture Notes of the Unione Matematica Italiana*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xxii + 470 pp. ISSN-L 1862-9113. ISBN 3-642-05194-4. ISBN-13 978-3-642-05194-4. doi:10.1007/978-3-642-05195-1 {*2, 11, 50, 131, 150, 257, 258, 289, 339, 398*}

Thaler, A. E. and G. W. Milton 2014, December. Exact determination of the volume of an inclusion in a body having constant shear modulus. *Inverse Problems* **30**(**12**):125008 (16 pp.). ISSN-L 0266-5611. doi:10.1088/0266-5611/30/12/125008 {*143, 144, 398*}

Thaler, A. E. and G. W. Milton 2015. Bounds on the volume of an inclusion in a body from a complex conductivity measurement. *Communications in Mathematical Sciences* **13**(**4**):863–892. ISSN-L 1539-6746. doi:10.4310/CMS.2015.v13.n4.a2 {*18, 78, 129, 131, 339, 398*}

Ting, T. C. T. and V.-G. Lee 1997. The three-dimensional elastostatic Green's function for general anisotropic linear elastic solids. *Quarterly Journal of Mechanics and Applied Mathematics* **50**(**3**):407–426. ISSN-L 0033-5614. doi:10.1093/qjmam/50.3.407 {*358, 398*}

Tokarzewski, S., J. J. Telega, and A. Galka 2001. Torsional rigidities of cancellous bone filled with marrow: the application of multipoint Padé approximants. *Engineering transactions = Rozprawy Inżynierskie* **49**(**2–3**):135–153. ISSN-L 0867-888X. URL https://www.infona.pl/resource/bwmeta1.element.baztech-article-BPB2-0004-0067? {*153, 398*}

Torquato, S. 1980. Microscopic approach to transport in two-phase random media. Ph.D. thesis, State University of New York, Stony Brook, Stony Brook, NY. 296 pp. Advisor(s) George Stell. URL http://search.proquest.com/pqdtglobal/docview/303096772/ {*55, 67, 398*}

Torquato, S. 1997. Effective stiffness tensor of composite media. I. Exact series expansions. *Journal of the Mechanics and Physics of Solids* **45**(**9**):1421–1448. ISSN-L 0022-5096. doi:10.1016/S0022-5096(97)00019-7 {*257, 288, 398*}

Torquato, S. 2002. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*, Volume 16 of *Interdisciplinary Applied Mathematics*. Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. xxi + 701 pp. ISBN 0-387-95167-9. ISBN-13 978-0-387-95167-6. doi:10.1007/978-1-4757-6355-3 {*2, 50, 55, 131, 150, 257, 258, 288, 289, 339, 398*}

Uhlmann, G. and T. Zhou 2015. Inverse electromagnetic problems. In B. Engquist (ed.), *Encyclopedia of Applied and Computational Mathematics*, pp. 716–725 (of xxxiii + 1564). Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. ISBN 3-540-70528-7 (hardcover). ISBN-13 978-3-540-70528-4 (hardcover). doi:10.1007/978-3-540-70529-1 {*92, 96, 112, 398*}

Vinogradov, V. and G. W. Milton 2005. The total creep of viscoelastic composites under hydrostatic or antiplane loading. *Journal of the Mechanics and Physics of Solids* **53**(**6**):1248–1279. ISSN-L 0022-5096. doi:10.1016/j.jmps.2005.01.005 {*153, 399*}

Vladimirov, V. S. 2002. *Methods of the Theory of Generalized Functions*, Volume 6 of *Analytic Methods and Special Functions*. Boca Raton, FL, USA: Taylor and Francis. xiv + 311 pp. ISBN 0-415-27356-0. ISBN-13 978-0-415-27356-5. {*97, 119, 120, 399*}

