ON LOCAL MODE ANALYSIS IN MULTI-GRID METHODS

By

MARK ALLAN REIMERS

B. Sc., The University of Toronto, 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Mathematics)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1983

© Mark Allan Reimers, 1983

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 22 August 1983

DE-6 (3/81)

ABSTRACT

Local Mode Analysis was introduced by Achi Brandt as a heuristic practical tool for determining the expected convergence rate of a Multi-Grid algorithm. It is shown that this tool may be justified rigorously. Furthermore, computations analogous to actual relaxation processes yield results much like Local Mode Analysis predictions. This analysis is also useful for a heuristic understanding of certain relaxations independently of the Multi-Grid context.

TABLE OF CONTENTS

Abstract.....ii

List of Tables.....v

List of Figures.....vi

Acknowledgement.....vii

Chapter 1 - Introduction

1.1 The Subject.....1

1.2 Outline of this Thesis.....4

Chapter 2 - Background

2.1 Finite Difference Solution of Elliptic Partial Differential Equations.....6

2.2 The Multi-Grid (MG) Method.....11

2.3 Local Mode Analysis.....17

Chapter 3 - A Justification of Local Mode Analysis

3.1 The Meaning of Local Mode Analysis.....21

3.2 Accuracy of Local Mode Analysis.....23

3.3 Examples.....35

3.4 Comments and Miscellaneous Consequences.....41

Chapter 4 - A Heuristic Look at Relaxations using Local Mode Analysis

4.1 The SOR Method.....46

4.2 Alternating Direction Implicit Iteration.....51

4.3 Smoothing Property of Relaxations.....55

Chapter 5 - Discussion

5.1 Applicability of Local Mode Analysis.....57

5.2 The Potential of LMA Outside the MG Context..58

5.3 Possible Developments of the Theorem.....59

5.4 Summary.....61

Bibliography.....62

Appendix A.....64

Appendix B.....67

TABLES

Table I - Attrition of Fourier Components of Gauss-Seidel Relaxation of Laplace's Equation in the Unit Square.....38

Table II - Line Relaxation.....39

Table III - Distributive Relaxation of the Cauchy-Riemann Equations.....41

Table IV - Best Values of ω.....50

FIGURES

Figure 1 - Orderings: (a) Lexicographic (b) Red-Black.10

Figure 2 - Relation of Fine (f) and Coarse (C) Grids..13

Figure 3 - Numbering of Grid Points.....24

Figure 4 - Entries in R1 Corresponding to Grid Points on the Boundary.....32

Figure 5 - Contour Lines of $|\mu(\theta)|$ for SOR Relaxation.48

ACKNOWLEDGMENT

I would like to thank Robert Miura for many encouraging talks, even though my work was outside his field of research, and Uri Ascher, who first exposed me to Multi-Grid, and Marc Lemieux, a very friendly office-mate. Most of all I would like to thank Mary Regester who put up with me during all the consummate aggravation of producing this work.

CHAPTER ONE

INTRODUCTION

1.1 The Subject

Numerical techniques for the solution of partial differential equations (PDEs) are finding increasing application in science and industry. Much current research is devoted to finding fast and robust algorithms for this purpose. Elliptic PDEs arise from steady state problems, and commonly the first step in their solution is the translation of the continuous problem into a large sparse system of algebraic equations by the use of a finite difference or finite element discretization. These systems are frequently solved by first approximating the solution very roughly, and then reducing the error in this aproximation by an iterative method (relaxation scheme).

In the Multi-Grid (MG) method, relaxation techniques are used, not to eliminate the error in an approximation to the solution, but only to smooth this error. (For a description of relaxation techniques and the MG method, the reader may refer to Chapter Two.) Since the practical introduction of the MG method by Achi Brandt in 1971, it has shown great early promise and is now being investigated systematically. A crucial problem for applications is the estimation of the rate of convergence of a MG algorithm. This requires an estimation of the efficiency of the smoothing performed by the relaxation.

Local Mode Analysis (LMA) or Local Fourier Analysis was introduced by Brandt (see Brandt(1977)) to realistically

estimate the convergence rate of the smoothing step in a Multi-Grid algorithm. Like the von Neumann stability analysis for parabolic systems, LMA treats a model system without boundaries, i.e., on a grid which is infinite in extent. LMA can be used to obtain good estimates of the (rapid) convergence rate of the MG algorithm as a whole. Brandt has not given a rigorous justification of this procedure (on this topic, see Brandt (1982)).

Other mathematicians have tried to develop rigourous convergence estimates for MG methods, but they have not found LMA useful for their theories. Most such rigourous convergence estimates have been too conservative (by several orders of magnitude) to be of any practical value. Therefore LMA remains the practical tool for realistic estimation and decision making with regard to MG methods.

Despite its practical success, the applicability of LMA to real problems has been neither rigourously justified nor disproven. It is not even clear how the analysis of an infinite model system ought to be brought to bear on the finite system at hand. Thus far mathematical work related to LMA has taken three directions.

Trottenberg and Stueben (1982) carried out a rigourous Fourier analysis of certain special MG situations (essentially Dirichlet problems in a square using Red-Black or Richardson relaxation for smoothing). They derived convergence rates for the smoothing step which turned out to be identical with convergence rates derived via an extension of LMA for the same problems. They claimed that the identity of these results gave credibility to the application of LMA in more general

settings. A second avenue of the mathematical work has been the elaboration of Brandt's original notion of an infinite grid model. Hemker (1980a) described in detail the Fourier analysis of infinite grid analogues of prolongation and restriction operators. Further Hemker (1980b, 1981), Mol (1981), and de Vries (1982) analysed an infinite grid model of the Incomplete L-U Decomposition (ILU) relaxation process. However, they did not rigourously justify the use of these estimates for finite grid problems.

It is worth noting that de Vries (1982) went further than Brandt's original intent regarding the range of applicabilty of LMA. de Vries computed the LMA convergence rate estimate for a certain low frequency Fourier component. He then compared this with the actual spectral radius of the relaxation operator on a grid whose mesh spacing was determined by the frequency. These two values were close, especially for fine meshes, which suggested that LMA might be used to estimate the overall performance of relaxation schemes. This idea extends the original horizons of the method, which was envisaged to be realistic only for high frequencies (see Brandt 1977, 1982).

A third mathematical direction related to LMA is mentioned by Trottenberg and Stueben (1982), Mol(1981), and Hackbusch (1983). LMA may be interpreted as a rigourous analysis of a finite discrete problem which arises from a differential equation problem in a rectangle with doubly periodic boundary conditions. These authors imply that the similarity between the periodic problem and the (more usual) Dirichlet problem justifies the use of LMA for the latter.

Hackbusch (1983) went on to relate the periodic and Dirichlet cases explicitly. He treated the MG solution of singular perturbation problems in one dimension, which are to be smoothed by Gauss-Seidel relaxation. He derived an equation involving the periodic relaxation operator, the Dirichlet relaxation operator, and a rank one perturbation matrix. He used this to link the size of the Dirichlet relaxation operator to the size of the periodic operator in a special norm related to the high frequency components. This estimate is further used in a rigourous estimate of the convergence rate of the MG algorithm for this problem. Hackbusch is the only author who has made use of LMA type analysis in a rigourous treatment of MG.

1.2 Outline of this Thesis

In this work I treat two-dimensional Dirichlet problems on grids which approximate arbitrary domains. The discrete systems arising from these problems are relaxed by any of a large class of iterative schemes to be described. In Chapter Three, I interpret LMA as an approximate description of the relaxation of Fourier components on the finite grids. This description is asymptotically correct in the limit of very fine grids. To show this, I perturb the matrix describing the actual relaxation to obtain a matrix which operates on Fourier components exactly as described by LMA. In the case of rectangular domains, this perturbed operator is identical with the relaxation operator for a problem with periodic boundary conditions. I show that this perturbation is asymptotically small in norm, and hence validate LMA on fine grids. This

validation is illustrated with several examples worked out in detail, to show the accuracy of LMA. LMA idealizations for selected Fourier components are compared with actual relaxations over a range of problems and meshes.

In Chapter Four, I investigate, as did de Vries (1982), the possibility of using LMA to describe the overall convergence rate of relaxations outside the MG context. The usefulness of LMA in predicting the optimum relaxation parameter for the SOR scheme is illustrated over a range of problems and meshes. LMA is also used to heuristically explain the efficiency of the ADI method applied to Laplace's equation and the degradation in performance for non-symmetric operators.

Chapter Five is a discussion of some of the ideas behind this work, as well as a proposed extension.

I hope this work will build confidence in LMA and suggest its possible uses as a theoretical tool.

CHAPTER TWO

BACKGROUND

2.1 Finite Difference Solution of Elliptic Partial Differential Equations

In practice the solution of elliptic Partial Differential Equations (PDEs) is frequently only feasible by numerical methods. Commonly the (continuum) PDE is translated into a system of discrete equations (discretization). These equations may be solved either by a direct method or by an iterative algorithm. A recent development has been the Multi-Grid (MG) method in which there is interaction between the iterative solution of the equations and the discretization itself. The MG method has proven quite effective in treating many different types of problems. In this work one facet of the MG technique, Local Mode Analysis (LMA), is justified and applied to the study of iterative methods.

Consider a two-dimensional finite domain Ω and an elliptic partial differential operator L. We will consider only linear operators L. The problem is to find a function u on Ω which satisfies

$$Lu = f(x,y), \tag{2-1}$$

where f is a prescribed function on Ω and u and/or its derivatives satisfy some prescribed linear boundary conditions on $\partial\Omega$, the boundary of Ω.

Our finite difference solution of this problem proceeds as follows. Place a discrete set of N points on Ω, spaced

evenly in x and in y to form a rectangular grid. Frequently the same spacing is used in both co-ordinate directions and this spacing is called the mesh size h. There are several ways to approximate the boundary of a non-rectangular domain. We will discuss one of them in Chapter Three. The grid points are ordered so that numbers assigned to them form the components of a vector U which we call a grid vector.

Now we approximate the continuum equation for u by a system of N algebraic equations which determine the values of U at the N grid points. The components of U are to approximate the values of u on the corresponding grid points. We construct an equation at each interior grid point (x,y) by approximating the derivatives that appear in Lu(x,y) by finite differences, e.g.,

$$(2\text{-}2)\quad \partial^2 u/\partial x^2(x,y) = (1/h^2)(u(x+h,y)-2u(x,y)+u(x-h,y)) + O(h^2).$$

These difference approximations to the derivatives are used to build up an approximation L*U(x,y) to Lu(x,y). The discrete equation corresponding to a grid point (x,y) is then

$$(2\text{-}3)\qquad L^*U(x,y) = f(x,y).$$

Thus for each interior grid point we have an equation involving the component of U corresponding to that point and the components of U corresponding to neighbouring points. At grid points (x,y) on or near the boundary of Ω, similar approximate equations are set up, but now they must take into account the boundary conditions imposed on u. We obtain a

linear N x N system of equations $AX = b$, whose solution $X = U$ approximates the solution u of the PDE. The approximation generally becomes better as N increases. For further information on the discretization process see Young (1971).

Now we must find an efficient algorithm for solving $AX = b$. This is the most costly part of the finite difference solution of the PDE. For the large systems ($N > 10^3$) that are desirable to ensure accuracy, the usual Gauss elimination (GE) algorithm turns out to be too space- and time-consuming. Other direct methods, based on the Fourier transform or on reduction (transformation of the large sparse system into a smaller dense system), may be used to obtain exact solutions of some special discrete systems. These methods are faster than GE and the development of large core computers has made them practical for the solution of larger systems.

In general large systems can be handled more quickly if they are solved only approximately using iterative methods (or relaxation techniques). A common class of relaxation schemes are the fixed point iterations. It is these methods with which we shall be concerned. A fixed point iteration for the system $AX = b$ obtains a matrix B which approximates A and which is easily inverted. Then the function

$$h(X) = X - B^{-1}AX + B^{-1}b \tag{2-4}$$

has a fixed point $X = U$ the desired discrete solution. Choose any grid vector X^0 and construct a sequence of iterates X^0, $X^1 = h(X^0)$, $X^2 = h(X^1)$,... Let e be the

difference between an iterate X and U, e = X - U. Then the difference e' between the next approximation X' and U is given by

(2-5) $e' = (I - B^{-1}A)e.$

If the spectral radius of the amplification matrix $I - B^{-1}A$ is less than 1 then clearly the errors in the sequence of iterates will go to 0 and the sequence will converge to X = U.

