A MULTIRESOLUTION STRATEGY FOR HOMOGENIZATION OF PARTIAL DIFFERENTIAL EQUATIONS

by

NICHOLAS ASHTON COULT

B. A., Carleton College, 1993

M. S., University of Colorado, 1995

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Applied Mathematics

1997

This thesis for the Doctor of Philosophy degree by

Nicholas Ashton Coult

has been approved for the

Department of

Applied Mathematics

by

____

Gregory Beylkin

____

Stephen McCormick

Date ______________________

Coult, Nicholas Ashton (Ph. D., Applied Mathematics)

A Multiresolution Strategy for Homogenization of Partial Differential Equations

Thesis directed by Professor Gregory Beylkin

Coefficients of PDE's are often changing across many spatial or temporal scales, whereas we might be interested in the behavior of the solution only on some relatively coarse scale. The multiresolution strategy for reduction and homogenization provides a method for finding an equation for the projection of the solution to a coarse scale. This equation explicitly incorporates the fine-scale behavior of the coefficients.

We present the multiresolution strategy for reduction and homogenization of differential equations, and apply it to linear wave equations in which the coefficients describe a layered medium (the problem reduces to a system of ordinary differential equations in this case) and to elliptic partial differential equations. For the layered-medium wave equations, we discuss and compare the multiresolution approach with classical techniques. For elliptic operators, it is known that the non-standard form has fast off-diagonal decay and the rate of decay is controlled by the number of vanishing moments of the wavelet basis. We prove that if an appropriate (e.g. high order) basis is used, the reduction procedure preserves the rate of decay over any finite number of scales and, therefore, results in sparse matrices for computational purposes. Furthermore, the reduction procedure approximately preserves small eigevenvalues of strictly elliptic operators. We also introduce a modified reduction procedure which preserves the small eigenvalues with greater accuracy than the usual reduction procedure and obtain estimates for the perturbation of those eigenvalues.

To Katie.

ACKNOWLEDGMENTS

I would like to give special thanks to my advisor, Gregory Beylkin, for offering me support and opportunities on all scales; Steve McCormick, for his many useful suggestions throughout my dissertation work; Lucas Monzon, for putting up with my frequent sanity checks and panicked proof-verification; David Trubatch, for his eager participation in spontaneous discussions on research topics; and the faculty, staff, and students of the Department of Applied Mathematics. Thanks also to the University of Colorado, for their partial financial support of my studies, and the National Science Foundation, which supported this work through a National Science Foundation Graduate Research Fellowship.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The problem of capturing the influence of fine and intermediate scales on a coarse scale is generally known as that of homogenization. For example, the Earth's crust contains complicated layers of rock ranging in size from sub-millimeters to meters or more in thickness, and long-wavelength (relative to the size of the smallest layers) waves exhibit behavior on the macroscale which is influenced by the microscale structure of the layers.

The motivation for studying problems of this kind is that, in the physical world, coarse scales may be easier or less costly to influence and observe than fine and intermediate scales and, in the mathematical models, solutions on coarse scales require significantly fewer operations to compute. Additionally, the parameters of interest may only be observable on the coarse scale, but the interactions which determine the values of these parameters may occur on many scales.

The mathematical difficulty of the homogenization problem is rooted in the fact that the coefficients on fine and intermediate scales affect the coarse-scale behavior of the solution in complicated, non-linear ways. Because of this, simple projection of the coefficients to the coarse scale of interest is usually not sufficient to incorporate the fine-scale behavior. Our goal is to write effective equations on the coarse scale which account for the fine and intermediate scales' influence on the coarse scale solution.

Typically, such problems were addressed by using asymptotic methods or weak limits, see for example, [8], [24], [36], [27], [13] and references therein. The basic limitation of these methods is that they require the fine scale behavior to be fairly well separated from the behavior on the coarser scales, so that small parameters may be found in the problem.

Recently, a multiresolution strategy for homogenization has been proposed in [10]. Using the notion of Multiresolution Analysis (MRA) (see Appendix A), the transition between two adjacent scales is considered explicitly. Namely, one obtains an equation for the projection of the solution on the coarser scale. This procedure (the so-called reduction) may then be repeated over many scales and, thus, does not require the small parameter assumptions typical for asymptotic methods.

The basic step of the reduction involves computing a Schur complement. (The use of the Schur complement in multilevel methods is not new and plays a role in multigrid and domain decomposition methods (see e.g. [18], [14], [33]). Steinberg and McCoy in [35] use the Schur complement for multiresolution effective medium computations. Additionally, Knapek in [26] has used the Schur complement for a multigrid-based homogenization technique. Two problems have to be addressed in order for the multiresolution strategy for homogenization to be a practical method. First, the transition between the two scales has to be computationally efficient. However, simply truncating the matrices, as suggested in some of the references mentioned above, is not satisfactory since there is no control of the quality of the approximation. Second, the form of equations must be preserved so that one can use the reduction step in a recursive manner. By the "form of the equations" we understand either algebraic form or some alternative algebraic structure. The only requirement is that it may be used recursively. The meaning of this remark will become clear further in this Thesis.

In [10] the multiresolution strategy for reduction and homogenization has been applied to a system of linear ordinary differential equations. It is observed in [10] that the transition between two consecutive scales may be achieved by eliminating variables locally and that a certain algebraic form of the equations is preserved, thus permitting a multiscale reduction.

We briefly outline the structure of this Thesis.

In Chapter 1, we introduce the multiresolution approach and compare it to existing approaches for homogenization. The results in Section 1.4 concerning homogenization of the acoustic equations are new.

Chapter 2 describes the multiresolution approach in the context of elliptic PDE's. Numerical as well as analytical results are presented. The results of this Chapter are new.

Chapter 3 deals with certain classes of matrices. The results of this chapter are provide a foundation for the results in Chapter 2. The theorems in this chapter are new in that they constitute an extension of the results obtained by P. Tchamitchian [37] and S. Jaffard [21].

Finally, Chapter 4 provides a summary of the results of this Thesis and describe directions for future work.

CHAPTER 1

MULTIRESOLUTION REDUCTION AND HOMOGENIZATION

In this chapter, we start by introducing the homogenization problem in the classical context and presenting some existing approaches to the problem. The classical approaches we consider are limited in their application to PDE's in that they require separation of fine and coarse scales and do not permit intermediate scales in the problem. We then introduce the multiresolution homogenization method, which allows for coefficients which vary on intermediate scales, and compare it to the existing classical approaches on a one-dimensional example. Since the multiresolution homogenization scheme allows one to incorporate intermediate scales and in this sense is more general than the classical approaches, the purpose of this comparison is to connect the multiresolution and classical approachs in parameter regimes where both are valid and show that classical results may be achieved using the multiresolution approach.

We demonstrate the classical techniques on a simple one-dimensional problem. In the parameter ranges where the classical methods are valid, the multiresolution approach gives the same results for the simple one-dimensional problem, but does not have the restrictions on separation of scales that the classical methods have. Thus, the result of the multiresolution method is more general in that we may apply the multiresolution technique to problems with a wider class of coefficients. We emphasize that the one-dimensional example is meant to provide a demonstration of the techniques we consider.

We also demonstrate this connection between the classical techniques and the multiresolution method for the layered-medium acoustic wave equation as well, a problem which has been studied in great detail for many years. We find that the multiresolution approach agrees with the classical approaches in parameter regimes where those approaches are valid.

1.1 Classical Homogenization

Homogenization in the context of differential equations refers to the problem posed by the presence of disparate scales in coefficients of the equation and its solutions. Computationally, this presents a difficulty because the microstructure of the coefficients may combine in a non-linear fashion to produce macroscale effects on the solution.

We use the term classical homogenization to refer to a limit procedure or asymptotic analysis of a differential equation in which the small parameter describes the size of the features in the coefficients relative to the scale of the solution. One then considers a limit in this small parameter, and determines the effective coefficients.

The classical techniques are in no way limited to one-dimensional problems, or differential equations (see e.g. [24], [8]). Also, periodicity of the coefficients is not strictly required for the classical techniques in one dimension. However, in multiple dimensions, the classical techniques we describe become difficult or impossible to apply if periodicity of the coefficients is not required. Therefore, we study a one-dimensional example with periodic coefficients in this section in order to illustrate the main idea.

We study the model problem

$$-\frac{d}{dx}(\tilde{a}(x)\frac{d}{dx}u(x)) = f(x), \quad \tilde{a}(x) > 0, \tag{1.1.1}$$

on the interval $[0, 1]$ with Dirichlet boundary conditions. If the function $\tilde{a}(x)$ is non-constant and periodic with period $\epsilon << 1$, then the fine-scale feature-size of the coefficients $\tilde{a}(x)$ may be thought of as ϵ. We can capture this property by considering a function $a(x)$ to be periodic on $[0, 1]$ and studying

$$-\frac{d}{dx}(a(\frac{x}{\epsilon})\frac{d}{dx}u^{\epsilon}(x)) = f(x), \tag{1.1.2}$$

where we have introduced $\frac{1}{\epsilon}$ into the coefficient a. We are essentially comparing two scales - the ϵ-sized fine scale of the coefficients and the fixed scale of the interval $[0, 1]$, representing the coarse scale of interest. For small ϵ, the coefficients $a(\frac{x}{\epsilon})$ are highly oscillatory (the frequency is given by $\frac{1}{\epsilon}$). These oscillations may introduce oscillatory components into all modes of the operator $-(\frac{d}{dx}a(\frac{x}{\epsilon})\frac{d}{dx})$, which means that even a right-hand side $f(x)$ in the span of only the lower Fourier modes will produce a solution with an oscillatory component. However, in the homogenization problem the oscillatory component of the solution is not of interest; it is only the gross or coarse-scale features of the solution which are. The goal is to find an effective coefficient a_0 so that, as $\epsilon \to 0$, the sequence u^ϵ will have a limit (in some sense) given by u^0 which solves the equation

$$-\frac{d}{dx}(a_0 \frac{d}{dx} u^0) = f(x). \tag{1.1.3}$$

In the following subsections, we present two typical methods for accomplishing this goal.

1.1.1 Weak Limit Method

One of the typical approaches (found in e.g. [8], [24]) to homogenization of (1.1.2) is to consider the weak limit of the family u^ϵ. Weak limits may be thought of as "canceling" oscillatory components. Thus, the weak limit u^0 of u^ϵ gives the essential non-oscillatory part of the family u^ϵ.

The goal is to find an equation of the same form as in (1.1.2) such that the solution of this equation is the weak limit in $H_0^1[0,1]$ of the sequence u^ϵ as $\epsilon \to 0$. In this analysis, the flow p^ϵ, given by

$$p^\epsilon(x) = -a(\frac{x}{\epsilon})\frac{du^\epsilon}{dx} \tag{1.1.4}$$

plays a role, and we consider its weak limit in $L^2[0,1]$.

This approach may be used to find the effective coefficients in (1.1.2) in the limit $\epsilon \to 0$ as the coefficients $a(x/\epsilon)$ become more oscillatory. In computing the various limits which are required in this analysis, the following Lemma from [24] is useful. For the sake of completeness we include its proof.

Lemma 1.1.1 *Let $g : \mathbf{R}^n \to \mathbf{C}$ be a periodic function on $\mathbf{R}^n$ whose period cell is a box B with lengths on each coordinate axis given by $l_1, \ldots, l_n$. If $g \in L^p(B)$, then $g(x/\epsilon)$ converges weakly to $\langle g \rangle$ in $L^p(\Omega)$, where Ω is an arbitrary bounded domain in $\mathbf{R}^n$ and*

$$\langle g \rangle = \frac{1}{|B|}\int_B g(x)dx. \tag{1.1.5}$$

Proof: We may restrict the proof to the case when Ω is a dilation of the unit cube C with ratio $s \geq 1$. Note that if $f \in L^p(C)$ and $\epsilon \leq 1$, then

$$\int_{\Omega} |f(\frac{x}{\epsilon})|^p dx = \epsilon^n \int_{s\epsilon^{-1}C} |f(x)|^p dx \leq \epsilon^n([s\epsilon^{-1}]+1)^n \langle |f|^p \rangle \leq c_0(\Omega)\langle |f|^p \rangle \tag{1.1.6}$$

where $[s\epsilon^{-1}]$ denotes the largest integer not larger than $s\epsilon^{-1}$. Given δ, choose a trigonometric polynomial w such that $\langle w \rangle = \langle g \rangle$ and $\langle |g-w|^p \rangle \leq \delta$. Then for $\epsilon \leq 1$, we see that

$$\int_{\Omega} |g(\frac{x}{\epsilon}) - w(\frac{x}{\epsilon})|^p dx \leq c_0(\Omega)\delta. \tag{1.1.7}$$

Of course, for the trigonometric polynomial w, Lemma 1.1.1 is true by the Riemann-Lebesgue Lemma. The estimate (1.1.7) shows that we may then extend the result to the function g since g is arbitrarily well-approximated by a trigonometric polynomial.

From (1.1.2), we extract the relation

$$-\int_0^1 \frac{du^\epsilon}{dx} dx = \int_0^1 a(\frac{x}{\epsilon})^{-1}(F(x) - c_\epsilon)dx = 0 \tag{1.1.8}$$

where $F(x) = \int_0^x f(x)dx$ and c_ϵ is constant in x. By Lemma 1.1.1, we determine that $\lim_{\epsilon \to 0} c_\epsilon = \int_0^1 F(x)dx$. Noting that

$$p^\epsilon(x) = -a(\frac{x}{\epsilon})\frac{du^\epsilon}{dx} = F(x) - c_\epsilon \tag{1.1.9}$$

and

$$\frac{du^\epsilon}{dx} = -a(\frac{x}{\epsilon})^{-1}(F(x) - c_\epsilon), \tag{1.1.10}$$

we compute the weak limits of p^ϵ and $\frac{du^\epsilon}{dx}$ in $L^2[0,1]$ as

$$p^0 = F(x) - \int_0^1 F(y)dy \tag{1.1.11}$$

and

$$\frac{du^0}{dx} = -\langle a^{-1} \rangle (F(x) - \int_0^1 F(y)dy). \tag{1.1.12}$$

The weak limit of u^ϵ is given by $u^0(x) = \int_0^x \frac{du^0}{dy}dy$. Also, by (1.1.11), (1.1.12), we see that

$$p^0 = -\langle a^{-1} \rangle^{-1} \frac{du^0}{dx}, \quad \frac{dp^0}{dx} = f \tag{1.1.13}$$

and, therefore, u^0 solves the equation

$$-\frac{d}{dx}(\langle a^{-1}\rangle^{-1}\frac{d}{dx}u^0) = f(x). \tag{1.1.14}$$

The quantity $\langle a^{-1}\rangle^{-1}$ is known as the harmonic mean (see e.g. [3]) of the function a. This quantity shows up quite often in everyday calculations. For example, if a particle is traveling in the positive direction on the real line with a velocity which depends on its position only, i.e. $v = v(x) > 0$, then its total travel time is given by $\langle v^{-1}\rangle = \int_0^1 (v(x))^{-1}dx$, and we compute its average velocity over this interval as $\langle v^{-1}\rangle^{-1}$, the spatial harmonic mean of its velocity.

To summarize, we use a weak limit analyis to obtain the homogenized equation (1.1.14) as an equation on the coarse scale. This equation is a constant-coefficient equation and its coefficients are found by computing the harmonic mean of the function $a(x)$.

1.1.2 Asymptotic Method of Multiple Scales

The method of multiple scales (see e.g. [7] for details) presents an alternative approach to the homogenization problem, and for (1.1.2) produces the same result as the weak limit method of the previous subsection.

The first step of the multiple-scales method is to identity the variables associated with different powers of the small parameter ϵ and treat them as independent variables. There are two scales present in the problem (1.1.2), identified with x and $\frac{x}{\epsilon}$. We consider the variable x as is and define a new variable $y = \frac{x}{\epsilon}$ which represents the fine-scale variable. We look for solutions of (1.1.2) of the form

$$u^\epsilon(x,y) = u^0(x) + \epsilon u^1(x,y). \tag{1.1.15}$$

We see that $\frac{d}{dx} = \frac{\partial}{\partial x} + \frac{1}{\epsilon}\frac{\partial}{\partial y}$. Thus,

$$\begin{aligned}
\frac{d}{dx}(a(y)\frac{d}{dx}u^\epsilon(x)) &= \epsilon^{-1}\left(\frac{\partial}{\partial y}\left(a(y)(\frac{\partial}{\partial x}u^0)\right) + \frac{\partial}{\partial y}\left(a(y)(\frac{\partial}{\partial y}u^1)\right)\right) \\
&+ a(y)(\frac{\partial}{\partial x})^2u^0 + \frac{\partial}{\partial x}\left(a(y)(\frac{\partial}{\partial y}u^1)\right) + \frac{\partial}{\partial y}\left(a(y)(\frac{\partial}{\partial x}u^1 1)\right) \\
&+ \epsilon(a(y)(\frac{\partial}{\partial x})^2u^1) \qquad (1.1.16) \\
&= f(x). \qquad (1.1.17)
\end{aligned}$$

We now collect terms in powers of ϵ. The first term on the right-hand side of (1.1.16) must be zero since f has no ϵ^{-1} term. This yields

$$\frac{\partial}{\partial y}\left(a(y)(\frac{\partial}{\partial y}u^1)\right) = -\left(\frac{d}{dy}a(y)\right)\left(\frac{d}{dx}u^0(x)\right). \tag{1.1.18}$$

We may consider this as an equation for u^1 with periodic boundary conditions and with the right-hand side having x as a parameter. Solutions of (1.1.18) satisfy the equation

$$u^1(x,y) = N(y)\frac{d}{dx}u^0, \tag{1.1.19}$$

where N is a solution of

$$\frac{d}{dy}(a(y)\frac{d}{dy}N) = -\frac{d}{dy}a(y). \tag{1.1.20}$$

We write (1.1.20) as

$$\frac{d}{dy}(a(y)(1+\frac{d}{dy}N)) = 0. \tag{1.1.21}$$

We also see that

$$u^\epsilon = u^0(x) + \epsilon N(y)\frac{d}{dx}u^0. \tag{1.1.22}$$

Thus, the $\mathcal{O}(1)$ term in (1.1.16) can be written as

$$\begin{gathered} a(y)(\frac{\partial}{\partial x})^2u^0 + \frac{\partial}{\partial x}\left(a(y)(\frac{\partial}{\partial y}u^1)\right) + \frac{\partial}{\partial y}\left(a(y)(\frac{\partial}{\partial x}u_1)\right) = \\ \frac{d^2u^0}{dx^2}\left(a(y) + \frac{d}{dy}(a(y)N(y)) + a(y)\frac{d}{dy}N(y)\right). \end{gathered} \tag{1.1.23}$$

To determine the homogenized operator, we average the right-hand side of (1.1.23) over y to obtain:

$$a_0\frac{d^2u^0}{dx^2} = \langle a(y) + a(y)\frac{dN(y)}{dy}\rangle\frac{d^2u^0}{dx^2}. \tag{1.1.24}$$

The solution of (1.1.20) is given by

$$N(y) = \langle a^{-1}\rangle^{-1}\int_0^y a(z)^{-1}dz, \tag{1.1.25}$$

so we see that

$$a_0 = \langle a(y) - a(y) + \langle a^{-1}\rangle^{-1}\rangle = \langle a^{-1}\rangle^{-1}. \tag{1.1.26}$$

This is of course the same as the result of subsection 1.1.1.

1.2 Multiresolution Reduction

In contrast to the classical approach to the homogenization problem, the multiresolution approach uses the algebraic transformation between scales provided by the multiresolution analysis to solve for the fine-scale behavior and explicitly eliminate it from the equation. This approach has the advantage that the coefficients may vary on arbitrarily many scales. The chain of subspaces

$$\ldots \subset \mathbf{V}_2 \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots \tag{1.2.1}$$

defines the hierarchy of scales that the multiresolution scheme uses. This chain of subspaces is defined in such a way that the space $\mathbf{V}_j$ is "finer" than the space $\mathbf{V}_{j+1}$, in the sense that (1) all of $\mathbf{V}_{j+1}$ is contained in $\mathbf{V}_j$, and (2) the component of $\mathbf{V}_j$ which is not in $\mathbf{V}_{j+1}$ consists of functions which resolve features on a scale finer than any function in $\mathbf{V}_{j+1}$ may resolve. The difference between succesive spaces in this chain is captured by the so-called wavelet space $\mathbf{W}_{j+1}$, defined to be the orthogonal complement of $\mathbf{V}_{j+1}$ in $\mathbf{V}_j$. An orthogonal basis for the wavelet space $\mathbf{W}_{j+1}$ is constructed which has vanishing moments, i.e. the basis elements are L^2-orthogonal to low-degree polynomials (see Appendix A for details). The existence of orthogonal wavelet bases with vanishing moments distiguishes the multiresolution approach from typical multi-scale discretizations provided by finite-element or hierarchical bases (see [5] for a description). If we are considering a multiresolution analysis defined on a bounded domain, then the hierarchy of scales defined by (1.2.1) has a coarsest scale (which we may call $\mathbf{V}_0$), and we write instead

$$\mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots \tag{1.2.2}$$

For more details, see Appendix A.

The multiresolution strategy for the reduction and homogenization of linear problems has been proposed in [10]. Let us briefly review here the reduction procedure (in its general form). Consider a bounded linear operator $\mathbf{S}_j : \mathbf{V}_j \to \mathbf{V}_j$. Since $\mathbf{V}_j$ is spanned by translations of the function $\phi(2^j x - k)$, we know that the operator $\mathbf{S}_j$ may be written as a matrix. If the multiresolution analysis is defined on a bounded domain, then this matrix is finite; otherwise it is an infinite matrix which we consider as an operator on l^2. Let us consider the equation

$$\mathbf{S}_j x = f. \tag{1.2.3}$$

The decomposition $\mathbf{V}_j = \mathbf{V}_{j+1} \oplus \mathbf{W}_{j+1}$ allows us to split the operator $\mathbf{S}_j$ into four pieces (recall that $\mathbf{W}_{j+1}$ is called the wavelet space and is the "detail" or fine-scale component of $\mathbf{V}_j$) and write

$$\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{C}_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} \begin{pmatrix} d_x \\ s_x \end{pmatrix} = \begin{pmatrix} d_f \\ s_f \end{pmatrix}, \tag{1.2.4}$$

where we have

$$\mathbf{A}_{\mathbf{S}_j} \quad : \mathbf{W}_{j+1} \to \quad \mathbf{W}_{j+1} \tag{1.2.5}$$

$$\mathbf{B}_{\mathbf{S}_j} \quad : \mathbf{V}_{j+1} \to \quad \mathbf{W}_{j+1} \tag{1.2.6}$$

$$\mathbf{C}_{\mathbf{S}_j} \quad : \mathbf{W}_{j+1} \to \quad \mathbf{V}_{j+1} \tag{1.2.7}$$

$$\mathbf{T}_{\mathbf{S}_j} \quad : \mathbf{V}_{j+1} \to \quad \mathbf{V}_{j+1}, \tag{1.2.8}$$

and $d_x, d_f \in \mathbf{W}_{j+1}$, $s_x, s_f \in \mathbf{V}_{j+1}$ are the L^2-orthogonal projections of x and f onto the $\mathbf{W}_{j+1}$ and $\mathbf{V}_{j+1}$ spaces. The projection s_x is thus the coarse-scale component of the solution x, and d_x is its fine-scale component.

Formally eliminating d_x from (1.2.4) by substituting $d_x = \mathbf{A}_{\mathbf{S}_j}^{-1}(d_f - \mathbf{B}_{\mathbf{S}_j} s_x)$ yields

$$(\mathbf{T}_{\mathbf{S}_j} - \mathbf{C}_{\mathbf{S}_j}\mathbf{A}_{\mathbf{S}_j}^{-1}\mathbf{B}_{\mathbf{S}_j})s_x = s_f - \mathbf{C}_{\mathbf{S}_j}\mathbf{A}_{\mathbf{S}_j}^{-1}d_f. \tag{1.2.9}$$

We call (1.2.9) the *reduced equation*, and the operator

$$\mathbf{R}_{\mathbf{S}_j} = \mathbf{T}_{\mathbf{S}_j} - \mathbf{C}_{\mathbf{S}_j}\mathbf{A}_{\mathbf{S}_j}^{-1}\mathbf{B}_{\mathbf{S}_j} \tag{1.2.10}$$

the *one-step reduction* of the operator $\mathbf{S}_j$, also known as the Schur complement (see [31]) of the block-matrix $\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{C}_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix}$.

Note that the solution s_x of the reduced equation is exactly $\mathbf{P}_{j+1}x$, where $\mathbf{P}_{j+1}$ is the projection onto $\mathbf{V}_{j+1}$ and x is the solution of (1.2.3)). Note that the reduced equation is not the same as the averaged equation, which is given by

$$\mathbf{T}_{\mathbf{S}_j}\tilde{s}_x = s_f. \tag{1.2.11}$$

Once we have obtained the reduced equation, it may formally be reduced again to produce an equation on $\mathbf{V}_{j+2}$, and the solution of this equation is just the $\mathbf{V}_{j+2}$-component of the solution of (1.2.3). Likewise, we may reduce these equations recursively n times (assuming that if the multiresolution analysis is on a bounded domain, then $j + n \leq 0$) to produce an equation on $\mathbf{V}_{j+n}$, the solution of which is the projection of the solution of (1.2.3) on $\mathbf{V}_{j+n}$.

We note that in the finite-dimensional case, if we are considering a multiresolution analysis defined on a domain in $\mathbf{R}$, the reduced equation (1.2.9) has half as many unknowns as the original equation (1.2.3). If the domain is in $\mathbf{R}^2$, then the reduced equation has one-fourth as many unknowns as the original equation. Reduction, therefore, preserves the coarse-scale behavior of solutions while reducing the number of unknowns.

1.3 Multiresolution Homogenization of Linear ODE's

In order to iterate the reduction step over many scales, we need to preserve the form of the equation as a way of deriving a recurrence relation. In (1.2.3) and (1.2.9), both $\mathbf{S}_j$ and $\mathbf{R}_{\mathbf{S}_j}$ are matrices, and thus the procedure may be repeated. However, just identifying the matrix structure is usually not sufficient. In particular, even though the typical matrix $\mathbf{A}$ for ODE's and PDE's is sparse, the $\mathbf{A}_{\mathbf{S}_j}^{-1}$ term may become dense, changing the equation from a local one to a global one. It is important to know under what, if any, circumstances the local nature of the differential operator may be (approximately) preserved. Furthermore, if the equation is of the form of

$$-\nabla \cdot (a(\mathbf{x}) \nabla u(\mathbf{x})) = f(x) \tag{1.3.1}$$

or some other variable-coefficient differential equation, we may wish to know if the reduction procedure preserves this form, so that we may find effective coefficients of the equation on the coarse scale. This process is the basic goal of homogenization techniques, and it extracts information from the reduced equation based on the form of the original equation. Thus, within the multiresolution approach, reduction and homogenization are closely related but have different goals: homogenization attempts to find effective equations and their coefficients on the coarse scale, whereas reduction merely finds a coarse-scale version of a given system of equations.

In this Section we describe the MRA homogenization procedure of [10] as applied to linear ODE's, and give an example of the the procedure applied to the one-dimensional version of (1.3.1).

1.3.1 The Homogenization Procedure

The MRA homogenization procedure is applied to systems of ODE's which may be written in the form

$$\mathbf{B}x + q + \lambda = \mathbf{K}(\mathbf{A}x + p). \tag{1.3.2}$$

In particular, we consider equations of the form

$$(\mathbf{I} + B(t))x(t) + q(t) + \lambda = \int_0^t (A(s)x(s) + p(s))\mathrm{d}s, \;\; t \in (0,1) \tag{1.3.3}$$

on $\mathbf{L}^2(0,1)$, where $B(t)$ and $A(t)$ are $n \times n$ matrix-valued functions, $p(t)$ and $q(t)$ are vector-valued forcing terms, and $x(t)$ is the solution vector. As a differential equation this is written as

$$\frac{d}{dt}\left((\mathbf{I} + B(t))x(t) + q(t)\right) = A(t)x(t) + p(t) \tag{1.3.4}$$

with the initial conditions $(\mathbf{I} + B(0))x(0) = -q(0) - \lambda$.

On $\mathbf{V}_j$, $j < 0$, the projection of equation (1.3.3) is written as

$$\mathbf{B}_j x_j + q_j + \lambda = \mathbf{K}_j(\mathbf{A}_j x_j + p_j), \tag{1.3.5}$$

or

$$\mathbf{S}_j x_j = f_j \tag{1.3.6}$$

where

$$\mathbf{S}_j = \mathbf{B}_j - \mathbf{K}_j \mathbf{A}_j, \quad f_j = \mathbf{K}_j p_j - q_j - \lambda, \quad x_j = \mathbf{P}_j x_j. \tag{1.3.7}$$

After one step of reduction, our goal is to have an equation on $\mathbf{V}_{j+1}$ of the form

$$\mathbf{B}_{j+1}^{(j)} x_{j+1}^{(j)} + q_{j+1}^{(j)} + \lambda = \mathbf{K}_{j+1}(\mathbf{A}_{j+1}^{(j)} x_{j+1}^{(j)} + p_{j+1}^{(j)}), \tag{1.3.8}$$

where $x_{j+1}^{(j)} = \mathbf{P}_{j+1}x$. We use the notation $\mathbf{B}_l^{(j)}$ to indicate that the equation is first projected to scale $\mathbf{V}_j$, and then the reduction procedure is applied $l-j$ times to obtain an equation on $\mathbf{V}_l$. In (1.3.8), this notation therefore indicates that the equation in (1.3.8) was obtained by reducing an equation of the same form on $\mathbf{V}_j$ one time to produce an equation on the coarser scale $\mathbf{V}_{j+1}$.

This allows one to establish a recurrence relation for $k = j, j+1, \ldots, 0$ between the operators and forcing terms $\mathbf{B}_k^{(j)}$, $\mathbf{A}_k^{(j)}$, $p_k^{(j)}$, and $q_k^{(j)}$ on $\mathbf{V}_k$ and the operators and forcing terms $\mathbf{B}_{k+1}^{(j)}$, $\mathbf{A}_{k+1}^{(j)}$, $p_{k+1}^{(j)}$, and $q_{k+1}^{(j)}$ on $\mathbf{V}_{k+1}$. It turns out that this task of finding the recurrence relations is simplified significantly if one uses a multiresolution analysis whose basis functions have non-overlapping support. We use the Haar basis, but a multiwavelet basis may be used if higher order elements are needed (see [2]).

In the Haar basis, the operators $\mathbf{B}_j$, $\mathbf{A}_j$, and $\mathbf{K}_j$ derived from equations of the form of (1.3.3) have a simple form. Each of these is an $(N_j n) \times (N_j n)$ matrix, where $N_j = 2^j$ is the number of unknowns on the scale $\mathbf{V}_j$, and n is the number of equations in the original system (1.3.2). Furthermore, $\mathbf{B}_j$ and $\mathbf{A}_j$ are both block-diagonal matrices. The diagonal blocks of $\mathbf{B}_j$ and $\mathbf{A}_j$ are $n \times n$ matrices. There are therefore N_j diagonal blocks, each of which is an $n \times n$ matrix. For $\mathbf{A}_j$ and $\mathbf{B}_j$, we denote their i-th diagonal blocks by $(B_j)_i$ and $(A_j)_i$. The reader should not confuse the indices j and i with the indices for the entries of these matrices. The matrices are given by the Haar coefficients of the $n \times n$ matrix-valued functions $B(x)$ and $A(x)$ on the scale $\mathbf{V}_j$. We write

$$\mathbf{B}_j = \mathrm{diag}\{\mathbf{I} + (B_j)_i\}_{i=0}^{i=2^j-1}, \tag{1.3.9}$$

$$\mathbf{A}_j = \mathrm{diag}\{(A_j)_i\}_{i=0}^{i=2^j-1}, \tag{1.3.10}$$

and

$$\mathbf{K}_j = \delta_j \begin{pmatrix} \frac{1}{2}\mathbf{I} & 0 & \cdots & 0 & 0 \\ \mathbf{I} & \frac{1}{2}\mathbf{I} & 0 & \cdots & 0 \\ \mathbf{I} & \mathbf{I} & \frac{1}{2}\mathbf{I} & \cdots & \\ \vdots & \vdots & & \ddots & \\ \mathbf{I} & \cdots & & \mathbf{I} & \frac{1}{2}\mathbf{I} \end{pmatrix} \tag{1.3.11}$$

where $\delta_j = 2^{-j}$, $\mathbf{I}$ is the $n \times n$ identity matrix, and $(B_j)_i$ and $(A_j)_i$ are the i-th Haar coefficients on scale V_j of the $n \times n$ matrix-valued functions $B(x)$ and $A(x)$.