Voigt, W. 1889. Über die Beziehung zwischen den beiden Elastizitätskonstanten isotroper Körper. (German) [On the relation between the two elasticity constants of isotropic bodies]. *Wiedemanns Annalen der Physik und Chemie (Leipzig)* **274**(**12**):573–587. doi:10.1002/andp.18892741206 {*64, 399*}

Voigt, W. 1910. *Lehrbuch der Krystallphysik. (German) [Textbook of Crystal Physics]*, Volume XXXIV of *Sammlung von Lehrbuchern auf dem Gebiete der mathematischen Wissenschaften mit Einschlüss ihrer Anwendungen*. Stuttgart, Germany / Leipzig, Germany: Teubner. xxiv + 964 pp. {*64, 399*}

Walpole, L. J. 1966, May. On bounds for the overall elastic moduli of inhomogeneous systems — I. *Journal of the Mechanics and Physics of Solids* **14**(**3**):151–162. ISSN-L 0022-5096. doi:10.1016/0022-5096(66)90035-4 {*54, 399*}

Wang, C.-Y. and J. D. Achenbach 1993. A new method to obtain 3-D Green's functions for anisotropic solids. *Wave Motion* **18**(**3**):273–289. ISSN-L 0165-2125. doi:10.1016/0165-2125(93)90076-R {*358, 399*}

Wang, C.-Y. and J. D. Achenbach 1995. Three-dimensional time-harmonic elastodynamic Green's functions for anisotropic solids. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **449**(**1937**):441–458. ISSN-L 0080-4630. doi:10.1098/rspa.1995.0052 {*358, 399*}

Welters, A. T. 2011a. On explicit recursive formulas in the spectral perturbation analysis of a Jordan block. *SIAM Journal on Matrix Analysis and Applications* **32**(**1**):1–22. ISSN-L 0895-4798. doi:10.1137/090761215 {*96, 399*}

Welters, A. T. 2011b. On the mathematics of slow light. Ph.D. Thesis, University of California, Irvine, Irvine, CA, USA. x + 200 pp. Advisor(s) Aleksandr Figotin. URL http://search.proquest.com/pqdtglobal/docview/884623051/ {*102, 399*}

Welters, A. T., Y. Avniel, and S. G. Johnson 2014, August. Speed-of-light limitations in passive linear media. *Physical Review A (Atomic, Molecular, and Optical Physics)* **90**(**2**):023847. ISSN-L 1050-2947. doi:10.1103/PhysRevA.90.023847 {*96, 110, 116, 399*}

Wiener, O. 1912. Die Theorie des Mischkörpers für das Feld des stationären Strömung. Erste Abhandlung die Mittelswertsätze für Kraft, Polarisation und Energie. (German) [The theory of composites for the field of steady flow. First treatment of mean value estimates for force, polarization and energy]. *Abhandlungen der mathematisch-physischen Klasse der Königlich Sächisischen Gesellschaft der Wissenschaften* **32**:509–604. {*64, 257, 399*}

Wightman, A. S. 1979. Introduction: Convexity and the notion of equilibrium state in thermodynamics and statistical mechanics. In *Convexity in the Theory of Lattice Gases*, Princeton Series in Physics, pp. ix–lxxxv. Princeton, New Jersey: Princeton University Press. ISBN 0-691-08209-X, 0-691-08216-2 (paperback). ISBN-13 978-0-691-08209-7, 978-0-691-08216-5 (paperback). {*2, 399*}

Willemse, M. W. M. and W. J. Caspers 1979, August. Electrical conductivity of polycrystalline materials. *Journal of Mathematical Physics* **20**(**8**):1824–1831. ISSN-L 0022-2488. doi:10.1063/1.524284 {*118, 399*}

Willis, J. R. 1965, November. The elastic interaction energy of dislocation loops in anisotropic media. *Quarterly Journal of Mechanics and Applied Mathematics* **18**(**4**):419–433. ISSN-L 0033-5614. doi:10.1093/qjmam/18.4.419 {*358, 399*}