There are simple sufficient conditions which are known to guarantee convergence of such an iteration scheme (see Young, 1971). For example: let A be split into A = B - (B - A); if B^{-1} and B - A are non-negative (a "regular" splitting), then the iteration will converge. Most of the common fixed point iterations produce regular splittings as outlined below.

In the Jacobi (or simultaneous displacement) method, B is the diagonal of A (see Young, 1971). In the Gauss-Seidel (G-S) method, B is the lower triangular part of A (including the diagonal). For line relaxation, B is the lower triangle and one of the non-zero superdiagonals of A. In Successive Over-Relaxation (SOR), B is $D + \omega L$ where D is the diagonal and L the lower triangle not including the diagonal of A. Incomplete L-U Decomposition (ILU) (see de Vries, 1982) finds matrices L, U, and R such that L is lower triangular, U is upper triangular and R is "small" in some sense and such that A = LU + R; B is LU and is therefore easily inverted. For sparse matrices A arising from PDEs L , U and R should be sparse. The Alternating-

Direction Implicit (ADI) method is not based on a simple splitting but rather on a sequence of two splittings. It will be discussed further later. The Conjugate-Gradient (CG) iteration is not of the type described above.

The practical implementation of a relaxation scheme entails more than just a knowledge of the splitting on which it is based. We will discuss the implementation of the simplest widely used relaxation, the Gauss-Seidel scheme. This method will be used frequently in examples.

The grid points are ordered in some manner. In that order each approximate solution value is adjusted to satisfy the one corresponding discrete equation. Depending on the ordering of the points, the determination of the value of the next iterate at a grid point may involve using the newly determined values at one or more neighbouring points as well as values of the

(a) (b)

Figure 1. Orderings: (a) Lexicographic and (b) Red-Black

current iterate. The most common orderings are lexicographic and red-black (see Figure 1). Lexicographic relaxation starts at one corner of the grid and proceeds along one grid line. When the end of the line is reached, relaxation starts again at the first position of the next line, and so on until all points on all grid lines have been modified. This constitutes

one iteration. Red-Black relaxation divides the grid points into two ordered subsets related to each other like the red squares and the black squares of a checkerboard. All the points of the red subset are relaxed in their order, and then all the points of the black subset.

2.2 The Multi-Grid (MG) Method

The best of the relaxation schemes mentioned above are better than Gauss Elimination, but they are still quite inefficient, and therefore still costly. Typically on the first few passes of a relaxation scheme, the error in the approximate solution (as measured by the size of the residuals $r = b - AX$) is decreased substantially (on the order of 10% to 50%). However, subsequent passes are not so efficient and long before the desired tolerance on the residuals is reached, they are being reduced by only a small fraction on each iteration. Very many relaxation sweeps are thus required to solve the equations.

For most of these schemes the theory shows that the spectral radius of the amplification matrix is of order $1 - O(h^k)$, where h is the mesh size. Depending on the scheme used and the specific problem, the power k of h varies from 3/2 to 5/2. In practice the convergence rate asymptotically approaches the spectral radius, since the components of the error along the eigenvectors of the amplification matrix, with the largest eigenvalues, gradually become dominant. That is, the quickly converging error components (small eigenvalues) are almost eliminated by the first few iterations, but the slowly converging components are worn down only gradually and

thus stand out after many passes. If we are to more efficiently solve the problem we need to find some other way to eliminate these components.

What do the slowly converging error components look like? The residuals after several applications of nearly any relaxation scheme are smoother than the residuals for the initial guess X^0. That is, there is more correlation between the residuals at neighbouring points, and fewer sign changes as one moves along a row of the grid. Thus it is natural to conjecture that the slowly converging error components are also smooth. Numerical computation of the eigenvalues of the amplification matrix for simple model problems bears out this intuition. The largest eigenvalues are generally associated with eigenvectors, whose values at neighbouring grid points are strongly correlated. On the other hand, the eigenvectors corresponding to small eigenvalues vary quickly along grid lines.

These observations suggest that the relaxation schemes might be better understood by decomposing the error $e = X - U$ in an iterate X, into a finite Fourier-type series. A heuristic analysis suggests, and this thesis makes it rigourous, that while most relaxation schemes are efficient at reducing the high frequency Fourier components of error, they are not efficient at reducing the low frequency components. Thus we cannot in general eliminate the slowly converging error components from one relaxation scheme simply by using a different scheme.

The idea behind the MG algorithm is to use the relaxation methods to do only what they do well - reducing the high

frequency components of the error, and to find some other method to eliminate the low frequency components. The problem then is to find a way to eliminate a "smooth" or slowly varying error in an approximate solution.

This may be approached as follows. Suppose several relaxation sweeps have been done on the discrete system $AX = b$ obtained from (2-1), yielding an approximate solution U^1, and suppose that the residuals $r^1 = b - AU^1$ are smooth. The determination of the exact discrete solution U is equivalent to the determination of the smooth error vector $e^1 = U^1 - U$ which solves $Ae^1 = r^1$. Consider $Ae^1 = r^1$ as the discretization of a PDE problem $Le = r$ in Ω, where e is to satisfy homogeneous boundary conditions. We might expect then that e will be smooth relative to the grid we are using for $Lu = f$. Then an accurate solution for e may be obtained with a much coarser grid. Thus we set up a finite difference system $A^*e^* = r^*$ by discretizing $Le = r$ on a coarser grid (see Figure 2). The solution e^* of this system should be a good

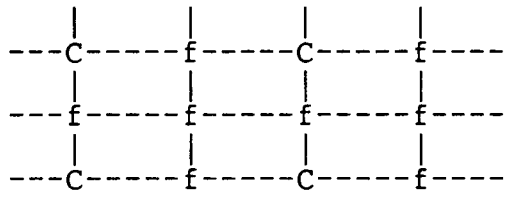


Figure 2. Relation of Fine (f) and Coarse (C) Grids

approximation on the points of the coarse grid to the smooth function e which is approximated on the fine grid by e^1. Since e is smooth, values of e on the fine grid points may be obtained accurately simply by interpolation from the coarse

grid points. Thus a good approximation for e^1 can be obtained from e^*, and this smooth error can be almost eliminated.

Why would this procedure (called Coarse Grid Correction (CGC)) be more efficient than simply eliminating e^1 by many more relaxations on the original fine grid? It is because each relaxation of $A^*e^* = r^*$ is much less expensive than a relaxation of $Ae^1 = r^1$ (which is equivalent to a further relaxation of $AX = b$). In addition fewer iterations will be required to solve the coarse grid system to a given tolerance than will be required to solve the fine grid system to the same tolerance. Thus a smooth error can be eliminated more quickly this way.

Of course the solution of the CGC problem may be accomplished in the same way; a few relaxation sweeps and then transfer to a yet coarser grid. Thus a recursive multiple-grid algorithm is devised. On each grid the rough component of the error is reduced by relaxation and the smooth (relative to the resolution of that grid) component is reduced via the next coarser grid.

These ideas may be implemented in several ways. The algorithm sketched below is referred to as an adaptive V-cycle correction scheme (see Brandt, 1977, 1982) Let the sequence of successively finer grids be $G^1, G^2, \ldots, G^n$; the desired solution is a grid vector on G^n. Let the mesh sizes be $h^1 > \ldots > h^n$; usually $h^m = 2h^{m+1}$. There are operators which transform grid vectors on one grid to grid vectors on another grid. Prolongation I^k_{k-1} acts on grid vectors on G^{k-1} and smoothly interpolates to yield values at all the grid points in G^k. The reverse procedure, restriction I^{k-1}_k, interpolates

values at grid points in G^{k-1} using the data supplied by the values of a grid vector on G^{k}. Both operations find a good representation on the target grid of a function which is represented by a vector on the original grid.

We solve a problem $A^{k} X^{k} = b^{k}$ on each grid G^{k} as follows. The system is relaxed several times until the error in the last approximate solution Z^{k} is judged to be "smooth". This is determined by measuring the rate of decrease of the residuals in some suitable norm. As long as the error has large non-smooth components, each relaxation should efficiently reduce the residuals. An efficiency criterion κ $(0 < \kappa < 1)$ must be set. As soon as the size of the residuals after a relaxation is no less than κ times the size before, then the error is judged to be smooth. Typically less than 5 iterations are done.

After these relaxation sweeps we set up the CGC problem on G^{k-1}. Let $r^{k} = b^{k} - A^{k} Z^{k}$, and $b^{k-1} = I^{k-1}_{k} r^{k}$. The system to be solved using grid G^{k-1} is then

$$A^{k-1} X^{k-1} = b^{k-1} . \tag{2-6}$$

The operator A^{k-1} used for G^{k-1} may be taken to be the finite difference approximation on G^{k-1} to the underlying differential operator L, or, (more suitably for theoretical purposes), as $A^{k-1} = I^{k-1}_{k} A^{k} I^{k}_{k-1}$.

On G^{k-1} (2-6) may be solved either by GE or by a sufficient number of relaxation sweeps. Using either method the amount of work done is small on the coarsest grid. On the other grids (2-6) is solved by smoothing steps and a CGC. It

is only necessary to solve the CGC problem to an accuracy commensurate with the size of the remaining non-smooth error on G^k.

When a solution $X^{k-1} = U^{k-1}$ is obtained on G^{k-1} then let $e^k = I^k_{k-1} U^{k-1}$. An improved solution Y^k on G^k is obtained via $Y^k = Z^k - e^k$. This eliminates most of the smooth part of the error in Z^k. Since the error $Y^k - U^k$ is now dominated by its non-smooth component, this error may be further efficiently reduced by relaxation steps.

Thus the non-smooth component of the error in the original approximation on G^n has been cut down by relaxations, and the smooth part has been almost eliminated by the CGC.

The Multi-Grid procedure outlined above is complicated and although the MG idea seems promising, there is still controversy among those who have applied it. Brandt first implemented the scheme in the early seventies (see Brandt, 1977) and suggested that it was an efficient (rapidly converging) algorithm for numerical solution of PDEs. He promoted it vigorously and many difficult problems were solved efficiently via MG. However, slow convergence of the method for other problems has led to much doubt regarding its robustness.

There is little theory which can help decide whether the poor performance of some MG applications is the result of inappropriate implementation, or whether it reflects a fundamental limitation of the method. Convergence proofs for the MG algorithm (given already in the early sixties by Bakhvalov (see Brandt, 1977)) show that a reduction of one

order of magnitude in the overall error can be achieved in O(N) operations (N is the number of grid points)[1]. However the constant in this O(N) is so large as to have no bearing on convergence rates in practice.

Brandt's claim that the MG algorithm will converge quickly is based not on rigorous theory but rather on a heuristic local Fourier analysis of relaxation (and also restriction and prolongation). With suitable restrictions and prolongations, this analysis suggests that MG will converge rapidly provided that the non-smooth error components on each grid are substantially reduced by relaxation. In this thesis we will be concerned with the application of this Local Mode Analysis (LMA) to relaxations.

2.3 Local Mode Analysis

For a given problem and a specified relaxation scheme LMA is used to estimate the rate of reduction of the various Fourier modes which comprise the error. Brandt's original idea was that relaxation at one grid point strongly affects the relaxation only at nearby grid points, i.e., "relaxation is a local process" (see Brandt, 1977). Thus he imagined a relaxation acting on an infinite grid and asked how relaxation would reduce a particular Fourier component.

We will illustrate LMA by considering Gauss-Seidel relaxation of Laplace's equation in some domain. The equation linking the new values X' to the old values X during a sweep over lexicographically ordered grid points is

---

[1] Straightforward use of a good relaxation will result in convergence in $O(N^2)$ (or possibly N to the power 3/2) operations.