For equation (1.3.3), the recursion relations are given by

$$(A^{(j)}_{k+1})_i = (S_A)_i - (D_A)_i F^{-1}((D_B)_i + \frac{\delta_k}{2}(S_A)_i) \quad (1.3.12)$$

$$(B^{(j)}_{k+1})_i = (S_B)_i - \frac{\delta_k}{2}(D_A)_i - ((D_B)_i - \frac{\delta_k}{2}(S_A)_i)F^{-1}((D_B)_i + \frac{\delta_k}{2}(S_A)_i) \quad (1.3.13)$$

$$(p^{(j)}_{k+1})_i = (S_p)_i - \frac{\delta_k}{2}(D_A)_i F^{-1}((D_q)_i + (S_p)_i) \quad (1.3.14)$$

$$(q^{(j)}_{k+1})_i = (S_q)_i - \frac{\delta_k}{2}(D_p)_i - \frac{\delta_k}{2}((D_B)_i - (S_A)_i)F^{-1}((D_q)_i + (S_p)_i), \quad (1.3.15)$$

where

$$(S_A)_i = \frac{1}{2}\left((A^{(j)}_k)_{2i} + (A^{(j)}_k)_{2i+1}\right) \quad (1.3.16)$$

$$(D_A)_i = \frac{1}{2}\left((A^{(j)}_k)_{2i} - (A^{(j)}_k)_{2i+1}\right) \quad (1.3.17)$$

$$(S_B)_i = \frac{1}{2}\left((B^{(j)}_k)_{2i} + (B^{(j)}_k)_{2i+1}\right) \quad (1.3.18)$$

$$(D_B)_i = \frac{1}{2}\left((B^{(j)}_k)_{2i} - (B^{(j)}_k)_{2i+1}\right) \quad (1.3.19)$$

$$(S_p)_i = \frac{1}{\sqrt{2}}\left((p^{(j)}_k)_{2i} + (p^{(j)}_k)_{2i+1}\right) \quad (1.3.20)$$

$$(D_p)_i = \frac{1}{\sqrt{2}}\left((p^{(j)}_k)_{2i} - (p^{(j)}_k)_{2i+1}\right) \quad (1.3.21)$$

$$(S_q)_i = \frac{1}{\sqrt{2}}\left((q^{(j)}_k)_{2i} + (q^{(j)}_k)_{2i+1}\right) \quad (1.3.22)$$

$$(D_q)_i = \frac{1}{\sqrt{2}}\left((q^{(j)}_k)_{2i} - (q^{(j)}_k)_{2i+1}\right) \quad (1.3.23)$$

$$F = \mathbf{I} + (S_B)_i + \frac{\delta_k}{2}(D_A)_i. \quad (1.3.24)$$

Note that the recurrence relations are local and can be carried out over many scales as needed (assuming the existence of F^{-1} at each scale).

Starting with equation (1.3.5) on $\mathbf{V}_{-j}$ and reducing j times yields on $\mathbf{V}_0$

$$\mathbf{B}^{(j)}_0 x^{(j)}_0 + q^{(j)}_0 + \lambda = \mathbf{K}_0(\mathbf{A}^{(j)}_0 x^{(j)}_0 + p^{(j)}_0), \quad (1.3.25)$$

where to compute $\mathbf{B}^{(j)}_0$, $\mathbf{A}^{(j)}_0$, $p^{(j)}_0$, and $q^{(j)}_0$ we use the recurrence relations j times.

Multiresolution homogenization is formulated as follows. First, we consider the limit of

(1.3.25) as $j \to -\infty$:

$$\mathbf{B}_0^{(-\infty)} x_0^{(-\infty)} + q_0^{(-\infty)} + \lambda = \mathbf{K}_0(\mathbf{A}_0^{(-\infty)} x_0^{(-\infty)} + p_0^{(-\infty)}). \tag{1.3.26}$$

This amounts to eliminating infinitely many fine scales from the equation. We call the matrices $\mathbf{B}_0^{(-\infty)}$ and $\mathbf{A}_0^{(-\infty)}$ the reduced coefficients of the equation (1.3.3).

We then look for the operators and forcing terms $B^h(t)$, $A^h(t)$, $q^h(t)$, and $p^h(t)$ with certain desired qualities (e.g. constant values) such that the equation,

$$(I + B^h(t))x(t) + q^h(t) + \lambda = \int_0^t (A^h(s)x(s) + p^h(s))\mathrm{d}s, \quad t \in (0,1), \tag{1.3.27}$$

when subject to the same reduction and limit procedure as (1.3.3), yields on $\mathbf{V}_0$ the same equation as in (1.3.26).

For (1.3.3), we usually require that A^h, B^h, p^h, and q^h be constant. The results of homogenization in this case are summarized as follows:

Proposition 1.3.1 *Given the equation (1.3.3), if the limits which determine the matrices* $\mathbf{B}_0^{(-\infty)}$ *and* $\mathbf{A}_0^{(-\infty)}$ *exist, then there exist constant matrices* B^h, A^h *and forcing terms* p^h, q^h, *such that the reduced coefficients and forcing terms of (1.3.27) are given by* $\mathbf{B}_0^{(-\infty)}$,$\mathbf{A}_0^{(-\infty)}$, $p_0^{(-\infty)}$, $q_0^{(-\infty)}$. *The homogenized coefficients* B^h *and* A^h *and forcing terms* p^h *and* q^h *are defined by*

$$A^h = A_0^{(-\infty)} \tag{1.3.28}$$

$$B^h = A^h \tilde{A}^{-1} - \mathbf{I} \tag{1.3.29}$$

$$p^h = p_0^{(-\infty)} \tag{1.3.30}$$

$$q^h = q_0^{(-\infty)} + (\mathbf{I} - \frac{1}{2}\tilde{A} - \tilde{A}(\exp(\tilde{A} - \mathbf{I})^{-1}A^h)^{-1}p^h, \tag{1.3.31}$$

where

$$\tilde{A} = \log(\mathbf{I} + \left(\mathbf{I} + B_0^{(-\infty)} - \frac{1}{2}A^h\right)^{-1} A^h). \tag{1.3.32}$$

Proof: Following [10], we see that for constant coefficients the recurrence relations (1.3.12) and (1.3.13) simplify to

$$A_{k+1}^h = A_k^h \tag{1.3.33}$$

$$B_{k+1}^h = B_k^h + \frac{\delta_k^2}{4} A_k^h(\mathbf{I} + B_k^h)^{-1} A_k^h. \tag{1.3.34}$$

Likewise, the recurrence relations for the forcing terms simplify to

$$p_{k+1}^h = p_k^h \tag{1.3.35}$$

$$q_{k+1}^h = q_k^h - \frac{\delta_k}{2} A_k^h (\mathbf{I} + B_k^h)^{-1} p_k^h. \tag{1.3.36}$$

Since the term A^h is unchanged by reduction, it is clear that $A^h = A_0^{(-\infty)}$. Similarly p^h is unchanged by reduction, so $p^h = p_0^{(-\infty)}$. The situation for B^h and q^h is more complicated. We solve for them analytically using the solution of (1.3.27).

Consider the case $p_0^{(-\infty)} = 0$. Clearly, then, it is the case that $q^h = q_0^{(-\infty)}$. The solution of (1.3.27) is therefore given by

$$x(t) = -\exp(\tilde{A}t)\tilde{q}, \tag{1.3.37}$$

where $\tilde{A} = (\mathbf{I} + B^h)^1 A^h$, $\tilde{q} = (\mathbf{I} + B^h)^{-1}(q^h + \lambda)$. The average of this solution must also solve (1.3.26) since (1.3.26) by definition is an equation for the average value of the solution. The average value of $x(t)$ in (1.3.37) on the interval $[0, 1]$ is given by

$$\langle x \rangle = \left(- \int_0^1 \exp(\tilde{A}t)dt \right) \tilde{q} = (\mathbf{I} - \exp(\tilde{A}))\tilde{A}^{-1}\tilde{q}. \tag{1.3.38}$$

The solution to (1.3.26) is given by

$$x_0^{(-\infty)} = -(\mathbf{I} + B_0^(-\infty) - \frac{1}{2} A_0^{(-\infty)})^{-1}(q_0^{(-\infty)} + \lambda). \tag{1.3.39}$$

The right-hand sides of (1.3.38) and (1.3.39) are shown in [10] to be equal for all λ; setting $\lambda = 0$ and solving for B^h yields the solution given in the statement of the proposition. $\square$

The case when $p_0^{(-\infty)} \neq 0$ proceeds similarly. We leave out the details of this case since our primary interest lies in the coefficients and not the forcing terms. Interested readers may see [10] for the complete proof.

Solutions of (1.3.27) have the same "average" or coarse-scale behavior as solutions of (1.3.3). Again, the details may be found in [10]. The main point is that this homogenization procedure allows for coefficients which vary on arbitrarily many intermediate scales. This is in contrast to the classical homogenization examples described in the previous Section, which did not allow for intermediate scales.

As formulated above, the multiresolution approach to homogenization requires the computation of $A_0^{(-\infty)}$ and $B_0^{(-\infty)}$, i.e. a limit over infinitely many scales. In practice, the multiresolution reduction algorithm is applied numerically over only finitely many scales. The typical practice is to compute successive $A_0^{(-J)}$ and $B_0^{(-J)}$ terms until finer approximations vary by less than some specified tolerance, and use these matrices as approximations to $A_0^{(-\infty)}$ and $B_0^{(-\infty)}$.

Besides establishing the general framework for multiresolution reduction and homogenization, it is observed in [10] that, for systems of linear ordinary differential equations, using the Haar basis (or a multiwavelet basis) provides a technical advantage. Since the functions of the Haar basis on a fixed scale do not have overlapping supports, the recurrence relations for the operators and forcing terms in the equation may be written as local relations and solved explicitly. Thus, for ODE's, an explicit local reduction *and* homogenization procedure is possible.

In the remainder of this section we consider the relationship between the multiresolution approach and the classical techniques. We demonstrate the connection with the one-dimensional example problem from Section 1.1.

1.3.2 An Example: MRA Homogenization of Second-Order Ordinary Differential Equations

Since multiresolution homogenization is a novel approach to the homogenization problem, we would like to place the multiresolution method in the proper context by comparing it to the existing methods described in Section 1.1.

We describe A. Gilbert's [19] demonstration of MRA homogenization applied to the equation

$$-\frac{d}{dx}(a(x)\frac{d}{dx}u(x)) = f(x), \qquad (1.3.40)$$

$x \in [0, 1]$, with initial conditions at $x = 0$.

Gilbert establishes a connection between the method of [8] and the multiresolution homogenization strategy of [10]. Equation (1.3.40) may be written as a first-order system,

$$\begin{cases} \frac{d}{dx}v(x) = -f(x) \\ \frac{d}{dx}u(x) = a(x)^{-1}v(x) \end{cases} . \qquad (1.3.41)$$

By writing (1.3.41) in an integral form, we have

$$\begin{pmatrix} u(x) \\ v(x) \end{pmatrix} - \begin{pmatrix} u(0) \\ v(0) \end{pmatrix} = \int_0^x \left(\begin{pmatrix} 0 & a(t)^{-1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} + \begin{pmatrix} 0 \\ -f(t) \end{pmatrix} \right) \mathrm{d}t. \qquad (1.3.42)$$

Thus, in the notation of (1.3.3), $B(x) = 0$, $A(x) = \begin{pmatrix} 0 & a(x)^{-1} \\ 0 & 0 \end{pmatrix}$, $\lambda = -\begin{pmatrix} u(0) \\ v(0) \end{pmatrix}$, and $q(t) = 0, p(t) = \begin{pmatrix} 0 \\ -f(t) \end{pmatrix}$.

Using the reduction procedure in the Haar basis for a system of linear differential equations (as in [10]), the goal is to find constant B^h, A^h, p^h, and q^h such that

$$(\mathbf{I} + B^h) \begin{pmatrix} u(x) \\ v(x) \end{pmatrix} + q^h + \lambda = \int_0^x \left(A^h \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} + p^h \right) \mathrm{d}t, \qquad (1.3.43)$$

after reduction to the scale $\mathbf{V}_0$ will be the same as (1.3.42) reduced to that scale. This is accomplished by solving the recursion relations between the operators in the reduced equations explicitly, element-by-element in each matrix. This is possible to do because of the non-overlapping supports of the Haar basis functions on a fixed scale. The result of [19] for the coefficients A^h and B^h is that

$$B^h = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \; A^h = \begin{pmatrix} 0 & M_1 - 2M_2 \\ 0 & 0 \end{pmatrix}, \qquad (1.3.44)$$

where

$$M_1 = \int_0^1 \frac{1}{a(t)} dt \; , \; M_2 = \int_0^1 \frac{(t - 1/2)}{a(t)} dt. \qquad (1.3.45)$$

Similar expressions for p^h and q^h can be found. Note that we have $p^h = q^h = 0$ if $f(x) = 0$ identically. Furthermore, in general B^h and A^h do not depend on p and q.

As a first-order system of ordinary differential equations, the homogenized equation yields

$$\begin{cases} \frac{d}{dx}u(x) = (M_1 - 2M_2)v(x) \\ \frac{d}{dx}v(x) = f^h \end{cases} . \tag{1.3.46}$$

The result is somewhat different than the classical result. We trace this difference to the fact that the multiresolution homogenization procedure allows the coefficients $a(x)$ to vary on arbitrarily many scales, whereas the classical approach in Section 1.1 allows only for coefficients of the form $a(x/\epsilon)$. In the multiresolution context this amounts to restricting the coefficients to an asymptotically fine scale. Let us apply the same limit in Section 1.1 to the coefficients in the multiresolution approach. We start with coefficients of the form $a(x/\epsilon)$. Applying the multiresolution homogenization scheme to the elliptic equation with these coefficients yields two terms, $M_1(\epsilon)$ and $M_2(\epsilon)$. If we take the limit as $\epsilon \to 0$, from Lemma 1.1.1 we find that

$$\lim_{\epsilon \to 0} M_1(\epsilon) = M_1 \tag{1.3.47}$$

and

$$\lim_{\epsilon \to 0} M_2(\epsilon) = 0 \tag{1.3.48}$$

Thus, the factor M_2 is present in the multiresolution context but not present in the classical context, and it is zero when the limit found in the classical approach is applied to the result of the multiresolution approach.

For the purposes of Chapter 2, we note here that Gilbert's results depend on the local solvability of the recurrence equations, a property that does not appear to hold for multi-dimensional elliptic problems. It is our observation that, for partial differential equations, the loss of locality (in solving for the variables in $\mathbf{W}_{j+1}$ in terms of the variables in $\mathbf{V}_{j+1}$) is a generic situation and the one-dimensional case is special due to the ability to express the problem using a first-order system of ordinary differential equations.

If the explicit locality of reduction cannot be achieved for partial differential equations, then one might as well consider the general scheme outlined in [10] where high order wavelets are used (since the Haar basis offers no special advantage). In fact, it turns out that in elliptic problems , only by using high order wavelets can we achieve approximate locality of the reduction procedure, and the order of the wavelets has important implications for the eigenvalue problem as well. We study these issues in Chapter 2.

1.4 Homogenization of Acoustic Wave Equations

The linear acoustic equations in a layered medium present an ideal model problem for comparison. These equations have been studied extensively for more than fifty years, allowing the multiresolution method for homogenization to be firmly placed in context and compared with existing approaches.

These equations essentially form a subset of the full linear equations of small motions of an elastic solid, in the sense that the wave solutions of the acoustic equations may be thought of as P-wave solutions of the elasticity equations. A wide variety of phenomena can be observed in the motion of waves through layered media, including dispersion, attenuation, and long-wavelength anisotropy. For a complete discussion of the existing approaches to these phenomena, see e.g. [11], [4], [29], [30], and references therein.

The general linear acoustic equations are given by

$$\begin{aligned} \rho(\mathbf{x}) v_t(\mathbf{x},t) + \nabla p(\mathbf{x},t) = 0 \\ \tfrac{1}{\rho(\mathbf{x})c(\mathbf{x})^2} p_t(\mathbf{x},t) + \nabla \cdot v(\mathbf{x},t) = 0, \end{aligned} \tag{1.4.1}$$

where $p(\mathbf{x},t)$ is pressure, $v(\mathbf{x},t)$ is particle velocity, $\rho(\mathbf{x})$ is density, and $c(\mathbf{x})$ is sound speed. Assume that the medium varies only in the vertical direction, i.e. that $\rho(\mathbf{x}) = \rho(x_3), c(\mathbf{x}) = c(x_3)$. We write (1.4.1) in component form:

$$\begin{aligned} \rho(x_3)\tfrac{\partial}{\partial t} v^{(1)}(\mathbf{x},t) + \tfrac{\partial}{\partial x_1} p(\mathbf{x},t) &= 0 \\ \rho(x_3)\tfrac{\partial}{\partial t} v^{(2)}(\mathbf{x},t) + \tfrac{\partial}{\partial x_2} p(\mathbf{x},t) &= 0 \\ \rho(x_3)\tfrac{\partial}{\partial t} v^{(3)}(\mathbf{x},t) + \tfrac{\partial}{\partial x_3} p(\mathbf{x},t) &= 0 \\ \tfrac{1}{\rho(x_3)c(x_3)^2}\tfrac{\partial}{\partial t} p(\mathbf{x},t) + \tfrac{\partial}{\partial x_1} v^{(1)}(\mathbf{x},t) + \tfrac{\partial}{\partial x_2} v^{(2)}(\mathbf{x},t) + \tfrac{\partial}{\partial x_3} v^{(3)}(\mathbf{x},t) &= 0. \end{aligned} \tag{1.4.2}$$

To reduce (1.4.2) to an ODE, assume its solutions are of the form

$$\begin{aligned} p(x_3,t) &= (e^{-i\omega t - i\xi_1 x_1 - i\xi_2 x_2})\hat{p}(x_3) \\ v^{(i)}(x_3,t) &= (e^{-i\omega t - i\xi_1 x_1 - i\xi_2 x_2})\hat{v}^{(i)}(x_3). \end{aligned} \tag{1.4.3}$$

In other words, we consider the t, x_1, and x_2 variables in the Fourier domain. By substituting (1.4.3) into (1.4.2), we derive two algebraic equations and two ordinary differential equations; substituting the solutions of the algebraic equations in the ODE's yields a 2×2 linear system of ODE's. We normalize the x_3 variable via $x = \frac{x_3}{h}$ to obtain

$$\frac{d}{dx}\begin{pmatrix} \hat{p}(x) \\ \hat{v}^{(3)}(x) \end{pmatrix} = ih\omega \begin{pmatrix} 0 & \rho(x) \\ \frac{1}{\rho(x)c(x)^2} - \frac{\eta^2}{\rho(x)} & 0 \end{pmatrix} \begin{pmatrix} \hat{p}(x) \\ \hat{v}^{(3)}(x) \end{pmatrix}, \tag{1.4.4}$$

where $\eta^2 = \frac{\xi_1^2+\xi_2^2}{\omega^2}$. The equation (1.4.4) defines a family of ODE's parameterized by η and ω.

The parameter ω is the temporal frequency of the solution in the form of (1.4.3). The parameter η is related to the angle relative to the horizontal of plane wave solutions of the homogeneous acoustic equations. To see this, note that if ρ and c are constant, then we have solutions (in $\hat{p}$) of the form

$$\hat{p}(x) = l_1 e^{ih\omega(\frac{1}{c^2}-\eta^2)^{\frac{1}{2}}x} + l_2 e^{-ih\omega(\frac{1}{c^2}-\eta^2)^{\frac{1}{2}}x}. \tag{1.4.5}$$

By substituting (1.4.5) into (1.4.3), we find that p is given by superpositions of

$$p(\mathbf{x},t) = e^{-i(\omega t+\xi_1 x_1+\xi_2 x_2+\omega\sqrt{\frac{1}{c^2}-\eta^2}x_3)} \tag{1.4.6}$$

and

$$p(\mathbf{x},t) = e^{-i(\omega t+\xi_1 x_1+\xi_2 x_2-\omega\sqrt{\frac{1}{c^2}-\eta^2}x_3)}. \tag{1.4.7}$$

The solution (1.4.6) is constant for values of t and (x_1,x_2,x_3) such that

$$-\omega t = (\xi_1,\xi_2,\omega\sqrt{\frac{1}{c^2}-\eta^2})\cdot(x_1,x_2,x_3)+d \tag{1.4.8}$$

for some constant d. For fixed t and d, the points (x_1,x_2,x_3) which satisfy (1.4.8) form a plane in $\mathbf{R}^3$. Thus, solutions p given by (1.4.6) are plane waves propagating with angle θ relative to the horizontal given by $\sin^2(\theta) = c^2\eta^2$. Similarly we see that (1.4.7) is a plane-wave solution propagating with the same angle θ but in the opposite direction. For plane waves propagating with vertical incidence, we set $\eta = 0$. In the case of a non-constant medium, such solutions are referred to (see e.g. [29]) as "pseudo plane-wave" solutions, the idea being that in a layered medium, the solution forms a plane wave locally inside each layer.

Now that we have introduced the notation and written the acoustic equations as a system of ODE's, we apply the multiresolution homogenization technique to this system and compare the results to existing approaches.

1.4.1 Multiresolution Homogenization of the Acoustic Equations

The homogenization procedure of [10] expects an integral equation of the form

$$(I + B(x))u(x) = \int_0^x A(s)u(s)ds. \tag{1.4.9}$$

If we set $B(x) = 0$, and $A(x) = ih\omega \begin{pmatrix} 0 & \rho(x) \\ \frac{1}{\rho(x)c(x)^2} - \frac{\eta^2}{\rho(x)} & 0 \end{pmatrix}$ then by integrating on both sides of (1.4.4) we may write it in the form of (1.4.9). Each value of ω and η defines a different $A(x)$ and, thus, for each value of ω and η we must apply the homogenization procedure separately.

When we apply multiresolution homogenization to the equation (1.4.4), we derive in the general case an equation of the form

$$\frac{d}{dx_3}(\mathbf{I} + B^h(\omega,\eta)) \begin{pmatrix} \hat{p}(x_3) \\ \hat{v}^{(3)}(x_3) \end{pmatrix} = A^h(\omega,\eta) \begin{pmatrix} \hat{p}(x_3) \\ \hat{v}^{(3)}(x_3) \end{pmatrix}. \tag{1.4.10}$$

The matrices B^h and A^h depend not only on ω and η, but also on the functions $\rho(x)$ and $c(x)$. However, we write $A^h(\omega,\eta)$ and $B^h(\omega,\eta)$ to emphasize that, even for a given medium defined by fixed functions $\rho(x)$ and $c(x)$, the homogenized matrices A^h and B^h may in general be dependent on ω and η.

Multiplying both sides of (1.4.10) by $(\mathbf{I} + B^h(\omega,\eta))^{-1}$ yields

$$\frac{d}{dx} \begin{pmatrix} \hat{p}(x) \\ \hat{v}^{(3)}(x) \end{pmatrix} = K^h(\omega,\eta) \begin{pmatrix} \hat{p}(x) \\ \hat{v}^{(3)}(x) \end{pmatrix}, \tag{1.4.11}$$

where $K^h(\omega,\eta) = (\mathbf{I} + B^h(\omega,\eta))^{-1}A^h(\omega,\eta) = \tilde{A}(\omega,\eta)$ (see (1.3.32)). The matrix $K^h(\omega,\eta)$ therefore may be considered as the *homogenized coefficients* of the ODE (1.4.4) with parameters ω and η. For constant ρ and c, the homogenization procedure (by construction) produces the original coefficients (1.4.4) as the homogenized coefficients.

In the case when ρ and c are not constant, the situation is more complicated. In this discussion, we consider only the case when $\eta = 0$, which implies vertical propagation of plane waves. We write $A^h(\omega,\eta) = A^h(\omega)$ and $B^h(\omega,\eta) = B^h(\omega)$. The matrix $K^h(\omega)$ in general has non-zero elements on the diagonal, which means that the equation is not of the form of the acoustic equations with the layered-medium assumption. Thus, the effective medium is not an acoustic medium. The eigenvalues λ_1 and λ_2 of this matrix can still be used to extract some information about the medium, however. In particular, if $\lambda_1 = \bar{\lambda}_2 = i\lambda$, then the medium permits wave solutions of the form

$$\begin{pmatrix} \hat{p}(x) \\ \hat{v}^{(3)}(x) \end{pmatrix} = e^{i\lambda x}\mathbf{c}_1 + e^{-i\lambda x}\mathbf{c}_2, \tag{1.4.12}$$

which may be written as

$$\begin{pmatrix} p(x_1,x_2,x_3,t) \\ v^{(3)}(x_1,x_2,x_3,t) \end{pmatrix} = e^{i(\omega t + \frac{\lambda}{h}x_3)}\mathbf{c}_1 + e^{i(\omega t - \frac{\lambda}{h}x_3)}\mathbf{c}_2. \tag{1.4.13}$$

These solutions are pressure/particle-velocity waves propagating with velocity $\frac{h\omega}{\lambda(\omega)}$. The effective velocity of this medium is therefore given by $\frac{h\omega}{\lambda(\omega)}$.

If the eigenvalues λ_1 and λ_2 are not pure imaginary, then the solutions of the homogenized ODE (1.4.11) may exhibit exponential growth or decay. Exponential decay of the solution as x_3 grows means that the plane wave solutions decay as they propagate downwards. Exponential growth of the solution as x_3 grows means that the plane wave solutions decay as they propagate upwards. If the eigenvalues λ_1 and λ_2 have non-zero imaginary part but are not complex conjugates of one another, then there are two different effective velocities of the medium. We show, however, that if the eigenvalues λ_1 and λ_2 have non-zero imaginary part, then they are always complex conjugates. This result is summarized in the following proposition:

Proposition 1.4.1 *The matrix $K^h(\omega,\eta)$ which results from multiresolution homogenization applied to (1.4.4) has eigenvalues λ_1, λ_2 such that $\lambda_1 = \bar{\lambda}_2$.*

Proof: The first part of this proof uses induction. Note that, for any j, $(B_j^{(j)})_l = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and $(A_j^{(j)})_l = \begin{pmatrix} 0 & ih\omega\rho_l \\ ih\frac{\omega}{\rho_l}(\frac{1}{c_l^2}-\eta^2) & \end{pmatrix}$. Thus, these matrices both have purely real diagonals and purely imaginary off-diagonals. Assume that $(A_k^{(j)})_l$ and $(B_k^{(j)})_l$ have this same form. For 2×2 matrices, it is clear that sums, products, and inverses of such matrices also have this form. So, by (1.3.12) and (1.3.13) we see that that $(A_{k+1}^{(j)})_l$ and $(B_{k+1}^{(j)})_l$ have this form as well. Therefore, $A_0^{(-\infty)}$ and $B_0^{(-\infty)}$ have purely real diagonals and purely imaginary off-diagonals.

From (1.3.28), we see that $A^h = A_0^{(-\infty)}$. By equation (1.3.32), it is clear that the matrix $\exp(\tilde{A}) = \mathbf{I} + (\mathbf{I} + \mathbf{B}_0^{(-\infty)} - \frac{1}{2})^{-1}A^h$ has purely real diagonal elements and purely imaginary off-diagonal elements. The characteristic polynomial of any 2×2 matrix of this form has all real coefficients, so it must be the case that, if the eigenvalues of $\exp(\tilde{A})$ have non-zero imaginary part, then they are complex conjugates. The eigenvalues of $\tilde{A}$ are given by the logarithm of the eigenvalues of $\exp(\tilde{A})$ and are therefore complex conjugates as well.□

Now consider homogenization of the acoustic equations when the medium consists of two distinct layers, inside each of which the medium is constant. We denote the width of the interval by h and the densities and sound speeds in each layer by ρ_0, ρ_1 and c_0, c_1. In this situation, it is possible to derive an analytic formula for the homogenized matrix $K^h(\omega)$. In the next subsection, we describe a technique for finding this exact formula for the homogenized matrix $K^h(\omega)$.

The imaginary part of the eigenvalues $\lambda_k(\omega)$ of $K^h(\omega)$ define the dispersion relation of the homogenized medium. These two eigenvalues are always complex conjugates if they have non-zero imaginary part, so we may define the dispersion relation in terms of a single value $\kappa(\omega)$, which is the absolute value of the imaginary part of $\lambda_k(\omega)$. Using the formula derived in the next subsection results in a rather long and complicated formula for $\kappa(\omega)$ which we will not display here. Such a formula is, of course, not necessary for computational purposes;

we derive it only to show that the multiresolution approach may be used to derive an exact formula for the dispersion relation.

1.4.2 Multiresolution Homogenization of a Two-Layer Medium

Although the multiresolution homogenization scheme is meant as a numerical procedure, we derive an exact formula for the homogenization procedure applied to a two-layer medium. This demonstrates that the multiresolution approach may be used to find an exact formula for the dispersion relation in a simple case.

We start with the ODE

$$\frac{d}{dz}\hat{u}(z) = i\omega\hat{M}(z)\hat{u}(z) \tag{1.4.14}$$

on the interval $[0,1]$. We assume that the medium consists of two uniform layers, so the matrix $\hat{M}z$ is defined as

$$\hat{M}z = \begin{cases} M_0 & 0 \leq z < \frac{1}{2} \\ M_1 & \frac{1}{2} \leq z \leq 1 \end{cases}. \tag{1.4.15}$$

By integrating (1.4.14) on both sides and setting $B(z) = 0, A(z) = i\omega\hat{M}(z)$, we may write (1.4.14) in the form of (1.3.3). The MRA homogenization procedure then may be applied; it yields matrices B^h, A^h which depend on ω, M_0, and M_1. The resulting equation of the form (1.3.27) may be written as

$$\frac{d}{dz}\hat{u}(z) = K^h\hat{u}(z) \tag{1.4.16}$$

by setting $K^h = (\mathbf{I} + B^h)^{-1}A^h$.

To derive the exact formula for K^h, first consider equation (1.4.14) in each layer separately. Denote by A_j^h, B_j^h the matrices obtained by applying the homogenization procedure to layer j (where $j = 0, 1$). Since the coefficients are constant inside each layer, we have by construction

$$B_j^h = B_j = 0, \quad A_j^h = A_j = i\omega M_j \tag{1.4.17}$$

By using (1.3.28), (1.3.29), and (1.3.32), we can find $(A_1^{(-\infty)})_j$ and $(B_1^{(-\infty)})_j$ in terms of A_j and B_j:

$$(A_1^{(-\infty)})_j = A_j, \quad (B_1^{(-\infty)})_j = \frac{A_j}{2}\left((\exp(\frac{A_j}{2}) - \mathbf{I})^{-1} + \frac{1}{2}\mathbf{I}\right) - \mathbf{I} \tag{1.4.18}$$

To compute K^h, we use the formula

$$\exp(K^h) = \mathbf{I} + \left(\mathbf{I} + B_0^{(-\infty)} - \frac{1}{2}A_0^{(-\infty)}\right)^{-1} A_0^{-\infty}. \tag{1.4.19}$$

As defined, $A_0^{(-\infty)}$ and $B_0^{(-\infty)}$ are computed by finding the limits of $A_0^{(-J)}$ and $B_0^{(-J)}$ as $J \to \infty$. The matrices $A_0^{(-J)}$ and $B_0^{(-J)}$ are defined in terms of $(A_1^{(-J+1)})_j$ and $(B_1^{(-J+1)})_j$ (via the recurrence relations (1.3.12) and (1.3.13)) as

$$B_0^{(-J)} = S_B - \frac{1}{4}D_A - (D_B - \frac{1}{4}S_A)F^{-1}(D_B + \frac{1}{4}S_A), \tag{1.4.20}$$

and

$$A_0^{(-J)} = S_A - D_A F^{-1}(D_B + \frac{1}{4}S_A), \tag{1.4.21}$$

where

$$S_A = \frac{1}{2}\left((A_1^{(-J+1)})_0 + (A_1^{(-J+1)})_1\right) \tag{1.4.22}$$

$$D_A = \frac{1}{2}\left((A_1^{(-J+1)})_0 - (A_1^{(-J+1)})_1\right) \tag{1.4.23}$$

$$S_B = \frac{1}{2}\left((B_1^{(-J+1)})_0 + (B_1^{(-J+1)})_1\right) \tag{1.4.24}$$

$$D_B = \frac{1}{2}\left((B_1^{(-J+1)})_0 - (B_1^{(-J+1)})_1\right) \tag{1.4.25}$$

$$F = \mathbf{I} + S_B + \frac{1}{4}D_A. \tag{1.4.26}$$

Because $(A_1^{(-J+1)})_j$ and $(B_1^{(-J+1)})_j$ converge element-by-element to $(A_1^{(-\infty)})_j$ and $(B_1^{(-\infty)})$ as $J \to \infty$, we may pass the limit in J through (1.4.20) and (1.4.21) to obtain

$$B_0^{(-\infty)} = S'_B - \frac{1}{4}D'_A - (D'_B - \frac{1}{4}S'_A)F^{-1}(D'_B + \frac{1}{4}S'_A) \tag{1.4.27}$$

and

$$A_0^{(-\infty)} = S'_A - D'_A F^{-1}(D'_B + \frac{1}{4}S'_A), \tag{1.4.28}$$

where

$$S'_A = \frac{1}{2}\left((A_1^{(-\infty)})_0 + (A_1^{(-\infty)})_1\right) \tag{1.4.29}$$

$$D'_A = \frac{1}{2}\left((A_1^{(-\infty)})_0 - (A_1^{(-\infty)})_1\right) \tag{1.4.30}$$

$$S'_B = \frac{1}{2}\left((B_1^{(-\infty)})_0 + (B_1^{(-\infty)})_1\right) \tag{1.4.31}$$

$$D'_B = \frac{1}{2}\left((B_1^{(-\infty)})_0 - (B_1^{(-\infty)})_1\right) \tag{1.4.32}$$

$$F = \mathbf{I} + S'_B + \frac{1}{4}D'_A. \tag{1.4.33}$$

This series of steps is summarized in the following theorem:

Theorem 1.4.1 *The homogenized matrix K^h for the two-layer wave equation given by (1.4.14) is defined by*

$$K^h = \log\left(\mathbf{I} + \left(\mathbf{I} + B_0^{(-\infty)} - \frac{1}{2}A_0^{(-\infty)}\right)^{-1} A_0^{-\infty}\right), \tag{1.4.34}$$

where $B_0^{(-\infty)}$ and $A_0^{(-\infty)}$ are defined by (1.4.27)-(1.4.33).

Thus, equations (1.4.27)-(1.4.33), together with equation (1.4.19), completely define the steps necessary to compute an exact formula for K^h. For any given equation, the matrix K^h may be a very complicated expression. However, the purpose of deriving this exact formula is not for computation but rather to demonstrate that the multiresolution approach can be used to compute analytic rather than numeric results, if desired. Towards this end, we will also compute the first few terms in the small ω expansion of K^h in a more general form in terms of the matrices M_0 and M_1.