Willis, J. R. 1977, June. Bounds and self-consistent estimates for the overall properties of anisotropic composites. *Journal of the Mechanics and Physics of Solids* **25**(**3**):185–202. ISSN-L 0022-5096. doi:10.1016/0022-5096(77)90022-9 {*257, 399*}

Willis, J. R. 1980, December. A polarization approach to the scattering of elastic waves. I. Scattering by a single inclusion. *Journal of the Mechanics and Physics of Solids* **28**(**5–6**):287–305. ISSN-L 0022-5096. doi:10.1016/0022-5096(80)90021-6 {*358, 400*}

Willis, J. R. 1981a. Variational and related methods for the overall properties of composites. *Advances in Applied Mechanics* **21**:1–78. ISSN-L 0065-2156. ISBN 0-12-002021-1. ISBN-13 978-0-12-002021-8. doi:10.1016/S0065-2156(08)70330-2 {*27, 29, 150, 257, 258, 288, 358, 400*}

Willis, J. R. 1981b. Variational principles for dynamic problems for inhomogeneous elastic media. *Wave Motion* **3**(**1**):1–11. ISSN-L 0165-2125. doi:10.1016/0165-2125(81)90008-1 {*27, 29, 400*}

Willis, J. R. 1982. Elasticity theory of composites. In H. G. Hopkins and M. J. Sewell (eds.), *Mechanics of Solids*, pp. 653–686 (of xxiii + 693). New York: Pergamon. ISBN 0-08-025443-8. ISBN-13 978-0-08-025443-2. {*358, 400*}

Willis, J. R. 1985. The non-local influence of density variations in a composite. *International Journal of Solids and Structures* **21**(**7**):805–817. ISSN-L 0020-7683. doi:10.1016/0020-7683(85)90084-8 {*27, 400*}

Willis, J. R. 1997. Dynamics of composites. In P. M. Suquet (ed.), *Continuum Micromechanics*, Volume 377 of *CISM courses and lectures*, pp. 265–290 (of 347). Berlin, Germany / Heidelberg, Germany / London, UK / etc.: Springer-Verlag. ISSN-L 0254-1971. ISBN 3-211-82902-4. ISBN-13 978-3-211-82902-8. doi:10.1007/978-3-7091-2662-2 {*27, 29, 400*}

Willis, J. R. and J. R. Acton 1976, May. The overall elastic moduli of a dilute suspension of spheres. *Quarterly Journal of Mechanics and Applied Mathematics* **29**(**2**):163–177. ISSN-L 0033-5614. doi:10.1093/qjmam/29.2.163 {*358, 400*}

Willot, F. 2015, March. Fourier-based schemes for computing the mechanical response of composites with accurate local fields. *Comptes Rendus M'ecanique* **343**(**3**):232–245. ISSN-L 1631-0721. doi:10.1016/j.crme.2014.12.005 {*78, 236, 400*}

Willot, F., B. Abdallah, and Y.-P. Pellegrini 2014, May 18. Fourier-based schemes with modified Green operator for computing the electrical response of heterogeneous media with accurate local fields. *International Journal for Numerical Methods in Engineering* **98**(**7**):518–533. ISSN-L 0029-5981. doi:10.1002/nme.464 {*78, 236, 288, 400*}

Young, L. C. 2000. *Lecture on the Calculus of Variations and Optimal Control Theory*. Providence, RI, USA: American Mathematical Society. xiii + 337 pp. Reprint of second corrected edition (Chelsea, NY: 1980). First edition (1969). ISBN 0-8218-2690-5. ISBN-13 978-0-8218-2690-4. {*14, 400*}

Zangwill, A. 2014. The education of Walter Kohn and the creation of density functional theory. *Archive for History of Exact Sciences* **68**(**6**):775–848. ISSN-L 0003-9519. doi:10.1007/s00407-014-0140-x {*x, 309, 400*}