(2-7) $4X'(i,j) - X'(i-1,j) - X'(i,j-1) - X(i+1,j) - X(i,j+1) = 0.$

Note that when the point (i,j) is reached, newly computed values X' at points (i-1,j) and (i,j-1) are already in place. The desired solution U also satisfies this equation and upon subtracting we obtain the error transformation equation,

(2-8) $4e'(i,j) - e'(i-1,j) - e'(i,j-1) - e(i+1,j) - e(i,j+1) = 0.$

Brandt (1977) suggested that one can "analyse it [2-8] in the interior of the grid by (locally) expanding the error in Fourier series. ... Thus to study the $\theta = (\theta^1,\theta^2)$ Fourier component of the error before and after the relaxation sweep, we put

(2-9) $e(i,j) = A(\theta)\exp(\iota(\theta^1 i+\theta^2 j))$ $(\iota = \sqrt{-1})$

and

$e'(i,j) = A'(\theta)\exp(\iota(\theta^1 i+\theta^2 j))$."

Here θ^1 and θ^2 must be interpreted as wave numbers in the x and y directions, respectively. Using the identity $\exp(\iota\theta(k\pm1)) = \exp(\iota\theta k)\exp(\pm\iota\theta)$ the error transformation equation (2-8) becomes

(2-10) $(4 - \exp(-\iota\theta^1) - \exp(-\iota\theta^2))A'(\theta)$

$+ (-\exp(\iota\theta^1) - \exp(\iota\theta^2))A(\theta) = 0.$

The convergence rate of the θ^1,θ^2 component is then given by the ratio

$$(2\text{-}11)\quad \mu(\theta) = A'(\theta)/A(\theta)$$
$$= (-\exp(\iota\theta^1) - \exp(\iota\theta^2)) \;/$$
$$(4 - \exp(-\iota\theta^1) - \exp(-\iota\theta^2)).$$

Brandt expected this estimate of the convergence rate to be good for the high frequencies (which are the only ones of interest in the MG context where the role of relaxation is smoothing only), because for these frequencies the interactions of the solution value at an interior point with the boundary conditions and with far away points are small. For the usual case where the mesh size in the finer grid is half the mesh size in the coarser grid, the boundary between "high frequencies" and "low frequencies" is taken as $\max(|\theta^1|,|\theta^2|) = \pi/2$, since the (θ^1,θ^2) component is indistinguishable from the $(\theta^1-\pi,\theta^2-\pi)$ component if only values at every second grid point in the fine grid are considered. The exact form of $\mu(\theta)$ is usually ignored and an estimate of the convergence rate of the high frequency components (called the smoothing factor) is taken as

$$(2\text{-}12)\quad \mu = \max\{|\mu(\theta)| \;;\; \max(|\theta^1|,|\theta^2|) \geq \pi/2\}.$$

This effective rate of smoothing can be used to estimate the efficiency of the algorithm as a whole, and to estimate the total work required for a MG solution. In practice it estimates accurately the high efficiency of many MG implementations.

Despite its practical success so far, many people who have worked on MG methods have found the LMA procedure

dubious. Brandt himself has never come forth with a justification for it or even a statement about when it might be reasonably expected to work, and when it might not.

CHAPTER THREE

A JUSTIFICATION OF LOCAL MODE ANALYSIS

In this chapter we will consider the validity of Local Mode Analysis (LMA) for problems with Dirichlet boundary conditions. We will take the point of view that LMA should approximately describe the transformation of a Fourier mode error under relaxation. We will show that this approximation becomes better as the grid is refined. We will give several concrete examples of how good the approximation is in practical situations.

3.1 The Meaning of Local Mode Analysis.

In explanations of Local Mode Analysis (e.g., Brandt, 1977) the error is supposed to be written as a "local Fourier series" near a grid point (i,j). This notion of "local Fourier series" arises from an attempt to make LMA look plausible in the case of general boundary conditions and variable coefficients, by ignoring the boundary conditions and the variation of coefficients. However we must first dispose of the notion of a "local Fourier series" in order to make a rigourous, global, interpretation of the analysis.

We first discuss this in the context of a one-dimensional continuum, and later we will discuss the multi-dimensional discrete case by analogy. Consider a function f(x) defined on some interval I containing a point x = a. Consider what would be required of a "local Fourier series"

$$(3\text{-}1) \qquad \sum_{\theta} d(\theta)\exp(\iota\theta(x-a))$$

for f(x) near x = a. We might require it to converge in a neighbourhood of x = a and to describe the behaviour of f close to a, perhaps converging at a to the value f(a). But these 'local' requirements actually come from the idea of a Taylor series. Furthermore, how would the coefficients $d(\theta)$ be determined? If we determine them over the neighbourhood (a-c,a+c) of x = a, then the coefficient of the θ component is given by

$$(3\text{-}2)\quad d(\theta) = \int_{a-c}^{a+c} \exp(-\theta \iota x) f(x)\,dx \Big/ \int_{a-c}^{a+c} |\exp(-\iota\theta x)|^2\,dx.$$

Note that

$$(3\text{-}3)\quad \lim_{c\to 0} d(\theta) = \exp(-\iota\theta a) f(a)$$

which is independent of the behaviour of f(x) near a, so that the limit of $d(\theta)$ as c goes to 0 is not a determination of the local importance of the θ-component. We must then pick some c > 0, but which one? This brings us to the heart of the problem. A Fourier series can be defined only for a function and an interval, not a function and an indefinite neighbourhood of a point.

In the case of a discrete point set in real N-space, e.g., a grid covering a domain[1], consider a grid vector v and a specific point a. The coefficient of a Fourier mode in a

---

[1] The discrete Fourier modes are then vectors with components equal to the continuum Fourier functions evaluated at points of this ordered set. The inner product < , > on such a grid is pointwise complex conjugate multiplication of two vectors with summation over all grid points.
$\langle v,w \rangle = \sum v(i,j)w(i,j)$

"local expansion" of v near x = a will depend on which points are considered to form the neighbourhood of a. This neighbourhood is not well-defined. It is also not clear which discrete frequencies are to be selected for the expansion. A discrete Fourier series can only be defined for values of a grid vector corresponding to a specific rectangular subset of the grid.

There is then, no unique way to form a "local Fourier series" in the usual sense that 'local' objects are defined in mathematics. The concept of "local Fourier series" is really a conflation of the idea of a Fourier series and that of a Taylor series. While the existence of such a "local Fourier series", and of "local Fourier modes", is suggested by the language in some of the papers written on MG methods, this is misleading. We suggest that LMA be reinterpreted as a local analysis of the action of a relaxation scheme on Fourier modes over the whole grid. We will show that the attrition rate of errors having the form of a Fourier component (over the whole grid) is accurately computed by LMA for fine meshes. The interest of these special forms of error is only that they will be used to give an acceptable interpretation of LMA.

3.2 Accuracy of Local Mode Analysis

We develop an interpretation of LMA and justify it in the following situation. We consider a second order linear elliptic PDE problem in two dimensions with constant coefficients of the form

$$(3\text{-}4)\quad Lu = a\partial^2u/\partial x^2 + b\partial^2u/\partial y^2 + c\partial u/\partial x + d\partial u/\partial y + eu = f$$

in a domain Ω with smooth boundary $\partial\Omega$. The solution u is assumed to satisfy Dirichlet boundary conditions on $\partial\Omega$. A grid G covers Ω and without loss of generality we assume that the grid spacings in the x direction and in the y direction are equal. We consider zero-order approximations to the boundary and to the boundary conditions. That is, the domain is approximated by line segments along grid lines, and the Dirichlet data are assigned to grid points on these lines.

Grid points will be doubly indexed (i,j) (in 2D) as if the grid were part of a larger rectangular grid running parallel to the co-ordinate axes (see Figure 3). The first

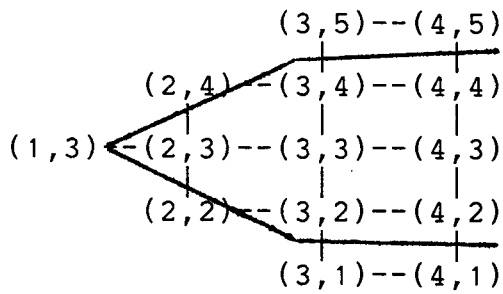


Figure 3 - Numbering of Grid Points

index i numbers the values of x and j numbers the values of y. The equations are doubly indexed (i,j) like the grid points, so that equation (i,j) involves values of the solution at points (i,j), (i-1,j), (i+1,j), (i,j-1), (i,j+1) (if all these are in the grid interior). This notation makes discussion of LMA easier.

If the usual centered differences are used to approximate the derivatives in this PDE, then only the values at a point and its four nearest neighbours are involved in the difference equation corresponding to that point. Thus a "five point

formula" may be used in the interior of Ω. A system of finite difference equations is obtained, $AX = f$, where each equation corresponds to a grid point. In case the equation at a point involves boundary points, (at most three), the fixed values at those points are to be included in the grid vector f. We will suppose that the equations have been multiplied through by h^2, where h is the mesh size.

We will justify LMA for linear relaxations of $AX = f$ which meet the following conditions. First the relaxations must be of SPLITTING TYPE, i.e., the coefficient matrix A is rewritten as $A = L - R$ with L non-singular. We find the next iterate X' from the current iterate x by solving

(3-5) $LX' = RX + f.$

This splitting need not be regular (L^{-1} and R non-negative), and it need not be the same at each iteration. Secondly, we will only discuss splittings for which each split equation involves only the same five points as the original equation. This means

(3-6) $a(m,n) = 0 \quad \Rightarrow l(m,n) = r(m,n) = 0.$

Finally, in order that LMA itself be feasible we must require that every line in $LX' = RX + f$ which corresponds to a point which is not next to a boundary point is of the form

$$(3\text{-}7)\quad l^1X'(i,j) + l^2X'(i-1,j) + l^3X'(i,j-1) + l^4X'(i+1,j) + l^5X'(i,j+1) = r^1X(i,j) + r^2X(i-1,j) + r^3X(i,j-1) + r^4X(i+1,j) + r^5X(i,j+1) + f(i,j),$$

where $l^1,\ldots,l^5,r^1,\ldots,r^5$ are constant coefficients. This condition will be subsequently referred to as 'homogeneity'.

Many common linear relaxation schemes satisfy these conditions, including Jacobi, Gauss-Seidel, SOR, line versions of these, and ADI (see Chapter Two). One iteration of ADI consists of two steps such as are considered here, with L and R interchanged. The ILU method does not satisfy the homogeneity condition and cannot, strictly speaking, be analysed by LMA. In practice an idealization of the ILU method is analysed this way, but we will not justify that here. Red-Black and Zebra relaxations do not satisfy these conditions and it is known (S. McCormick and K. Stueben, private communication) that LMA applied simple-mindedly yields misleading estimates of smoothing rates in these cases.

In justifying LMA we will consider the reduction of error from one iterate X to the next iterate X' which is given by (see notation in Chapter Two)

$$(3\text{-}8)\quad Le' = Re.$$

It is too complicated to consider an arbitrary error e, and write a Fourier series for it (which amounts to taking a finite Fourier transform), and then ask what happens to the values of the Fourier transform under a relaxation step. Rather we suppose that the error consists of a multiple of

only one Fourier component $\psi(\theta)$, as defined below, and show that after a relaxation step the new error ψ' is $\mu(\theta)\psi(\theta)$ plus something small. On a grid with the index convention, the value of a discrete exponential Fourier component, $\psi(\theta)$ (with frequencies $\theta = (\theta^1,\theta^2)$, in two dimensions), at a grid point (i,j) is given by

(3-9) $\psi(\theta)(i,j) = \exp\{\iota(i\theta^1 + j\theta^2)\}$.

We need only consider $-\pi < \theta^1,\theta^2 < \pi$.

As mentioned in Chapter Two, LMA treats the Fourier component $\psi(\theta)$ as an eigenvector of the amplification matrix of the relaxation. As noted in Chapter One, this is only the case for certain relaxations of problems with periodic boundary conditions on a rectangle. However, as will be seen, for many relaxations, the Fourier components are in fact "close to being eigenvectors", because they are eigenvectors of an idealized relaxation that is "close" to the actual relaxation. The idealized relaxation will be identical with the actual relaxation except at points next to a boundary point. Since the proportion of such points in the grid becomes smaller as the mesh becomes finer, the idealized relaxation gives a good approximation to the actual relaxation in the limit of fine grids. The estimate of the accuracy of LMA which will be developed makes such arguments rigourous.