First, we compute $(B_1^{(-\infty)})_j$ and $(A_1^{(-\infty)})_j$ in terms of M_0 and M_1. We see that if $A = i\omega M$, then

$$\begin{aligned}
\exp\left(\frac{1}{2}A\right) - \mathbf{I} &= \exp\left(\frac{i\omega}{2}M\right) - \mathbf{I} && (1.4.35)\\
&= \frac{i\omega}{2}M - \frac{\omega^2}{8}M^2 - i\frac{\omega^3}{48}M^3 + \mathcal{O}(\omega^4) && (1.4.36)\\
&= \frac{i\omega}{2}M\left(\mathbf{I} + i\frac{\omega}{4}M - \frac{\omega^2}{24} + \mathcal{O}(\omega^3)\right). && (1.4.37)
\end{aligned}$$

Thus,

$$\begin{aligned}
(\exp\left(\frac{1}{2}A\right) - \mathbf{I})^{-1} &= \left(\mathbf{I} + i\frac{\omega}{4}M - \frac{\omega^2}{24}M^2 + \mathcal{O}(\omega^3)\right)^{-1}\left(\frac{i\omega}{2}M\right)^{-1} && (1.4.38)\\
&= \left(\mathbf{I} - i\frac{\omega}{4}M + \omega^2\left(\frac{1}{24} - \frac{1}{16}\right)M^2 + \mathcal{O}(\omega^3)\right)\frac{2}{i\omega}M^{-1}. && (1.4.39)
\end{aligned}$$

Therefore, by (1.4.18), we have

$$(B_1^{(-\infty)})_j = \frac{1}{2}A_j\left((\exp\left(\frac{1}{2}A_j\right) - \mathbf{I})^{-1} + \frac{1}{2}\mathbf{I}\right) - \mathbf{I} = -\frac{\omega^2}{48}M_j^2 + \mathcal{O}(\omega^4) \tag{1.4.40}$$

and of course

$$(A_1^{(-\infty)})_j = i\omega M_j. \tag{1.4.41}$$

This leads to

$$S'_A = \omega \bar{S}_A = \omega(\frac{1}{2}i(M_0 + M_1)) \tag{1.4.42}$$

$$D'_A = \omega \bar{D}_A = \omega(\frac{1}{2}i(M_0 - M_1)) \tag{1.4.43}$$

$$S'_B = \omega^2 \bar{S}_B = \omega^2(-\frac{1}{96}(M_0^2 + M_1^2)) + \mathcal{O}(\omega^4) \tag{1.4.44}$$

$$D'_B = \omega^2 \bar{D}_B = \omega^2(-\frac{1}{96}(M_0^2 - M_1^2)) + \mathcal{O}(\omega^4). \tag{1.4.45}$$

We have

$$F = \mathbf{I} + S'_B + \frac{1}{4}D'_A = \mathbf{I} + \frac{1}{4}\omega \bar{D}_A + \omega^2 \bar{S}_B \tag{1.4.46}$$

and, thus,

$$F^{-1} = \mathbf{I} - \frac{1}{4}\omega \bar{D}_A + \omega^2 \left(\frac{1}{16}(\bar{D}_A)^2 - \bar{S}_B\right) + \mathcal{O}(\omega^3). \tag{1.4.47}$$

Using (1.4.27) and (1.4.28), we derive

$$\begin{aligned} B_0^{(-\infty)} &= \omega^2 \bar{S}_B - \frac{\omega}{4}\bar{D}_A - (\omega^2 \bar{D}_B - \frac{\omega}{4}\bar{S}_A)(\mathbf{I} - \frac{\omega}{4}\bar{D}_A + \mathcal{O}(\omega^2))(\omega^2 \bar{D}_B + \frac{\omega}{4}\bar{S}_A) \\ &= -\frac{\omega}{4}\bar{D}_A + (\bar{S}_B + \frac{1}{16}(\bar{S}_A)^2)\omega^2 + \mathcal{O}(\omega^3) \end{aligned} \tag{1.4.48}$$

and

$$A_0^{(-\infty)} = \omega \bar{S}_A - \omega \bar{D}_A(\mathbf{I} - \frac{\omega}{4}\bar{D}_A + \mathcal{O}(\omega^2))(\omega^2 \bar{D}_B + \frac{\omega}{4}\bar{S}_A) \tag{1.4.49}$$

$$= \omega \bar{S}_A - \frac{\omega^2}{4}\bar{D}_A \bar{S}_A + \mathcal{O}(\omega^3). \tag{1.4.50}$$

Finally, after some more algebra, this yields

$$\exp(K^h) = \mathbf{I} + (\mathbf{I} + B_0^{(-\infty)} - \frac{1}{2}A_0^{(-\infty)})^{-1} x A_0^{(-\infty)} \tag{1.4.51}$$

$$= \mathbf{I} + \omega \bar{S}_A + \frac{\omega^2}{2}\bar{S}_A^2 + f_3\omega^3 + \mathcal{O}(\omega^4) \tag{1.4.52}$$

and thus

$$K^h = \omega \bar{S}_A + \omega^3(f_3 - \frac{1}{6}\bar{S}_A^3) + \mathcal{O}(\omega^4) \tag{1.4.53}$$

The term f_3 is left unexpanded because it is rather complicated. The formula (1.4.53) does not necessarily tell us that the eigenvalues of K^h have an asymptotic expansion in ω of the same form. This will depend on the particular form of the matrices involved. We do immediately see, however, that the leading order behavior in ω of K^h may be found simply by averaging M_0 and M_1.

Also, since the multiresolution homogenization procedure gives the same result for all values of the initial conditions, we may extend the exact-homogenization formula for the two-layer problem to a periodically layered medium with two distinct, repeated layers of equal size. In the case where we have many non-periodic layers of equal size, the leading order behavior in ω of K^h may be found simply by averaging $\hat{M}$ over all the layers. This can be seen by starting with the two-layer asymptotic expansion and applying the formula recursively.

1.4.3 Comparison to Existing Theory

In this subsection, we give a brief description of existing methods for computing the dispersion relation for layered media, then compare them with the multiresolution approach.

Many techniques exist for studying qualitative behavior of solutions to the layered-medium acoustic equations. These techniques range from asymptotic expansions for small ω (see e.g. [11], [29], [30]) to exact solutions for periodic two-layer media (Floquet theory, see e.g. [12]). As we showed above, such results may be obtained from the multiresolution approach as well.

However, the multiresolution approach is intended as a numerical procedure for computing the homogenized coefficients of an equation which may have many layers. The methods described in e.g. [11], [29], and [30] may not be practically applied when there are many layers or frequency bands that we wish to study.

The only alternative we are aware of that may achieve the same result as the multiresolution method is to use the fact that, in the case of ODE's, there is a "preferred direction." For this preferred direction, an exact solution may be obtained by using propagator matrices for each layer (see e.g. [1] for a description of this technique). Once such a solution is computed, one can easily obtain its projection on all scales.

The multiresolution homogenization method simply finds an equation on each scale which has as its solution the projection of the exact solution to that scale. Thus, multiresolution homogenization and other exact techniques, such as Floquet theory or the propagator matrix technique, will give the same results for a periodic layered medium in one dimension. In this context, the multiresolution approach may be viewed as an alternative to computing the exact solution directly.

As far as we are aware, the propagator matrix technique has not been generalized to fully two or three dimensional problems. The multiresolution approach, on the other hand, does generalize to higher dimensional problems. This generalization is the subject of study in Chapter 2.

CHAPTER 2

REDUCTION OF ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

In this chapter, we apply the multiresolution reduction procedure of Section 1.2 to matrices which arise as projections of linear partial differential equations into a multiresolution analysis.

Let $\mathbf{S}$ be the linear operator on $\mathbf{L}^2(\mathbf{R}^d)$ $(d = 1, 2)$ given by a PDE with appropriate boundary conditions. We formulate the conditions and prove our results in terms of the matrix $\mathbf{S}_j$, which is the projection of the operator $\mathbf{S}$ onto a space $\mathbf{V}_j$ of the multiresolution analysis (where $\mathbf{V}_j$ is some sufficiently fine scale). We comment where appropriate on the connection between properties of the matrix $\mathbf{S}_j$ and properties of the operator $\mathbf{S}$. Thus, our results are proved for infinite matrices representing the operator restricted to a finite (but arbitrary) number of scales.

In Section 2.1, we start by showing that matrices $\mathbf{S}_j$ which are a projection of the second derivative $-\frac{d^2}{dx^2}$ into the multiresolution analysis have their form preserved under the reduction procedure, in the sense that the matrix $\mathbf{R}_{\mathbf{S}_j}$ (see (1.2.10)) is an approximation to the second derivative of the same order as the matrix $\mathbf{S}_j$. We also describe the results of [17], which show that matrices in discrete divergence form have this divergence form preserved (in a certain sense) under the reduction procedure if the Haar basis is used as the multiresolution analysis.

In Section 2.2, which contains the main results of this Thesis, we are concerned primarily with second-order elliptic operators which have a sparse approximation in the wavelet basis. However, some of the results are applicable to a wider class of operators.

In order to clarify the discussion, let us explain the use of the terms "elliptic" and "sparse" here. We use the term "second-order elliptic operator to mean"

- a symmetric partial differential operator of the form $-\nabla(a(\mathbf{x})\nabla)$, where $a(\mathbf{x}) > 0$

or, more generally,

- a symmetric second-order elliptic pseudodifferential operator (see [32] for details), i.e. a symmetric pseudodifferential operator with a symbol $a(\mathbf{x}, \xi)$ of order 2 which satisfies the condition that there exists an $\epsilon > 0$ such that $|a(x, \xi)| \geq \epsilon|\xi|^2$.

The term "sparse" is used in this Thesis in a somewhat non-traditional fashion. Namely, let us define, for an operator $\mathbf{T}$, the restriction of $\mathbf{T}$ to a "banded form" $\mathbf{T}_B$, with a finite bandwidth B. The bandwidth B typically refers to the number of non-zero elements per row. We say that the operator $\mathbf{T}$ has a sparse approximation if for any $\epsilon > 0$ there is a bandwidth B such that $\|\mathbf{T} - \mathbf{T}_B\| < \epsilon$, where $\|\|$ is some operator norm (usually the l^∞ operator norm). This definition is used primarily for infinite matrices $\mathbf{T}$; in the case of a finite matrix $\mathbf{T}$, the bandwidth B should be made reasonably small compared to the size of the matrix, if possible. If the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the blocks of the non-standard form of the operator $\mathbf{S}$ (see [9] or Appendix A for a description) and have a sufficient rate of decay away from the diagonal, then for practical purposes elements smaller than a given threshold may be set to zero. After this thresholding procedure, the matrices are sparse in the traditional sense, in that there are few non-zero elements.

We now summarize the results of Section 2.2. We show first that the spectral bounds of symmetric and positive matrices $\mathbf{S}_j$ are preserved under the reduction procedure. We then show that the reduction procedure applied to the non-standard form of an operator (which has a sparse approximation) preserves the rate of off-diagonal decay in the matrices of the non-standard form. This means simply that, for practical purposes, the sparsity of the matrices is preserved under the reduction procedure. We then describe the results of [20], which show that, for such sparse matrices, the reduction procedure may be performed in a numerically efficient manner. Next, we show that, if $\mathbf{S}_j$ is the projection of a second-order elliptic operator with a compact inverse, then the small eigenvalues of the reduced matrix $\mathbf{R}_{\mathbf{S}_j}$ are good approximations to the small eigenvalues of the matrix $\mathbf{S}_j$. Finally, we give some numerical examples which demonstrate the eigenvalue results in practice for second-order elliptic operators in one and two dimensions.

2.1 Preservation of Form Under Reduction

As we demonstrated in Chapter 1, the multiresolution strategy for homogenization of ODE's preserves an explicit form of the equations under the reduction procedure, provided a basis with non-overlapping supports is used for the reduction step.

In this section we consider second-order elliptic operators, and in particular operators of the form

$$-\nabla \cdot (a(\mathbf{x})\nabla) \tag{2.1.1}$$

on $\mathbf{L}^2(\mathbf{R}^1)$ and $\mathbf{L}^2(\mathbf{R}^2)$. We would like to find out whether the form of the matrix $\mathbf{S}_j$, which is the projection of the operator (2.1.1) into a multiresolution analysis, is preserved under reduction. This would provide us with an analogy to the recursion relation for the coefficients in the multiresolution homogenization scheme for ODE's.

2.1.1 Reduction of Differential and Convolution Operators

For matrices $\mathbf{S}_j$ which are projections of operators on $\mathbf{L}^2(\mathbf{R})$, the reduction procedure simplifies greatly if the matrix $\mathbf{S}_j$ is a convolution matrix. We show that if $\mathbf{S}_j$ is a convolution matrix which approximates the second derivative to some order, then the reduced matrix $\mathbf{R}_{\mathbf{S}_j}$ is a convolution matrix as well, and approximates the second derivative to the same order as $\mathbf{S}_j$.

Let $\mathbf{S}_j$ be a bounded convolution matrix on $\mathbf{V}_j \subset \mathbf{L}^2(\mathbf{R})$. There are two ways to represent application of this operator to an element x of $\mathbf{V}_j$. The first is simply to perform matrix-vector multiplication, $(\mathbf{S}_j x)_k = \sum_l (\mathbf{S}_j)_{(k,l)} x_l$. The second is to form two 2π-periodic functions

$$s_j(\xi) = \sum_{n \in \mathbf{Z}} (\mathbf{S}_j)_{(0,n)} e^{in\xi} \tag{2.1.2}$$

$$\hat{x}(\xi) = \sum_{n \in \mathbf{Z}} x_n e^{in\xi}$$

and then consider their product $s_j(\xi)\hat{x}(\xi)$. The coefficient of $e^{ik\xi}$ in this product is exactly $(\mathbf{S}_j x)_k$. Thus, as is well known, we have a way to represent application of convolution matrices as multiplication by a function in the Fourier domain. Furthermore, if a convolution has an inverse then we may represent application of its inverse as division by a function in the Fourier domain.

The $\mathbf{A}_{\mathbf{S}_j}, \mathbf{B}_{\mathbf{S}_j}, \mathbf{C}_{\mathbf{S}_j}$, and $\mathbf{T}_{\mathbf{S}_j}$ blocks are also convolution operators, and we use the representation

$$\begin{aligned} t_j(2\xi) &= |m_0(\xi)|^2 s_j(\xi) + |m_1(\xi)|^2 s_j(\xi+\pi) \\ a_j(2\xi) &= |m_1(\xi)|^2 s_j(\xi) + |m_0(\xi)|^2 s_j(\xi+\pi) \\ b_j(2\xi) &= m_1(\xi)\overline{m_0(\xi)} s_j(\xi) + m_1(\xi+\pi)\overline{m_0(\xi+\pi)} s_j(\xi+\pi) \\ c_j(2\xi) &= m_0(\xi)\overline{m_1(\xi)} s_j(\xi) + m_0(\xi+\pi)\overline{m_1(\xi+\pi)} s_j(\xi+\pi)), \end{aligned} \tag{2.1.3}$$

where the functions m_0 and m_1 are the filters associated with the multiresolution analysis. (For a description and review of their properties, see Section A.1 of the Appendix and also [15]). We see that, in the Fourier domain, the analogue of the formula $\mathbf{R}_{\mathbf{S}_j} = \mathbf{T}_{\mathbf{S}_j} - \mathbf{C}_{\mathbf{S}_j} \mathbf{A}_{\mathbf{S}_j}^{-1} \mathbf{B}_{\mathbf{S}_j}$ may be written as

$$r_j(\xi) = t_j(\xi) - \frac{b_j(\xi) c_j(\xi)}{a_j(\xi)}. \tag{2.1.4}$$

We desire $r_j(\xi)$ to be bounded on $[-\pi, \pi]$, so we must choose the discretization $\mathbf{S}_j$ in such a way that either

$$a_j(\xi) \neq 0 \tag{2.1.5}$$

or

$$\lim_{\xi' \to \xi} \frac{b_j(\xi') c_j(\xi')}{a_j(\xi')} < \infty \tag{2.1.6}$$

for all $\xi \in [-\pi, \pi]$.

The condition $a_j(\xi) \neq 0$ for $\xi \in [-\pi, \pi]$ is not unreasonable for a projection of $-\frac{d^2}{dx^2}$ on $\mathbf{L}^2(\mathbf{R})$. (In fact, this condition is analagous to the ellipticity condition for pseudodifferential symbols; see [32] and the introduction to this Chapter for a description). Note that $|m_1(\xi)|^2 = 1 - |m_0(\xi)|^2$ and $m_0(0) = 1$. Thus, if $\mathbf{S}_j$ is constructed so that $s_j(\xi) = 0 \Rightarrow \xi = 0$ and $s_j(\xi)$ is real and positive for $0 < \xi \leq \pi, -\pi \leq \xi < 0$, we see that

$$a_j(\xi) = |m_1(\xi/2)|^2 s_j(\xi/2) + |m_0(\xi/2)|^2 s_j(\xi/2 + \pi/2) > 0 \tag{2.1.7}$$

for all $\xi \in [-\pi, \pi]$.

Additionally, we see that if $\mathbf{S}_j$ is symmetric and positive, then $s_j(\xi)$ is real and non-negative. For the second-derivative operator, then, in order to ensure the existence of $\mathbf{A}_{\mathbf{S}_j}^{-1}$, we construct $\mathbf{S}_j$ so that it is symmetric and positive, and 0 is an element of the spectrum of $\mathbf{S}_j$ only if $s_j(\xi)$ is non-zero everywhere except at $\xi = 0$.

Now, assume that we have constructed $\mathbf{S}_j$ as a discretization of $-\frac{d^2}{dx^2}$ in such a way that we assure the existence of $\mathbf{A}_{\mathbf{S}_j}^{-1}$ (as above). We use properties of m_0 and m_1 (see Appendix A) to arrive at

$$r_j(2\xi) = t_j(2\xi) - \frac{b_j(2\xi)c_j(2\xi)}{a_j(2\xi)} = \ldots = \frac{s_j(\xi)s_j(\xi+\pi)}{|m_1(\xi)|^2 s_j(\xi) + |m_0(\xi)|^2 s(\xi+\pi)}. \tag{2.1.8}$$

(In fact the above equation holds for any $s_j(\xi)$ as long as $\mathbf{A}_{\mathbf{S}_j}^{-1}$ exists.)

We summarize our results as

Proposition 2.1.1 *Let $\mathbf{V}_j$ and $\mathbf{V}_{j+1}$ be successive subspaces of a multiresolution analysis with $m-1$ vanishing moments. Suppose $\mathbf{S}_j$ is a symmetric and positive convolution matrix on $\mathbf{V}_j \subset \mathbf{L}^2(\mathbf{R})$ such that $s_j(\xi)$ (as defined by (2.1.2)) is non-zero everywhere except at $\xi = 0$, and*

$$s_j(\xi) \sim \frac{1}{(h_j)^2}\xi^2 + \mathcal{O}(\xi^q) \tag{2.1.9}$$

for $\xi \ll 1$ (where $2 < q \leq 2m$). Then $\mathbf{R}_{\mathbf{S}_j}$ is a symmetric and positive convolution operator on $\mathbf{V}_{j+1} \subset \mathbf{L}^2(\mathbf{R})$ such that $r_j(\xi)$ is non-zero everywhere except at $\xi = 0$, and

$$r_j(\xi) \sim \frac{1}{(h_{j+1})^2}\xi^2 + \mathcal{O}(\xi^q), \tag{2.1.10}$$

where $h_k = \frac{2\pi}{2^k}$ is the step-size of the discretization.

Proof: We already have that $\mathbf{R}_{\mathbf{S}_j}$ is a convolution matrix. If $\mathbf{S}_j$ is symmetric then it is clear that $s_j(\xi)$ is an even function about $\xi = 0$. Therefore, $s_j(-\frac{\xi}{2} - \pi) = s_j(\frac{\xi}{2} + \pi) = s_j(\frac{\xi}{2} - \pi)$ since s_j is 2π periodic. Since $|m_0(\xi)|^2$ and $|m_1(\xi)|^2$ are both even functions, it is clear that $r_j(\xi)$ is an even function as well. We may write the entry $(\mathbf{R}_{\mathbf{S}_j})_{0,k}$ in terms of the function $r_j(\xi)$ as

$$(\mathbf{R}_{\mathbf{S}_j})_{0,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} r_j(\xi)e^{-ik\xi}d\xi = \frac{1}{2\pi}\int_{-\pi}^{\pi} r_j(\xi)\cos(k\xi)d\xi, \tag{2.1.11}$$

from which we deduce

$$(\mathbf{R}_{\mathbf{S}_j})_{0,-k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} r_j(-\xi)\cos(-k\xi)d\xi = \frac{1}{2\pi}\int_{-\pi}^{\pi} r_j(\xi)\cos(k\xi)d\xi = (\mathbf{R}_{\mathbf{S}_j})_{0,k}. \tag{2.1.12}$$

Therefore, $\mathbf{R}_{\mathbf{S}_j}$ is a symmetric convolution operator. Using (2.1.8), we see that $r_j(\xi) \neq 0$ except at $\xi = 0$.

Furthermore, if we are using wavelets with $m-1$ vanishing moments, then for small ξ we know that

$$|m_0(\xi)|^2 \sim 1 + \mathcal{O}(\xi^{2m}) \tag{2.1.13}$$

and

$$|m_1(\xi)|^2 \sim \mathcal{O}(\xi^{2m}). \tag{2.1.14}$$

We can rewrite (2.1.8) as

$$r_j(2\xi) = \frac{s_j(\xi)}{|m_1(\xi)|^2 \frac{s_j(\xi)}{s_j(\xi+\pi)} + |m_0(\xi)|^2}.$$

Assuming $\xi \ll 1$ gives us

$$r_j(2\xi) \sim \frac{s_j(\xi)}{\mathcal{O}(\xi^{2m})(\mathcal{O}(\xi^2) + \mathcal{O}(\xi^q)) + 1 + \mathcal{O}(\xi^{2m})} \sim s_j(\xi) + \mathcal{O}(\xi^q) \tag{2.1.15}$$

which yields

$$r_j(\xi) \sim s_j(\frac{\xi}{2}) + \mathcal{O}(\xi^q) \sim \frac{1}{(h_j)^2}\frac{\xi^2}{4} + \mathcal{O}(\xi^q) \sim \frac{1}{(h_{j+1})^2}\xi^2 + \mathcal{O}(\xi^q). \tag{2.1.16}$$

The proof of Proposition 2.1.1 is therefore complete. □

Thus, we see that for the second derivative operator in one dimension, the reduction procedure preserves the "form" of projection $\mathbf{S}_j$ of this operator in the sense that the reduced operator $\mathbf{R}_{\mathbf{S}_j}$ is symmetric and positive, and it approximates the second derivative to the same order as the matrix $\mathbf{S}_j$.

2.1.2 Preservation of Divergence Form

M. Dorobantu and B. Engquist in [17] study the multiresolution reduction procedure applied to projections of operators of the form

$$-\nabla \cdot (a(\mathbf{x})\nabla). \tag{2.1.17}$$

In one dimension, this operator may be written in the Haar basis as

$$\mathbf{S}_j = -h_j^{-2}\Delta_+\text{diag}(a_j)\Delta_-, \tag{2.1.18}$$

where h_j is the step-size on scale $\mathbf{V}_j$, Δ_+ is the standard forward-difference operator whose stencil is $(-1, 1)$, and $\Delta_- = \Delta_+^*$. The operator $\text{diag}(a_j)$ represents multiplication by the function a and in the case of the Haar basis is simply a diagonal matrix whose entries are the coefficients of the vector $a_j = \mathbf{P}_j a(x)$ in the Haar basis. The form (2.1.18) is called discrete divergence form.

In the Haar basis, the operators Δ_+ and Δ_- have a relatively simple decomposition, given by the following:

$$\begin{pmatrix} \mathbf{A}_{\Delta_+} & \mathbf{B}_{\Delta_+} \\ \mathbf{C}_{\Delta_+} & \mathbf{T}_{\Delta_+} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} M & -\Delta_+ \\ \Delta_+ & \Delta_+ \end{pmatrix}, \tag{2.1.19}$$

where M is convolution with the stencil $(-3, -1)$. Also, the operator $\text{diag}(a_j)$ is decomposed as

$$\begin{pmatrix} \mathbf{A}_{\text{diag}(a_j)} & \mathbf{B}_{\text{diag}(a_j)} \\ \mathbf{C}_{\text{diag}(a_j)} & \mathbf{T}_{\text{diag}(a_j)} \end{pmatrix} = \begin{pmatrix} \text{diag}(s_a) & \text{diag}(d_a) \\ \text{diag}(d_a) & \text{diag}(s_a) \end{pmatrix}, \tag{2.1.20}$$

where s_a and d_a are vectors given by $(s_a)_k = \frac{(a_j)_{2k}+(a_j)_{2k+1}}{2}$ and $(d_a)_k = \frac{(a_j)_{2k}-(a_j)_{2k+1}}{2}$. This simple structure in the Haar basis can be exploited so that $\mathbf{R}_{\mathbf{S}_j}$ is written in a simpler form. In [17] this form is given by

$$\mathbf{R}_{\mathbf{S}_j} = (h_{j+1})^{-2}\Delta_+ H\Delta_-. \tag{2.1.21}$$

The operator H is called the homogenized coefficient matrix. It is in general not a diagonal matrix and so does not directly represent a multiplication operator in the Haar basis. Its general form is given by

$$H = 2(\text{diag}(s_a)+\text{diag}(d_a))+(\text{diag}(s_a)+\text{diag}(d_a))(\Delta_-+M^*)\mathbf{A}_{\mathbf{S}_j}^{-1}(M-\Delta_+)(\text{diag}(s_a)+\text{diag}(d_a)), \tag{2.1.22}$$

where

$$\mathbf{A}_{\mathbf{S}_j} = \frac{1}{\sqrt{2}}(\Delta_+\text{diag}(s_a)\Delta_- - M\text{diag}(s_a)M^* + \Delta_+\text{diag}(d_a)M^* - M\text{diag}(d_a)\Delta_-. \tag{2.1.23}$$

The analysis in [17] does not make it clear if in general the homogenized coefficient matrix H has a form more particular than that given by (2.1.22). Numerical experiments in [17] indicate that H has a faster rate of decay in the entries away from the diagonal than the matrix $\mathbf{R}_{\mathbf{S}_j}$, which may have quite slow decay. When H has this property, it may be truncated and compressed more effectively than $\mathbf{R}_{\mathbf{S}_j}$. Additionally, since the matrix H is not diagonal, it does not appear that the form given by (2.1.21) may be used to find a recursion relation for H across many scales.

However, we may say more about H if we make some restrictions on the coefficients, as is done in [17]. The analysis of [17] yields:

Proposition 2.1.2 *Suppose $a(x) = \bar{a} + \tilde{a}(x)$, where $\tilde{a}(x) \in \mathbf{W}_{j+1}$, the finest wavelet space in the domain of $\mathbf{S}_j$, and $\tilde{a}(x)$ has constant amplitude such that $|\tilde{a}(x)| < \bar{a}$. Then, for any function $v(x) \in L^2([0,1])$ such that $v(x)$ has a continuous and bounded fourth derivative, we have*

$$\|\mathbf{R}_{\mathbf{S}_j}(\mathbf{P}_j v(x)) - \alpha \frac{1}{h_j^2}\Delta_+\Delta_-(\mathbf{P}_j v(x))\|_\infty \leq C h_j^2 \|v^{(4)}(x)\|_\infty, \tag{2.1.24}$$

where $\alpha = \langle a^{-1}\rangle^{-1}$ is the harmonic average of $a(x)$ on $[0,1]$.

We do not prove this proposition here; for the complete proof, see [17].

The basic idea of this proposition is that, for highly oscillatory coefficients which are resolved only by the finest scale of $\mathbf{V}_j$, the reduction procedure applied to projections of operators of the form $-\frac{d}{dx}(a(x)\frac{d}{dx})$ yields the same result as classical homogenization, plus a small perturbation term. (Given the results of Section 1.3, this is not unexpected.) We note that this is the case even if the oscillatory part of $a(x)$ is set to zero (i.e. $a(x)$ is a constant). This corresponds with our analysis of the previous section for purely differential operators.

Of perhaps more interest in the analysis of [17] is the observation that the divergence form in two dimensions is preserved as well. In particular, if we have the operator $\mathbf{S}_j$ on $\mathbf{V}_j$ defined by

$$\mathbf{S}_j = \frac{1}{(h_j)^2}(\Delta_+^x A^{(11)}\Delta_-^x + \Delta_+^x A^{(21)}\Delta_- + \Delta_+^y A^{(21)}\Delta_-^x + \Delta_+^y A^{(22)}\Delta_-^y) \tag{2.1.25}$$

where Δ_+^x, Δ_-^x, Δ_+^y, and Δ_-^y are the forward and backward difference operators in the x and

y directions with stencils $(-1, 1)$ and $(1, -1)$, then the reduced operator $\mathbf{R}_{\mathbf{S}_j}$ has the form

$$\mathbf{R}_{\mathbf{S}_j} = \frac{1}{(h_{j+1})^2} \Delta_+^x H^{(11)} \Delta_-^x + \Delta_+^x H^{(21)} \Delta_- + \Delta_+^y H^{(21)} \Delta_-^x + \Delta_+^y H^{(22)} \Delta_-^y). \tag{2.1.26}$$

The matrices $H^{(ij)}$ are the homogenized coefficient matrices. No comment is made in [17] on the form or structure of these matrices, unlike for the one-dimensional case. It is important that a form such as (2.1.26) may be extracted from $\mathbf{R}_{\mathbf{S}_j}$, where $\mathbf{S}_j$ is in the discrete divergence form of (2.1.25). However, without insight into properties of the matrices $H^{(ij)}$, this form may not be exploited to yield a recursion relation over many scales. Additionally, the proof in [17] that this form is preserved relies heavily on special properties of Δ_+ and Δ_- in the Haar basis.

It does not appear that forms such as (2.1.25) are preserved when the reduction procedure is performed using higher order wavelets. Even for Haar wavelets, the form does not give insight into any other properties which may be preserved over more than one scale of reduction. In the next section of this chapter, we investigate which properies of such operators are preserved under reduction, and how the order of wavelets used in the reduction procedure influences these properties.

2.2 Multiresolution Reduction of Elliptic Equations Using High Order Wavelets

The use of high order wavelets to perform multiresolution reduction is desirable for two distinct reasons which we will explore in this section, namely, the sparsity of reduced operators and the preservation of small eigenvalues.

We show that, under the reduction procedure, the rate of the off-diagonal decay of the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks of the reduced operators remains the same. Also, the spectral bounds are preserved as well as (approximately) small eigenvalues and the corresponding eigenvectors. We introduce a modified reduction procedure which better approximates the small eigenvalues. The accuracy of the approximation of small eigenvalues as well as the number of eigenvalues which are preserved with a given accuracy strongly depends on the order of wavelets (and some other properties of the basis).

We briefly consider computational issues since, in (1.2.9), computing the matrix $\mathbf{A}_{S_j}^{-1}$ may appear to present some computational difficulty. Using an algorithm from [20], the operator $\mathbf{R}_{S_j}$ may be computed without computing $\mathbf{A}_{S_j}^{-1}$ directly.

2.2.1 Preservation of Spectral Bounds

An important observation made in [16] is that the reduction procedure preserves the lower bound of the spectrum. The proof is very simple and we present a slightly more general result here, using some relations from [31]. This result is well-known in the field of domain-decomposition methods, where the Schur complement plays a prominent role.

Theorem 2.2.1 (Preservation of spectral bounds) *Let $\mathbf{S}_j$ be a self-adjoint positive-definite operator on $\mathbf{V}_j$,*

$$m\,\|x\|^2 \leq (\mathbf{S}_j x, x) \leq M\,\|x\|^2, \tag{2.2.1}$$

for all $x \in \mathbf{V}_j$, where $0 < m \leq M \leq \infty$.

Then

$$\mathbf{R}_{\mathbf{S}_j} = \mathbf{R}_{\mathbf{S}_j}^* \tag{2.2.2}$$

and

$$m\,\|x\|^2 \leq (\mathbf{R}_{\mathbf{S}_j} x, x) \leq M\,\|x\|^2, \tag{2.2.3}$$

for all $x \in \mathbf{V}_{j+1}$.