Zeller, R. and P. H. Dederichs 1973, February. Elastic constants of polycrystals. *Physica Status Solidi. B, Basic Research* **55**(**2**):831–842. ISSN-L 0370-1972. doi:10.1002/pssb.2220550241 {*257, 288, 400*}

Zettl, A. 2005. *Sturm–Liouville Theory*, Volume 121 of *Mathematical Surveys and Monographs*. Providence, RI, USA: American Mathematical Society. xi + 328 pp. ISBN 0-8218-3905-5, 0-8218-5267-1, 1-4704-1348-5 (e-book). ISBN-13 978-0-8218-5267-5 (softcover), 978-0-8218-3905-8, 978-1-4704-1348-4 (e-book). doi:10.1090/surv/121 {*103, 104, 400*}

Zhang, D. and E. Cherkaev 2009, August 20. Reconstruction of spectral function from effective permittivity of a composite material using rational function approximations. *Journal of Computational Physics* **228**(**15**):5390–5409. ISSN-L 0021-9991. doi:10.1016/j.jcp.2009.04.014 {*125, 258, 400*}

Zhang, D., M. P. Lamoureux, G. F. Margrave, and E. Cherkaev 2011, January. Rational approximation for estimation of quality Q factor and phase velocity in linear, viscoelastic, isotropic media. *Computational Geosciences* **15**(**1**):117–133. ISSN-L 1420-0597. doi:10.1007/s10596-010-9201-7 {*151, 400*}

Zhikov, V. V. 2000. On an extension and an application of the two-scale convergence method. *Matematicheskii Sbornik* **191**(**7**):31–72. English translation in Sbornik: Mathematics, **191**(**7**), 973–1014 (2000). ISSN-L 1064-5616. doi:10.1070/SM2000v191n07ABEH000491 {*21, 401*}

Zhikov, V. V., S. M. Kozlov, and O. A. Oleinik 1994. *Homogenization of Differential Operators and Integral Functionals*. Berlin / Heidelberg / London / etc.: Springer-Verlag. xi + 570 pp. Translated from the Russian by G. A. Yosifian [G. A. Iosif'yan]. ISBN 3-540-54809-2 (Berlin), 0-387-54809-2 (New York). ISBN-13 978-3-540-54809-6 (Berlin), 978-0-387-54809-8 (New York). doi:10.1007/978-3-642-84659-5 {*11, 48, 49, 270, 401*}

Author/editor index

Primary authors are shown in Small Caps, and secondary authors in roman. Page numbers of citations of primary authors are in **bold**.