For a specific Fourier component, $\psi(\theta)$ given by (3-9), we will construct a system $L^*\psi^* = R^*\psi(\theta)$, similar to $L\psi' = R\psi(\theta)$, for which $\psi(\theta)$ is an eigenvector of both L^* and R^*. First note that every line in $L\psi' = R\psi(\theta)$ which

corresponds to a point which is not next to a boundary point is of the form

$$(3\text{-}10)\quad l^1\psi'(i,j) + l^2\psi'(i-1,j) + l^3\psi'(i,j-1) + l^4\psi'(i+1,j) + l^5\psi'(i,j+1) = r^1\psi(\theta)(i,j) + r^2\psi(\theta)(i-1,j) + r^3\psi(\theta)(i,j-1) + r^4\psi(\theta)(i+1,j) + r^5\psi(\theta)(i,j+1),$$

(cf. (3-7)). Using (3-9) the right hand side simplifies to

$$(3\text{-}11)\quad (r^1 + r^2\exp(-\iota\theta^1) + r^3\exp(-\iota\theta^2) + r^4\exp(\iota\theta^1) + r^5\exp(\iota\theta^2))\psi(\theta)(i,j) = r^*\psi(\theta)(i,j).$$

Now if $\psi(\theta)$ is substituted for ψ' on the left side of (3-10) then it simplifies to

$$(3\text{-}12)\quad (l^1 + l^2\exp(-\iota\theta^1) + l^3\exp(-\iota\theta^2) + l^4\exp(\iota\theta^1) + l^5\exp(\iota\theta^2))\psi(\theta)(i,j) = l^*\psi(\theta)(i,j).$$

Now we construct R^* so that values of $[R^*\psi(\theta)](i,j)$ which correspond to points (i,j) next to a boundary are equal to $r^*\psi(\theta)(i,j)$, like values corresponding to other interior points. We construct L^* likewise so that for every interior point (i,j) $[L^*\psi(\theta)](i,j) = l^*\psi(\theta)(i,j)$. We do this by including an appropriate multiple of $\exp(\pm\iota\theta^1)$ or $\exp(\pm\iota\theta^2)$, for every equation involving a boundary point, to

make it like the equation for points away from the boundary[2]. Specifically, suppose we are constructing an equation at a left edge point (i,j) where a term $l^2\psi'(i-1,j)$ and a term $r^2\psi(\theta)(i-1,j)$ are missing in (3-10). Then to form the (i,j) equation in $L^*\psi^* = R^*\psi(\theta)$, add $l^2\exp(-\iota\theta^1)\psi^*(i,j)$ and $r^2\exp(-\iota\theta^1)\psi(\theta)(i,j)$ to the left and right sides, respectively.[3]

If we denote these extra elements along the diagonal by $L1 = L^* - L$ and $R1 = R^* - R$, then $L1$ is 0 except for diagonal entries corresponding to boundary points.

LEMMA 1: The Fourier component $\psi(\theta)$ is an eigenvector of the system $\mu L^*e = R^*e$ with eigenvalue $\mu(\theta)$.

PROOF: If (i,j) is an interior grid point the (i,j) entries in $L^*\psi(\theta)$ and $L\psi(\theta)$ are both equal to $l^*\psi(\theta)(i,j)$. At a point next to a boundary point the additional entries in L^* are chosen to make $L^*\psi(\theta)(i,j)$ equal to $l^*\psi(\theta)(i,j)$. Similarly, for interior points and points near the boundary, the (i,j) entry in $R^*\psi(\theta)$ is $r^*\psi(\theta)(i,j)$.

If we do a LMA and put

---

[2] If Ω is a rectangle, and the values of θ^1, θ^2 are restricted to those values that would occur in a Fourier transform, then the matrices L^* and R^* describe the relaxation of a problem with periodic boundary conditions. However the construction used here is not restricted to rectangular domains.

[3] E.g., for the Gauss-Seidel relaxation of Laplace's equation, the left edge equation which has the form

$4e'(i,j) - e'(i,j-1) = e(i+1,j) + e(i,j+1)$, becomes,

$$(4 - \exp(-\iota\theta^1))e^*(i,j) - e^*(i,j-1) = e(i+1,j) + e(i,j+1).$$

(3-13) $v = A(\theta)\exp(\iota(\theta^1 i + \theta^2 j))$, and

$v' = A'(\theta)\exp(\iota(\theta^1 i + \theta^2 j))$,

where v and v' are the errors before and after relaxation, then

(3-14) $A'(\theta)(l^1 + l^2\exp(-\iota\theta^1) + l^3\exp(-\iota\theta^2)$

$+ l^4\exp(\iota\theta^1) + l^5\exp(\iota\theta^2))\exp(\iota(\theta^1 i + \theta^2 j))$

$= A(\theta)(r^1 + r^2\exp(-\iota\theta^1) + r^3\exp(-\iota\theta^2)$

$+ r^4\exp(\iota\theta^1) + r^5\exp(\iota\theta^2))\exp(\iota(\theta^1 i + \theta^2 j))$.

Thus,

(3-15) $\mu(\theta) = A'(\theta)/A(\theta) = r^*/l^*$

and therefore

(3-16) $L^*\mu(\theta)\psi(\theta) = R^*\psi(\theta)$. QED

Now let us consider the difference between $\psi^* = \mu(\theta)\psi(\theta)$ which solves

(3-17) $L^*\psi^* = R^*\psi(\theta)$

and the grid vector ψ' which is the actual solution of

(3-18) $L\psi' = R\psi(\theta)$.

The difference between L and L^* occurs only at points next to a boundary point. For this difference to be small then, we need a lemma to guarantee a small number of grid points next to a boundary point.

LEMMA 2: Let the domain Ω be finite with piecewise smooth boundary $\partial\Omega$. Then for h sufficiently small there is a number a, such that for any grid of mesh size h, the number $N(b)$ of

grid points in the interior of Ω, one of whose four nearest neighbours lies on or outside $\partial\Omega$, is bounded by $a\sqrt{N}$, where N is the total number of grid points in intΩ.

PROOF: This is intuitively obvious although there are subtle points in a rigourous justification. For technical details see Appendix A.

For ease of analysis we will consider only relaxations where, in the equation corresponding to an interior grid point (i,j), the values at each of the neighbouring points (i-1,j),(i,j-1),(i+1,j),(i,j+1) occur either on the left side or on the right side of this equation, but not both. Formally this says that only one of the members of each of the pairs $(l^2,r^2),\ldots,(l^5,r^5)$ is not zero. All of the common relaxation schemes satisfy this condition.

Now we are ready to consider the relationship between equations (3-17) and (3-18).

THEOREM 1: If $\partial\Omega$ is smooth and the induced 2-norm of L^{-1} is bounded independently of the number of grid points N, then there is a constant A independent of the grid such that

$$(3\text{-}19)\quad |\psi' - \mu(\theta)\psi(\theta)| \leq A|\psi(\theta)|/\sqrt[4]{N}$$

PROOF: From (3-16) we find that

$$(3\text{-}20)\quad R\psi(\theta) = (R^* - R1)\psi(\theta)$$
$$= \mu(\theta)(L + L1)\psi(\theta) - R1\psi(\theta).$$

Thus from (3-18)

$$(3\text{-}21)\quad \psi' = L^{-1}R\psi(\theta)$$
$$= L^{-1}(\mu(\theta)L1\psi(\theta) + \mu(\theta)L1\psi(\theta)\ ) - L^{-1}R1\psi$$
$$= \mu(\theta)\psi(\theta) + L^{-1}(\mu(\theta)L1\psi(\theta) - R1\psi(\theta)\ ).$$

This equation shows that $\psi(\theta)$ is "almost an eigenvector of the relaxation" if the term $L^{-1}(\mu(\theta)L1\psi(\theta) - R1\psi(\theta)\ )$ is of small norm. We will show this by showing that $L1\psi(\theta)$ and $R1\psi(\theta)$ are of small norm.

Consider the entries of $R1\psi(\theta)$. This grid vector will be zero at grid points not next to a boundary. Corresponding to a grid point exactly one of whose four nearest neighbours is on the boundary, the vector will have one of the following values: $r^2\exp(-\iota\theta^1)$, $r^3\exp(-\iota\theta^2)$, $r^4\exp(\iota\theta^1)$, or $r^5\exp(\iota\theta^2)$. For the case of a point with two nearest neighbours on the boundary, e.g., with points to the right and on the bottom, one obtains the entry $r^3\exp(-\iota\theta^2) + r^4\exp(\iota\theta^1)$ (see Figure 4). At other 'corner' points and at points with three nearest

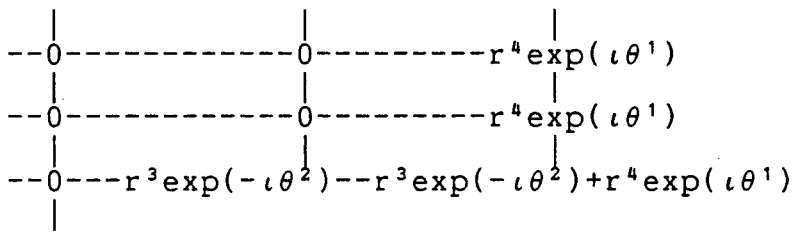


Figure 4 - Entries in R1 Corresponding to Grid Points on the Boundary

neighbours on the boundary, the entry will similarly be the sum of two (or three) of the preceeding numbers. $L1\psi(\theta)$ will be similar. Let N(e), N(c) and N(w) be the number of points with one (edges), two (corners), and three (wedges) neighbouring boundary points, respectively. Let

$a^0 = \max(|r^2|,\ldots,|r^5|,|l^2|,\ldots,|l^5|)$. Then in the grid vector $\mu(\theta)L1\psi(\theta) - R1\psi(\theta)$ there are N(e) entries of modulus bounded by a^0 (since $|\mu(\theta)| \leq 1$), N(c) entries of modulus bounded by $2a^0$, and N(w) entries of modulus no larger than $3a^0$. Thus taking the 2-norm by summing the squares of the entries,

$$(3\text{-}22)\quad | \mu(\theta)L1\psi(\theta) - R1\psi(\theta) | \leq a^0\sqrt{(N(e) + 4N(c) + 9N(w))}.$$

From Lemma 2 there is a number a such that for any grid with h small enough,

$$(3\text{-}23)\quad N(e) + N(c) + N(w) < a\sqrt{N}.$$

Thus from (3-20) and (3-22)

$$(3\text{-}24)\quad |\psi' - \mu(\theta)\psi(\theta)| \leq |L^{-1}|a^0{}^4\sqrt{N}\ 3\sqrt{a}.$$

If N is the number of interior grid points then the norm of $\psi(\theta)$, each of whose entries has modulus one, is $\sqrt{N}$. Since $3|L^{-1}|\ a^0\ \sqrt{a}$ is bounded independently of N the result follows. QED

This is a bound on the relative norm of the difference between an actual relaxation and the LMA idealization.

We may also obtain simple lower bounds on this difference, if for some $b(e) > 0$, $N(e) \geq b(e)\sqrt{N}$ for N sufficiently large. This occurs for circular and rectangular domains. Obviously for rectangular domains which have been rotated through $\pi/4$ radians there are no edge points so that the following simple argument won't go through.