Proof: Note that, using (A.2.2), we can write

$$\mathbf{C}_{\mathbf{S}_j} = \mathbf{P}_{j+1}\mathbf{S}_j\mathbf{Q}_{j+1} = (\mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1})^* = \mathbf{B}_{\mathbf{S}_j}^*, \tag{2.2.4}$$

$$\mathbf{T}_{\mathbf{S}_j} = \mathbf{P}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1} = (\mathbf{P}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1})^* = \mathbf{T}_{\mathbf{S}_j}^*, \tag{2.2.5}$$

and

$$\mathbf{A}_{\mathbf{S}_j} = \mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{Q}_{j+1} = (\mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{Q}_{j+1})^* = \mathbf{A}_{\mathbf{S}_j}^*. \tag{2.2.6}$$

Therefore, we have

$$\mathbf{R}_{\mathbf{S}_j}^* = \mathbf{T}_{\mathbf{S}_j}^* - (\mathbf{B}_{\mathbf{S}_j}^*\mathbf{A}_{\mathbf{S}_j}^{-1}\mathbf{B}_{\mathbf{S}_j})^* = \mathbf{T}_{\mathbf{S}_j} - \mathbf{B}_{\mathbf{S}_j}^*\mathbf{A}_{\mathbf{S}_j}^{-1}\mathbf{B}_{\mathbf{S}_j} = \mathbf{R}_{\mathbf{S}_j}. \tag{2.2.7}$$

Since $\mathbf{S}_j$ is positive definite, so is $\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix}$ and, thus, it follows that $\mathbf{A}_{\mathbf{S}_j}$ is positive definite and $\mathbf{A}^{-1}_{\mathbf{S}_j}$ exists. Let us consider the operator

$$\mathbf{Z} = \begin{pmatrix} \mathbf{I} & -\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{B}_{\mathbf{S}_j} \\ 0 & \mathbf{I} \end{pmatrix}.$$

Then we have

$$\mathbf{Z}^* \begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} \mathbf{Z} = \begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & 0 \\ 0 & \mathbf{R}_{\mathbf{S}_j} \end{pmatrix},$$

and

$$(\mathbf{R}_{\mathbf{S}_j} x, x) = \left(\mathbf{Z}^* \begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} \mathbf{Z} \begin{pmatrix} 0 \\ x \end{pmatrix}, \begin{pmatrix} 0 \\ x \end{pmatrix} \right).$$

For the lower bound we obtain

$$\begin{aligned} (\mathbf{R}_{\mathbf{S}_j} x, x) &= \left(\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} \mathbf{Z} \begin{pmatrix} 0 \\ x \end{pmatrix}, \mathbf{Z} \begin{pmatrix} 0 \\ x \end{pmatrix} \right) \\ &\geq m(\|\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{B}_{\mathbf{S}_j} x\|^2 + \|x\|^2) \\ &\geq m\|x\|^2. \end{aligned} \tag{2.2.8}$$

If $M < \infty$, then to estimate the upper bound, we use $\mathbf{R}_{\mathbf{S}_j} + \mathbf{B}^*_{\mathbf{S}_j}\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{B}_{\mathbf{S}_j} = \mathbf{T}_{\mathbf{S}_j}$ and positive definiteness of $\mathbf{A}^{-1}_{\mathbf{S}_j}$ to obtain

$$(\mathbf{R}_{\mathbf{S}_j} x, x) \leq (\mathbf{T}_{\mathbf{S}_j} x, x).$$

Since $\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix}$ satisfies the same spectral bounds as $\mathbf{S}_j$, we have

$$(\mathbf{T}_{\mathbf{S}_j} x, x) \leq M\,\|x\|^2.$$

This completes the proof.□

We note that since we have made no assumptions (other than orthogonality) about the multiresolution analysis, the properties (2.2.3) and (2.2.2) do not depend on dimension or the choice of wavelet basis.

The estimate of (2.2.3) raises the important question of whether it is possible (and under which conditions) to have exactly or approximately the lower eigenvalues of $\mathbf{S}_j$ as eigenvalues of $\mathbf{R}_{\mathbf{S}_j}$. We will consider these questions in Section 2.2.4 below.

2.2.2 Rate of Off-diagonal Decay and Sparsity of Reduced Operators

In this section, we show that, for matrices whose $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks have a certain rate of decay in the magnitude of their entries away from the diagonal, the reduction scheme preserves the rate of the off-diagonal decay in the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks of the reduced matrix over any finite number of scales. This rate is affected by the number of vanishing moments of the wavelet function.

As shown in [9], the elliptic operators considered in this chapter (and their Green's functions) are compressible in wavelet bases. Recall that by compressible we mean that the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks of its non-standard form have fast decay away in the magnitude of their entries away from the diagonal. In fact the result of this subsection is valid for projections of operators from a wider class than just elliptic operators.

We represent the operators $\mathbf{A}_j, \mathbf{B}_j, \mathbf{C}_j, \mathbf{T}_j$ by the matrices $\alpha^j, \beta^j, \gamma^j, s^j$, where

$$\alpha^j_{k,k'} = \int\int K(x,y)\psi_{j,k}(x)\psi_{j,k'}(y)dxdy, \tag{2.2.9}$$

$$\beta^j_{k,k'} = \int\int K(x,y)\psi_{j,k}(x)\phi_{j,k'}(y)dxdy, \tag{2.2.10}$$

$$\gamma^j_{k,k'} = \int\int K(x,y)\phi_{j,k}(x)\psi_{j,k'}(y)dxdy, \tag{2.2.11}$$

$$s^j_{k,k'} = \int\int K(x,y)\phi_{j,k}(x)\phi_{j,k'}(y)dxdy, \tag{2.2.12}$$

and $K(x,y)$ is the kernel of a Calderón-Zygmund or a pseudo-differential operator $\mathbf{S}$. We assume that K satisfies the conditions

$$|K(x,y)| \leq \frac{1}{|x-y|}, \tag{2.2.13}$$

$$|\partial_x^M K(x,y)| + |\partial_y^M K(x,y)| \leq \frac{C_0}{|x-y|^{1+M}}. \tag{2.2.14}$$

We also assume either that the kernel K defines a bounded operator on L^2, or that, if it defines an unbounded operator, then the kernel K satisfies a weaker condition (called the "weak cancellation condition")

$$\left| \int_{I \times I} K(x,y)\,dxdy \right| \leq C|I|, \tag{2.2.15}$$

for all dyadic intervals I. Under these conditions, we have (see [9])

Theorem 2.2.2 *If the wavelet basis has M vanishing moments, then, for any kernel K satisfying the conditions (2.2.13), (2.2.14), and (2.2.15), the matrices α^j, β^j, γ^j satisfy the estimate*

$$|\alpha^j_{k,l}| + |\beta^j_{k,l}| + |\gamma^j_{k,l}| \leq C^j_M \ (1 + |k-l|)^{-M-1}, \tag{2.2.16}$$

for all integer k, l.

This theorem has a straightforward higher dimensional analogue.

Bi-infinite matrices $\{m_{kl}\}_{k,l \in \mathbf{Z}}$ which satisfy estimates of the form (2.2.16) fit into the more general class of matrices which decay away from the diagonal according to the estimate

$$|m_{kl}| < C\,(1 + |k-l|)^{-r}, \tag{2.2.17}$$

where $r > 1$ is a parameter and C is a constant. The following elegant theorem dealing with the algebra of invertible matrices $\{m_{kl}\}_{k,l \in \mathbf{Z}}$ has been communicated to us by Ph. Tchamitchian. This theorem is an enhancement of the result presented in [37] (following [21]).

Theorem 2.2.3 *If the matrix $\{m_{kl}\}_{k,l \in \mathbf{Z}}$ is invertible on l^2, then*

$$|m^{-1}_{k,l}| < C'(1 + |k-l|)^{-r}. \tag{2.2.18}$$

The proof uses relations between commutators of an unbounded operator X on l^2 defined by $X(y_k) = \{k\,y_k\}$ and operators $M = \{m_{k,l}\}_{k,l \in \mathbf{Z}}$ and $M^{-1} = \{m^{-1}_{k,l}\}_{k,l \in \mathbf{Z}}$; it is quite elaborate and we refer to [37] for details.

We prove a two dimensional analogue of this theorem in Chapter 3. Although in higher dimensions, the object $\mathbf{S}_j$ and its blocks are actually tensors, we refer to them as matrices. We state it here:

Theorem 2.2.4 *If a matrix* $\{m_{k,k',l,l'}\}_{k,k',l,l'\in\mathbf{Z}}$ *satisfies*

$$|m_{k,k',l,l'}| < C(1+|k-k'|+|l-l'|)^{-2-\alpha} \tag{2.2.19}$$

(where $\alpha \in \mathbf{Z}, \alpha \geq 2$*) and if the matrix is invertible on* l^2*, then*

$$|m^{-1}_{k,k',l,l'}| < C''(1+|k-k'|+|l-l'|)^{-2-\alpha}. \tag{2.2.20}$$

See Figure 2.1 for an example of such a matrix after truncation of elements below a given threshold. Matrices which satisfy (2.2.19) also form an algebra under multiplication; for a proof of this see Chapter 3.

We use Theorems 2.2.3 and 2.2.4 to show that, at all stages of the reduction procedure in both one and two dimensions, the matrices representing the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks of the reduced operators (1.2.10) satisfy the same off-diagonal decay estimate (2.2.16) as the blocks of the non-standard form in Theorem 2.2.2 and its two-dimensional analogue. In other words, the reduction procedure preserves sparsity for a wide class of operators. In this sense, the form (or structure) is preserved under the reduction procedure which allows us to apply it over a finite number of scales. The following theorem applies to the one-dimensional case, but analagous results for two dimensions can be proved using Theorem 2.2.4.

Theorem 2.2.5 (Preservation of structure over finitely many scales) *Assume that the operator* $\mathbf{S}$ *and the wavelet basis satisfy the conditions of Theorem 2.2.2. Let* $\mathbf{R}_j$ *be the reduced operator on some scale* j*, where reduction started at some scale* n, $n \leq j$, $n, j \in \mathbf{Z}$, *and let* $\mathbf{A}_{\mathbf{R}_j}$, $\mathbf{B}_{\mathbf{R}_j}$ *and* $\mathbf{C}_{\mathbf{R}_j}$ *be its blocks. Then the bi-infinite matrices* $\alpha^{r,j}$, $\beta^{r,j}$ *and* $\gamma^{r,j}$ *representing these blocks satisfy*

$$|\alpha^{r,j}_{k,l}| + |\beta^{r,j}_{k,l}| + |\gamma^{r,j}_{k,l}| \leq C^{n,j}_M \; (1+|k-l|)^{-M-1}, \tag{2.2.21}$$

for all integers k, l.

Proof: Our starting point is the operator $\mathbf{S}_n$ and its blocks, $\mathbf{A}_{\mathbf{S}_n}$, $\mathbf{B}_{\mathbf{S}_n}$, $\mathbf{C}_{\mathbf{S}_n}$ and $\mathbf{T}_{\mathbf{S}_n} = \mathbf{S}_{n+1}$. Matrices representing these blocks satisfy the estimate of Theorem 2.2.2. Since $\mathbf{S}_n$ is positive definite, so is $\mathbf{A}_{\mathbf{S}_n}$ (see Section 2.2.1) and, thus, $\mathbf{A}^{-1}_{\mathbf{S}_n}$ exists and, according to Theorem 2.2.3, satisfies the estimate in (2.2.16). Since $\mathbf{B}_{\mathbf{S}_n}$ and $\mathbf{C}_{\mathbf{S}_n}$ satisfy the same estimate (2.2.16), the product $\mathbf{C}_{\mathbf{S}_n}\,\mathbf{A}_{\mathbf{S}_n}^{-1}\,\mathbf{B}_{\mathbf{S}_n}$ satisfies it as well. The reduced operator $\mathbf{R}_{n+1}$,

$$\mathbf{R}_{n+1} = \mathbf{R}_{\mathbf{S}_n} = \mathbf{T}_{\mathbf{S}_n} - \mathbf{C}_{\mathbf{S}_n}\,\mathbf{A}_{\mathbf{S}_n}^{-1}\,\mathbf{B}_{\mathbf{S}_n}, \tag{2.2.22}$$

consists of the difference of two terms,

$$\mathbf{R}_{n+1} = \mathbf{S}_{n+1} - \mathbf{F}_{n+1}, \tag{2.2.23}$$

where

$$\mathbf{F}_{n+1} = \mathbf{C}_{\mathbf{S}_n}\,\mathbf{A}_{\mathbf{S}_n}^{-1}\,\mathbf{B}_{\mathbf{S}_n}. \tag{2.2.24}$$

The operator $\mathbf{S}_{n+1}$ is the projection on the scale $n+1$ of the operator $\mathbf{S}$ and the operator $\mathbf{F}_{n+1}$ has fast decay and satisfies the estimate (2.2.16). The blocks $\mathbf{A}_{\mathbf{R}_{n+1}}$, $\mathbf{B}_{\mathbf{R}_{n+1}}$, $\mathbf{C}_{\mathbf{R}_{n+1}}$, and $\mathbf{T}_{\mathbf{R}_{n+1}}$ of the operator $\mathbf{R}_{n+1}$ may be written as a difference of the corresponding blocks of these two terms. Theorem 2.2.2 guarantees that the contribution from $\mathbf{S}_{n+1}$ has the proper decay. On the other hand, the contributions from $\mathbf{F}_{n+1}$ have at least the same rate of decay as $\mathbf{F}_{n+1}$ itself since the blocks are obtained by a wavelet transform.

We prove Theorem 2.2.5 by induction, assuming that on some scale j we have

$$\mathbf{R}_j = \mathbf{S}_j - \mathbf{F}_j, \tag{2.2.25}$$

where $\mathbf{S}_j$ is the projection on the scale j of the kernel K and $\mathbf{F}_j$ satisfies the estimate in (2.2.16). The induction step is a repeat of the considerations above with the additional use of Theorem 2.2.1 (preservation of spectral bounds) in order to ensure the invertibility of the $\mathbf{A}_{\mathbf{R}_j}$ block.□

Remark 1. There are (more narrow) classes of operators for which the rate of the off-diagonal decay is faster than that in Theorems 2.2.2 and 2.2.5. For example, if we consider strictly elliptic pseudo-differential operators of order n with symbols satisfying

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x,\xi)| \le C(\alpha,\beta)|\xi|^{n-|\alpha|+|\beta|},$$

then the rate of the off-diagonal decay is faster than that in (2.2.16), namely (using a wavelet basis with all vanishing moments),

$$|\alpha_{k,l}^j| + |\beta_{k,l}^j| + |\gamma_{k,l}^j| \le C_m\, 2^{nj}\,(1+|k-l|)^{-m}, \tag{2.2.26}$$

for all integer k, l and m. Matrices $\{m_{kl}\}_{k,l\in\mathbf{Z}}$ which are invertible on l^2 and which satisfy for all integer m the inequality

$$|m_{k,l}^{-1}| < \tilde{C}_m(1+|k-l|)^{-m}, \tag{2.2.27}$$

form an algebra (see [37]). We may thus repeat the above considerations to prove a version of Theorem 2.2.5 with the decay condition replaced by a decay condition of the form of (2.2.27).

Remark 2. It is clear that Theorem 2.2.5 may be viewed as a combination of the results of Tchamitchian [37] and Beylkin-Coifman-Rokhlin [9]. We conjecture that (for a narrower class of operators) this theorem can be extended to reduction over an infinite number of scales, thus showing that the constant $C_M^{n,j}$ does not depend on $n-j$. Such an extension requires more precise estimates to replace (2.2.18) and (2.2.20). In particular, in both estimates, the constants on the right-hand side need to be bounded more precisely; currently, they are merely shown to be finite.

2.2.3 A Fast Method for Computing the Reduced Operator

In practical application of the reduction procedure (1.2.9), one of the critical issues is the cost of computing the reduced operator (1.2.10). The sparsity of the operators involved in the reduction is assured by Theorem 2.2.5, but an algorithm for computing the reduced matrix is still needed. It turns out that a multiresolution LU decomposition algorithm may be used for this purpose [20]. The multiresolution LU decomposition is performed with respect to the product of non-standard forms rather than the ordinary matrix product. It has complexity $\mathcal{O}(N)$ for a fixed relative error ϵ and provides a direct solver for linear systems written using the non-standard form.

The algorithm in [20] provides an alternative to the computation of $\mathbf{A}_{\mathbf{S}_j}^{-1}$ by noting that the decomposition

$$\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{C}_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{A}}_{\mathbf{S}_j} & 0 \\ \hat{\mathbf{C}}_{\mathbf{S}_j} & \hat{\mathbf{T}}_{\mathbf{S}_j} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{A}}_{\mathbf{S}_j} & \tilde{\mathbf{B}}_{\mathbf{S}_j} \\ 0 & \tilde{\mathbf{T}}_{\mathbf{S}_j} \end{pmatrix} \tag{2.2.28}$$

implies that

$$\mathbf{R}_{\mathbf{S}_j} = \mathbf{T}_{\mathbf{S}_j} + \hat{\mathbf{C}}_{\mathbf{S}_j}\tilde{\mathbf{B}}_{\mathbf{S}_j}. \tag{2.2.29}$$

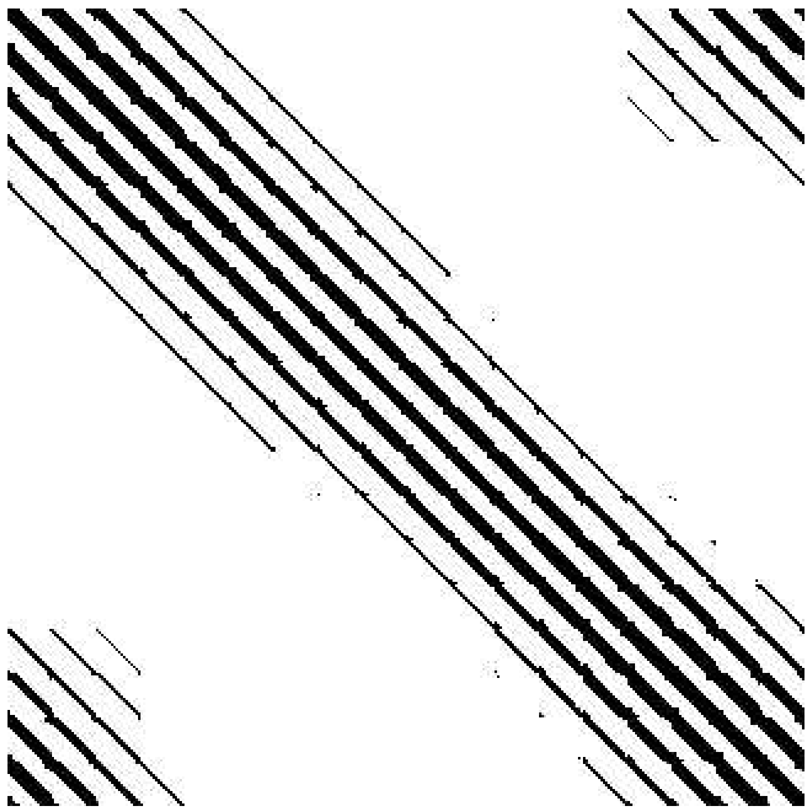

Figure 2.1: $\mathbf{R}_{\mathbf{S}_j}$ after truncation of entries smaller than $0.02 * \|\mathbf{R}_{\mathbf{S}_j}\|_\infty$. $\mathbf{S}_j$ is the projection of $-\nabla \cdot (a(x,y)\nabla)$ on the unit square with periodic boundary conditions into the multiwavelet basis with two vanishing moments. The space $\mathbf{V}_j$ has 5184 total unknowns.

In the one-dimensional case, if $\mathbf{A}_{\mathbf{S}_j}$ is banded with bandwidth m, then its LU-factors will also be banded with bandwidth m, so they may be computed in $\mathcal{O}(Nm^2)$. If $\mathbf{B}_{\mathbf{S}_j}$ is also banded with this same bandwidth, then we may solve for $\tilde{\mathbf{B}}_{\mathbf{S}_j}$ in $\mathcal{O}(Nm^2)$; and similarly for $\hat{\mathbf{C}}_{\mathbf{S}_j}$. For fixed relative accuracy ϵ (and, hence, fixed bandwidth m), this leads directly to the $\mathcal{O}(N)$ procedure for computing $\mathbf{R}_{\mathbf{S}_j}$ via the sparse incomplete block LU decomposition given by (2.2.28).

The two-dimensional case is more complicated. Each of the blocks on the left-hand side of (2.2.28) will in general exhibit a multi-banded structure, as seen in Figure 2.1. Thus, one may expect the LU-factors of $\mathbf{A}_{\mathbf{S}_j}$ to fill in between the bands. Indeed, this is the case, but the fill-in which occurs is observed in practice to be fill-in with rapid decay, so that truncating to a given accuracy as we compute the LU factors results in a fast method for computing the reduction (as in the one-dimensional case).

There are many details involved in the description of the multiresolution LU decomposition, and we refer to [20] for a full treatment of them. We note finally that, by virtue of this algorithm the reduction procedure requires $\mathcal{O}(N)$ operations.

2.2.4 Eigenvalues and Eigenvectors of the Reduced Operators

In this section, we further investigate the relations between the spectra of the operators $\mathbf{S}_j$ and $\mathbf{R}_{\mathbf{S}_j}$. In Section 2.2.1, we established relations between the spectral bounds of these operators. Here, we consider relations between the small eigenvalues and corresponding eigenvectors of the operators $\mathbf{S}_j$ and $\mathbf{R}_{\mathbf{S}_j}$.

We view $\mathbf{S}_j$ as the projection of a positive definite self-adjoint elliptic operators with a compact inverse; this class includes variable-coefficient elliptic operators. For such an operator $\mathbf{S}$, the spectrum consists of isolated eigenvalues with finite multiplicity and the only accumulating point is at infinity. The eigenvalues may be ordered according to

$$0 < \lambda_0 \leq \lambda_1 \leq \lambda_2 \ldots .$$

The eigenvectors of such operators form an orthonormal basis in the Hilbert space $\mathcal{H}$, and each eigenspace is a finite-dimensional subspace. Heuristically, e.g. in numerical literature, it is always assumed for elliptic operators that the eigenvectors which correspond to small eigenvalues are less oscillatory than those which correspond to large eigenvalues and the number of oscillations increases as $\lambda_n \to \infty$. For example, such statements typically form the basis for the heuristic justification of multigrid methods. There are many other examples of theorems where this property is a subject of consideration, see e.g. [23]. Here we formulate a simple, general proposition capturing this property for the purposes of this chapter.

Definition. Let $\mathcal{S}$ be a subspace of the Hilbert space $\mathcal{H}$. We will say that the subspace $\mathbf{V}_n$ of MRA is an ϵ-approximating subspace for $\mathcal{S}$ if any function in $\mathcal{S}$ may be approximated by functions from $\mathbf{V}_n$ with relative error ϵ.

We denote by $\mathcal{S}_l$ the span of eigenvectors of $\mathbf{T}$ which correspond to all eigenvalues λ_k, $\lambda_k \leq \lambda_l$. Clearly,

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \mathcal{S}_2 \subset \ldots$$

Proposition 2.2.1 *For any ϵ there exists a monotone sequence $k_l \geq 0$, $k_l \in \mathbf{Z}$, such that the subspaces $\mathbf{V}_{k_l}$ of the MRA,*

$$\mathbf{V}_{k_0} \subset \mathbf{V}_{k_1} \subset \mathbf{V}_{k_2} \subset \ldots$$

are each ϵ-approximating subspaces for $\mathcal{S}_l$.

Proof: The proof of this proposition is straightforward; since each $\mathcal{S}_l$ is finite-dimensional, we need only approximate each function in the basis of $\mathcal{S}_l$ by a function in some (sufficiently fine) space $\mathbf{V}_n$ to accuracy ϵ. Since there are finitely many basis functions we may choose a finest $\mathbf{V}_n$ to approximate all of them with relative error ϵ.$\Box$

This proposition captures the relationship between larger eigenvalues and more oscillatory eigenfunctions in the sense that the eigenfunctions corresponding to larger eigenvalues may require a finer space in the multiresolution analysis to be resolved to accuracy ϵ.

In [28] a similar but stronger statement is made for a narrower class of operators; in e.g. [23] this topic is approached in terms of nodal lines of eigenfunctions.

As stated above, the proposition is quite meaningless for practical purposes. By choosing a fine enough scale, we always may use the MRA to approximate any finite-dimensional space to any accuracy ϵ. The only point of the proposition is that the MRA may be used to approximate the eigenspaces in a natural sequence, proceeding from less oscillatory to more oscillatory. For practical purposes, however, we have to construct the MRA very carefully if we want to achieve this property for the first few scales that are involved. For example, it is clear that in order to have good approximating properties, the basis functions of the MRA have to satisfy the boundary conditions. For the same reasons, in choosing an MRA for equations where the coefficients have singularities, it makes good sense to incorporate appropriate singularities into the basis.

We now illustrate our approach by a simple example. Suppose that $\lambda > 0$ is an eigenvalue and x an eigenvector of the self-adjoint positive definite operator $\mathbf{S}_j$, $x \in \mathbf{V}_j$ and $\mathbf{Q}_{j+1}x = 0$ (in other words, $x \in \mathbf{V}_{j+1}$). Then we have

$$\mathbf{S}_j x = \lambda x, \tag{2.2.30}$$

and $\mathbf{Q}_{j+1}x = 0$ implies that $\mathbf{P}_{j+1}x = x$, so that

$$\mathbf{T}_{\mathbf{S}_j} x = \mathbf{P}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1}x = \mathbf{P}_{j+1}\mathbf{S}_j x = \mathbf{P}_{j+1}\lambda x = \lambda x \tag{2.2.31}$$

and

$$\begin{aligned} \mathbf{R}_{\mathbf{S}_j} x &= \mathbf{T}_{\mathbf{S}_j} x - \mathbf{B}^*_{\mathbf{S}_j}\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{B}_{\mathbf{S}_j} x \\ &= \lambda x - \mathbf{B}^*_{\mathbf{S}_j}\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1}x \\ &= \lambda x - \lambda\mathbf{B}^*_{\mathbf{S}_j}\mathbf{A}^{-1}_{\mathbf{S}_j}\mathbf{Q}_{j+1}x \\ &= \lambda x. \end{aligned} \tag{2.2.32}$$

In other words, eigenvectors of $\mathbf{S}_j$ which are exactly represented on the subspace $\mathbf{V}_{j+1}$ will be preserved (with the same eigenvalue) under the reduction step.

The condition $\mathbf{Q}_{j+1}x = 0$ is certainly too stringent for a general elliptic operator. However, the ϵ-approximating property of the MRA guarantees that we can attain $\|\mathbf{Q}_{j+1}x\| < \epsilon$ if we consider the eigenvalue problem on a fine enough scale. If we accept that eigenvectors x corresponding to small eigenvalues are not very oscillatory, then the ϵ-approximating property may be achieved by a relatively coarse scale in the MRA. More precisely, we will show if the MRA is chosen so that a set of eigenvectors may be well approximated at some coarse scale, then, up to that scale, the eigenvalues corresponding to these eigenvectors will not be significantly affected by the reduction procedure.

Given the eigenvalue problem

$$\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{B}^*_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} = \lambda \begin{pmatrix} d \\ s \end{pmatrix}, \tag{2.2.33}$$

we use the same approach as in deriving (1.2.9). Solving for d in terms of s and assuming that $(\mathbf{A}_{\mathbf{S}_j} - \lambda\mathbf{I})^{-1}$ exists, we obtain

$$(\mathbf{T}_{\mathbf{S}_j} - \mathbf{B}^*_{\mathbf{S}_j}(\mathbf{A}_{\mathbf{S}_j} - \lambda\mathbf{I})^{-1}\mathbf{B}_{\mathbf{S}_j})s = \lambda s. \tag{2.2.34}$$

The existence of

$$\mathbf{G}(\lambda) \;=\; (\mathbf{A}_{\mathbf{S}_j} - \lambda\mathbf{I})^{-1} \tag{2.2.35}$$

is assured if we consider (2.2.34) for λ smaller than the lower bound of $\mathbf{A}_{\mathbf{S}_j}$.

We now consider approximations of the left-hand side of (2.2.34) and the accuracy of solutions based on these approximations. We will use the following simple lemma.

Lemma 2.2.1 *For a normal matrix* $\mathbf{M}$*, if*

$$\mathbf{M}x = \lambda x + \xi, \tag{2.2.36}$$

then there exists an eigenvalue $\lambda_{\mathbf{M}}$ *of* $\mathbf{M}$ *such that*

$$|\lambda - \lambda_{\mathbf{M}}| \leq \frac{\|\xi\|}{\|x\|}. \tag{2.2.37}$$

Proof: The proof is straightforward. Let $\mathbf{G} = \mathbf{M} - \lambda\mathbf{I}$. Then there is a singular value σ_0 of $\mathbf{G}$ such that

$$\sigma_0 = \inf_{\|y\|\neq 0} \frac{(\mathbf{G}^*\mathbf{G}y, y)^{\frac{1}{2}}}{\|y\|} \leq \frac{\|\mathbf{G}x\|}{\|x\|} = \frac{\|\xi\|}{\|x\|}. \tag{2.2.38}$$

Since $\mathbf{G}$ is normal, it is diagonalizable by a unitary matrix $\mathbf{Q}$. Therefore, the singular values of $\mathbf{G}$ are given by the absolute values of its eigenvalues. Since at least one singular value of $\mathbf{G}$ satisfies (2.2.38), the estimate (2.2.37) follows.□

From (2.2.33), it is clear that

$$d = -\mathbf{G}(\lambda)\mathbf{B}_{\mathbf{S}_j}s. \tag{2.2.39}$$

We rewrite (2.2.34) as

$$\mathbf{T}_{\mathbf{S}_j}\, s = \lambda\, s + \mathbf{B}^*_{\mathbf{S}_j}d. \tag{2.2.40}$$

Using

$$\mathbf{G}(\lambda) - \mathbf{G}(0) = \lambda \mathbf{G}(\lambda)\mathbf{G}(0), \tag{2.2.41}$$

where $\mathbf{G}(0) = \mathbf{A}_{\mathbf{S}_j}^{-1}$, we obtain from (2.2.40) that

$$(\mathbf{T}_{\mathbf{S}_j} - \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-1}\mathbf{B}_{\mathbf{S}_j})s = \lambda\, s + \lambda\, \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-1}d, \tag{2.2.42}$$

or

$$\mathbf{R}_{\mathbf{S}_j} s = \lambda\, s + \lambda\, \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-1}d. \tag{2.2.43}$$

Applying (2.2.41) one more time, we obtain from (2.2.43) that

$$\mathbf{R}_{\mathbf{S}_j} s = \lambda \left(\mathbf{I} + \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-2}\mathbf{B}_{\mathbf{S}_j}\right) s + \lambda^2\, \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-2}d. \tag{2.2.44}$$

We approximate the eigenvalue problems in (2.2.40), (2.2.43), and (2.2.44) by

$$\mathbf{T}_{\mathbf{S}_j} s = \lambda\, s, \tag{2.2.45}$$

$$\mathbf{R}_{\mathbf{S}_j} s = \lambda\, s, \tag{2.2.46}$$

and

$$\mathbf{R}_{\mathbf{S}_j} s = \lambda \left(\mathbf{I} + \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-2}\mathbf{B}_{\mathbf{S}_j}\right) s. \tag{2.2.47}$$

The last equation gives rise to what we call the modified reduction procedure. As in the case (2.2.46), we would like to iterate the modified reduction procedure over many scales for (2.2.47) so that this form is preserved. To this end, we factor the operator $\mathbf{I} + \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-2}\mathbf{B}_{\mathbf{S}_j}$ by using the Cholesky decomposition and obtaining

$$\mathbf{I} + \mathbf{B}_{\mathbf{S}_j}^{*}\mathbf{A}_{\mathbf{S}_j}^{-2}\mathbf{B}_{\mathbf{S}_j} = \mathbf{L}_{\mathbf{S}_j}\mathbf{L}_{\mathbf{S}_j}^{*}. \tag{2.2.48}$$

We rewrite (2.2.47) as

$$\mathbf{L}_{\mathbf{S}_j}^{-1}\mathbf{R}_{\mathbf{S}_j}(\mathbf{L}_{\mathbf{S}_j}^{*})^{-1} z = \lambda z \tag{2.2.49}$$

where

$$z = \mathbf{L}_{\mathbf{S}_j}^{*} s, \tag{2.2.50}$$

and define

$$\mathbf{Y}_{\mathbf{S}_j} = \mathbf{L}_{\mathbf{S}_j}^{-1}\mathbf{R}_{\mathbf{S}_j}(\mathbf{L}_{\mathbf{S}_j}^{*})^{-1}. \tag{2.2.51}$$

The equations (2.2.48), (2.2.50), and (2.2.51) represent the modified reduction procedure. The operator $\mathbf{Y}_{\mathbf{S}_j}$ is self-adjoint and positive definite; to iterate the modified reduction procedure we compute the $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ blocks of the operator $\mathbf{Y}_{\mathbf{S}_j}$ and obtain (2.2.47) on the next scale. Note that, in the modified reduction procedure, we have to keep track of the projections of the eigenvector since at each step they are modified via (2.2.50).