C

D

E

F

G

H

I

J

K

N

O

P

Q

R

S

T

U

V

W

Y

Z

Index

C

D

E

F

M

N

O

P

Q

R

S

T

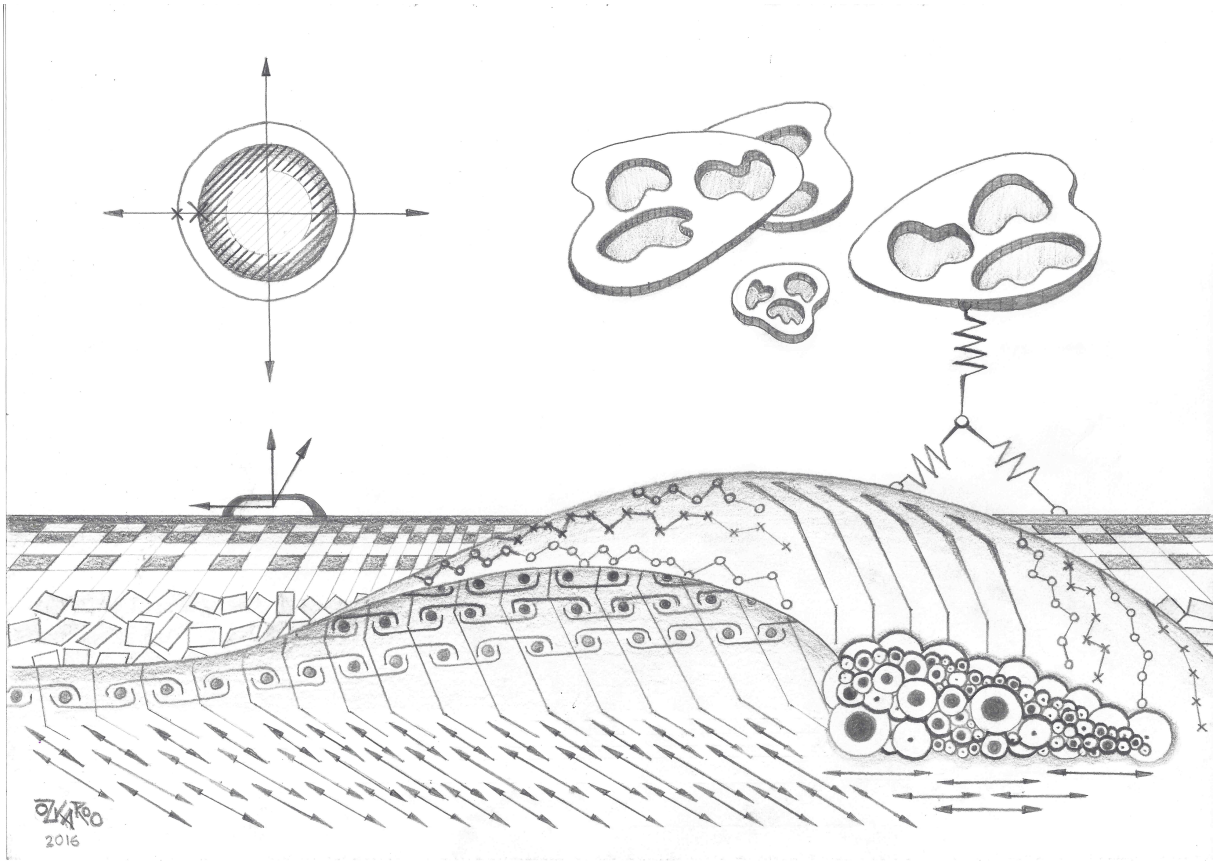
U

V

W

Y

Z



This book changes the landscape of may problems in science, ranging from reformulating Schrödinger's equation (of importance to Chemistry, Physics, and Materials Science) leading to new algorithms for solving multi-electron systems, casting in a new light inverse problems, where one seeks to determine what is inside a body from boundary measurements, that could lead to new methods of imaging, generalizing conservation laws to boundary field equalities and inequalities, deriving integral representation formulae for the response of bodies, bounding the transient response of composites, and even introducing a new type of function, called a superfunction, where the basic object is a subspace collection, plus many other groundbreaking ideas.

Graeme W. Milton is a distinguished professor of mathematics at the University of Utah, and one of the world's leading experts in the field of composites having written one of the definitive texts on the subject, titled "The Theory of Composites". He is a SIAM fellow, and held Sloan and Packard fellowships, was winner of the Society for Industrial and Applied Mathemaics Ralph E. Kleinman Prize " "for his many and deep contributions to the analysis of composite materials", winner of the Prager Medal of the Society for Engineering Science for "ground breaking mathematical analyses of heterogeneous media", and winner of the Rolf Landauer Medal of the International ETOPIM Association "for research excellence in the field of composite science". He has published more than 150 papers in journals spanning the scientific disciplines. He is best known for his work in the theory of composite and metamaterials, but has also made significant contributions to the field of inverse problems and helped pioneer the science of invisibility. The book is produced in collaboration with his colleagues Maxence Cassier, Ornella Mattei, Moti Milgrom, and Aaron Welters who coauthored four chapters, and with help from Hervé Moulinec and Pierre Suquet on testing the new Fast Fourier Transformation algorithm for composites.

The above illustration is by his brother Ian Milton as a representation of the world of composite materials.