Let b^0 be the minimum of the non-zero coefficients $|r^2|,\ldots,|r^5|,|l^2|,\ldots,|l^5|$. Then the grid vector

$\mu(\theta)L1\psi(\theta) - R1\psi(\theta)$ has at least N(e) entries of size at least $|\mu(\theta)|b^0 > 0$ (since each of the four nearest neighbours is involved in the relaxation at a point). Thus since $N(e) \geq b(e)\sqrt{N}$

$$(3\text{-}25)\quad |\mu(\theta)L1\psi(\theta) - R1\psi(\theta)| \geq |\mu(\theta)|b^0\sqrt{N(e)} \geq |\mu(\theta)|b^0\sqrt{b(e)}\sqrt[4]{N}.$$

Now

$$(3\text{-}26)\quad \min |L^{-1}x|/|x| = (\max |Lx|/|x|)^{-1} = 1/|L|.$$

Thus from (3-20) and (3-26),

$$(3\text{-}27)\quad |\psi' - \mu(\theta)\psi(\theta)| \geq (1/\sqrt[4]{N})(|\mu(\theta)|b^0\sqrt{b(e)}/|L|)|\psi(\theta)|.\text{QED}$$

In interpreting inequalities (3-24) and (3-27) it is important to realize that the estimate above is NOT the difference between the norm of the result of relaxing the vector ψ, and the norm of $\mu(\theta)\psi$. These norms may be closer than this estimate would suggest, as will be seen in the tables. It is rather the norm of the vector difference between these two which is considered. This difference may have a significant component normal to ψ, which means that this component "feeds" other Fourier components during relaxation, especially on small grids. It seems from some sample calculations that usually the projection of this vector difference $\psi' - \mu(\theta)\psi(\theta)$, is largest on Fourier components whose first frequency lies in the range $(\theta^1-O(h),\theta^1+O(h))$ and whose second frequency lies in the range $(\theta^2-O(h),\theta^2+O(h))$.

(O(h) means of bounded by ch where c is a positive constant.)

Except for simple cases, some of which will be discussed later, the estimates (3-24) and (3-27) above are too cumbersome to be of much use in practice. However (3-24) can be used to justify LMA asymptotically.[4] This gives some psychological grounds for confidence in its practical use on fine grids. On the other hand, (3-27) suggests that LMA is not such a good estimate for very coarse grids. The examples will illustrate how good LMA is in some typical situations.

3.3 Examples

3.3.1 Gauss-Seidel Relaxation of Laplace's Equation in the Unit Square with Dirichlet Boundary Conditions.

The grid is as follows. We may suppose that the boundaries coincide with grid lines. The mesh in x and y will be $h = 1/(n+1)$, and the values of x and y which occur at grid points will be indexed from 0 through n+1. Points not on the grid boundary will have both indices i,j between 1 and n.

The finite difference system consists of n^2 equations, where equation (i,j) has the form

$$(3\text{-}28)\quad 4u(i,j) - u(i-1,j) - u(i,j-1) - u(i+1,j) - u(i,j+1) = 0,$$

---

[4] One unexpected inference we may draw from (3-24) and (3-27) above is that LMA works well for giving an asymptotic estimate of the attrition rate of the low frequency Fourier components. This is contrary to Brandt's expectation (1977) that LMA predictions would only be good for high frequency components. This fact suggests the use of LMA for the understanding of the overall performance (as distinct from smoothing rate) of relaxation schemes. This will be discussed later.

where i,j run from 1 to n. If any indices in this equation take on values 0 or n+1 then the values will be understood to be taken from the boundary data.

Relaxation will proceed from the bottom left corner to the right (lexicographic ordering). The equation to be solved to obtain the next iterate x' from the current iterate x is

(3-29) $4x'(i,j) - x'(i-1,j) - x'(i,j-1) = x(i+1,j) + x(i,j+1)$.

Then the generic equation transforming the errors from one iteration to the next is

(3-30) $4e'(i,j) - e'(i-1,j) - e'(i,j-1) = e(i+1,j) + e(i,j+1)$.

We estimate the difference between the result of relaxing a Fourier component $\psi(\theta)$ and the LMA estimate $\mu(\theta)\psi(\theta)$ by following through the steps of the theorem. First we estimate $|L1\psi(\theta)|$ and $|R1\psi(\theta)|$. The nxn grid vector $L1\psi(\theta)$ is 0, except for 2(n-1) entries of modulus 1 along the top and right sides, and an entry of modulus between 0 and 2 in the bottom left corner. $R1\psi(\theta)$ is 0 except for 2(n-1) entries of modulus one along the top and right sides, and one of modulus between 0 and 2 in the top right corner. Then $\mu(\theta)L1\psi(\theta) + R1\psi(\theta)$ will have entries of modulus $|\mu(\theta)|$ along the bottom and left sides and of modulus 1 along the top and right sides. The sum of the moduli of the corner entries is bounded by $6(1+|\mu(\theta)|)$. Thus

$$(3\text{-}31)\quad (1+|\mu(\theta)|)\sqrt{(2n-4)} \le |\mu(\theta)L1\psi(\theta) + R1\psi(\theta)| \le (1+\mu(\theta)|)\sqrt{(2n+2)}.$$

The 2-norm of L is bounded by 6 and that of L^{-1} by 1/3. These estimates are obtained in Appendix B. The estimates of the difference between ψ' and $\mu(\theta)\psi(\theta)$ become

$$(3\text{-}32)\quad (1/6n)\sqrt{(2n-4)} \le |\psi' - \mu(\theta)\psi(\theta)|/|\psi(\theta)|(1+|\mu(\theta)|) \le (1/2n)\sqrt{(2n+2)}.$$

For an n=100 by 100 grid (10,000 interior points),

$$(3\text{-}33)\quad .0233 \le |\psi'-\mu(\theta)\psi(\theta)|/|\psi(\theta)| \le .1421.$$

This should be compared with Table I.

The values in Table I were obtained as follows. For an (n+2)x(n+2) grid, values were assigned to the nxn interior points as per the definition of $\psi(\theta)$ for various θ^1,θ^2. One Gauss-Seidel relaxation sweep was performed, taking the values of boundary points to be 0. The norm of the relaxed vector was computed, as well as its projection on the original Fourier component . These were both normalized, using $|\psi(\theta)| = n$.

Table I illustrates that as n gets larger, the LMA estimate better approximates the relaxed vector.

Table I - Attrition Rates of Fourier Components of Gauss-Seidel Relaxation of Laplace's Equation in the Unit Square

Grid (nxn)	$(\theta^1,\theta^2)/\pi$	Relative norm $\lvert\psi'\rvert/\lvert\psi(\theta)\rvert$	Projection $(\psi',\psi(\theta))/\lvert\psi(\theta)\rvert$	$\lvert\mu(\theta)\rvert^2$
5x5				
	1,1	.6174	.5859	.7529
	1/5,3/5	.3829	.3453	.4232
	1/5,1	.1839	.1122	.1460
	3/5,3/5	.3625	.3303	.4004
	1,1	.3028	.2731	.3333
20x20				
	4,4	.7201	.7104	.7529
	1/5,1/2	.4864	.4768	.5000
	1/5,1	.1589	.1369	.1460
	1/2,1/2	.4360	.4279	.4472
	1,1	.3256	.3187	.3333
100x100				
	1/5,1/5	.7464	.7444	.7529
	1/5,1/2	.4973	.4953	.5000
	1/5,1	.1488	.1442	.1460
	1/2,1/2	.4449	.4433	.4472
	1,1	.3318	.3304	.3333

3.3.2 Line Relaxation of $.01(\partial^2u/\partial x^2) + (\partial^2u/\partial y^2) = 0$ in the Unit Square.

Table II was obtained like the preceeding table using y-line Gauss-Seidel relaxation on a 100x100 grid. In a sweep of y-line relaxation, the values of the new iterate along an entire grid line of points whose second index j is the same are determined simultaneously, using values of the new iterate at points whose second index is j-1 and values of the current iterate for points whose j index is j+1.

The LMA estimate is less accurate here than for the point G-S relaxation on a comparable grid. This is because the norm of the matrix L^{-1} is much larger (although still bounded independently of h).

Table II - Line Relaxation

$(\theta^1,\theta^2)/\pi$	Norm after relaxation of Fourier mode	$\mu(\theta)$
.01,.01	.7767	.9094
.50,.01	.3907	.4301
.50,.05	.2121	.2187
.50,.50	.0051	.0049
1.0,1.0	.0025	.0025

3.3.3 Accuracy of Local Mode Analysis Estimates for Distributive Gauss-Seidel (DGS) Relaxation of the Cauchy-Riemann Equations

We seek solutions u , and v to

$$\partial u/\partial x = \partial v/\partial y, \tag{3-34}$$
$$-\partial u/\partial y = \partial v/\partial x$$

in the unit square.

The domain on which a solution is sought is parcelled up into square cells of side h. Values of the discrete variable u are sought at the centers of vertical cell sides and values of v are sought at the centers of horizontal cell sides. This placing of discrete variables ensures second order accuracy of the discrete equations

$$u(x+h/2,y) - u(x-h/2,y) = v(x,y+h/2) - v(x,y-h/2) \tag{3-35}$$
$$u(x,y-h/2) - u(x,y+h/2) = v(x+h/2,y) - v(x-h/2,y).$$

The first equation is located at cell centers, and the second is located at cell corners. There is no natural correspondence between variables and equations.

To solve these equations a sweep is made over all the first equations and then over the second equations. To relax

each equation each of the four variables that are involved in that equation are adjusted. The details of the distributive relaxation may be found in Brandt(1979). In that paper he calculates a smoothing rate for residuals. The second column in Table III corresponds to the actual attrition rate of these residuals for the 40x40 grid. In the LMA idealization, the amplitudes a^1, a^2 of Fourier components

$r^1 = a^1(\theta)\exp(\iota(i\theta^1+j\theta^2)$, and $r^2 = a^2(\theta)\exp(\iota(i\theta^1+j\theta^2)$

before relaxation are related linearly to the amplitudes $a^{1'}(\theta)$ and $a^{2'}(\theta)$ after the sweep. Thus a matrix $A(\theta)$ may be constructed specifying the linear linkage between the amplitudes of r^1 and r^2 before and after the sweep. In reality a significant component orthogonal to the original vector arises. The norms of the relaxed vectors are compared below with the idealization $A(\theta)$. The first column of each matrix records the relative norms of the relaxed residuals if the original residuals were $r^1=\psi(\theta), r^2=0$, the second column if they were $r^1=0, r^2=\psi(\theta)$. One sweep of DGS was performed taking the initial and boundary values of u and v to be 0. After both phases of the relaxation sweep, the residuals were recomputed and the norms of the two residuals taken. These norms were normalized by the vector norms of the original residuals, $|r^1|=N$, $|r^2| = N-1$. The third column records the actual transformation of errors, which is too difficult to obtain via LMA. To obtain these values, vectors u and v were constructed, one of which was given values as per $\psi(\theta)$, and the other was set to 0. DGS relaxation was done, taking f1=f2=0, and the norms of the resulting vectors were taken and normalized.

As may be seen from Table III the norms of the residuals

Table III - Distributive Relaxation of the Cauchy-Riemann Equations

θ^1,θ^2	LMA Estimate	Actual Residual	Actual Error
π,π	.333 0 0 .333	.110 0 0 .107	.548 .038 .038 .548
$\pi/2,\pi/2$	.440 0 0 .440	.721 0 0 .720	.444 .022 .022 .444
$\pi/40,\pi$	.020 0 0 .020	.709 0 0 .716	.119 .035 .025 .079
$\pi/40,\pi/2$	.460 0 0 .460	.727 0 0 .722	.465 .021 .020 .454
$\pi/40,\pi/40$	.997 0 0 .997	.997 0 0 .968	.973 .012 .012 .973

after relaxation are considerably different from what would be predicted on the basis of LMA. In fact the LMA prediction for the residuals seems to conform more to the norm of the errors after relaxation though the difference is still marked. In this situation then, LMA does not yield very accurate estimates of the effect of relaxation on the various Fourier components. For this weakly elliptic system, the relaxation is complicated and cannot be described in the terms used to prove Theorem I.

3.4 Comments and Miscellaneous Consequences

3.4.1 Alternate Fourier Components

In the above discussion only exponential Fourier components have been used. There is no reason a priori why we could not have embodied the various frequencies as sine Fourier components. With the usual indexing convention such a component corresponding to frequencies θ^1 and θ^2 would be

defined as

(3-36) $\omega(\theta)(i,j) = \sin(\theta^1 i)\sin(\theta^2 j)$.