We now use Lemma 2.2.1 to estimate the accuracy of the approximations given by (2.2.45), (2.2.46), and (2.2.47). The lemma allows us to use the size of the neglected terms to bound the perturbation of the eigenvalues. For the term neglected in the approximation of (2.2.40) given by (2.2.45), we have

$$\|\mathbf{B}^*_{\mathbf{S}_j} d\|_2 \leq \|\mathbf{B}_{\mathbf{S}_j}\|_2 \|d\|_2. \tag{2.2.52}$$

By introducing the spectral bounds of the operator $\mathbf{A}_{\mathbf{S}_j}$,

$$m^j_{\mathbf{A}} \|x\|^2 \leq (\mathbf{A}_{\mathbf{S}_j} x, x) \leq M^j_{\mathbf{A}} \|x\|^2,$$

for (2.2.46) we obtain the estimate

$$\|\lambda\, \mathbf{B}^*_{\mathbf{S}_j} \mathbf{A}^{-1}_{\mathbf{S}_j} d\|_2 \leq \frac{\lambda}{m^j_{\mathbf{A}}} \|\mathbf{B}_{\mathbf{S}_j}\|_2 \ \|d\|_2. \tag{2.2.53}$$

For the term neglected in (2.2.47), we follow the above considerations for the modified reduction procedure. After multiplying by $\mathbf{L}^{-1}_{\mathbf{S}_j}$ on the left and substituting (2.2.50) in (2.2.44), we have

$$\mathbf{Y}_{\mathbf{S}_j} z = \lambda z + \lambda^2 \mathbf{L}^{-1}_{\mathbf{S}_j} \mathbf{B}^*_{\mathbf{S}_j} \mathbf{A}^{-2}_{\mathbf{S}_j} d, \tag{2.2.54}$$

to which Lemma 2.2.1 may be applied. Let $\mathbf{Z}_{\mathbf{S}_j} = \mathbf{I} + \mathbf{B}^*_{\mathbf{S}_j} \mathbf{A}^{-2}_{\mathbf{S}_j} \mathbf{B}_{\mathbf{S}_j}$. The lower spectral bound of $\mathbf{Z}_{\mathbf{S}_j}$ is clearly bounded below by one (since $\mathbf{B}^*_{\mathbf{S}_j} \mathbf{A}^{-2}_{\mathbf{S}_j} \mathbf{B}_{\mathbf{S}_j}$ is positive definite), so $\|\mathbf{Z}^{-\frac{1}{2}}_{\mathbf{S}_j}\|_2 \leq 1$. Furthermore, there exists a unitary $\mathbf{Q}$ such that $\mathbf{Z}^{\frac{1}{2}}_{\mathbf{S}_j} = \mathbf{Q}^* \mathbf{L}_{\mathbf{S}_j}$, where $\mathbf{L}_{\mathbf{S}_j}$ is the Cholesky factor of $\mathbf{Z}_{\mathbf{S}_j}$; thus, $\mathbf{L}^{-1}_{\mathbf{S}_j} = \mathbf{Z}^{-\frac{1}{2}}_{\mathbf{S}_j} \mathbf{Q}^*$ and

$$\|\mathbf{L}^{-1}_{\mathbf{S}_j}\|_2 \leq \|\mathbf{Z}^{-\frac{1}{2}}_{\mathbf{S}_j}\|_2 \leq 1. \tag{2.2.55}$$

This yields (from (2.2.54))

$$\|\lambda^2 \mathbf{L}^{-1}_{\mathbf{S}_j} \mathbf{B}^*_{\mathbf{S}_j} \mathbf{A}^{-2}_{\mathbf{S}_j} d\|_2 \leq \left(\frac{\lambda}{m^j_{\mathbf{A}}}\right)^2 \|\mathbf{B}_{\mathbf{S}_j}\|_2 \ \|d\|_2. \tag{2.2.56}$$

Lemma 2.2.1 in conjunction with (2.2.52), (2.2.53), and (2.2.56) yields the following result.

Theorem 2.2.6 *Given an eigenvector x of $\mathbf{S}_j$ such that $\mathbf{S}_j x = \lambda x$, $\|x\|_2 = 1$, $d = \mathbf{Q}_{j+1}x$, and $\|d\|_2^2 << \frac{1}{2}$, there exist real $\lambda_{\mathbf{T}}$, $\lambda_{\mathbf{R}}$, and $\lambda_{\mathbf{Y}}$ which solve (2.2.45), (2.2.46), and (2.2.47), respectively, such that*

$$|\lambda_{\mathbf{T}} - \lambda| \leq C_d \|\mathbf{B}_{\mathbf{S}_j}\|_2 \|d\|_2 \tag{2.2.57}$$

$$|\lambda_{\mathbf{R}} - \lambda| \leq C_d \|\mathbf{B}_{\mathbf{S}_j}\|_2 \|d\|_2 \left(\frac{\lambda}{m^j_{\mathbf{A}}}\right) \tag{2.2.58}$$

$$|\lambda_{\mathbf{Y}} - \lambda| \leq C_d \|\mathbf{B}_{\mathbf{S}_j}\|_2 \|d\|_2 \left(\frac{\lambda}{m^j_{\mathbf{A}}}\right)^2, \tag{2.2.59}$$

where $1 \leq C_d \leq \sqrt{2}$.

We may now identify two factors that affect the estimate: $\|d\|_2$ and the ratio $\lambda/m^j_{\mathbf{A}}$. In order for $\|d\|_2$ to be small, we have to assume that the eigenvector of the problem in (2.2.33) has a small projection on the subspace $\mathbf{W}_{j+1}$. If the subspace $\mathbf{V}_{j+1}$ is ϵ-approximating for the subspace of eigenvectors, then $\|d\|_2 \leq \epsilon$ and the perturbation of the corresponding eigenvalues is small. The eigenvalue problem (2.2.45) is merely the projection of the original eigenvalue problem to the next coarsest scale, and reflects the current practice in setting up eigenvalue problems. The estimate (2.2.52) simply shows that it is safe to project the eigenvalue problem to a coarser scale as long as the eigenvectors are represented on that scale to the desired accuracy.

The reduction and modified reduction procedures improve the eigenvalue estimate with the additional factor $\lambda/m^j_{\mathbf{A}}$. This ratio is small (in a generic situation) since the operator $\mathbf{A}_{\mathbf{S}_j}$ is typically well-conditioned (see Table 2.1) and captures the "high-frequency" component of the operator $\mathbf{S}_j$. Thus, for the lower frequency modes with smaller eigenvalues, we expect that $m^j_{\mathbf{A}} >> \lambda$. We show later in numerical examples that this factor makes a significant difference.

Of great importance is the fact that all of the considerations in this section are *independent* of dimension; the guarantee of the ϵ-approximating property in arbitrary dimensions provides this. However, in higher dimensions, the consideration of optimizing the MRA for a given operator becomes the chief practical difficulty.

Remark. In Section 2.2.3, we outlined an $\mathcal{O}(N)$ procedure for computing the reduced operator to relative accuracy ϵ. For small eigenvalues, however, it might be necessary to maintain absolute rather than relative accuracy while performing the reduction. This puts an additional computational burden on the reduction procedure in the case of ill-conditioned operators.

In particular, if we compute $\hat{\mathbf{A}}_{\mathbf{S}_j}$ and $\hat{\mathbf{C}}_{\mathbf{S}_j} = \tilde{\mathbf{B}}^*_{\mathbf{S}_j}$ to some absolute accuracy δ, and from this compute $\mathbf{R}'_{\mathbf{S}_j}$, it is clear (from (2.2.29)) that

$$\|\mathbf{R}'_{\mathbf{S}_j} - \mathbf{R}_{\mathbf{S}_j}\| < \delta \|\hat{\mathbf{C}}_{\mathbf{S}_j}\|. \tag{2.2.60}$$

In the worst case, the eigenvalues of $\mathbf{R}'_{\mathbf{S}_j}$ will approximate the eigenvalues of $\mathbf{R}_{\mathbf{S}_j}$ with accuracy no better than $\delta\|\hat{\mathbf{C}}_{\mathbf{S}_j}\|$ (see e.g. [25]). For a typical second-order elliptic operator $\mathbf{S}$, the norms of each of the blocks $\mathbf{A}_{\mathbf{S}_j}$ and $\mathbf{C}_{\mathbf{S}_j} = \mathbf{B}^*_{\mathbf{S}_j}$ behave like $\mathcal{O}(h_j^{-2})$ (where h_j is the step size of the discretization). Furthermore, in the Cholesky decomposition, the norm of the lower triangular factor is equal to the square root of the norm of the matrix. Therefore, if we compute the LU factorization defined in Section 2.2.3 to absolute accuracy δ, then the resulting matrix $\mathbf{R}'_{\mathbf{S}_j}$ approximates $\mathbf{R}_{\mathbf{S}_j}$ to absolute accuracy δh_j^{-1}, as can easily be seen from (2.2.60).

In other words, to compute $\mathbf{R}'_{\mathbf{S}_j}$ so that its eigenvalues approximate the small eigenvalues of $\mathbf{R}_{\mathbf{S}_j}$ with absolute accuracy ϵ, it is necessary to compute the multiresolution LU decomposition with working precision $\epsilon{\cdot}h_j$. For a given accuracy δ, the bandwidth m of matrices which satisfy (2.2.16) (or its two-dimensional analogue) is given by $m \sim (C\delta)^{-\frac{1}{M}}$, where M is the number of vanishing moments of the wavelet basis (see e.g. [9] for details). Thus, as h_j decreases (and the scale becomes finer) it is necessary to keep a wider band in the LU decomposition. This thickening of the band as the scale becomes finer means that, for the purposes of eigenvalue computations with fixed absolute accuracy, the reduction procedure is $\mathcal{O}(N^{1+\frac{4}{M}})$ rather than $\mathcal{O}(N)$.

This estimate is obtained if we choose the number of vanishing moments M based on the desired accuracy ϵ. A typical choice is $M \sim -\log(\epsilon)$. With this choice we have the bandwidth $m \sim (\epsilon^{-1}(h_j)^{-2})^{\frac{1}{M}}$. For matrices with bandwidth m in n dimensions (where n=1, 2) the multiresolution LU decomposition requires $\mathcal{O}(Nm^{2n})$ operations. But $h_j = N^{-\frac{1}{n}}$, so we see

Table 2.1: Condition numbers and lower bounds of $\mathbf{A}_{\mathbf{S}_j}$.

N	$\kappa(\mathbf{A}_j)$	$m_{\mathbf{A}_j}$	$\kappa(\mathbf{S}_j)$
256	6.18	1.4577×10^3	1.16×10^2
1024	8.06	5.0363×10^3	6.26×10^2
2304	10.48	1.0043×10^4	1.53×10^3
4096	11.66	1.5948×10^4	2.86×10^3
5184	13.06	1.9077×10^4	3.66×10^3

Table 2.1: Condition numbers and lower bounds for the $\mathbf{A}$-block of the operator $-\nabla \cdot (a(x,y)\nabla)$ on the unit square with periodic boundary conditions. Here, N is the number of unknowns in the two-dimensional spatial grid. Multiwavelets with two vanishing moments are used, and the coefficients $a(x,y)$ are set to $a(x,y) = 2 + \cos(16\pi x)\cos(16\pi y)$, which provides a moderate amount of oscillation in the coefficients. The condition number depends only weakly on the scale, unlike the condition number of the original matrix (denoted in the table as $\kappa(\mathbf{S}_j)$), which for second-order elliptic operators scales as h^{-2} (where h is the step-size of the discretization). Note that $m_{\mathbf{A}_j}$ also scales as h^{-2}.

that the multiresolution LU decompositon requires $\mathcal{O}(N N^{\frac{4n}{nM}}) = \mathcal{O}(N^{1+\frac{4}{M}})$ operations to compute the matrix $\mathbf{R}'_{\mathbf{S}_j}$ so that its eigenvalues approximate the eigenvalues of $\mathbf{R}_{\mathbf{S}_j}$ to absolute accuracy ϵ. This means, for example, that when $M = 2$, the computational complexity could be $\mathcal{O}(N^3)$, which is as bad as computing the Cholesky decomposition exactly. However, we see in Table 2.2 that in practice, even with $M = 2$ things may be better than this.

2.2.5 Numerical Experiments

In this section, we present preliminary results of numerical experiments. The goal of these experiments is to study the influence of the number of vanishing moments of the wavelet bases and the effect of using different reduction procedures on the preservation of small eigenvalues.

In our first example, we consider the operator $\mathbf{S} = \frac{d}{dx} a(x) \frac{d}{dx}$ on $[0,1]$ with periodic boundary conditions and its discretization $\mathbf{S}_0 = \mathbf{D} M \mathbf{D}^T$, where $\mathbf{D}$ is a fifth-order forward-

Table 2.2: Run times for the sparse versus full reduction procedure.

N	T_{exact}	$T_{\epsilon h_j}$	$\|\mathbf{R}_{\mathbf{S}_j} - \mathbf{R}'_{\mathbf{S}_j}\|_\infty$
576	5.21	4.33	3.3×10^{-3}
1296	51.69	25.54	4.1×10^{-3}
2304	287.52	66.04	4.9×10^{-3}
3600	18.3 min*	121.67	5.2×10^{-3}
5184	54.6 min*	195.74	5.9×10^{-3}
9216	5.1 hrs*	417.47	$7.8 \times 10^{-3*}$
16384	28.7 hrs*	901.21	†

Table 2.2: Run times for exact versus truncated computation of $\mathbf{R}_{\mathbf{S}_j}$ for various scales. The operator $\mathbf{S}_j$ is the projection of $-\nabla \cdot (a(x,y)\nabla)$ on the unit square with periodic boundary conditions to a space $\mathbf{V}_j$ in an MRA on the unit square. Multiwavelets with two vanishing moments are used. Here, N is the number of unknowns in the two-dimensional spatial grid for the scale. T_{exact} is the computation time in seconds (except where noted) required to compute $\mathbf{R}_{\mathbf{S}_j}$ exactly using no truncation in the Cholesky decomposition. $T_{\epsilon h_j}$ is the computation time in seconds required to compute $\mathbf{R}_{\mathbf{S}_j}$ using a truncation threshold of ϵh_j, where $\epsilon = 0.001$ and h_j is the step-size for the scale. The $*$ indicates estimated figures. $\|\mathbf{R}_{\mathbf{S}_j} - \mathbf{R}'_{\mathbf{S}_j}\|_\infty$ is the absolute error between the two versions of $\mathbf{R}_{\mathbf{S}_j}$. The error for the last row (indicated by the †) was not computed due to memory constraints for the exact computation. All computations were performed on an SGI O2 175 Mhz workstation.

difference approximation to the first derivative, M is a diagonal matrix with uniform samples of $a(x)$ on the diagonal, and the step size $h = \frac{1}{1024}$. Although we could have computed the proper projection of this operator, we prefer to use the finite difference discretization in our experiments since we have in mind using our method as a linear algebra tool and want to demonstrate robustness.

We examine eigenvalues of the reduced operators for an $a(x)$ which is pseudo-random (and, hence, highly-oscillatory). The first reduction technique is simply to consider the $\mathbf{T}$ block of $\mathbf{S}_0$ on the coarse scale. The second is to use the reduced operator $\mathbf{R}_{\mathbf{S}_0}$ (1.2.10). Finally, we consider the modified reduced operator defined by (2.2.49). Figure 2.2 compares the performance of these three techniques after one reduction step using compactly supported wavelets with 12 vanishing moments. Experiments clearly show the advantages of using the reduced and modified reduced operators.

In Figures 2.3 and 2.4, we perform reduction over 4 scales so that the reduced matrix is of size 64×64 (the original matrix is of size 1024×1024), and we compare the 64 smallest eigenvalues of the original matrix with eigenvalues of the reduced 64×64 matrix. The three curves correspond to using compactly supported wavelets with different number of vanishing moments. Figure 2.3 was obtained by using the reduced operator $\mathbf{R}_{\mathbf{S}_j}$ after four steps of reduction. Figure 2.4 demonstrates the performance of the modified reduction procedure. For some regimes of the spectrum, we observe that, as expected, increasing the number of vanishing moments increases the accuracy of the approximation.

Our second example illustrates some preliminary two-dimensional results. We consider the operator $\mathbf{S} = -\nabla \cdot (a(x, y)\nabla)$ on the unit square with periodic boundary conditions; we define $a(x, y) = 2 + \cos(32\pi x)$. We discretize this operator in a multiwavelet basis (see [2]) with two vanishing moments, on an interval grid of size 32 by 32. (This results in 4096 unknowns for the fine-scale problem.) Figure 2.5 shows the relative error for the three techniques after one step of reduction, which reduces the number of unknowns to 1024.

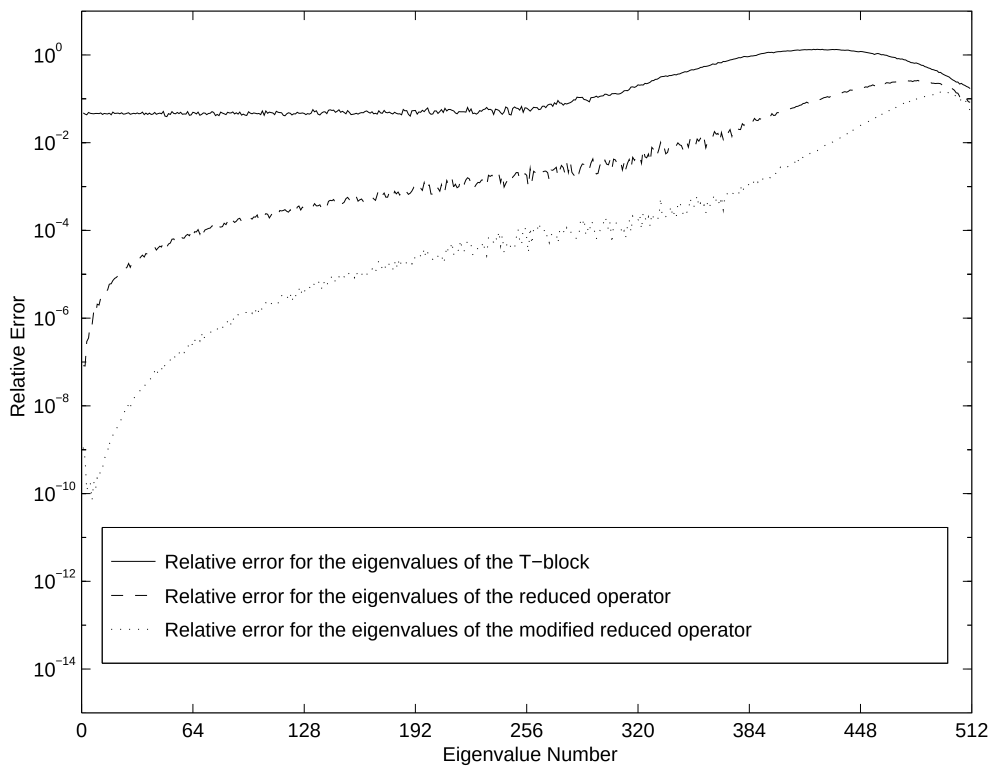

Figure 2.2: Relative error of eigenvalues of the coarse scale operators obtained by different methods compared to the eigenvalues of the original operator.

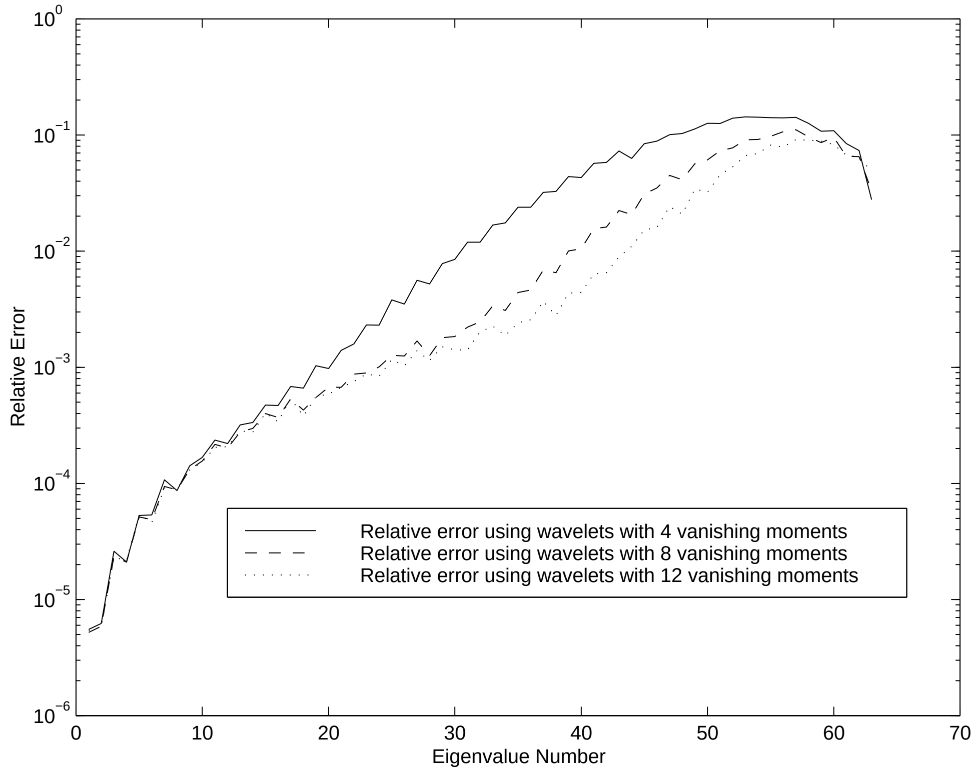

Figure 2.3: Relative error of eigenvalues of the one-dimensional example operator reduced over four scales, using wavelets with four, eight, and twelve vanishing moments.

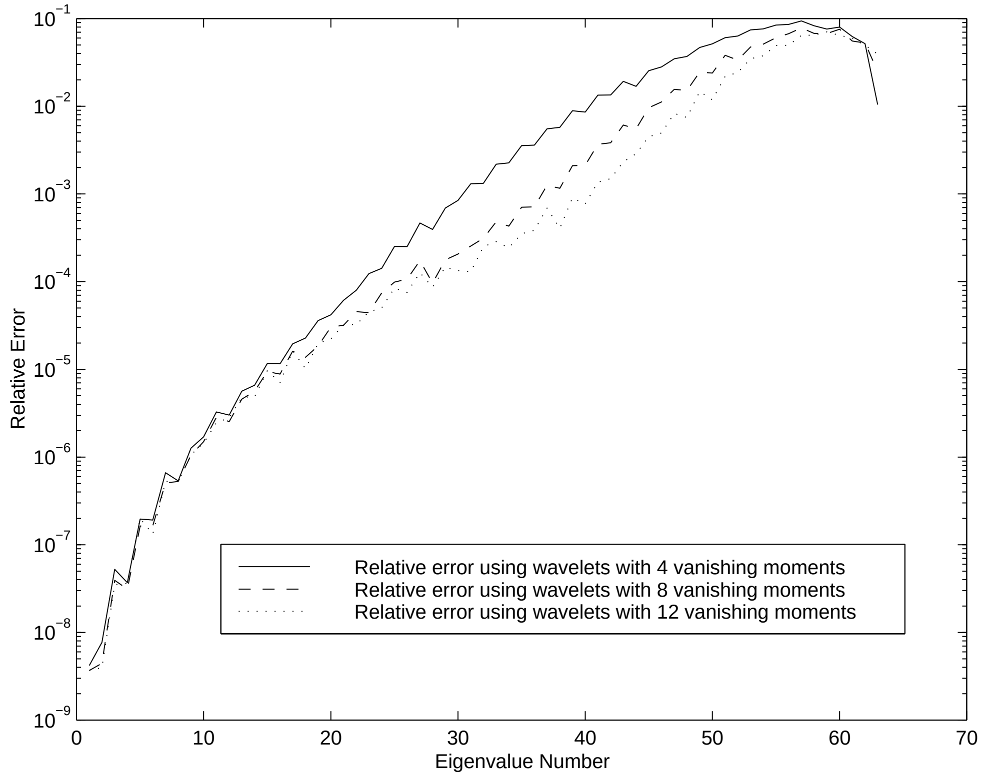

Figure 2.4: Relative error of eigenvalues of the one-dimensional example operator reduced via the modified reduction procedure over four scales, using wavelets with four, eight, and twelve vanishing moments.

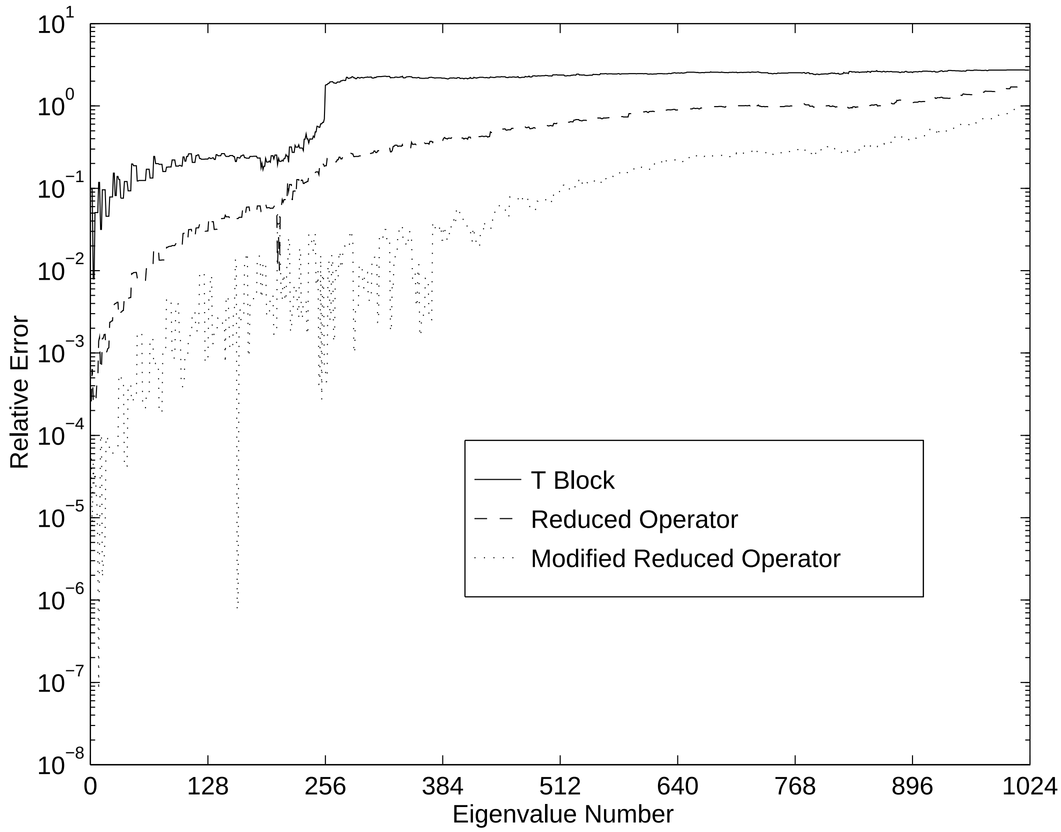

Figure 2.5: Relative error of eigenvalues of the operator $-\nabla(2+\cos(32\pi x))\nabla$ using the three techniques. Multiwavelets with two vanishing moments are used.

CHAPTER 3

CLASSES OF MATRICES

In practice, one of the chief benefits of using wavelets for numerical analysis of differential and integral equations is that the matrix representations of the operators have sparse approximations. The structure of these sparse approximations typically takes the form of fast decay in the magnitude of the elements away from the diagonal, so that most of the elements of the matrix may be truncated to zero. For matrices of finite size, it is impossible to quantify the rate of decay of the elements away from the diagonal for a class of matrices since such a statement is inherently a limit statement. Statements about the rate of decay apply only to the infinite-dimensional (which we also call bi-infinite) matrices. In this chapter, we consider several classes of bi-infinite matrices, and prove some results which were used in Chapter 2. As the experimental results of Chapter 2 demonstrate, the decay rate for matrices of finite size is sufficient to render the algorithms presented in that chapter useful in practice.

3.1 Preliminary Considerations

In this chapter, we consider bounded linear operators on $\mathbf{L}_2(\mathbf{Z}^d)$. If $d = 1$ these operators are represented as infinite-dimensional matrices with two indices; for arbitrary d, they are infinite-dimensional matrices with $2d$ indices. We represent such an operator $\mathbf{M}$ by a matrix $\{m_{\mathbf{k},\mathbf{l}}\}$, where $\mathbf{k}$ and $\mathbf{l}$ are multi-indices with d elements each. (We use indices in **bold face** to denote multi-indices; otherwise we assume they are scalars.) In the case $d = 1$, $\mathbf{k}$ and $\mathbf{l}$ are simply integer indices. If $d = 2$, then they are each a pair of indices, and we may write

$$m_{\mathbf{k},\mathbf{l}} = m_{k_1,l_1,k_2,l_2}. \tag{3.1.1}$$

where the number $m_{\mathbf{k},\mathbf{l}}$ is the $\mathbf{k},\mathbf{l}$ entry of the matrix $\mathbf{M}$. We may also denote this entry by $(\mathbf{M})_{\mathbf{k},\mathbf{l}}$, so that we may write $(\mathbf{M}_1\mathbf{M}_2)_{\mathbf{k},\mathbf{l}}$ for the $\mathbf{k},\mathbf{l}$ entry of the matrix defined by the matrix-matrix product of $\mathbf{M}_1$ and $\mathbf{M}_2$.

In the following sections, we study matrices which have various types of decay in the magnitude of the elements as one moves away from the diagonal. We define the distance between multi-indices in terms of the absolute values of the differences of their components, i.e.

$$|\mathbf{k}-\mathbf{l}| = \sum_{n=1}^{d} |k_n - l_n|. \tag{3.1.2}$$

As defined above, $|\mathbf{k}-\mathbf{l}|$ satisfies the triangle inequality, as well as other properties of the usual notion of distance. The diagonal of a matrix is defined by $|\mathbf{k}-\mathbf{l}| = 0$, or $\mathbf{k} = \mathbf{l}$. Operations on matrices and vectors, such as transposes and multiplication, are defined using this notation in the usual way, for example: $\mathbf{M}^* = \{m_{\mathbf{k},\mathbf{l}}\}^* = \{m_{\mathbf{l},\mathbf{k}}\}$.

3.2 Matrices with Exponential Decay

If $d = 1$, we define (following [37] and [22]) the class of matrices $\mathcal{X}$ with exponential decay away from the diagonal as follows:

$$\mathcal{X} = \left\{ \{\mathbf{M} = m_{k,l}\}_{k,l \in \mathbf{Z}} \,|\, \text{there exists } C(\mathbf{M}), \epsilon \text{ s.t. } |m_{k,l}| < C(\mathbf{M}) e^{-\epsilon(|k-l|)} \right\}. \tag{3.2.1}$$

This class is closed under addition and multiplication. Additionally, it is shown in [37] and [22] that if $\mathbf{M} \in \mathcal{X}$ and $\mathbf{M}^{-1}$ exists on l^2 then $\mathbf{M} \in \mathcal{X}$.

For the purposes extending (3.2.1) to $d = 2$, we define the class $\mathcal{X}$ of bounded linear operators on l^2 by

$$\mathcal{X} = \left\{ \{m_{\mathbf{k},\mathbf{l}}\}_{\mathbf{k},\mathbf{l} \in \mathbf{Z}^2} \,|\, \text{there exists } C(\mathbf{M}), \epsilon \text{ s.t. } |m_{\mathbf{k},\mathbf{l}}| < C(\mathbf{M}) e^{-\epsilon|\mathbf{k}-\mathbf{l}|} \right\}. \tag{3.2.2}$$

Unless stated otherwise, we assume for the rest of this section that $d = 2$ and use $\mathcal{X}$ to denote the class defined by (3.2.2).

We define the space l^p, $1 \leq p \leq \infty$, to be the set of all sequences $\{x_{\mathbf{k}}\}_{\mathbf{k} \in \mathbf{Z}^d}$ such that

$$\|x\|_p = \left(\sum_{\mathbf{k} \in \mathbf{Z}^d} |x_{\mathbf{k}}|^p \right)^{\frac{1}{p}} < \infty \tag{3.2.3}$$

for $1 \le p < \infty$, and

$$\|x\|_\infty = \sup_{\mathbf{k}\in\mathbf{Z}^d} |x_\mathbf{k}| < \infty. \tag{3.2.4}$$

Similarly, for a linear transformation $\mathbf{S}$ with domain l^p and range l^q, we define $\|\mathbf{S}\|_{p,q} = \sup_{x\in l^p, \|x\|_p=1} \|\mathbf{S}x\|_q$. We note that $\|\mathbf{S}\|_{2,2} < \infty$ provides a uniform bound on the elements of the matrix $\mathbf{S}$, since

$$\begin{aligned} |(\mathbf{S})_{\tilde{\mathbf{k}},\tilde{\mathbf{l}}}| &\le \left(\sum_{\mathbf{k}\in\mathbf{Z}^d} \left[(\mathbf{S})_{\mathbf{k},\tilde{\mathbf{l}}}\right]^2\right)^{\frac{1}{2}} && (3.2.5)\\ &\le \sup_{\|x\|_2=1} \left(\sum_{\mathbf{k}\in\mathbf{Z}^d}\left(\sum_{\mathbf{l}\in\mathbf{Z}^d} (\mathbf{S})_{\mathbf{k},\mathbf{l}} x_\mathbf{l}\right)^2\right)^{\frac{1}{2}} && (3.2.6)\\ &= \|\mathbf{S}\|_{2,2} && (3.2.7)\end{aligned}$$

We will prove that each element in the class $\mathcal{X}$ defines a bounded operator on l^1 and l^∞, that this class is closed under addition and multiplication, and that if $\mathbf{M} \in \mathcal{X}$ and $\mathbf{M}^{-1}$ is a bounded operator on l^2 then $\mathbf{M}^{-1} \in \mathcal{X}$.

At this point, we state two important theorems from [34] which will be useful in our considerations. Note that we define the convolution of two sequences x and y as

$$(x * y)_\mathbf{l} = \sum_{\mathbf{k}\in\mathbf{Z}^d} x_{\mathbf{l}-\mathbf{k}} y_\mathbf{k}. \tag{3.2.8}$$

The sequence which bounds the rows of matrices in the class $\mathcal{X}$ is denoted by $\mathcal{E}(\epsilon) = \{e^{-\epsilon|\mathbf{k}|}\}_{\mathbf{k}\in\mathbf{Z}^d}$.

Theorem 3.2.1 (Young's Inequality) *If $f \in l^p$ and $g \in l^r$, then*

$$\|f * g\|_q \le \|f\|_p \|g\|_r \tag{3.2.9}$$

whenever $\frac{1}{q} = \frac{1}{p} + \frac{1}{r} - 1$.