In fact, sine Fourier components like ω above are the eigenvectors of the Jacobi relaxation of the model problem, whereas exponential components ψ are not eigenvectors for any scheme. However, one serious difficulty arises; we cannot do LMA for sine Fourier components for lack of appropriate simple identities connecting values at adjacent points. As it turns out the rates of attrition of sine Fourier components are quite different from exponential Fourier components for the same relaxation on the same problem. This casts some doubt on the interchangeability of the two types of Fourier components.

3.4.2 Alternate Norms

We have used the 2-norm throughout, because this arises from an inner product which must be used to compute orthogonal projections. The reader may check that if the 1-norm were used then the result of the main theorem (whose developement is independent of the norm used) is

(3-37) $|\psi' - \mu(\theta)\psi(\theta)| = O(|\psi(\theta)|/\sqrt{N})$.

If the supremum norm is used no useful bound is obtained, since the contribution of boundary values to the norm of a grid vector is not in proportion to their numbers.

3.4.3 Residuals

Using LMA it is possible to justify the intuitively

reasonable claim that the residuals

$$(3\text{-}38)\quad r = b - AX = -Ae,$$

reflect local errors in the current iterate X more than global errors, so that the size of residuals alone is not a good estimate of the error in an approximate solution. We must also know whether the error is smooth or oscillatory.

First we prove a lemma which justifies the use of LMA type computations of residuals, and then we use these computations to show that the low frequencies in the error contribute much less to the residuals than the high frequencies do.

LEMMA 3: Suppose the difference equation

$$(3\text{-}39)\quad a^1U(i,j) + a^2U(i-1,j) + a^3U(i,j-1) + a^4U(i-1,j) + a^5U(i,j+1) = h^2b(i,j)$$

is to be satisfied at every interior point (i,j) of a grid G covering a convex domain Ω, and that zero order approximations to the boundary of Ω and to Dirichlet boundary conditions are made. Let $\psi(\theta)$ be as before. Then as h -> 0 the residuals associated with an error of the form $\psi(\theta)$ are asymptotically given by

$$(3\text{-}40)\quad r(i,j) = (a^1 + a^2\exp(-\iota\theta^1) + a^3\exp(-\iota\theta^2) + a^4\exp(\iota\theta^1) + a^5\exp(\iota\theta^2))\ \psi(\theta)(i,j).$$

PROOF: Consider the system of equations

(3-41) (I + A)X = 0.

Suppose we relax this system by solving an equation of the form

(3-42) IX' = -AX.

This iteration satisfies all the conditions of Theorem I. If the error in the current iterate X is $\psi(\theta)$ then the error in the next iterate X' is r, the residual for AU = h^2b. Theorem I then says that

(3-43) | r - $\mu(\theta)\psi(\theta)$ | = $O(1/\sqrt[4]{N})$.

But $\mu(\theta)$ for this relaxation is given by (3-40). QED

In our description of residual computation we assume that the sum $a^1+a^2+a^3+a^4+a^5$ is 0. Full generality is restored by adding a term $a^0U(i,j)$ if the function u itself appears in the differential equation which underlies (3-39). Consider what happens to a single Fourier component of the form $\exp(\iota(\theta^1 i+\theta^2 j)$ under the mapping e->r:

(3-44) $r(i,j) = \exp(\iota(\theta^1 i + \theta^2 j))(a^0 + a^1 + a^2\exp(-\iota\theta^1)$
$+ a^3\exp(-\iota\theta^2) + a^4\exp(\iota\theta^1) + a^5\exp(\iota\theta^2))$

If θ^1 and θ^2 are both small, then $\exp(\pm\iota\theta^1)$and $\exp(\pm\iota\theta^2)$ are close to 1. At an interior point (i,j) therefore the modulus of r(i,j) is close to $|a^0|$. However if θ^1 and θ^2 are both close to π then $\exp(\pm\iota\theta^1)$ and $\exp(\pm\iota\theta^2)$ are all nearly -1, and

$|D(e)(i,j)|$ is close to $|a^0 + 2a^1|$. Lemma 3 insures that this indicates the size of the global residuals. Thus unless $-a^0$ is of the same order as $2a^1$ (it is typically much smaller in practice), the high frequency amplitudes are very much over-represented in the residual vis-a-vis the low frequency components. Since the lowest frequencies are on the order of the mesh size h, and exp(O(h)) is 1 + O(h), we see from (3-44) that the amplitude of a very low frequency residual is only O(h) of the amplitude of the corresponding low frequency error. On the other hand the amplitude of a very high frequency residual can be many times (for the Gauss-Seidel relaxation of the model problem $a^0 + 2a^1 = 8$) the amplitude of the high frequency error.

CHAPTER FOUR

A HEURISTIC LOOK AT RELAXATIONS USING LOCAL MODE ANALYSIS

In Chapter Three it was shown that LMA could be justified in the limit of very fine grids. It was also shown that the validity of the function $\mu(\theta)$ was not restricted to values of θ^1, θ^2 bounded away from 0. Here we compute the function $\mu(\theta)$ for two common schemes: the SOR method and the ADI method. The form of $\mu(\theta)$ will help us understand heuristically what these schemes do. The algebraic theory of these schemes for special cases will be adduced to confirm the understanding using LMA. LMA will also be used to discuss why it is that most relaxations which meet the conditions outlined in Chapter Three do not efficiently reduce smooth errors.

4.1 The SOR Method

The SOR (Successive Over-Relaxation) method has been commonly used for constant coefficient problems with some success. A parameter ω sets the amount of over-relaxation performed. A crucial question is how to pick ω $(0 < \omega < 2)$ in order to optimize the performance of the scheme over many iterations, i.e., minimize the spectral radius of the amplification matrix. Here we study the behaviour of the SOR method applied to the Laplace equation in a domain Ω which has been discretized on a grid with mesh size h.

The discrete Laplace equation at a point (i,j) is given by

$$(4\text{-}1)\quad U(i,j) = (1/4)(U(i-1,j) + U(i,j-1) + U(i+1,j) + U(i,j+1)).$$

Applying the SOR method, the equation of relaxation at a point (i,j) is

$$(4\text{-}2)\quad X'(i,j) = (1-\omega)X(i,j) + (\omega/4)(X'(i-1,j) + X'(i,j-1) + X(i+1,j) + X(i,j+1)).$$

Subtracting (4-1) from (4-2) we obtain the transformation of errors

$$(4\text{-}3)\quad e'(i,j) = (1-\omega)e(i,j) + (\omega/4)(e'(i-1,j) + e'(i,j-1) + e(i+1,j) + e(i,j+1)).$$

Let $e(i,j) = A(\theta)\exp(\iota(\theta^1 i+\theta^2 j))$ and let $e'(i,j) = A'(\theta)\exp(\iota(\theta^1 i + \theta^2 j))$.
Then (4-3) becomes

$$(4\text{-}4)\quad A'(\theta) = (1-\omega)A(\theta) + (\omega/4)(A'(\theta)\exp(-\iota\theta^1) + A'(\theta)\exp(-\iota\theta^2) + A(\theta)\exp(\iota\theta^1) + A(\theta)\exp(\iota\theta^2)).$$

The LMA estimate is then given by

$$(4\text{-}5)\quad \mu(\theta) = (\exp(\iota\theta^1)+\exp(\iota\theta^2)-4(\omega-1)/\omega) / (4/\omega - \exp(-\iota\theta^1) - \exp(-\iota\theta^2)).$$

The maximum of $|\mu(\theta)|$ occurs at $\theta^1 = \theta^2 = 0$ and has the value 1. The minimum occurs at $(0,\pi)$ or $(\pi,0)$ and has the value $\omega - 1$. The profile of the surface $\mu(\theta)$ resembles that for Gauss-Seidel relaxation, with a local maximum at (π,π), see Figure 5.

If ω is near 1 then the values of $\mu(\theta)$ are well

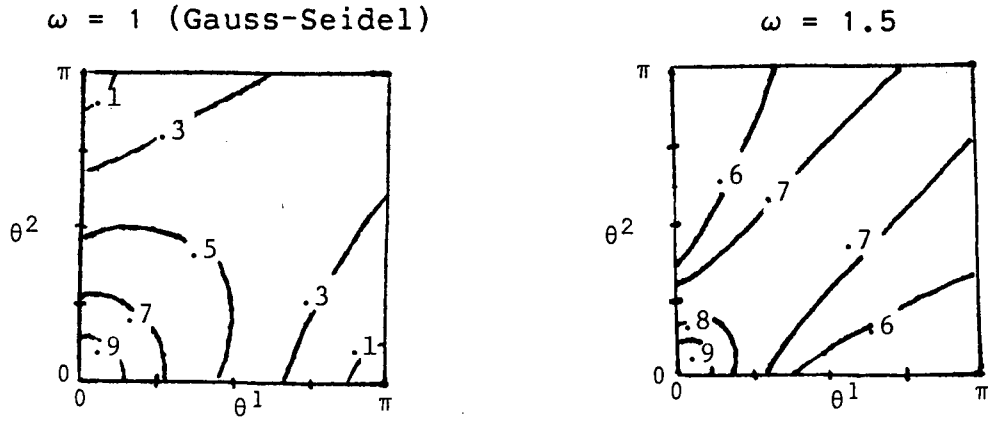


Figure 5 - Contour Lines of $|\mu(\theta)|$ for SOR Relaxation

distributed, but as ω increases to 2 the values of $\mu(\theta)$ cluster together. For model problems it is known (Young, 1971) that the eigenvalues of the amplification matrix behave similarily.

Can $\mu(\theta)$ be used to estimate the rate of convergence of the SOR relaxation? Unlike de Vries (1982), I found that $\mu(\pi h, \pi h)$ was not a good estimate of κ the spectral radius of the amplification matrix. There is no other clear choice of Fourier component which should be analogous to the smoothest eigenvector of the relaxation (since $\mu(0,0) = 1$ always, it cannot be used). Typically $1 - \kappa$ was two or three times bigger than $1 - \mu(\pi h, \pi h)$ for problems in the unit square. Nevertheless, in certain cases it is possible to use LMA to estimate the best value of the parameter ω as described below.

We consider Laplace's equation in the following domains: Ω^1 is the unit square; Ω^2 is the circle $(x-1/2)^2 + (y-1/2)^2 < 1/4$; Ω^3 is the curved region $\{(x,y) | \; y>0, \; 4/5-(2x-1)^2 < y < 1-(2x-1)^2\}$. Finite difference systems were set up for these problems for several values of h, as outlined in Table IV. It was not clear

which frequency ought to be analogous to the smoothest eigenvector for the domains Ω^2 and Ω^3. Since the number of grid points is readily computed, the value of μ was computed for $\Omega = (\pi/\sqrt{N}, \pi/\sqrt{N})$. This is nearly equivalent to $(\pi h, \pi h)$ in the unit square. The value ω' of ω which minimizes $|\mu|$ is recorded in Table IV. For comparison purposes the value ω^* of ω which minimizes the actual radius of the amplification matrix is included. For the unit square ω^* is known from a complete analysis of this model problem (see Young(1971)), and is given by

$$(4\text{-}6)\quad \omega^* = 2/(1+\sin(\pi h)).$$

For the other problems, κ' an approximation to κ, was determined using the successive iterates $X^1, X^2, \ldots$. It is defined by

$$(4\text{-}7)\quad \kappa' = |X^{52} - X^{51}| \;/\; |X^{51} - X^{50}|.$$

An ω^* was computed which minimized κ' as a function of ω.

We see that ω' is a good estimate of ω^* particularly, for large N. For Ω^3 the ratio of the number of grid points next to the boundary to the total number of interior grid points is rather larger than for the other domains with comparable N. The validity of LMA depends on this ratio being small, and so it is not surprising that estimates for Ω^3 were not as good as those for Ω^1 and Ω^2.

Can LMA be used to choose parameters for more general

Table IV - Best values of ω

	h	N	ω'	ω^*
Ω^1	.2	16	1.236	1.260
	.1	81	1.524	1.528
	.05	361	1.729	1.730
	.02	2401	1.882	1.882
Ω^2	.02	1847	1.86	1.89
Ω^3	.02	419	1.73	1.58
	.01	1787	1.86	1.79
	.005	7371	1.93	1.93

problems? I considered extension to two classes of equations; asymmetric constant coefficient equations and equations with one variable coefficient.

I encountered two difficulties with asymmetric operators of the form $-\partial^2/\partial x^2 - \partial^2/\partial y^2 - a\partial/\partial x$. The function $\mu(\theta)$ computed by LMA for such an operator may have a maximum at $(\pi,0)$ rather than at $(0,0)$ for some values of ω. There is then even less plausiblity in computing μ for a low frequency component. Secondly, the eigenfunctions of the asymmetric operator are skewed exponentially ($u(x,y) = \exp(-ax/2)\sin(n\pi x)\sin(m\pi y)$ in the unit square, for integers m,n). For h of order 1/a or larger, the discrete eigenfunctions are quite dissimilar to the Fourier components, and thus we cannot expect LMA to be valid at all. Several methods to estimate optimal parameter values using LMA were tried and none approached the experimentally determined optimal value.

Extension to problems with one variable coefficient was more successful. Numerical experiments were done using the following problem:

(4-8) $(1/10 + x^2 + y^2)\partial^2u/\partial x^2 + \partial^2u/\partial y^2 = 0,$

with Dirichlet boundary conditions on [0,1]X[0,1]. A mesh size of .02 was used. The determination of ω' was insensitive to the coefficient of $\partial^2u/\partial x^2$ over the range .1 to 2.; $\omega' = 1.88$. The value of ω^* was determined experimentally, as outlined previously, to be $\omega^* = 1.90$. Similar comparisons where a term $-u/10h^2$ was included in the discrete operator failed to give such good agreement between ω' and ω^*.

Apparently LMA has some limited usefulness with respect to choice of parameter for solution of Laplace type equations by SOR relaxation. This is particularly true for domains with a small ratio of boundary length to area, approximated by fine grids. However, there is no obvious way to use LMA to estimate the actual rate of convergence with the optimal parameter.

4.2 Alternating-Direction Implicit Iteration

For ADI iteration on an equation $AX = h^2b$, the matrix A is split into matrices H^1 and V^1. For any grid vector X, and any grid point (i,j), $(H^1X)(i,j)$ involves only the values of X at the points (i,j), (i-1,j), and (i+1,j), and $V^1X(i,j)$ involves only the values of X at the points (i,j), (i,j-1), and (i,j+1). Thus H^1 and V^1 represent the finite-differencing in the x and y directions, respectively. A parameter ρ, which may vary from one iteration to the next, determines the form of the relaxation. Optimal choice of values for ρ is important in ensuring quick convergence.

One iteration of ADI consists of solving

$$(4\text{-}9)\quad (H^1 + \rho h^2 I\ )X^* = (\rho h^2 I - V^1\ )X + h^2 b,$$

and then solving

$$(4\text{-}10)\quad (V^1 + \rho h^2 I\ )X' = (\rho h^2 I - H^1\ )X^* + h^2 b,$$

to obtain the next iterate X' from the current iterate X, using X* as an intermediate.

If the two matrices H^1 and V^1 commute, then the algebraic theory guarantees rapid convergence of the iterates (see Young 1971). The matrices commute only for the special case of separable equations of the form

$$(4\text{-}11)\quad (\partial/\partial x)(f(x)\partial u/\partial x) + (\partial/\partial y)(g(y)\partial u/\partial y) + cu = h(x,y)$$

on rectangular domains. If these conditions are not met, (for example, if large first-order derivative terms are present) then convergence is frequently slow in practice. An heuristic explanation of this phenomenon using LMA is now given. Theorem 1 of Chapter 2 provides some justification for this non-rigorous approach.

The local mode analysis for the ADI method applied to Laplace's equation runs as follows. The relaxation at grid point (i,j) consists of solving

$$(4\text{-}12)\quad ((2 + \rho h^2)X^*(i,j) - X^*(i-1,j) - X^*(i+1,j))$$
$$= ((\rho h^2 - 2)X(i,j) + X(i,j-1) + X(i,j+1)\ ) + h^2 b(i,j),$$

and then solving

(4-13) $((2 + \rho h^2)X'(i,j) - X'(i,j-1) - X'(i,j+1))$

$= ((\rho h^2 - 2)X^*(i,j) + X^*(i-1,j) + X^*(i+1,j)) + h^2 b(i,j).$

Let e^* be the error in X^*. Then the error transformation corresponding to (4-13) is

(4-14) $((2 + \rho h^2)e^*(i,j) - e^*(i,j-1) - e^*(i,j+1))$

$= ((\rho h^2 - 2)e(i,j) + e(i,j-1) + e(i,j+1))$.

Let $e(i,j) = A(\theta)\exp(\iota(\theta^1 i + \theta^2 j))$, and
let $e^*(i,j) = A^*(\theta)\exp(\iota(\theta^1 i + \&^2 j))$. Then (4-15) becomes

(4-15) $A^*(\theta)(2 + \rho h^2 - 2\cos\theta^1) = A(\theta)(-2 + \rho h^2 + 2\cos\theta^2)$.

Using similar arguments we arrive also at

(4-16) $A'(\theta)(2 + \rho h^2 - 2\cos\theta^2) = A^*(\theta)(-2 + \rho h^2 + 2\cos\theta^1)$.

Thus

(4-17) $\mu(\theta) = \frac{v - s}{v + s} \cdot \frac{v - t}{v + t}$,

where $v = \rho h^2$, $s = 2 - 2\cos\theta^1$, and $t = 2 - 2\cos\theta^2$.

Notice that it is possible to make $\mu(\theta)$ equal to 0 for any θ by suitable choice of ρ. For small ρ the components with small θ^1 or small θ^2 values are being eliminated since $1 - \cos(O(h)) = O(h^2)$. As ρ is increased, the zero of $\mu(\theta)$

occurs at increasingly higher frequencies, until for $\rho = 4/h^2$ the zero is (π,π). This is the same range for ρ that was arrived at by more difficult calculations in Young (1971).

Now consider the PDE

$$(4\text{-}18)\quad \partial^2u/\partial x^2 + \partial^2u/\partial y^2 - a\partial u/\partial x - b\partial u/\partial y = 0.$$

As mentioned above, ADI methods are not expected to work as well for this equation. The finite difference form of the equation at grid point (i,j) is

$$(4\text{-}19)\quad -4U(i,j) + (1+ah)U(i-1,j) + (1+bh)U(i,j-1) + (1-ah)U(i+1,j) + (1-bh)U(i,j+1) = 0.$$

Then LMA yields

$$(4\text{-}20)\quad \mu(\theta) = \frac{v - s - 2\imath ah\sin\theta^1}{v + s + 2\imath ah\sin\theta^1} \cdot \frac{v - t - 2\imath bh\sin\theta^2}{v + t + 2\imath bh\sin\theta^2}.$$

Notice that it is not possible to choose ρ in order to make $\mu(\theta)$ equal to 0 for any given θ. While for small ρ the minimum of $|\mu(\theta)|$ still occurs at low frequencies, its value is of order 1 if ah and bh are O(1). For large r the minimum occurs at high frequencies and although its value is not 0, it will be small (O(h)) provided that ah and bh are not $O(h^{-1})$.

Thus there is no choice of ρ which will effectively reduce the smooth (low frequency) error components. Thus although the ADI method will still effectively smooth the error its overall performance will be significantly degraded from its

performance on the model problem.

4.3 Smoothing Property of Relaxations

Suppose we have a difference equation at every interior point of a grid covering a domain Ω;

$$(4\text{-}21)\quad a^1U(i,j) + a^2U(i-1,j) + a^3U(i,j-1) + a^4U(i+1,j) + a^5U(i,j+1) = h^2b(i,j),$$

where $a^1,\ldots,a^5$ are constant coefficients. Suppose there is no term of degree zero in u in the underlying PDE so that

$$(4\text{-}22)\quad a^1 + a^2 + a^3 + a^4 + a^5 = 0.$$

Suppose we have a splitting $A = L - R$ which yields a relaxation scheme for (4-22),

$$(4\text{-}23)\quad l^1X'(i,j) + l^2X'(i-1,j) + l^3X'(i,j-1) + l^4X'(i+1,j) + l^5X'(i,j+1) = r^1X(i,j) + r^2X(i-1,j) + r^3X(i,j-1) + r^4X(i+1,j) + r^5X(i,j+1) + h^2b(i,j).$$

Then since $l^* - r^* = a^*$, $* = 1,\ldots,5$, we must have

$$(4\text{-}24)\quad l^1+l^2+l^3+l^4+l^5 = r^1+r^2+r^3+r^4+r^5.$$

The reader will recall from Chapter Three that the function $\mu(\theta)$ for the relaxation scheme (4-24) is

$$(4\text{-}25)\quad \mu(\theta) = \left(r^1 + r^2\exp(-\iota\theta^1) + r^3\exp(-\iota\theta^2) + r^4\exp(\iota\theta^1) + r^5\exp(\iota\theta^2) \right) / \left(l^1 + l^2\exp(-\iota\theta^1) + l^3\exp(-\iota\theta^2) + l^4\exp(\iota\theta^1) + l^5\exp(\iota\theta^2) \right).$$

Thus, if

$$(4\text{-}26)\quad l^1 + l^2 + l^3 + l^4 + l^5 \neq 0,^1$$

$$(4\text{-}27)\quad \mu(\theta) \to 1, \text{ as } \theta^1, \theta^2 \to 0.$$

The limit (4-28) says that the function $\mu(\theta)$ is near 1 if θ^1, θ^2 are near 0. Thus for the relaxation of a constant coefficient equation by a scheme which satisfies the conditions outlined in Chapter Three, the low frequency Fourier components in the error are not efficiently reduced. The best we can hope for in general is efficient smoothing, but not efficient reduction of the error overall.

---

[1] Only the ADI relaxation of the model problem with a small value of the parameter r comes close to violating this condition. The reader will recall that low frequency errors could be efficiently relaxed in this case.

CHAPTER FIVE

DISCUSSION

5.1 Applicability of Local Mode Analysis

The results in section 2.2 justify LMA under the following conditions. The relaxation for which LMA is done must be a linear fixed point iteration which is based on a splitting of the coefficient matrix into $A = L - R$. Alternately each relaxation step could be a sequence of several procedures each of which is based on a splitting. Although the splitting(s) need not be regular, they must be homogeneous, i.e., each point must be in the same position relative to the set of previously relaxed points. Thus red-black relaxations are excluded. A further condition, which we imposed for ease of analysis, was that in the relaxation equation at a given point, no other neighbouring points occur than those which occur in the original difference equation at that point. We considered only second order difference schemes based on five point formulae, which excluded relaxations for which the relaxation at point (i,j) required values at points other than the four nearest neighbours. This condition may not, in fact, be neccessary for the results but no relaxation scheme is known to us for which it is not true.

The condition that the relaxation be based on a splitting of the coefficient matrix was intended to ensure that Fourier analysis was meaningful. We do not think that Fourier analysis is relevant to the understanding of schemes such as the Conjugate Gradient Method.

The homogeneity condition guaranteed that the

cross-coupling of Fourier components undergoing relaxation was asymptotically negligible; i.e., that each Fourier component behaved like an eigenvector. A local Fourier analysis is possible for inhomogeneous schemes such as red-black relaxation; then however, because cross-coupling of the Fourier components is important, they must be arranged in invariant subspaces rather than being treated like eigenvectors. Under red-black ordered Gauss-Seidel relaxation, Fourier components with frequencies (θ^1,θ^2), $(\theta^1,\pi-\theta^2)$, $(\pi-\theta^1,\theta^2)$, and $(\pi-\theta^1,\pi-\theta^2)$ feed into each other significantly for all mesh sizes, and must be treated as a four-dimensional invariant subspace (see Trottenberg and Stueben, 1982).

Although the theorem in Chapter Three was established only for relaxations of scalar problems with Dirichlet boundary conditions which met the requirements described above, some extensions may be contemplated. Scalar problems with Neumann or mixed boundary conditions or higher order approximations to Dirichlet boundary conditions may be treated using the method in Chapter Three, however, the algebra may be much more complicated. It seems straightforward to extend the result to strongly elliptic systems. Extension to weakly elliptic systems may be possible, but we anticipate difficulties with distributive relaxation schemes. Of course a theorem like that in Chapter Three is not meaningful for variable coefficient problems or non-linear problems.