Theorem 3.2.2 (M. Riesz Convexity Theorem) *For a linear operator $\mathbf{T}$, if $\|\mathbf{T}\|_{p_i,q_i} < \infty$ for $i = 0, 1$, and if $1/p_t = (1-t)/p_0 + t/p_1$, $1/q_t = (1-t)/q_0 + t/q_1$ with $0 \le t \le 1$, then*

$$\|\mathbf{T}\|_{p_t,q_t} < \|\mathbf{T}\|_{p_0,q_0}^{1-t} \|\mathbf{T}\|_{p_1,q_1}^t. \tag{3.2.10}$$

We start by showing the following:

Lemma 3.2.1 *If* $\mathbf{M} \in \mathcal{X}$*, then* $\mathbf{M}$ *is a bounded operator on* l^1 *and* l^∞*.*

Proof: Via Young's Inequality, we see that

$$\begin{aligned}
\|\mathbf{M}\|_{1,1} &= \sup_{\|x\|_1=1} \|\mathbf{M}x\|_1 \\
&= \sup_{\|x\|_1=1} \sum_{\mathbf{k}\in\mathbf{Z}^2} \left| \sum_{\mathbf{l}\in\mathbf{Z}^2} m_{\mathbf{k},\mathbf{l}} x_{\mathbf{l}} \right| \\
&\leq C(\mathbf{M}) \sup_{\|x\|_1=1} \sum_{\mathbf{k}\in\mathbf{Z}^2} \sum_{\mathbf{l}\in\mathbf{Z}^2} |\mathcal{E}(\epsilon)_{\mathbf{k}-\mathbf{l}}||x_{\mathbf{l}}| \\
&= C(\mathbf{M}) \sup_{\|x\|_1=1} \|\mathcal{E}(\epsilon) * x\|_1 \\
&\leq C(\mathbf{M})\|\mathcal{E}(\epsilon)\|_1 < \infty.
\end{aligned}$$

A similar proof shows the result for $\|\mathbf{M}\|_{\infty,\infty}$.□

If $d = 1$, it was shown in [37] and [22] that the class $\mathcal{X}$ is an algebra under matrix-matrix multiplication. We show the same for $\mathcal{X}$ if $d = 2$, following the proof of [22] very closely.

Lemma 3.2.2 *If* $\mathbf{M}_1 \in \mathcal{X}$ *and* $\mathbf{M}_2 \in \mathcal{X}$*, then* $\mathbf{M}_1\mathbf{M}_2 \in \mathcal{X}$*. Additionally, if the* $\mathbf{k},\mathbf{l}$ *entry of* $\mathbf{M}_i$ *is denoted by* $m^{(i)}_{\mathbf{k},\mathbf{l}}$*, and satisfies* $|m^{(i)}_{\mathbf{k},\mathbf{l}}| \leq C_i e^{-\epsilon_i|\mathbf{k}-\mathbf{l}|}$ *for* $i = 1, 2$*, and* $\epsilon_1 < \epsilon_2$*, then the* $\mathbf{k},\mathbf{l}$ *entry of the product* $\mathbf{M}_3 = \mathbf{M}_1\mathbf{M}_2$ *satisfies*

$$|m^{(3)}_{\mathbf{k},\mathbf{l}}| \leq C_1C_2Ae^{-\epsilon_1|\mathbf{k}-\mathbf{l}|}, \tag{3.2.11}$$

where $A = A(\epsilon_2 - \epsilon_1)$*.*

Proof: To prove (3.2.11), consider the product element-by-element:

$$\begin{aligned}
|m^{(3)}_{\mathbf{k},\mathbf{l}}| &= |\sum_{\mathbf{j}\in\mathbf{Z}^2} m^{(1)}_{\mathbf{k},\mathbf{j}} m^{(2)}_{\mathbf{j},\mathbf{l}}| & (3.2.12)\\
&\leq \sum_{\mathbf{j}\in\mathbf{Z}^2} |m^{(1)}_{\mathbf{k},\mathbf{j}}||m^{(2)}_{\mathbf{j},\mathbf{l}}| & (3.2.13)\\
&\leq C_1C_2 \sum_{\mathbf{j}\in\mathbf{Z}^2} e^{-\epsilon_1|\mathbf{k}-\mathbf{j}|} e^{-\epsilon_2|\mathbf{j}-\mathbf{l}|} & (3.2.14)\\
&\leq C_1C_2 \sum_{\mathbf{j}\in\mathbf{Z}^2} e^{-\epsilon_1(|\mathbf{k}-\mathbf{l}|-|\mathbf{l}-\mathbf{j}|)} e^{-\epsilon_2|\mathbf{j}-\mathbf{l}|} & (3.2.15)\\
&= C_1C_2e^{-\epsilon_1|\mathbf{k}-\mathbf{l}|} \sum_{\mathbf{j}\in\mathbf{Z}^2} e^{-(\epsilon_2-\epsilon_1)|\mathbf{j}-\mathbf{l}|} & (3.2.16)\\
&= C_1C_2A(\epsilon_2 - \epsilon_1)e^{-\epsilon_1|\mathbf{k}-\mathbf{l}|}, & (3.2.17)
\end{aligned}$$

where $A(\epsilon) = \sum_{\mathbf{j}\in\mathbf{Z}^2} e^{-\epsilon|\mathbf{j}|} < \infty$. Thus the function $A(\epsilon)$ depends only on ϵ, and ,furthermore, $A(\epsilon) > 1$. We see that if $\epsilon_1 < \epsilon_2$, then (3.2.11) implies that $\mathbf{M}_1\mathbf{M}_2 \in \mathcal{X}$. If $\epsilon_1 = \epsilon_2$ we note that we may always adjust ϵ_1 so that it is smaller than ϵ_2. Thus, for any values of $\epsilon_1, \epsilon_2 > 0$, we have $\mathbf{M}_1\mathbf{M}_2 \in \mathcal{X}$.□

It is shown in [37] and [22] if $d = 1$, then invertible elements of the class $\mathcal{X}$ have inverses which are elements of $\mathcal{X}$. We use Lemma 3.2.2 to prove that the same is true if $d = 2$. Our proof follows that of [37] and [22] very closely.

Theorem 3.2.3 *If $\mathbf{M} \in \mathcal{X}$, and $\mathbf{M}$ is invertible on l^2, then $\mathbf{M}^{-1} \in \mathcal{X}$.*

Proof: First, consider the case where $\mathbf{M} = \mathbf{I} - \mathbf{U}$ and $\|\mathbf{U}\|_{2,2} < 1$. Then we have $\mathbf{M}^{-1} = \sum_{n=0}^{\infty} \mathbf{U}^n$. We denote the $\mathbf{k},\mathbf{l}$ entry of $\mathbf{U}^n$ by $(u^{(n)})_{\mathbf{k},\mathbf{l}}$. Since $\mathbf{U} \in \mathcal{X}$, we write $|(u^{(1)})_{\mathbf{k},\mathbf{l}}| \leq Ce^{-\epsilon|\mathbf{k}-\mathbf{l}|}$. Using Lemma 3.2.2 and noting that $|(u^{(1)})_{\mathbf{k},\mathbf{l}}| \leq Ce^{-\epsilon|\mathbf{k}-\mathbf{l}|} \leq Ce^{-\frac{\epsilon}{2}|\mathbf{k}-\mathbf{l}|}$, we estimate

$$|(u^{(n)})_{\mathbf{k},\mathbf{l}}| \leq C^n A(\frac{\epsilon}{2})^{n-1} e^{-\frac{\epsilon}{2}|\mathbf{k}-\mathbf{l}|}. \tag{3.2.18}$$

Since $A > 1$, we then estimate

$$\left|\sum_{n=0}^{N} (\mathbf{U}^n)_{\mathbf{k},\mathbf{l}}\right| = \left|\sum_{n=0}^{N} (u^{(n)})_{\mathbf{k},\mathbf{l}}\right| \leq \sum_{n=0}^{N} (AC)^n e^{-\frac{\epsilon}{2}(|\mathbf{k}-\mathbf{l}|)}. \tag{3.2.19}$$

For the remainder, we write

$$\sum_{n=N+1}^{\infty} \|\mathbf{U}^n\|_{2,2} \leq \sum_{n=N+1}^{\infty} \|\mathbf{U}\|_{2,2}^n \frac{\|\mathbf{U}\|_{2,2}^{N+1}}{1 - \|\mathbf{U}\|_{2,2}}. \tag{3.2.20}$$

Since we have

$$\sum_{n=0}^{N} (AC)^n = \frac{1 - (AC)^{N+1}}{1 - AC}, \tag{3.2.21}$$

and the norm $\|\mathbf{U}^n\|_{2,2}$ provides a uniform bound on the elements of the matrix $\mathbf{U}^n$, we obtain

$$\begin{aligned} \left|\left(\sum_{n=0}^{\infty} \mathbf{U}^n\right)_{\mathbf{k},\mathbf{l}}\right| &\leq \left|\sum_{n=0}^{N} (\mathbf{U}^n)_{\mathbf{k},\mathbf{l}}\right| + \left|\left(\sum_{n=N+1}^{\infty} \mathbf{U}^n\right)_{\mathbf{k},\mathbf{l}}\right| & (3.2.22)\\ &\leq \frac{1 - (AC)^{N+1}}{1 - AC} e^{-\frac{\epsilon}{2}|\mathbf{k}-\mathbf{l}|} + \frac{\|\mathbf{U}\|_{2,2}^{N+1}}{1 - \|\mathbf{U}\|_{2,2}} & (3.2.23)\\ &= \frac{1 - e^{(N+1)\log(AC)}}{1 - AC} e^{-\frac{\epsilon}{2}|\mathbf{k}-\mathbf{l}|} + \frac{e^{\log(\|\mathbf{U}\|_{2,2})(N+1)}}{1 - \|\mathbf{U}\|_{2,2}}. & (3.2.24) \end{aligned}$$

In order to obtain exponential decay in $|\mathbf{k}-\mathbf{l}|$ on the right-hand side of this expression, we have to choose an N which depends on $|\mathbf{k}-\mathbf{l}|$ in such a way that

$$(N+1)\log(AC) - \frac{\epsilon}{2}|\mathbf{k}-\mathbf{l}| < -\epsilon'|\mathbf{k}-\mathbf{l}|, \tag{3.2.25}$$

for some $\epsilon' > 0$, and also that

$$\log(\|\mathbf{U}\|_{2,2})(N+1) < -\epsilon'|\mathbf{k}-\mathbf{l}|. \tag{3.2.26}$$

If we choose $\epsilon' = \frac{\epsilon}{2}\frac{\log(\|U\|_{2,2}^{-1})}{\log(AC)+\log(\|U\|_{2,2}^{-1})}$, then we see that

$$\frac{\epsilon'}{\log(\|U\|_{2,2}^{-1})} = \frac{\frac{\epsilon}{2}-\epsilon'}{\log(AC)} \tag{3.2.27}$$

which means that

$$\frac{\epsilon'}{\log(\|U\|_{2,2}^{-1})} < \frac{\frac{\epsilon}{2}-\epsilon'}{\log(AC)} + \frac{1}{|\mathbf{k}-\mathbf{l}|} \tag{3.2.28}$$

and, therefore,

$$\frac{\epsilon'}{\log(\|U\|_{2,2}^{-1})}|\mathbf{k}-\mathbf{l}| < \frac{\frac{\epsilon}{2}-\epsilon'}{\log(AC)}|\mathbf{k}-\mathbf{l}| + 1. \tag{3.2.29}$$

We may always choose a non-negative integer N so that $N+1$ fits in between the lower and upper bounds of the inequality (3.2.29), which implies that $N+1$ satisfies the inequalities (3.2.25) and (3.2.26).

Therefore, under the assumption that $\mathbf{M} = \mathbf{I}-\mathbf{U}$ and $\|\mathbf{U}\|_{2,2} < 1$, i.e. $\mathbf{M}$ is a perturbation of the identity, we have shown that $\mathbf{M}^{-1} \in \mathcal{X}$. Following the proof in [37], the extension to the general case is simple. First we show that the theorem is true for symmetric positive-definite operators. If $\mathbf{L} \in \mathcal{X}$ is symmetric and positive definite, then there exist constants $A, B > 0$ such that

$$A\|x\|_2 \leq (\mathbf{L}x, x) \leq B\|x\|_2, \tag{3.2.30}$$

for all $x \in l^2$. Thus, it follows that

$$\|\mathbf{L} - \left(\frac{B+A}{2}\right)\mathbf{I}\|_{2,2} \leq \frac{B-A}{2} \tag{3.2.31}$$

from which we can deduce

$$\mathbf{L} = \left(\frac{A+B}{2}\right)(\mathbf{I}-\mathbf{U}), \tag{3.2.32}$$

where $\|\mathbf{U}\|_{2,2} \leq \frac{B-A}{B+A} < 1$. Thus, if $\mathbf{L} \in \mathcal{X}$ is symmetric and positive-definite, it may be written as a scalar multiple of a perturbation of the identity, for which Theorem 3.2.3 has been proved.

Now, we see that if $\mathbf{M} \in \mathcal{X}$ has a bounded inverse on l^2 and $\mathbf{L} = \mathbf{M}\mathbf{M}^*$, then we have $\mathbf{L}^{-1} \in \mathcal{X}$, which implies (using Lemma 3.2.2) that $\mathbf{M}^{-1} = (\mathbf{M}^*\mathbf{M})^{-1}\mathbf{M}^* \in \mathcal{X}$. □

3.3 Matrices with Polynomial Decay

In this section we consider bi-infinite matrices with polynomial decay in the magnitude of the elements away from the diagonal. We follow the techniques and results of [37] and [21] very closely, and extend the results of those papers from the case $d = 1$ to the case $d = 2$.

The class $\mathcal{I}_\alpha$, $\alpha > 0$, is defined in [37] as follows:

$$\mathcal{I}_\alpha = \left\{ \{m_{k,l}\}_{k,l\in\mathbf{Z}} \,|\, \text{there exists } C(\mathbf{M}) \text{ s.t. } |m_{k,l}| < C(\mathbf{M})(1+|k-l|)^{-1-\alpha} \right\}. \tag{3.3.1}$$

It is shown that this class is closed under addition and multiplication, and that if $\mathbf{M} \in \mathcal{I}_\alpha$ and $\mathbf{M}$ is invertible on l^2, then $\mathbf{M}^1 \in \mathcal{I}_\alpha$ for all positive real α. For matrices which are the blocks of a non-standard form of a Calderon-Zygmund operator, the parameter α is the number of vanishing moments of the wavelet basis (see [9]).

If $d = 2$, we define the analogue of $\mathcal{I}_\alpha$ (which we also denote by $\mathcal{I}_\alpha$) as:

$$\mathcal{I}_\alpha = \left\{ \{m_{\mathbf{k},\mathbf{l}}\}_{\mathbf{k},\mathbf{l}\in\mathbf{Z}^2} \,|\, \text{there exists } C(\mathbf{M}) \text{ s.t. } |m_{\mathbf{k},\mathbf{l}}| < C(\mathbf{M})(1+|\mathbf{k}-\mathbf{l}|)^{-2-\alpha} \right\}. \tag{3.3.2}$$

The goal is to show that the class $\mathcal{I}_\alpha$ is closed under multiplication, and that if $\mathbf{M} \in \mathcal{I}_\alpha$ and $\mathbf{M}$ is invertible on l^2 then $\mathbf{M}^{-1} \in \mathcal{I}_\alpha$. This result is essential in showing the result of Chapter 2 that the reduction procedure preserves sparsity.

Theorem 3.3.1 *If $\alpha > 0$, the class $\mathcal{I}_\alpha$ is closed under multiplication and addition.*

Proof: It is obvious that $\mathcal{I}_\alpha$ is closed under addition. For multiplication, we write

$$\begin{aligned} |(\mathbf{M}_1\mathbf{M}_2)_{\mathbf{k},\mathbf{l}}| &= \left| \sum_{\mathbf{j}\in\mathbf{Z}^2} m^{(1)}_{\mathbf{k},\mathbf{j}} m^{(2)}_{\mathbf{j},\mathbf{l}} \right| && (3.3.3) \\ &\leq C_1 C_2 \sum_{\mathbf{j}\in\mathbf{Z}^2} (1+|\mathbf{k}-\mathbf{j}|)^{-2-\alpha}(1+|\mathbf{j}-\mathbf{l}|)^{-2-\alpha} && (3.3.4) \end{aligned}$$

$$= \quad C_1 C_2 \sum_{\mathbf{j}\in\mathbf{Z}^2} (1+|(\mathbf{k}-\mathbf{l})-\mathbf{j}|)^{-2-\alpha}(1+|\mathbf{j}|)^{-2-\alpha}. \tag{3.3.5}$$

We set $\tilde{\mathbf{k}} = \mathbf{k} - \mathbf{l}$ and split $\mathbf{Z}^2$ into two sets:

$$R_1 = \left\{\mathbf{j} \in \mathbf{Z}^2 | m(\tilde{\mathbf{k}}, \mathbf{j}) \geq m(\mathbf{j}, 0)\right\} \tag{3.3.6}$$

and

$$R_2 = \mathbf{Z}^2 \setminus R_1, \tag{3.3.7}$$

where m is the usual euclidean distance $m(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$. We note that $\frac{1}{\sqrt{2}}|\mathbf{p} - \mathbf{q}| \leq m(\mathbf{p}, \mathbf{q}) \leq |\mathbf{p} - \mathbf{q}|$, which implies that

$$(1+|\mathbf{p}-\mathbf{q}|)^{-2-\alpha} \leq (1+m(\mathbf{p},\mathbf{q}))^{-2-\alpha} \leq 2^{1+\frac{\alpha}{2}}(1+|\mathbf{p}-\mathbf{q}|)^{-2-\alpha}. \tag{3.3.8}$$

Additionally, it is clear that if $\mathbf{j} \in R_1$, then

$$m(0, \tilde{\mathbf{k}}) \leq m(0, \mathbf{j}) + m(\tilde{\mathbf{k}}, \mathbf{j}) \leq 2m(\tilde{\mathbf{k}}, \mathbf{j}), \tag{3.3.9}$$

so, for the sum over R_1, we may write

$$\sum_{\mathbf{j}\in R_1} (1+|\tilde{\mathbf{k}}-\mathbf{j}|)^{-2-\alpha}(1+|\mathbf{j}|)^{-2-\alpha} \leq C(1+|\tilde{\mathbf{k}}|)^{-2-\alpha} \sum_{\mathbf{j}\in R_1} (1+|\mathbf{j}|)^{-2-\alpha} \leq C'(1+|\tilde{\mathbf{k}}|)^{-2-\alpha} \tag{3.3.10}$$

for $|\tilde{\mathbf{k}}|$ large enough.

We note that $\mathbf{j} \in R_2$ implies that $m(\mathbf{j}, 0) > m(\tilde{\mathbf{k}}, \mathbf{j})$, from which we deduce $m(\mathbf{j} - \tilde{\mathbf{k}}, -\tilde{\mathbf{k}}) > m(0, \mathbf{j} - \tilde{\mathbf{k}})$, which means that $\tilde{\mathbf{k}} - \mathbf{j} \in R_1$. For the sum over R_2, we write

$$\sum_{\mathbf{j}\in R_2} (1+|\tilde{\mathbf{k}}-\mathbf{j}|)^{-2-\alpha}(1+|\mathbf{j}|)^{-2-\alpha} = \sum_{\mathbf{j}'\in R_1} (1+|\mathbf{j}'|)^{-2-\alpha}(1+|\tilde{\mathbf{k}}-\mathbf{j}'|)^{-2-\alpha} \leq C'(1+|\tilde{\mathbf{k}}|)^{-2-\alpha} \tag{3.3.11}$$

for $|\tilde{\mathbf{k}}|$ large enough, where $\mathbf{j}' = \tilde{\mathbf{k}} - \mathbf{j}$.

We arrive at the inequality

$$|(\mathbf{M}_1\mathbf{M}_2)_{\mathbf{k},\mathbf{l}}| \leq C(1+|\mathbf{k}-\mathbf{l}|)^{-2-\alpha}. \tag{3.3.12}$$

Thus the operator $\mathbf{M}_1\mathbf{M}_2$ is an element of $\mathcal{I}_\alpha$.□

Now, our goal is to prove that if $\alpha \in \mathbf{Z}$, $\alpha \geq 2$ and $\mathbf{M} \in l^2$ has a bounded inverse on l^2, then $\mathbf{M}^{-1} \in l^2$. At the time of writing, we do not have a proof for all positive real α. The condition $\alpha \geq 2$ restricts us to using wavelets with at least two vanishing moments.

Theorem 3.3.2 *If* $\mathbf{M} \in \mathcal{I}_\alpha$, $\alpha \in \mathbf{Z}$, $\alpha \geq 2$ *and* $\mathbf{M}$ *is invertible on* l^2, *then* $\mathbf{M}^{-1} \in \mathcal{I}_\alpha$.

The rest of this chapter is devoted to the proof of this theorem. The proof is quite involved and requires a number of steps.

First, define the unbounded operators $\mathbf{X}_1(\{c_\mathbf{k}\}) = \{k_1 c_\mathbf{k}\}$, and $\mathbf{X}_2(\{c_\mathbf{k}\}) = \{k_2 c_\mathbf{k}\}$. Note that the commutator $[\mathbf{X}_i, \mathbf{M}]$ can be written as the matrix $(k_i - l_i)m_{\mathbf{k},\mathbf{l}}$, and that the order n commutator of $\mathbf{X}_i$ with $\mathbf{M}$, defined by

$$[\mathbf{X}_i, \mathbf{M}]_n = \underbrace{[\mathbf{X}_i, [\mathbf{X}_i, \ldots, [\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}]] \cdots]]}_{n \text{ commutators}} \tag{3.3.13}$$

may be written as the matrix $(k_i - l_i)^n m_{\mathbf{k},\mathbf{l}}$. Additionally, we see that if $\mathbf{S} \in \mathcal{I}_\alpha$ then $[\mathbf{X}_i, \mathbf{S}]_m \in \mathcal{I}_{\alpha - m}$.

Now, we note that, given a bounded linear operator $\mathbf{S} : l^1 \to l^\infty$, the norm $\|\mathbf{S}\|_{1,\infty}$ of this operator provides a uniform bound on the magnitudes of the elements of the matrix which represents the operator:

$$\begin{aligned}
\|\mathbf{S}\|_{1,\infty} &= \sup_{\|x\|_1 = 1} \|\mathbf{S}x\|_\infty && (3.3.14)\\
&= \sup_{\|x\|_1 = 1} \sup_{\mathbf{l} \in \mathbf{Z}^2} \sum_{\mathbf{k} \in \mathbf{Z}^2} |s_{\mathbf{l},\mathbf{k}} x_\mathbf{k}| && (3.3.15)\\
&\geq \sup_{\mathbf{l} \in \mathbf{Z}^2} |s_{\mathbf{l},\tilde{\mathbf{k}}}| && (3.3.16)\\
&\geq |s_{\tilde{\mathbf{l}},\tilde{\mathbf{k}}}| && (3.3.17)
\end{aligned}$$

for any $\tilde{\mathbf{k}}, \tilde{\mathbf{l}}$.

Thus, if the commutator $[\mathbf{X}_i, \mathbf{M}^{-1}]_{2+\alpha}$, $i = 1, 2$ is a bounded linear operator from l^1 to l^∞, then

$$|k_i - l_i|^{2+\alpha} |m^{-1}_{\mathbf{k},\mathbf{l}}| \leq \|[\mathbf{X}_i, \mathbf{M}]_{2+\alpha}\|_{1,\infty} \tag{3.3.18}$$

for $i = 1, 2$. It is easily shown that

$$(|k| + |l|)^q \leq 2^q(|k|^q + |l|^q), \tag{3.3.19}$$

so, by adding together (3.3.18) for $i = 1$ and $i = 2$, we obtain

$$\begin{aligned}
2^{-(1+\alpha)}(|\mathbf{k} - \mathbf{l}|)^{2+\alpha} |m^{-1}_{\mathbf{k},\mathbf{l}}| &\leq (|k_1 - l_1|^{2+\alpha} + |k_2 - l_2|^{2+\alpha}) |m^{-1}_{\mathbf{k},\mathbf{l}}| && (3.3.20)\\
&\leq \|[\mathbf{X}_i, \mathbf{M}]_{2+\alpha}\|_{1,\infty}. && (3.3.21)
\end{aligned}$$

Since $(|\mathbf{k}-\mathbf{l}|)^{2+\alpha} \leq (1+|\mathbf{k}-\mathbf{l}|)^{2+\alpha}$, we may divide both sides of the above inequality by $(1+|\mathbf{k}-\mathbf{l}|)^{2+\alpha}$ to obtain

$$|m^{-1}_{\mathbf{k},\mathbf{l}}| \leq C(\mathbf{M},\alpha)(1+|\mathbf{k}-\mathbf{l}|)^{-2-\alpha}, \tag{3.3.22}$$

which implies that $\mathbf{M}^{-1} \in \mathcal{I}_\alpha$. In order to complete the proof of Theorem 3.3.2, we need to show that

Proposition 3.3.1 *If $\mathbf{M} \in \mathcal{I}_\alpha$ is invertible on l^2 and $\alpha \in \mathbf{Z}$, $\alpha \geq 2$, then the commutator $[\mathbf{X}_i, \mathbf{M}]_{2+\alpha}$ is a bounded linear operator from l^1 to l^∞.*

Proof: The proof of this proposition hinges on the following:

Proposition 3.3.2 *If $\mathbf{M} \in \mathcal{I}_\alpha$ is invertible on l^2 and $\alpha \geq 2$, then $\mathbf{M}$ is invertible on l^p, $1 \leq p \leq \infty$.*

We prove Proposition 3.3.2 at the end of this chapter.

The proof of Proposition 3.3.1 also requires the following lemma (based on a lemma from [37]).

Lemma 3.3.1 *If $\mathbf{A}$ is a bounded linear operator on l^2 such that*

$$|a_{\mathbf{k},\mathbf{l}}| \leq C(1+|\mathbf{k}-\mathbf{l}|)^{-s}, \quad s \in (0,2], \tag{3.3.23}$$

and $\frac{1}{p} - \frac{1}{q} > 1 - \frac{s}{2}$, then $\mathbf{A}$ is continuous from l^p to l^q.

Proof: We define the sequence $\mathcal{P}(\gamma)$ by $(\mathcal{P}(\gamma))_{\mathbf{k}} = (1+|\mathbf{k}|)^{-\gamma}$. Suppose $x \in l^p$ and $\|x\|_p = 1$. Then

$$\|\mathbf{A}x\|_q \leq C\|\mathcal{P}(s) * x\|_q \leq C\|\mathcal{P}(s)\|_r \|x\|_p \tag{3.3.24}$$

if $\frac{1}{q} = \frac{1}{r} + \frac{1}{p} - 1$ (by Young's inequality). But $\|\mathcal{P}(s)\|_r < \infty$ iff $sr > 2$, so if $\frac{1}{p} - \frac{1}{q} > 1 - \frac{s}{2}$ we can choose an r such that $sr > 2$ and $\frac{1}{q} = \frac{1}{r} + \frac{1}{p} - 1$.□

Now, we continue with the proof of Proposition 3.3.1. We use the commutators of $\mathbf{X}_i$ with $\mathbf{M}^{-1}$ to obtain all but the last two degrees of decay in the entries of $\mathbf{M}^{-1}$. We make use of the commutator identities

$$[\mathbf{X}_i, \mathbf{M}^{-1}] = -\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1} \tag{3.3.25}$$

and

$$[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}^{-1}]] = \mathbf{M}^{-1}(-[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}]] - 2[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}])\mathbf{M}^{-1} \tag{3.3.26}$$

and their higher-order analogues. That is, $[\mathbf{X}_i, \mathbf{M}^{-1}]_\alpha$ is a sum of products of $\mathbf{M}^{-1}$ and commutators of $\mathbf{X}_i$ and $\mathbf{M}$ of order no greater than α (see [37], [6]). By Lemma 3.3.1, it is clear that each of these commutators is a bounded linear operator from l^p to $l^{p+\epsilon}$ for any $p > 1$, $\epsilon > 0$, so we may string together these commutators with $\mathbf{M}^{-1}$, which by Proposition 3.3.2 is bounded on all l^p spaces, in any combination of products to obtain a bounded linear operator from l^1 to l^∞. This implies that $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha-2}$.

We obtain the last two degrees of decay using a bootstrapping proof following [37]. The following lemma provides the first step:

Lemma 3.3.2 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *(* $\alpha \in \mathbf{Z}$, $\alpha \geq 2$ *) and* $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha+\gamma-2}$, *where* $0 \leq \gamma \leq 1$, *then*

$$[\mathbf{X}_i, \mathbf{M}^{-1}]_\alpha = -\mathbf{M}^{-1}([\mathbf{X}_i, \mathbf{M}]_\alpha + [\mathbf{X}_i, \mathbf{M}]_{\alpha-1}\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}] + [\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]_{\alpha-1} + \mathbf{K}_\alpha)\mathbf{M}^{-1} \tag{3.3.27}$$

where $\mathbf{K}_\alpha \in \mathcal{I}_\gamma$.

Proof: First, consider the case where $\alpha = 2$. We see that the identity (3.3.26) implies that $\mathbf{K}_2 = 0$ and so the result is trivially true.

Now consider the case where $\alpha > 2$. which we prove in this case by induction. As the first step in the proof, consider the identity

$$\begin{aligned} [\mathbf{X}_i, \mathbf{M}^{-1}]_3 &= \mathbf{M}^{-1}(-[\mathbf{X}_i, \mathbf{M}] - [\mathbf{X}_i, \mathbf{M}]_2\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}] - [\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]_2 \\ &+ 4[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}] \\ &- 2[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]])\mathbf{M}^{-1}, \end{aligned} \tag{3.3.28}$$

and set

$$\mathbf{K}_3 = 4[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}] - 2[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]]. \tag{3.3.29}$$

Note that $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha-2+\gamma}$, and $[\mathbf{X}_i, \mathbf{M}] \in \mathcal{I}_{\alpha-1}$. Since $\mathcal{I}_n$ forms a multiplicative algebra if $n > 0$, and $\alpha > 2$, it is clear that any combination of products of these terms is an element of $\mathcal{I}_{\alpha-2+\gamma}$. The commutator of $\mathbf{X}_i$ with any element of $\mathcal{I}_{\alpha-2+\gamma}$ is an element of $\mathcal{I}_{\alpha-3+\gamma}$, so from this we see that $\mathbf{K}_3 \in \mathcal{I}_{\alpha-3+\gamma}$.

Now, assume that

$$[\mathbf{X}_i,\mathbf{M}^{-1}]_n = -\mathbf{M}^{-1}([\mathbf{X}_i,\mathbf{M}]_n+[\mathbf{X}_i,\mathbf{M}]_{n-1}\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]+[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]_{n-1}+\mathbf{K}_n)\mathbf{M}^{-1} \tag{3.3.30}$$

where $\mathbf{K}_n \in \mathcal{I}_{\alpha-n+\gamma}$ and n is an integer strictly less than α. We wish to show that $[\mathbf{X}_i,\mathbf{M}^{-1}]_{n+1}$ has this same form and $\mathbf{K}_{n+1} \in \mathcal{I}_{\alpha-(n+1)+\gamma}$.

By the definition of the commutator, we have

$$[\mathbf{X}_i,\mathbf{M}^{-1}]_{n+1} = \mathbf{X}_i[\mathbf{X}_i,\mathbf{M}^{-1}]_n - [\mathbf{X}_i,\mathbf{M}^{-1}]_n\mathbf{X}_i, \tag{3.3.31}$$

which leads us to the relation

$$\begin{aligned}[\mathbf{X}_i,\mathbf{M}^{-1}]_{n+1} \quad = \quad & \mathbf{M}^{-1}\big(-[\mathbf{X}_i,\mathbf{M}]_{n+1} \\ & -[\mathbf{X}_i,\mathbf{M}]_n\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}] \\ & -[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]_n + \mathbf{K}_{n+1}\big)\mathbf{M}^{-1}\end{aligned} \tag{3.3.32}$$

where

$$\begin{aligned}\mathbf{K}_{n+1} \quad = \quad & [\mathbf{X}_i,\mathbf{M}]_{n-1}\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}] && (3.3.33) \\ + \quad & 2[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]_{n-1}\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}] && (3.3.34) \\ + \quad & [\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]_{n-1} && (3.3.35) \\ + \quad & -[\mathbf{X}_i,[\mathbf{X}_i,\mathbf{M}]_{n-1}\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]] - [\mathbf{X}_i,[\mathbf{X}_i,\mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i,\mathbf{M}]_{n-1}] && (3.3.36) \\ + \quad & \mathbf{M}[\mathbf{M}^{-1}\mathbf{K}_n\mathbf{M}^{-1},\mathbf{X}_i]\mathbf{M}. && (3.3.37)\end{aligned}$$

It remains to show that $\mathbf{K}_{n+1} \in \mathcal{I}_{\alpha-(n+1)+\gamma}$. Since $[\mathbf{X}_i,\mathbf{M}]_{n-1} \in \mathcal{I}_{\alpha-n+1}$, it is clear that each of the three terms (3.3.33), (3.3.34), and (3.3.35) is an element of $\mathcal{I}_{\alpha-n+1}$. Additionally, we see that the term (3.3.36) is an element of $\mathcal{I}_{\alpha-n}$. The last term, (3.3.37), is an element of $\mathcal{I}_{\alpha-n+\gamma-1}$ since $\mathbf{K}_n \in \mathcal{I}_{\alpha-n+\gamma}$ and $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha-2}$. Thus, we have shown that $\mathbf{K}_{n+1} \in \mathcal{I}_{\alpha-(n+1)+\gamma}$ for $n < \alpha$. In particular, we may set $n = \alpha - 1$ to obtain $\mathbf{K}_\alpha \in \mathcal{I}_\gamma$, and we have completed the proof of Lemma 3.3.2.□

In the next step of the bootstrap proof of Proposition 3.3.1, we use Lemma 3.3.2. The technique we use is a slight extension of that of [37].

Lemma 3.3.3 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *(* $\alpha \in \mathbf{Z}$*,* $\alpha \geq 2$ *) and* $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha+\gamma-2}$*, where* $0 \leq \gamma \leq 1$*, then the matrix whose* $\mathbf{k}, \mathbf{l}$ *element is defined by*

$$|k_i - l_i|^{\beta+\alpha} |m^{-1}_{\mathbf{k},\mathbf{l}}| \tag{3.3.38}$$

is a bounded linear operator from l^1 *to* l^∞ *if* $\beta \leq 2$ *and* $\beta < 2 + \gamma$*.*

Proof: We note that $|k_i - l_i|^{\beta+\alpha} |m^{-1}_{\mathbf{k},\mathbf{l}}| = |k_i - l_i|^{\beta} ([\mathbf{X}_i, \mathbf{M}^{-1}]_\alpha)_{\mathbf{k},\mathbf{l}}$. We use equation (3.3.27) to break the right-hand side of this equation into three parts:

$$|k_i - l_i|^{\beta} (\mathbf{M}^{-1} [\mathbf{X}_i, \mathbf{M}]_\alpha \mathbf{M}^{-1})_{\mathbf{k},\mathbf{l}}, \tag{3.3.39}$$

$$|k_i - l_i|^{\beta} (\mathbf{M}^{-1} [\mathbf{X}_i, \mathbf{M}]_{\alpha-1} \mathbf{M}^{-1} [\mathbf{X}_i, \mathbf{M}] + [\mathbf{X}_i, \mathbf{M}] \mathbf{M}^{-1} [\mathbf{X}_i, \mathbf{M}]_{\alpha-1} \mathbf{M}^{-1})_{\mathbf{k},\mathbf{l}}, \tag{3.3.40}$$

and

$$|k_i - l_i|^{\beta} (\mathbf{M}^{-1} \mathbf{K}_\alpha \mathbf{M}^{-1})_{\mathbf{k},\mathbf{l}}, \tag{3.3.41}$$

Consider the matrices whose $\mathbf{k}, \mathbf{l}$ element is defined by one of these three expressions.

The matrices defined by (3.3.39) and (3.3.41) require some care. In both cases we make use of the inequality

$$|k - l|^\beta \leq C(|k - p|^\beta + |p - q|^\beta + |q - l|^\beta) \tag{3.3.42}$$

for $0 \leq \beta \leq 2$. Using this inequality, we see that (3.3.39) yields

$$|k_i - l_i|^\beta \sum_{\mathbf{p},\mathbf{q}} |m^{-1}_{\mathbf{k},\mathbf{p}}| |(p_i - q_i)^\alpha m_{\mathbf{p},\mathbf{q}}| |m^{-1}_{\mathbf{q},\mathbf{l}}| \leq D \sum_{\mathbf{p},\mathbf{q}} (a_{\mathbf{p},\mathbf{q}} + b_{\mathbf{p},\mathbf{q}} + c_{\mathbf{p},\mathbf{q}}) \tag{3.3.43}$$

where

$$a_{\mathbf{p},\mathbf{q}} = |k_i - q_i|^\beta |m^{-1}_{\mathbf{k},\mathbf{p}}| |(p_i - q_i)^\alpha m_{\mathbf{p},\mathbf{q}}| |m^{-1}_{\mathbf{q},\mathbf{l}}|, \tag{3.3.44}$$

$$b_{\mathbf{p},\mathbf{q}} = |m^{-1}_{\mathbf{k},\mathbf{p}}| |p_i - q_i|^{\alpha+\beta} |m_{\mathbf{p},\mathbf{q}}| |m^{-1}_{\mathbf{q},\mathbf{l}}|, \tag{3.3.45}$$

and

$$c_{\mathbf{p},\mathbf{q}} = |m^{-1}_{\mathbf{k},\mathbf{p}}| |(p_i - q_i)^\alpha m_{\mathbf{p},\mathbf{q}}| |q_i - l_i|^\beta |m^{-1}_{\mathbf{q},\mathbf{l}}|. \tag{3.3.46}$$

We show (under certain restrictions on β) that each of the matrices $A = \left\{\sum_{\mathbf{p},\mathbf{q}} a_{\mathbf{p},\mathbf{q}}\right\}$, $B = \left\{\sum_{\mathbf{p},\mathbf{q}} b_{\mathbf{p},\mathbf{q}}\right\}$, and $C = \left\{\sum_{\mathbf{p},\mathbf{q}} c_{\mathbf{p},\mathbf{q}}\right\}$ is a bounded linear operator from l^1 to l^∞. Thus, we may use the inequality (3.3.43) to show that the matrix defined by (3.3.39) is a bounded linear operator from l^1 to l^∞.

Each of the three matrices A, B, and C is a product of three of the following matrices. The matrix defined by

- $|m^{-1}_{\mathbf{k},\mathbf{p}}|$ is bounded on l^1 and l^∞.
- $|k_i - p_i|^\beta |m^{-1}_{\mathbf{k},\mathbf{p}}|$ is an element of $\mathcal{I}_{\alpha-2+\gamma-\beta}$ and therefore (1) is bounded from l^1 to l^r if $(\alpha+\gamma-\beta) > \frac{2}{r}$, by Young's inequality; and (2) is bounded from l^r to l^∞ if $\frac{2}{r} > 2 - (\alpha+\gamma-\beta)$, again by Young's inequality.
- $|p_i - q_i|^{\beta+\alpha} |m_{\mathbf{p},\mathbf{q}}|$ is an element of $\mathcal{I}_{-\beta}$ and if $\beta \leq 2$ it is bounded from l^1 to l^∞.
- $|p_i - q_i|^{\alpha} |m_{\mathbf{p},\mathbf{q}}|$ is an element of $\mathcal{I}_0$ and so is bounded from l^r to $l^{r+\epsilon}$ for any $\epsilon > 0$.

Therefore, we see that the matrix A is bounded from l^1 to l^∞ if there exists an $r > 1$ such that $\frac{2}{r} > 2 - (\alpha+\gamma-\beta)$. Since $\alpha \geq 2$, we can only guarantee this if $\frac{1}{r} > \frac{\beta-\gamma}{2}$. Since $r > 1$, this implies that we require $\beta < 2 + \gamma$. Likewise, the matrix C is bounded from l^1 to l^∞ if there exists an $r \geq 1$ such that $\frac{2}{r} < \alpha+\gamma-\beta$. Since $\gamma \geq 0$, we can guarantee this if $\beta < \alpha+\gamma$. Since $\alpha \geq 2$, this is already covered by the condition for the matrix A. Additionally, the matrix B is bounded from l^1 to l^∞ for any $\beta \leq 2$.

Therefore, we have shown that the matrix defined by (3.3.39) is a bounded operator from l^1 to l^∞ if $\beta \leq 2$ and $\beta < 2 + \gamma$.

We show the matrix defined by (3.3.41) to be a bounded linear operator from l^1 to l^∞ with the same restrictions on β. If $\alpha = 2$, then this is trivially true since $\mathbf{K}_2 = 0$. For $\alpha > 2$, the technique is again to write

$$|k_i - l_i|^\beta \sum_{\mathbf{p},\mathbf{q}} |m^{-1}_{\mathbf{k},\mathbf{p}}||(\mathbf{K}_\alpha)_{\mathbf{p},\mathbf{q}}||m^{-1}_{\mathbf{q},\mathbf{l}}| \leq D \sum_{\mathbf{p},\mathbf{q}} \left(\tilde{a}_{\mathbf{p},\mathbf{q}} + \tilde{b}_{\mathbf{p},\mathbf{q}} + \tilde{c}_{\mathbf{p},\mathbf{q}}\right), \qquad (3.3.47)$$

where

$$\tilde{a}_{\mathbf{p},\mathbf{q}} = |k_i - q_i|^\beta |m^{-1}_{\mathbf{k},\mathbf{p}}||(\mathbf{K}_\alpha)_{\mathbf{p},\mathbf{q}}||m^{-1}_{\mathbf{q},\mathbf{l}}|, \qquad (3.3.48)$$

$$\tilde{b}_{\mathbf{p},\mathbf{q}} = |m^{-1}_{\mathbf{k},\mathbf{p}}||p_i - q_i|^\beta |(\mathbf{K}_\alpha)_{\mathbf{p},\mathbf{q}}||m^{-1}_{\mathbf{q},\mathbf{l}}|, \tag{3.3.49}$$

and

$$\tilde{c}_{\mathbf{p},\mathbf{q}} = |m^{-1}_{\mathbf{k},\mathbf{p}}||(\mathbf{K}_\alpha)_{\mathbf{p},\mathbf{q}}||q_i - l_i|^\beta |m^{-1}_{\mathbf{q},\mathbf{l}}|. \tag{3.3.50}$$

We use the fact that $\mathbf{K}_\alpha \in \mathcal{I}_\gamma$ to derive the same restrictions on β that were derived for the matrix (3.3.39). The details are very similar to those of the previous case, so we do not display them here.

The matrix defined by (3.3.40) is dealt in two different ways depending on the value of α. First, consider the case where $\alpha > 2$. If $\beta \leq 2$, then we see that the matrix defined by (3.3.40) is at worst an element of $\mathcal{I}_{-1}$, so there exists a uniform bound on its elements. Any such matrix is a bounded linear operator from l^1 to l^∞, as follows from Young's inequality.

If $\alpha = 2$, then at best we can say only that $\mathbf{M}^{-1} \in \mathcal{I}_0$. The class $\mathcal{I}_0$ does not form a multiplicative algebra, so for $\alpha = 2$ the matrix (3.3.40), given by

$$|k_i - l_i|^\beta (\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1}[\mathbf{X}_i, \mathbf{M}]\mathbf{M}^{-1})_{\mathbf{k},\mathbf{l}}, \tag{3.3.51}$$

requires a similar trick to that of (3.3.39). Namely, the factor $|k_i - l_i|^\beta$ is "spread" across the five matrices in the product. The technique is virtually identical to that which has been previously described, and it yields the same restrictions on β. We do not repeat the argument here. Therefore, we have shown that the matrix defined by (3.3.38) is a bounded linear operator from l^1 to l^∞ if $\beta \leq 2$ and $\beta < 2 + \gamma$, and thus have proved Lemma 3.3.3.□

Now, we use Lemmas 3.3.2 and 3.3.3 in sequence to prove Proposition 3.3.1. We have already shown that $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha-2}$. In terms of Lemmas 3.3.2 and 3.3.3, this means that we may set $\gamma = 0$. The two lemmas then imply that the matrices defined by $|k_i - l_i|^{\beta+\alpha}|m^{-1}_{\mathbf{k},\mathbf{l}}|$ $(i = 1, 2)$ are both bounded linear operators from l^1 to l^∞ if $\beta < 2$. In order to use Lemmas 3.3.2 and 3.3.3 again, we choose $\beta = 1$ and see that, therefore, $\mathbf{M}^{-1} \in \mathcal{I}_{\alpha-1}$. Thus, in the application of Lemma 3.3.2 and Lemma 3.3.3 again, we may now set $\gamma = 1$. This allows us to choose $\beta \leq 2$, from which we see that $\mathbf{M}^{-1} \in \mathcal{I}_\alpha$, which yields Theorem 3.3.2.□

The remaining task for completion of the proof of Theorem 3.3.2 is to prove Proposition 3.3.2, which says that if $\mathbf{M} \in \mathcal{I}_\alpha$ and $\mathbf{M}$ is invertible on l^2, then $\mathbf{M}^{-1}$ has a bounded inverse on all l^p spaces.

Proof of Proposition 3.3.2. The main idea of our proof comes from [37]. First we define the matrix $\mathbf{M}_{(N)}$ by

$$(\mathbf{M}_{(N)})_{\mathbf{k},\mathbf{l}} = \begin{cases} 0 & |\mathbf{k}-\mathbf{l}| > N \\ m_{\mathbf{k},\mathbf{l}} & |\mathbf{k}-\mathbf{l}| \leq N \end{cases}. \tag{3.3.52}$$

We define the remainder as $\mathbf{R} = \mathbf{M} - \mathbf{M}_{(N)}$. Then, by making estimates on the norm of $\mathbf{M}_{(N)}$, $\mathbf{M}_{(N)}^{-1}$, and $\mathbf{M} - \mathbf{M}_{(N)}$, we will show that, by making the bandwidth of $\mathbf{M}_{(N)}$ large enough we can make the perturbation $\mathbf{M} - \mathbf{M}_{(N)}$ small enough so that the inverse of $\mathbf{M}$ can be shown to exist. We construct the proof in a series of lemmas.

First, we claim that

Lemma 3.3.4 *If* $\mathbf{M} \in \mathcal{I}_\alpha$, *then* $\|\mathbf{R}\|_{1,1} + \|\mathbf{R}\|_{\infty,\infty} \leq C(\mathbf{M},\alpha)N^{-\alpha}$.

To prove this, we see that

$$\|\mathbf{R}\|_{1,1} \leq C \sup_{\|x\|_1=1} \sum_{\mathbf{k}\in\mathbf{Z}^2} \sum_{|\mathbf{k}-\mathbf{l}|>N} (1+|\mathbf{k}-\mathbf{l}|)^{-2-\alpha}|x_\mathbf{l}| \tag{3.3.53}$$

$$\leq \sum_{\mathbf{l}\in\mathbf{Z},|\mathbf{l}|>N} (1+|\mathbf{l}|)^{-2-\alpha} \tag{3.3.54}$$

by Young's inequality. We then write

$$\sum_{\mathbf{l}\in\mathbf{Z},|\mathbf{l}|>N} (1+|\mathbf{l}|)^{-2-\alpha} = \sum_{|l_1|>N} \sum_{l_2\in\mathbf{Z}} (1+|l_1|+|l_2|)^{-2-\alpha} + \tag{3.3.55}$$

$$\sum_{|l_1|\leq N} \sum_{|l_2|>N-|l_1|} (1+|l_1|+|l_2|)^{-2-\alpha}. \tag{3.3.56}$$

We estimate

$$\begin{aligned}
\sum_{|l_1|>N} \sum_{l_2\in\mathbf{Z}} (1+|l_1|+|l_2|)^{-2-\alpha} &= \sum_{|l_1|>N} \left((1+|l_1|)^{-2-\alpha} + 2\sum_{l_2=1}^{\infty} (1+|l_1|+|l_2|)^{-2-\alpha} \right) \\
&\leq \sum_{|l_1|>N} \left((1+|l_1|)^{-2-\alpha} + 2\int_0^\infty (1+|l_1|+y)^{-2-\alpha} dy \right) \\
&= \sum_{|l_1|>N} \left((1+|l_1|)^{-2-\alpha} + \frac{2}{1+\alpha}(1+|l_1|)^{-1-\alpha} \right) \\
&\leq 2\int_N^\infty (1+x)^{-2-\alpha} + \frac{2}{1+\alpha}(1+x)^{-1-\alpha} dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{1+\alpha}(1+N)^{-1-\alpha} + \frac{4}{\alpha+\alpha^2}(1+N)^{-\alpha} \\
&\leq CN^{-\alpha},
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{|l_1|\leq N}\sum_{|l_2|>N-|l_1|}(1+|l_1|+|l_2|)^{-2-\alpha} &= 4\sum_{0\leq l_1\leq N}\sum_{l_2<l_1-N}(1+l_1-l_2)^{-2-\alpha} \\
&\leq 4\sum_{0\leq l_1\leq N}\int_{-\infty}^{l_1-N}(1+l_1-y)^{-2-\alpha}dy \\
&= 4\sum_{0\leq l_1\leq N}\frac{1}{1+\alpha}(1+N)^{-1-\alpha} \\
&\leq CN^{-\alpha},
\end{aligned}
$$

and so we arrive at

$$
\|\mathbf{R}\|_{1,1} \leq CN^{-\alpha}. \tag{3.3.57}
$$

Additionally, $\|\mathbf{R}\|_{\infty,\infty} = \|\mathbf{R}^*\|_{1,1}$, and $\|\mathbf{R}^*\|_{1,1}$ satisfies an inequality of the same form since each step in the derivation of the inequality is also valid for $\mathbf{R}^*$, so the result of Lemma 3.3.4 is proved.

We use the above Lemma and Theorem 3.2.2 to prove the following lemma:

Lemma 3.3.5 *If $\mathbf{M} \in \mathcal{I}_\alpha$ is invertible on l^p and $l^{p'}$, where $1/p + 1/p' = 1$, and $p \leq p'$, then for N large enough, $\mathbf{M}_{(N)}$ is invertible on l^p and $l^{p'}$.*

Proof: By Theorem 3.2.2 and Lemma 3.3.4,

$$
\|\mathbf{R}\|_{q,q} \leq \|\mathbf{R}\|_{1,1}^{1/q}\|\mathbf{R}\|_{\infty,\infty}^{1-1/q} \leq C(\mathbf{M},\alpha)N^{-\alpha}, \tag{3.3.58}
$$

where $q = p$ and $q = p'$, so, for N large enough, we know that

$$
\|\mathbf{R}\|_{q,q} < (\|\mathbf{M}^{-1}\|_{q,q})^{-1}. \tag{3.3.59}
$$

Thus,

$$
\|\mathbf{R}\mathbf{M}^{-1}\|_{q,q} \leq \|\mathbf{R}\|_{q,q}\|\mathbf{M}^{-1}\|_{q,q} < 1, \tag{3.3.60}
$$

which means that the series $\sum_{n=0}^{\infty}(\mathbf{R}\mathbf{M}^{-1})^n$ is convergent in the (q,q)-norm. But, then,

$$
\mathbf{M}_{(N)}^{-1} = (\mathbf{M}-\mathbf{R})^{-1} = \mathbf{M}^{-1}(\mathbf{I}-\mathbf{R}\mathbf{M}^{-1})^{-1} = \mathbf{M}^{-1}\sum_{n=0}^{\infty}(\mathbf{R}\mathbf{M}^{-1})^n, \tag{3.3.61}
$$

which means that $\mathbf{M}_{(N)}^{-1}$ exists in the (p,p)-norm and the (p',p')-norm.□

Lemma 3.3.6 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *is invertible on* l^2, *then* $\mathbf{M}_{(N)}$ *is invertible on* l^1 *and* l^∞.

Proof: Clearly, $\mathbf{M}_{(N)} \in \mathcal{X}$ and $\mathbf{M}_{(N)}$ has an inverse so $\mathbf{M}_{(N)}^{-1} \in \mathcal{X}$. But all elements of $\mathcal{X}$ are bounded operators on l^1 and l^∞.

Now, we estimate the norm of $\mathbf{M}_{(N)}$ in terms of N and p.

Proposition 3.3.3 *If* $\mathbf{M} \in \mathcal{I}_\alpha$, *and* $\mathbf{M}$ *is invertible on* l^p, *and* $l^{p'}$ *(where* $1/p + 1/p' = 1$, $1 \leq p \leq 2$*), then* $\|\mathbf{M}_{(N)}^{-1}\|_{1,1} + \|\mathbf{M}_{(N)}^{-1}\|_{\infty,\infty} \leq C(\mathbf{M},p)N^{3-2/p}\log N$.

Proof: The proof of this proposition involves many parts and each will be proved as a separate lemma. Throughout, we set $s = 3 - 2/p + \alpha$.

Lemma 3.3.7 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *and* $\mathbf{M}$ *is invertible on* l^p, $1 \leq p \leq 2$, *then*

$$|(k_i - k_i')^2(\mathbf{M}_{(N)})_{k_1,k_1',k_2,k_2'}| \leq C(p)N^{3-2/p}\left(1 + |k_1 - k_1'| + |k_2 - k_2'|\right)^{-s} \tag{3.3.62}$$

$$|(k_i - k_i')(\mathbf{M}_{(N)})_{k_1,k_1',k_2,k_2'}| \leq C(p)N^{2-2/p}\left(1 + |k_1 - k_1'| + |k_2 - k_2'|\right)^{-s} \tag{3.3.63}$$

Proof: First, consider integers k and l such that $1 \leq k \leq N$, $1 \leq l \leq N$. Clearly,

$$(k+l)^{3-\frac{2}{p}} \leq C(p)N^{3-\frac{2}{p}}. \tag{3.3.64}$$

Furthermore,

$$(1 + \frac{1}{k+l}) \leq 2, \tag{3.3.65}$$

so

$$(k+l)^{3-\frac{2}{p}}(1 + \frac{1}{k+l})^{1-\frac{2}{p}} \leq C(p)N^{3-\frac{2}{p}}. \tag{3.3.66}$$

Thus,

$$(k+l)^2(1+k+l)^{1-\frac{2}{p}} \leq C(p)N^{3-\frac{2}{p}}, \tag{3.3.67}$$

which yields

$$(k+l)^2(1+k+l)^{-2-\alpha} \leq C(p)N^{3-\frac{2}{p}}(1+k+l)^{-(3-\frac{2}{p}+\alpha)}. \tag{3.3.68}$$

Of course,

$$|(k_i - k_i')^2| \leq |(k_1 - k_1')^2 + (k_2 - k_2')^2| \leq (|k_1 - k_1'| + |k_2 - k_2'|)^2, \tag{3.3.69}$$

so combining (3.3.68) and (3.3.69) yields

$$|(k_i - k_i')^2 (\mathbf{M}_{(N)})_{k_1,k_1',k_2,k_2'}| \leq C(p) N^{3-\frac{2}{p}} (1 + |k_1 - k_1'| + |k_2 - k_2'|)^{-s} \tag{3.3.70}$$

provided $k_1 \neq k_1'$, $k_2 \neq k_2'$; this case may be handled by adjusting the constant $C(p)$. Thus, we have proved the inequality (3.3.62); the inequality (3.3.63) is proved similarly.□

Lemma 3.3.8 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *and* $\mathbf{M}$ *is invertible on* l^p, $1 \leq p \leq 2$, *then, given* $\epsilon \geq 0$, $[\mathbf{X}_i, \mathbf{M}_{(N)}]$ *and* $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]$ *are continuous from* l^q *to* $l^{q+\epsilon}$, *where* $q \geq 1$.

Proof: Since we have assumed $\alpha \geq 2$ we see that $s \geq 3$. Via Young's inequality (Theorem 3.2.1), we obtain

$$\begin{aligned} \|[\mathbf{X}_i, \mathbf{M}_{(N)}]\|_{q,q} &= \sup_{\|x\|_q=1} \|[\mathbf{X}_i, \mathbf{M}_{(N)}]x\|_q \\ &\leq C(p) N^{2-2/p} \sup_{\|x\|_q=1} \|\{(1+|k_1|+|k_2|)^{-s}\} * x\|_q \\ &\leq C(p) N^{2-2/p} \sup_{\|x\|_q=1} \|\{(1+|k_1|+|k_2|)^{-s}\}\|_1 \|x\|_q \\ &\leq C(\mathbf{M}, p) N^{2-2/p}. \end{aligned} \tag{3.3.71}$$

Likewise, we can show that $\|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]\|_{q,q} \leq C(\mathbf{M}, p) N^{3-2/p}$ if $s > 2$. The result is proved by noting that $\|\mathbf{T}\|_{q,q+\epsilon} \leq \|\mathbf{T}\|_{q,q}$ for all operators $\mathbf{T}$.□

Lemma 3.3.9 *If* $\mathbf{M} \in \mathcal{I}_\alpha$ *and* $\mathbf{M}$ *is invertible on* l^p, $1 \leq p \leq 2$, *and* $1/p + 1/p' = 1$, *then* $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]$ *is continuous from* l^p *to* $l^{p'}$ *and*

$$|k_i - k_i'|^2 |(\mathbf{M}_{(N)})^{-1}_{k_1,k_1',k_2,k_2'}| \leq C(\mathbf{M}, \alpha) N^{3-2/p}. \tag{3.3.72}$$

Proof: Since $\mathbf{M}$ is invertible on l^2, we know from Lemma 3.3.5 that $\mathbf{M}_{(N)}$ is invertible on l^2; we are assuming that $\mathbf{M}$ is invertible on l^p and $l^{p'}$, so Lemma 3.3.5 gives us invertibility of $\mathbf{M}_{(N)}$ on l^p and $l^{p'}$ also. We have the commutator identity

$$[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]] = \mathbf{M}_{(N)}^{-1} (-[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]] + 2[\mathbf{X}_i, \mathbf{M}_{(N)}] \mathbf{M}_{(N)}^{-1} [\mathbf{X}_i, \mathbf{M}_{(N)}]) \mathbf{M}_{(N)}^{-1}. \tag{3.3.73}$$

By Lemma 3.3.8, $[\mathbf{X}_i, \mathbf{M}_{(N)}]$ is continuous from l^p to l^2 and from l^2 to $l^{p'}$. Thus, the continuity of $\mathbf{M}_{(N)}^{-1}$ on l^2 assures the continuity of $[\mathbf{X}_i, \mathbf{M}_{(N)}] \mathbf{M}_{(N)}^{-1} [\mathbf{X}_i, \mathbf{M}_{(N)}]$ from l^p to $l^{p'}$. Furthermore, Lemma 3.3.8 shows that $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]$ is continuous from l^p to $l^{p'}$. Thus, the identity (3.3.73) shows that $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]$ is continuous from l^p to $l^{p'}$.

The identity (3.3.73) shows that

$$\begin{aligned} \|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]\|_{p,p'} \quad \leq \quad & \|\mathbf{M}_{(N)}^{-1}\|_{p,p}\|\mathbf{M}_{(N)}^{-1}\|_{p',p'} \times \\ & \big(\|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]\|_{p,p'} \\ & + \|\mathbf{M}_{(N)}^{-1}\|_{2,2}\|[\mathbf{X}_i, \mathbf{M}_{(N)}]\|_{p,2}\|[\mathbf{X}_i, \mathbf{M}_{(N)}]\|_{2,p'}\big). \end{aligned} \tag{3.3.74}$$

By the inequalities in the proof of Lemma (3.3.9), we see that

$$\|[\mathbf{X}_i, \mathbf{M}_{(N)}]\|_{p,2}\|[\mathbf{X}_i, \mathbf{M}_{(N)}]\|_{2,p'} \leq CN^{4-4/p} \tag{3.3.75}$$

and

$$\|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]\|_{p,p'} \leq CN^{3-2/p}, \tag{3.3.76}$$

which imply

$$\|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]\|_{p,p'} \leq \|\mathbf{M}_{(N)}^{-1}\|_{p,p}\|\mathbf{M}_{(N)}^{-1}\|_{p',p'} \left(CN^{3-2/p} + C\|\mathbf{M}_{(N)}^{-1}\|_{2,2}N^{4-4/p}\right). \tag{3.3.77}$$

Clearly $\|\mathbf{M}_{(N)}^{-1}\|_{q,q}$ has a uniform upper bound in N since $\|\mathbf{M}_{(N)}^{-1}\|_{q,q} \to \|\mathbf{M}^{-1}\|_{q,q}$ as $N \to \infty$ when $q = p, p', 2$. Since $1 \leq p \leq 2$, we know that $N^{4-4/p} \leq N^{3-2/p}$, so we have

$$\|[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]\|_{2,2} \leq CN^{3-2/p}, \tag{3.3.78}$$

which immediately implies the inequality (3.3.72).□

Lemma 3.3.10 *If $\mathbf{M} \in \mathcal{I}_\alpha$ and $\mathbf{M}$ is invertible on l^p, $1 \leq p \leq 2$, then for any $0 \leq \epsilon \leq 1$,*

$$|(\mathbf{M}_{(N)})^{-1}_{\mathbf{k},\mathbf{l}}| \leq C(\mathbf{M}, \alpha)N^{3-\frac{2}{p}}N^{((J-3)+\frac{2}{p})\epsilon}(1 + |\mathbf{k} - \mathbf{l}|)^{-2-\epsilon}, \tag{3.3.79}$$

where $J \geq 3$.

Proof: By construction, we have that

$$|(\mathbf{M}_{(N)})_{\mathbf{k},\mathbf{l}}| \leq CN^3(1 + |\mathbf{k}, \mathbf{l}|)^{-5-\alpha} \tag{3.3.80}$$

which implies that $[\mathbf{X}_i, \mathbf{M}_{(N)}]$, $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]$, and $[\mathbf{X}_i, [\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]]$ are continuous on l^2 with norms less than CN^3. Therefore, since the $[\mathbf{X}_i, [\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}^{-1}]]]$ may be written as a finite sum of products between $\mathbf{M}_{(N)}^{-1}$ and the commutators $[\mathbf{X}_i, \mathbf{M}_{(N)}]$, $[\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]$, and $[\mathbf{X}_i, [\mathbf{X}_i, [\mathbf{X}_i, \mathbf{M}_{(N)}]]]$, it is continous on l^2 and its norm is less than $C(\mathbf{M})N^J$, where $J \geq 3$. Since the norm of an operator on l^2 is also a bound on the elements of the matrix which represents the operator, we know that

$$|(\mathbf{M}_{(N)})_{\mathbf{k},\mathbf{l}}^{-1}| \leq C(\mathbf{M}, \alpha) N^{27} (1 + |\mathbf{k} - \mathbf{l}|)^{-3}. \tag{3.3.81}$$

Now we can combine the inequalities (3.3.72) and (3.3.81) by raising one to the $1 - \epsilon$ power and the other to the ϵ power and taking their product; this yields

$$|(\mathbf{M}_{(N)})_{\mathbf{k},\mathbf{l}}^{-1}| \leq C(\mathbf{M}, \alpha) N^{(3-\frac{2}{p})(1-\epsilon)} N^{J\epsilon} (1 + |\mathbf{k} - \mathbf{l}|)^{-2(1-\epsilon)-3\epsilon}, \tag{3.3.82}$$

which after simplification becomes the inequality (3.3.79).□

The proof of Proposition 3.3.3 is almost complete. To complete it, consider that

$$\begin{aligned} \|\mathbf{M}_{(N)}^{-1}\|_{1,1} &\leq C(\mathbf{M}, \alpha) N^{3-\frac{2}{p}} N^{(J-3+\frac{2}{p})\epsilon} \sup_{\|x\|_1=1} \|\{\mathcal{P}(2+\epsilon) * x\|_1 \\ &\leq C(\mathbf{M}, \alpha) N^{3-\frac{2}{p}} N^{(J-3+\frac{2}{p})\epsilon} \sup_{\|x\|_1=1} \|\mathcal{P}(2+\epsilon)\|_1 \|x\|_1 \\ &\leq C(\mathbf{M}, \alpha) N^{3-\frac{2}{p}} N^{(J-3+\frac{2}{p})\epsilon} \|\mathcal{P}(2+\epsilon)\|_1 \end{aligned} \tag{3.3.83}$$

by Young's Inequality. The term $\|\mathcal{P}(2+\epsilon)\|_1$ can be bounded by $4 + \frac{4}{(\epsilon+\epsilon^2)} + \frac{2}{1+\epsilon}$ by an integral comparison. Now we set $\epsilon = (\log N)^{-1}$ in (3.3.83) using the bound
$\|\mathcal{P}(2+\epsilon)\|_1 \leq 4 + \frac{4}{(\epsilon+\epsilon^2)} + \frac{2}{1+\epsilon}$; after some further estimates, this yields

$$\|\mathbf{M}_{(N)}^{-1}\|_{1,1} \leq C(\mathbf{M}, \alpha) N^{3-\frac{2}{p}} e^{J-3+\frac{2}{p}} \log N \tag{3.3.84}$$

for N large enough. Of course, this estimate is also true for $(\mathbf{M}_{(N)}^{-1})^*$, and since $\|\mathbf{M}_{(N)}^{-1}\|_{\infty,\infty} = \|(\mathbf{M}_{(N)}^{-1})^*\|_{1,1}$ it is also true for $\|\mathbf{M}_{(N)}^{-1}\|_{\infty,\infty}$. Thus the inequality of Proposition 3.3.3 is proved.□

Lemma 3.3.11 *Let* $p_1 \in [1, p]$, $\quad t \in [0, 1]$ *such that*

$$\frac{1}{p_1} = t + \frac{1-t}{p}. \tag{3.3.85}$$

If $\mathbf{M} \in \mathcal{I}_\alpha$ *and* $\mathbf{M}$ *is invertible on* l^p *and* $l^{p'}$*, where* $1 \leq p \leq 2$ *and* $1/p + 1/p' = 1$*, then*

$$\|\mathbf{M}_{(N)}^{-1}\|_{p_1,p_1} + \|\mathbf{M}_{(N)}^{-1}\|_{p_1',p_1'} \leq C(\mathbf{M}, \alpha) N^{(3-\frac{2}{p})t} (\log N)^t \tag{3.3.86}$$

where $\frac{1}{p_1} + \frac{1}{p_1'} = 1$.

Proof: By Theorem 3.2.2,

$$\|\mathbf{M}_{(N)}^{-1}\|_{p_1,p_1} \le \|\mathbf{M}_{(N)}^{-1}\|_{p,p}^{t} \|\mathbf{M}_{(N)}^{-1}\|_{1,1}^{t}. \tag{3.3.87}$$

However, $\|\mathbf{M}_{(N)}^{-1}\|_{p,p}$ is uniformly bounded in N, so

$$\|\mathbf{M}_{(N)}^{-1}\|_{p_1,p_1} \le C(\mathbf{M},\alpha) \|\mathbf{M}_{(N)}^{-1}\|_{1,1}^{t}. \tag{3.3.88}$$

Now,

$$\|\mathbf{M}_{(N)}^{-1}\|_{p_1',p_1'} \le C(\mathbf{M},\alpha) \|\mathbf{M}_{(N)}^{-1}\|_{\infty,\infty}^{t} \tag{3.3.89}$$

by Theorem 3.2.2, so combining Proposition 3.3.3 with the inequalities (3.3.88) and (3.3.89) yields the result of this lemma.□

Lemma 3.3.12 $\mathbf{M}$ *is invertible on* l^{p_1} *and* $l^{p_1'}$ *as soon as* $(3 - \frac{2}{p})t < \alpha$.

Proof: The inequality

$$\|\mathbf{M}_{(N)}^{-1}\mathbf{R}\|_{p_1,p_1} + \|\mathbf{M}_{(N)}^{-1}\mathbf{R}\|_{p_1',p_1'} \le C(\mathbf{M},\alpha) N^{(3-\frac{2}{p})t-\alpha} (\log N)^t \tag{3.3.90}$$

follows immediately from the inequality (3.3.58) and Lemma 3.3.11. Thus, if $(3 - \frac{2}{p})t < \alpha$, then, for large N, $C(\mathbf{M},\alpha) N^{(3-\frac{2}{p})t-\alpha} (\log N)^t$ will be less than 1. Therefore, the series $\sum_{n=0}^{\infty} (-1)^n (\mathbf{M}_{(N)}^{-1} R)^n$ will be convergent in the (p_1, p_1) and (p_1', p_1') norms. Now, from Lemma 3.3.6 and the inequality (3.3.58), we know that $\mathbf{M}_{(N)}$ is invertible on all $1 \le q \le \infty$. So, on l^{p_1} and $l^{p_1'}$ we can write :

$$\mathbf{M}^{-1} = (\mathbf{M}_{(N)} + \mathbf{R})^{-1} = (\mathbf{I} + \mathbf{M}_{(N)}^{-1}\mathbf{R})^{-1} \mathbf{M}_{(N)}^{-1} = \left(\sum_{n=0}^{\infty} (-1)^n (\mathbf{M}_{(N)}^{-1}\mathbf{R})^n \right) \mathbf{M}_{(N)}^{-1} \tag{3.3.91}$$

which is convergent in the (p_1, p_1) and (p_1', p_1') norms as long as $(3 - \frac{2}{p})t < \alpha$.□

Finally, the proof of Proposition 3.3.2 is now possible. Starting from the fact that $\mathbf{M}$ is invertible on l^2, we know that it is invertible on l^{p_1} and $l^{p_1'}$ as long as $2t < \alpha$, where t and p_1 satisfy the equation (3.3.85). Since $\alpha \ge 2$, we may continue choosing successive p's which satisfy (3.3.85) and reach any $1 \le p < 2$ in a finite number of steps; therefore, $\mathbf{M}$ is invertible on all l^p spaces, and the result of Proposition 3.3.2 is proved.□

CHAPTER 4

CONCLUSIONS AND FURTHER DIRECTIONS

In Chapter 1, we described the multiresolution homogenization approach in the context of linear ODE's. We compared the multiresolution approach with existing approaches and found that classical results may be obtained using the multiresolution technique.

In Chapter 2, we discussed the generalization of the multiresolution approach to partial differential equations. We showed that the multiresolution reduction procedure preserves the sparsity of operators which are compressible in the wavelet basis, and also approximately preserves small eigenvalues of elliptic operators.

In Chapter 3, we proved results concerning algebras of bi-infinite matrices and tensors. We showed that bi-infinite matrices or tensors with exponential or polynomial decay away from the diagonal form an algebra under inversion, and that, if a matrix in either class is invertible, then its inverse is in that class as well. These results were used in Chapter 2.

In this chapter, we describe directions for future research.

4.1 Multiresolution Reduction of Hyperbolic and Parabolic Partial Differential Equations

This section outlines further work and explains the importance of the fact that the reduction procedure preserves small eigenvalues of elliptic equations. In particular, there are implications for solving hyperbolic and parabolic initial value problems.

Consider, for example, the differential equation

$$u_{tt}(x,t) + \mathbf{S}u(x,t) = 0, \tag{4.1.1}$$

where $\mathbf{S}$ is a second-order elliptic operator with variable coefficients, supplemented with some boundary conditions and the initial conditions

$$u(x,0) = g(x), \quad u_t(x,t)|_{t=0} = 0. \tag{4.1.2}$$

This equation describes (for example) wave propagation in a medium with variable velocity. Consider a problem where the velocity is changing very rapidly (one may think of a highly stratified rock structure) but the initial condition $g(x)$ has relatively low wavenumbers, i.e. the wavelength of the initial condition is large compared with the typical length over which the velocity changes. A space-discretization of this problem would typically require a step-size smaller than the smallest length over which the velocity changes, which may be prohibitively expensive in practical applications. The key point of this section is that this difficulty may be overcome by replacing $\mathbf{S}$ by the reduced operator on some scale.

4.1.1 Reduction of Hyperbolic PDE's

Let us project $\mathbf{S}$ onto $\mathbf{V}_j$ and write, as usual, $\mathbf{S}_j = \mathbf{P}_j\mathbf{S}\mathbf{P}_j$. We assume that the space $\mathbf{V}_j$ has fine-enough resolution to capture smallest features of the behavior of the coefficients of $\mathbf{S}$. Consider the eigenvalue problem

$$\mathbf{S}_j v_n^j(x) = \lambda_n^j v_n^j(x), \tag{4.1.3}$$

where $v_n^j(x)$ satisfies the boundary conditions. The eigenvalues $\{\lambda_n^j\}$ of $\mathbf{S}_j$ are all real and positive, and we enumerate them in ascending order. The eigenvectors $\{v_n^j(x)\}$ of $\mathbf{S}_j$ form an orthonormal basis for $\mathbf{V}_j$.

We may, therefore, look for solutions of the equation

$$\left(\frac{\partial}{\partial t}\right)^2 u^j(x,t) + \mathbf{S}_j u^j(x,t) = 0 \tag{4.1.4}$$

in the form

$$u^j(x,t) = \sum_n (a_n \cos(\sqrt{\lambda_n^j}\, t) + b_n \sin(\sqrt{\lambda_n^j}\, t))\, v_n^j(x). \tag{4.1.5}$$

Satisfying the initial conditions

$$u^j(x,0) = g^j(x), \quad \frac{\partial}{\partial t} u^j(x,t)|_{t=0} = 0, \tag{4.1.6}$$

we obtain

$$u^j(x,t) = \sum_n a_n \cos((\lambda_n^j)^{\frac{1}{2}}\, t)\, v_n^j(x), \tag{4.1.7}$$

where the coefficients a_n are obtained from $\mathbf{P}_j g(x) = g^j(x) = \sum_n a_n v_n^j(x)$.

We observe that there is no mixing between the eigenfunctions $\{v_n^j(x)\}$ over time. In particular, if $g^j(x) = \sum_{n=0}^{K} a_n v_n^j(x)$ and $\mathbf{V}_{j+k}$ is an ϵ-approximation of the span of $\{v_n^j(x)\}_{n=0}^{n=K}$, then we may approximate solutions of (4.1.4) projected onto $\mathbf{V}_{j+k}$ by solutions of

$$\left(\frac{\partial}{\partial t}\right)^2 u^{j+k}(x,t) + \mathbf{R}_{j+k} u^{j+k}(x,t) = 0, \tag{4.1.8}$$

where $\mathbf{R}_{j+k}$ is the k-step reduction of $\mathbf{S}_j$, with the initial conditions

$$u^{j+k}(x,0) = g^{j+k}(x), \quad \frac{\partial}{\partial t} u^{j+k}(x,t)|_{t=0} = 0. \tag{4.1.9}$$

Solving (4.1.8) on $\mathbf{V}_{j+k}$ is less expensive than solving (4.1.4) on $\mathbf{V}_j$ since, in a compact domain, there are $2^{d\,k}$-times as many degrees of freedom in $\mathbf{V}_j$ than in $\mathbf{V}_{j+k}$, where d is the spatial dimension.

4.1.2 Reduction of Parabolic PDE's

The considerations for the hyperbolic case above also apply in the parabolic case

$$u_t(x,t) + \mathbf{S}u(x,t) = 0 \tag{4.1.10}$$

(with boundary and initial conditions). The situation for (4.1.10) is even more favorable. Namely, on $\mathbf{V}_j$ we may write the solution of (4.1.10) in the form

$$u^j(x,t) = \sum_n a_n \mathrm{e}^{-\lambda_n^j t}\, v_n^j(x), \tag{4.1.11}$$

where $\lambda_n^j > 0$. Similar considerations as in the hyperbolic case apply. In addition, if we are interested in the long-time solution, (due to the factor $\mathrm{e}^{-\lambda_n^j t}$) only those eigenvectors corresponding to small eigenvalues will contribute, and we may replace $\mathbf{S}_j$ by $\mathbf{R}_{j+k}$ *regardless* of which eigenvectors constitute the initial condition $g(x)$.

The efficiency of the reduction procedure in the hyperbolic and parabolic case depends on the quality of the approximations of the eigenvectors of $\mathbf{S}_j$ by functions in subspaces of the

MRA. Since the eigenvectors of $\mathbf{S}_j$ satisfy the boundary conditions, it is very important to use an MRA where the scaling functions on coarse scales satisfy the same boundary conditions; further work is required in this direction.

Procedures for reduction of wave propagation models have been extensively studied (see e.g. [11]), but most results are concerned with situations where there is a preferred direction, thus enabling the use of methods suitable for ODE's. As far as we know, no method has been proposed that addresses the problem where all directions of wave propagation are allowed.

BIBLIOGRAPHY

[1] K. Aki and P. Richards. *Quantitative Seismology: Theory and Methods.* W.H. Freeman, San Francisco, 1980.

[2] B. Alpert. A Class of Bases in l^2 for the Sparse Representation of Integral Operators. *SIAM J. Math. Anal*, 24(1):246–262, 1993.

[3] H. Alzer. An inequality for arithmetic and harmonic means. *Aequationes Math.*, 46(3):257–263, 1993.

[4] G. E. Backus. Long-wave elastic anistropy produced by horizontal layering. *J. Geophys. Res.*, 67(11):4427–4440, 1962.

[5] R. E. Bank, T. F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numer. Math.*, 52(4):427–458, 1988.

[6] R. Beals. Characterization of pseudodifferential operators and applications. *Duke Math. J.*, 44(1):45–57, March 1977.

[7] C. Bender and S. Orszag. *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill Book Co., New York, 1978.

[8] A. Bensoussan, J.L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures.* North-Holland Pub. Co., New York, 1978.

[9] G. Beylkin, R. R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Comm. Pure and Appl. Math.*, 44:141–183, 1991. Yale University Technical Report YALEU/DCS/RR-696, August 1989.

[10] M. E. Brewster and G. Beylkin. A Multiresolution Strategy for Numerical Homogenization. *ACHA*, 2:327–349, 1995. PAM Report 187, 1994.

[11] R. Burridge, G. S. Papanicolaou, and B. S. White. One-dimensional wave propagation in a highly discontinuous medium. *Wave Motion*, 10(1):19–44, 1988.

[12] R.M. Christensen. *Mechanics of Composite Materials.* John Wiley and Sons, Inc., 1979.

[13] R. Coifman, P.L. Lions, Y. Meyer, and S. Semmes. Compensated compactness and hardy spaces. *J. Math. Pures Appl.*, 72:247–286, 1993.

[14] W. Dahmen and L. Elsner. Algebraic multigrid methods and the Schur complement. In *Robust multi-grid methods (Kiel, 1988)*, Notes Numer. Fluid Mech., 23, pages 58–68. Vieweg, Braunschweig, 1989.

[15] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Series in Applied Mathematics. SIAM, 1992.

[16] M. Dorobantu. *Wavelet-based Algorithms for Fast PDE Solvers.* PhD thesis, Royal Institute of Technology, Stockholm University, 1995.

[17] M. Dorobantu and B. Engquist. Wavelet-based numerical homogenization. Preprint, 1996.

[18] H. Foerster, K. Stüben, and U. Trottenburg. *Elliptic Problem Solvers*, chapter "Non-standard multigrid techniques using checkered relaxation and intermediate grids". Academic Press, New York, 1981.

[19] A. Gilbert. A comparison of multiresolution and classical homogenization schemes. *To appear in Applied and Computational Harmonic Analysis.*, 1997.

[20] D.L. Gines, G. Beylkin, and J. Dunn. LU factorization of non-standard forms and direct multiresolution solvers. *PAM Report 278*, 1996.

[21] S. Jaffard. Propriétés des matrices *bien localisees* près de leur diagonale et quelques applications. *Annales de l'IHP, Series Problemes non linéaires*, 7(5):461–476, 1990.

[22] S. Jaffard and Y. Meyer. Bases d'ondelettes dans des ouverts de $\mathbf{r}^n$. *J. Math. Pures Appl. (9)*, 68(1):95–108, 1989.

[23] D. Jerison. The diameter of the first nodal line of a convex domain. *Annals of Mathematics*, 141:1–33, 1995.

[24] V.V. Jikov, S.M. Kozlov, and O.A. Oleinik. *Homogenization of Differential Operators and Integral Functionals.* Springer-Verlag, NY, 1994.

[25] T. Kato. *Perturbation Theory for Linear Operators.* Springer-Verlag New York, Inc., New York, 1996.

[26] S. Knapek. Matrix-dependent multigrid-homogenization for diffusion problems. *Preprint*, 1996.

[27] F. Murat. Compacité par compensation. *Ann. Scuola Norm. Sup. Pisa Cl. Sci*, 5(4):485–507, 1978.

[28] A.R. Nahmod. Generalized uncertainty principles on spaces of homogeneous type. *Journal of Functional Analysis*, 119:171–209, 1994.

[29] A. Norris. Waves in Periodically Layered Media: a Comparison of Two theories. *SIAM J. Appl. Math.*, 53:1195–1209, 1993.

[30] A. Norris and F. Santosa. Shear wave propagation in a periodically layered medium - an asymptotic theory. *Wave Motion*, 16:33–55, 1992.

[31] V.V. Prasolov. *Problems and Theorems in Linear Algebra*, volume 134 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, R.I., 1994.

[32] X. Saint Raymond. *Elementary Introduction to the Theory of Pseudodifferential Operators.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1991.

[33] B. F. Smith, P. E. Bjørstad, and W. D. Gropp. *Domain decomposition.* Cambridge University Press, Cambridge, 1996. Parallel Multilevel Methods for Elliptic Partial Differential Equations.

[34] E.M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, 1971.

[35] B. Z. Steinberg and J. J. McCoy. Towards effective parameter theories using multiresolution decomposition. *J. Acoust. Soc. Am.*, 96:1130–1143, 1994.

[36] L. Tartar. *Compensated Compactness and Applications to Partial Differential Equations.* Heriot-Watt Sympos. vol. IV. Pitman, NY, 1979.

[37] P. Tchamitchian. *Wavelets: Theory and Application*, chapter Wavelets, Functions, and Operators, pages 83–181. ICASE/LaRC Series in Computational Science and Engineering. Oxford University Press, New York, 1996. Edited by G. Erlebacher, M.Y. Hussaini and L. Jameson.

Appendix A

WAVELETS, MULTIRESOLUTION ANALYSES, AND OPERATORS

A.1 Wavelets and Multiresolution Analyses

In this section, we set our notation and give a brief description of the concept of multiresolution analysis (MRA) and wavelets. For details, we refer to e.g. [15].

A.1.1 Notation and Preliminary Considerations

As usual, we consider a chain of subspaces

$$\ldots \subset \mathbf{V}_2 \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots \tag{A.1.1}$$

such that

$$\bigcap_j \mathbf{V}_j = \{0\} \text{ and } \overline{\bigcup_j \mathbf{V}_j} = \mathbf{L}^2(\Omega) \tag{A.1.2}$$

where Ω is some domain in $\mathbf{R}^d$. If the domain Ω is bounded, then there is a coarsest space $\mathbf{V}_0$ and, instead of (A.1.1), we write

$$\mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \mathbf{V}_{-2} \subset \ldots \tag{A.1.3}$$

The subspace $\mathbf{V}_j$ is spanned by an orthonormal basis $\{\phi^j_k(x) = 2^{-j/2}\phi(2^{-j}x - k)\}_{k\in\mathbf{Z}}$. The function ϕ is called the scaling function, and it satisfies the two-scale difference equation

$$\phi(x/2) = \sqrt{2}\sum_k h_k \phi(x-k). \tag{A.1.4}$$

We consider the space V_j to specify a *scale* or *resolution* of the space of $\mathbf{L}^2$ functions on Ω, and use the index j to identify the scale. As $j \to \infty$, the scale grows "coarser," and as

$j \to -\infty$, the scale grows "finer." Functions in $\mathbf{L}^2(\Omega)$ which are smooth or slowly-varying may be represented on a coarser scale of the MRA than those which are highly-oscillatory or have steep gradients.

We denote by $\mathbf{W}_j$ the orthogonal complement of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$, $\mathbf{V}_{j-1} = \mathbf{V}j \oplus \mathbf{W}j$ and use $\mathbf{P}_j$ and $\mathbf{Q}_j$ to denote the orthogonal projection operators onto $\mathbf{V}_j$ and $\mathbf{W}_j$. Note that $\mathbf{Q}_{j+1} = \mathbf{P}_j - \mathbf{P}_{j+1}$. If $x \in \mathbf{V}_j$, we write $s_x = \mathbf{P}_{j+1}x$ and $d_x = \mathbf{Q}_{j+1}x$, where $s_x \in \mathbf{V}_{j+1}$ and $d_x \in \mathbf{W}_{j+1}$. If $d = 1$, then the subspace $\mathbf{W}_j$ is spanned by an orthonormal basis $\{\psi^j_k(x) = 2^{-j/2}\psi(2^{-j}x - k)\}_{k\in\mathbf{Z}}$. The function ψ is called the wavelet, and it may be computed using the scaling function ϕ via the two-scale relation

$$\psi(x/2) = \sqrt{2}\sum_k g_k \phi(x-k). \tag{A.1.5}$$

The space $\mathbf{W}_{j+1}$ represents the "detail" component of the space $\mathbf{V}_j$, and the function $\psi^j_k(x)$ captures the highly-oscillatory, quickly-varying component of functions in $\mathbf{V}_j$.

From (A.1.2), we see that

$$\mathbf{L}^2(\Omega) = \bigoplus_j \mathbf{W}_j. \tag{A.1.6}$$

If $d \geq 2$, then (for rectangular domains) the basis in the subspace $\mathbf{W}_j$ may be constructed using products of wavelets and scaling functions. For example, if $d = 2$, then functions $\{\psi^j_k(x)\psi^j_{k'}(y), \phi^j_k(x)\psi^j_{k'}(y), \psi^j_k(x)\phi^j_{k'}(y),\}_{k,k'\in\mathbf{Z}}$ form an orthonormal basis of $\mathbf{W}_j$.

An important property of the wavelet is vanishing moments, i.e. orthogonality to low-degree polynomials:

$$\int \psi(x)x^k\,dx = 0 \quad \text{for} \quad k = 0, 1, \ldots, M-1. \tag{A.1.7}$$

If the above property holds, then we say that the wavelet has M vanishing moments.

An example of an MRA in which the wavelets have 1 vanishing moment is the Haar system, defined by

$$\phi(x) = \begin{cases} 1 & \text{if} \quad 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{A.1.8}$$

and

$$\psi(x) = \begin{cases} 1 & \text{if} \quad 0 \le x < \frac{1}{2} \\ -1 & \text{if} \quad \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{A.1.9}$$

There are many examples of wavelets with more vanishing moments. The Haar basis is the only (anti) symmetric orthogonal wavelet basis with compact support. Daubechies in [15] constructed compactly supported wavelets with more vanishing moments.

The sequences h and g in (A.1.4) and (A.1.5) may be either finite or infinite. Given a function

$$f^j(x) = \sum_k f_k^j \phi_k^j(x) \tag{A.1.10}$$

in $\mathbf{V}_j$, we may compute the coefficients of s_f and d_f in the bases ϕ^{j+1} and ψ^{j+1}. Using the relations (A.1.4) and (A.1.5), we derive

$$s_f^j(x) = \sum_k (s_f^j)_k \phi_k^{j+1}(x), \quad (s_f^j)_k = \sum_l h_{l-2k} f_l \tag{A.1.11}$$

and

$$d_f^j(x) = \sum_k (d_f^j)_k \phi_k^{j+1}(x). \quad (d_f^j)_k = \sum_l g_{l-2k} f_l \tag{A.1.12}$$

The projection operations represented by equations (A.1.11) and (A.1.12) may be performed by convolving the sequences h and g with the sequence f and keeping only the even-numbered elements, a process which is called convolution-decimation. We call the mapping of f^j to s_f^j and d_f^j the wavelet decomposition. We may iterate it over many scales to obtain

$$f^j = \sum_{j'=j_0}^{j} d_f^{j'} + s_f^{j_0}. \tag{A.1.13}$$

If $d = 1$ and Ω is a bounded interval, then the vector of coefficients of f^j will be finite in length (which we denote by N). The vectors of coefficients of s^j and d^j will then both be of length $N/2$. Thus, the finite-dimensional wavelet decomposition takes a vector of length N and produces two vectors, each of length $N/2$. In this situation, we may define the coordinate transformation and projection from $\mathbf{V}_j$ to $\mathbf{V}_{j+1}$ by a $N/2 \times N$ matrix H. Likewise, we may define the matrix G to be the coordinate transformation and projection from $\mathbf{V}_j$ to $\mathbf{W}_{j+1}$.

In this context, we represent the projection $s_f = \mathbf{P}_{j+1} f$ as application of the matrix H to the vector of coefficients of f.

Just as we have defined the wavelet decomposition, we may also define the wavelet reconstruction (or synthesis). Given functions $s(x) \in \mathbf{V}_{j+1}$ and $d(x) \in \mathbf{W}_{j+1}$ with coefficients s_k and d_k, we may write

$$f(x) = \sum_k f_k \phi_k^j(x) = s(x) + d(x), \tag{A.1.14}$$

where

$$f_k = \sum_l \left(h_{k-2l} s_l + g_{k-2l} d_l \right). \tag{A.1.15}$$

A.1.2 Fourier Analysis

The sequences h and g are *filters* which are applied to sequences. In the Fourier domain, we represent these filters as trigonometric polynomials defined by

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_k h_k e^{-ik\xi} \tag{A.1.16}$$

and

$$m_1(\xi) = \frac{1}{\sqrt{2}} \sum_k g_k e^{-ik\xi}. \tag{A.1.17}$$

We define $\hat{g}$ to be the Fourier transform of the function g as follows:

$$\hat{g}(\xi) = \frac{1}{2\pi} \int e^{-i\xi x} g(x)\, dx. \tag{A.1.18}$$

The relations (A.1.4) and (A.1.5) may be recast in the Fourier domain as

$$\hat{\phi}(\xi) = m_0(\xi/2)\hat{\phi}(\xi/2) \tag{A.1.19}$$

and

$$\hat{\psi}(\xi) = m_1(\xi/2)\hat{\phi}(\xi/2). \tag{A.1.20}$$

The functions m_0 and m_1 (and hence the sequences h and g) define a *quadrature-mirror filter*. The function m_0 is a "low-pass" filter and captures low-frequency components in the Fourier domain; m_1 is a "high-pass" filter and captures the high-frequency components in the Fourier domain. We list here some properties of the functions m_0 and m_1:

1. $|m_0(\xi)|^2 + |m_1(\xi)|^2 = 1$

2. $m_0(\xi) = 1$

3. If the wavelet has M vanishing moments, then

$$\left(\frac{d}{d\xi}\right)^k m_1(\xi)\Big|_{\xi=0} = 0 \quad \text{for} \quad k = 0, \dots, M-1 \tag{A.1.21}$$

$$\left(\frac{d}{d\xi}\right)^k m_0(\xi)\Big|_{\xi=0} = 0 \quad \text{for} \quad k = 1, \dots, M-1 \tag{A.1.22}$$

The convolution-decimation operations in (A.1.11) and (A.1.12) may be represented in the Fourier domain. Note that, via (A.1.10), we derive

$$\hat{f}(\xi) = \tilde{f}(\xi)\hat{\phi}(\xi). \tag{A.1.23}$$

where $\tilde{f}(\xi) = \sum_k f_k^j e^{-ik\xi}$. From this we obtain

$$\hat{s}_f(\xi) = \tilde{s_f}(\xi)\hat{\phi}(\xi/2) \tag{A.1.24}$$

and

$$\hat{d}_f(\xi) = \tilde{d}_f(\xi)\hat{\phi}(\xi/2), \tag{A.1.25}$$

where

$$\tilde{s}_f(2\xi) = m_0(\xi)\tilde{f}(\xi) + m_1(\xi)\tilde{f}(\xi + \pi) \tag{A.1.26}$$

and

$$\tilde{d}_f(2\xi) = m_1(\xi)\tilde{f}(\xi) + m_0(\xi)\tilde{f}(\xi + \pi). \tag{A.1.27}$$

Equations (A.1.26) and (A.1.27) are the Fourier-domain equivalent of convolution-decimation of the sequence f with the sequences h and g, respectively. Figure A.1 shows typical low- and high-pass filters m_0 and m_1. Both functions are 2π-periodic. The function $m_0(\xi)$ in general has most of its mass between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$; $m_1(\xi)$ is likewise centered around $\xi = \pi$.

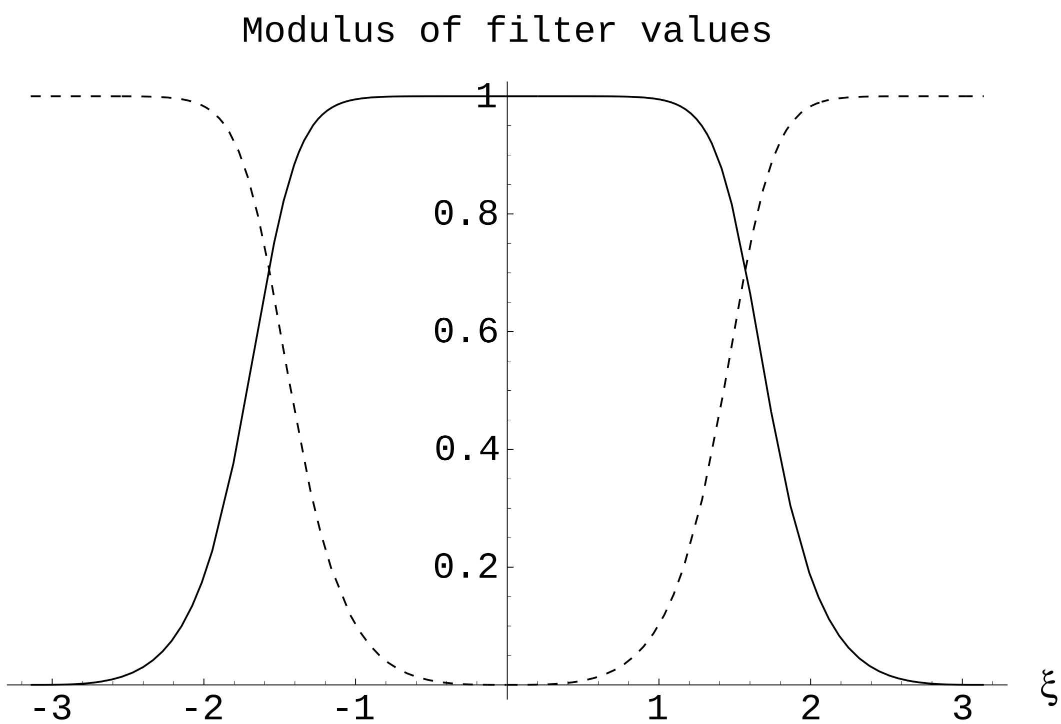

Figure A.1: Plots on $[\pi, \pi]$ of the modulus of the low-pass filter $m_0(\xi)$ (solid line) and the high-pass filter $m_1(\xi)$ (dashed line). In this example the filters were derived from non-compact orthogonal spline wavelets of degree 2. Clearly, m_0 low-frequency and m_1 is high-frequency.

A.2 Operators in the Wavelet Basis

In the wavelet basis, certain classes of operators may be represented using matrices with relatively few significant coefficients, so that fast application of matrices and matrices and matrices to vectors may be achieved (see [9] for more details).

In this section we set the notation for representation of operators in the wavelet basis and describe some features of the "non-standard form."

A.2.1 Notation and Preliminary Considerations

Given a bounded linear operator $\mathbf{S}$ on $\mathbf{L}^2(\mathbf{R}^d)$, consider its projection $\mathbf{S}_j$ on $\mathbf{V}_j$, $\mathbf{S}_j = \mathbf{P}_j\mathbf{S}\mathbf{P}_j$. Since $\mathbf{V}_j$ is a subspace spanned by translations of ϕ^j, we may represent the operator $\mathbf{S}_j$ as a (possibly infinite) matrix in that basis. With a slight abuse of notation, we will use the same symbol $\mathbf{S}_j$ to represent both the operator and its matrix. Since $\mathbf{V}_j = \mathbf{V}_{j+1} \oplus \mathbf{W}_{j+1}$, we may also write $\mathbf{S}_j : \mathbf{V}_j \to \mathbf{V}_j$ in a block form

$$\mathbf{S}_j = \begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{C}_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix} : \mathbf{V}_{j+1} \oplus \mathbf{W}_{j+1} \to \mathbf{V}_{j+1} \oplus \mathbf{W}_{j+1}, \tag{A.2.1}$$

where

$$\begin{aligned} \mathbf{A}_{\mathbf{S}_j} &= \mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{Q}_{j+1}, \\ \mathbf{B}_{\mathbf{S}_j} &= \mathbf{Q}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1}, \\ \mathbf{C}_{\mathbf{S}_j} &= \mathbf{P}_{j+1}\mathbf{S}_j\mathbf{Q}_{j+1}, \\ \mathbf{T}_{\mathbf{S}_j} &= \mathbf{P}_{j+1}\mathbf{S}_j\mathbf{P}_{j+1}. \end{aligned} \tag{A.2.2}$$

We note that $\mathbf{T}_{\mathbf{S}_j} = \mathbf{S}_{j+1}$. Each of the operators in (A.2.2) may be considered as a matrix. We note, however, that in the matrix form the transition from $\mathbf{S}_j$ in (A.2.1) to $\begin{pmatrix} \mathbf{A}_{\mathbf{S}_j} & \mathbf{B}_{\mathbf{S}_j} \\ \mathbf{C}_{\mathbf{S}_j} & \mathbf{T}_{\mathbf{S}_j} \end{pmatrix}$ requires application of the wavelet transform. We will use the operator notation throughout this Thesis and comment, if necessary, on the required numerical computations. For example, if $d = 1$ and $\mathbf{S}_j$ is finite and of size N by N, then each operator block in (A.2.2) is of size $\frac{N}{2}$ by $\frac{N}{2}$. In terms of linear algebra, when $d = 1$ and $\mathbf{S}_j$ is finite, we compute the matrices $\mathbf{A}_{\mathbf{S}_j}$, $\mathbf{B}_{\mathbf{S}_j}$, $\mathbf{C}_{\mathbf{S}_j}$, and $\mathbf{T}_{\mathbf{S}_j}$ by applying the wavelet decomposition first to the rows the matrix $\mathbf{S}_j$, and then apply the wavelet decomposition again to the columns of that matrix.

The operators (and their matrix representations) in (A.2.2) are referred to as the $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{T}$ *blocks* of $\mathbf{S}_j$. Also, for an operator $\mathbf{Z}$, we use notation $\mathbf{A_Z}$, $\mathbf{B_Z}$, $\mathbf{C_Z}$, and $\mathbf{T_Z}$ to indicate its blocks.

A.2.2 The Standard and Non-Standard Forms

In the standard form, the idea is to represent the operator $\mathbf{S}_j$ in the coordinates of the decomposition (A.1.13). If $d = 1$, this is done by applying the wavelet decomposition over several scales to the rows and columns of the matrix $\mathbf{S}_j$. Thus, the standard form is simply a representation of the original matrix $\mathbf{S}_j$ in a different basis.

The non-standard form (see e.g. [9]) is an alternative representation of the operator which is not (strictly speaking) a representation of the matrix in a different basis. We start with the telescoping series

$$\mathbf{S} = \sum_j \mathbf{P}_{j-1}\mathbf{S}\mathbf{P}_{j-1} - \mathbf{P}_j\mathbf{S}\mathbf{P}_j. \tag{A.2.3}$$

We note that $\mathbf{Q}_j = \mathbf{P}_{j-1} - \mathbf{P}_j$ and rewrite this series as

$$\mathbf{S} = \sum_j \mathbf{A}_j + \mathbf{B}_j + \mathbf{C}_j \tag{A.2.4}$$

where $\mathbf{A}_j = \mathbf{Q}_j\mathbf{S}\mathbf{Q}_j$, $\mathbf{B}_j = \mathbf{Q}_j\mathbf{S}\mathbf{P}_j$, and $\mathbf{C}_j = \mathbf{P}_j\mathbf{S}\mathbf{Q}_j$. If $\mathbf{S}_{j_0}$ is a bounded operator on $\mathbf{V}_{j_0}$, then we see that

$$\mathbf{S}_{j_0} = \left(\sum_{j=j_0}^{j_1} \mathbf{A}_j + \mathbf{B}_j + \mathbf{C}_j \right) + \mathbf{P}_{j_1}\mathbf{S}_{j_0}\mathbf{P}_{j_1}. \tag{A.2.5}$$

This equation shows that we may think of the operator $\mathbf{S}_{j_0}$ as a coarse-scale component $\mathbf{P}_{j_1}\mathbf{S}_{j_0}\mathbf{P}_{j_1}$ together with interactions between succesively finer scales. Slightly more work is required to apply the operator in this form to another operator or a vector.

If the operator $\mathbf{S}$ is an integral operator with kernel $K(x, y)$, then the entries of the matrices which represent the operators $\mathbf{A}_j$, $\mathbf{B}_j$, and $\mathbf{C}_j$ are given by

$$a^j_{k,k'} = \int_\Omega \int_\Omega K(x, y)\psi^j_k(x)\psi^j_{k'}(y)\, dx\, dy \tag{A.2.6}$$

$$b^j_{k,k'} = \int_\Omega \int_\Omega K(x,y)\psi^j_k(x)\phi^j_{k'}(y)\,dx\,dy \tag{A.2.7}$$

$$c^j_{k,k'} = \int_\Omega \int_\Omega K(x,y)\phi^j_k(x)\psi^j_{k'}(y)\,dx\,dy \tag{A.2.8}$$