5.2 The Potential of LMA Outside the MG Context

The theorem in Chapter Three justified our explorations

in Chapter Four. We saw that LMA could be used to explain the behaviour of the SOR and ADI methods, and also the fact that linear relaxations reduce smooth errors poorly. We speculate that LMA has the potential to be more widely used in order to gain further heuristic understanding of such methods.

Frequently a scientist considers the use of a particular relaxation scheme for the solution of a problem and wishes to know how appropriate it is or how to pick relevant parameters. A rigourous algebraic answer would involve finding the eigenvalues and the eigenspaces of the amplification matrix of the relaxation. This is generally more work than is needed to solve the original problem. On the other hand LMA is a simple analytical tool and it can yield much the same information.

5.3 Possible Developments of the Theorem

The theorem in Chapter Three establishes that LMA is a meanigful tool and imparts confidence in its use for MG problems. However, we have only shown that LMA yields a good description of the results of relaxation where the error is a multiple of a single Fourier component. As will be discussed below it would be very useful to have a result of this kind for arbitrary errors.

Suppose we have an arbitrary error $e = x-u$ and decompose it as

$$(5\text{-}1)\quad e = \sum_{\theta} A(\theta)\psi(\theta).$$

We ask how closely does e' the error after relaxation resemble e^* the LMA estimate, which is given by

(5-2) $e^* = \sum_{\theta} \mu(\theta) A(\theta) \psi(\theta)$.

An argument along the lines of Chapter Three might be made that the difference $e' - e^*$ is asymptotically small. We will sketch this here. We will consider only rectangular regions so that the Fourier decomposition (5-1) is well defined and the coefficients $A(\theta)$ form the finite Fourier transform of e, where θ^1 and θ^2 take the values $2\pi k/N$ for $k = 0,1,\ldots,N-1$. Note that the low frequency region is $(0,\pi/2)U(3\pi/2,2\pi)$ with these choices of frequencies. This is equivalent to the previous definition. Now matrices R^* and L^* may be constructed independently of θ so that

(5-3) $R^*\psi(\theta) = r^*(\theta)\psi(\theta)$ and $L^*\psi(\theta) = l^*(\theta)\psi(\theta)$,

for $\theta^1,\theta^2 = 2\pi k/N$. Then since $r^*(\theta)/l^*(\theta) = \mu(\theta)$, formula (3-17) reads

(5-4) $e' = e^* + L^{-1}(L1e^* - R1e)$,

where again $L1 = L^* - L$ and $R1 = R^* - R$. Now however we have no guarantee that $L1e^*$ and $R1e$ are asymptotically small relative to e. In fact if e is non zero only on points next to the boundary, $L1e^*$ and $R1e$ can be large for any N. Thus the estimate obtained in Chapter Three may not be extended to arbitrary errors.

Does this fact invalidate convergence estimates of MG algorithms that rely on LMA for estimating the smoothing factor for any error? I think not. The smoothing rate is taken as $\mu=\max|\mu(\theta)|$ for one of θ^1,θ^2 in the range $(\pi/2,3\pi/2)$. Most

of the high frequency Fourier components are reduced by a factor much smaller than μ. I believe that it is still true that the non-smooth part er [1] of an arbitrary error e is still reduced by a factor which is bounded by $\mu + O(1/\sqrt[4]{N})$.

If this could be proved then it would be possible to show convergence of the MG method in much more generality than previously. Trottenberg and Stuben (1981) have already shown that direct solution of the coarse grid problem (the coarse grid correction step) significantly reduces the smooth component es = e - er of the error in the fine grid problem. Thus a proof that the relaxation on the fine grid strongly reduces the non-smooth error is all that is needed to show convergence of the whole MG algorithm.

So far proofs of the convergence of the MG method have been difficult and restricted to a few simple cases. They have also been highly artificial. A proof via the LMA estimate which is used in practice would go a long way toward bringing together theory and application.

5.4 Summary

There seems to be potential for development of LMA beyond the practical role for which Brandt introduced it. As we have seen, under suitable conditions, local Fourier analysis can yield quantitative information about the performance of relaxation schemes outside the MG context. It can also be made rigourous and thus is possibly part of a sound mathematical basis for the MG method.

---

[1] This is the orthogonal projection of e onto the space spanned by $\{\psi(\theta) \mid \theta^1,\theta^2=2\pi k/N$, one of $\theta^1,\theta^2 \in (\pi/2,3\pi/2)\}$

BIBLIOGRAPHY

Brandt, A. "Multi-Level Adaptive Solutions to Boundary-Value Problems", Mathematics of Computation, 31 (1977), 333-390.

Brandt, A. "Guide to Multgrid Development", in Multigrid Methods, W. Hackbusch and U. Trottenberg, eds., Koeln-Porz, Nov 1981. LNM 960, Springer Verlag (1982), 220-312.

Hackbusch, W. "Multi-Grid Convergence for a Singular Perturbation Problem", Proceedings, Copper Mountain, Apr 1983. (to appear).

Hemker, P. W. "Fourier Analysis of Grid Functions, Prolongations and Restrictions", Report NW 93/80, Mathematical Centre, Amsterdam (1980).

Hemker, P. W. "The Incomplete LU-Decomposition as a Relaxation Method in Multi-Grid Algorithms", in Boundary and Interior Layers - Computational and Asymptotic Methods, J. Miller ed., Boole Press, Dublin (1980), 402-406.

Hemker, P. W. "An Accurate Method Without Directional Bias for the Numerical Solution of a 2-D Elliptic Singular Perturbation Problem", Report NW 117/81, Mathematical Centre, Amsterdam (1981).

Mol, W.J. "Smoothing and Coarse Grid Approximation Properties of Multigrid Methods", Report NW 110/81, Mathematical Centre, Amsterdam (1981).

Trottenberg, U. and K. Stueben. "Multigrid Methods", in Multigrid Methods, W. Hackbusch and U. Trottenberg, eds., Koeln-Porz, Nov 1981. LNM 960, Springer Verlag (1982), 1-176.

de Vries, H.B. "The Two-Level Algorithm for the Solution of the Initial Value Problem for Partial Differential Equations", Report NW 136/82, Mathematical Centre, Amsterdam (1982).

Young, D. M. Iterative Solution of Large Linear Systems. Academic Press, New York, 1971.

Appendix A

Proof of Lemma 3, Chapter Two

Let $\gamma(t) : [0,L] \to R^2$ be a continuous piecewise differentiable curve parametrized by arc length. For $0 < t < L$, let $H(t,h)$ be

(A-1) $H(t,h) = \{(x,y) \mid |(x,y)-\gamma(t^*)| < h \text{ for some } t^*, 0<t^*<t\}$

Define $A1(t)$ as the area of $H(t,h)$, denoted $A(H(t,h))$

Lemma A1: If γ is a straight line then $dA1/dt = 2h$.

Proof: Obvious.

Lemma A2 : $A(H(L,h)) = A1(L) \leq \pi h^2 + 2hL$

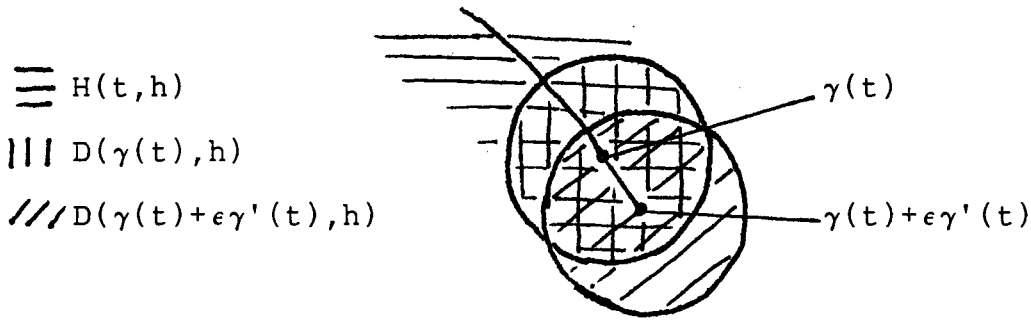
Proof: Let $D(a,h)$ be a disk of radius h centered at a point a in R^2. Then clearly $A1(t) = A(\cup D(\gamma(\tau),h), 0 \leq \tau \leq t)$. Clearly A1 is a continuous non-decreasing function of t. It is differentiable whenever γ is. Let t be a point where γ is differentiable; then (see Figure)

(A-2) $A1'(t) = \lim_{\epsilon \to 0} (A(H(t+\epsilon,h)) - A(H(t,h)))/\epsilon$

$= \lim_{\epsilon \to 0} (A(D(\gamma(t)+\epsilon\gamma'(t),h) \cup H(t,h)) - A(H(t,h))) / \epsilon$

$= \lim_{\epsilon \to 0} A(D(\gamma(t)+\epsilon\gamma'(t),h) \setminus H(t,h)) / \epsilon$

$\leq \lim_{\epsilon \to 0} A(D(\gamma(t)+\epsilon\gamma'(t),h) \setminus D(\gamma(t),h)) / \epsilon.$



Figure

This last limit is independent of γ (since $|\gamma'(t)| = 1$ was specified). If γ is a straight line the last inequality is an equality and for this case $A1'(t) = 2h$ by Lemma A1. Thus

(A-3) $A1'(t) \leq 2h$.

Since $A1(0) = \pi h^2$, the lemma follows. QED

Lemma A3: (Lemma 2 of Chapter Three): Let Ω be a domain with piecewise smooth boundary γ. There is a constant a such that if G is any sufficiently fine grid, and N is the number of grid points of G which lie interior to Ω, then the number M of such points which lie next to a grid point which is outside Ω is bounded by

(A-4) $M \leq a\sqrt{N}$

Proof: (I am indebted to David Boyd of the Math. Dept. UBC for his suggestion of using grid squares)

Let A be the area of Ω, L be the length of γ, and h be the mesh size of G. Let n be the number of grid squares (each of area h^2) which lie entirely in Ω. Let m be the number of squares which intersect γ, or whose boundary

intersects γ. Then

(A-5) $A \leq (m+n)h^2$.

The m squares which meet γ must lie entirely in the set $H(L,2\sqrt{2}h)$ defined above. Now by Lemma A2

(A-6) $mh^2 \leq A(H(L,2\sqrt{2}h) \leq 8\pi h^2 + 4\sqrt{2}hL$.

Choose $h < \sqrt{2}L/(8\pi)$. Then

(A-7) $mh^2 \leq 5\sqrt{2}Lh$.

Now choose $h < A/(10\sqrt{2}L)$. Clearly $A(H(L,2\sqrt{2}h)) \leq A/2$, thus by (A-7) and (A-9)

(A-8) $nh^2 \geq A/2$.

Thus

(A-9) $m \leq 5\sqrt{2}L/h$

$= (10L/\sqrt{A})(\sqrt{A}/\sqrt{2}h)$

$\leq (10L/\sqrt{A})\sqrt{n}$.

Now each of the n squares interior to Ω contribute at least one point to N (say the lower left corner). Thus $n \leq N$. Each of the M points which lie next to a point outside Ω, is a corner of one of the m squares. Clearly there are at most three contributions to M from each of the m squares. Thus

(A-10) $M \leq 3m \leq (30L/\sqrt{A})\sqrt{n} \leq (30L/\sqrt{A})\sqrt{N}$. QED

Appendix B

Norms of the matrices L and L^{-1} for the Gauss-Seidel Relaxation of the Model Problem

1. Form of L and L^{-1}

For an nxn grid with n^2 interior points the n^2 by n^2 matrix L has the form

$$
\text{(B-1)} \qquad
\begin{bmatrix}
\Lambda & 0 & 0 & \cdots & 0 \\
-I & \Lambda & 0 & & \vdots \\
0 & -I & \Lambda & & \vdots \\
\vdots & & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & -I & \Lambda
\end{bmatrix}
$$

where Λ is an nxn matrix of the form

$$
\text{(B-2)} \qquad
\begin{bmatrix}
4 & 0 & 0 & 0 & \cdots & 0 \\
-1 & 4 & 0 & 0 & & \vdots \\
0 & -1 & 4 & 0 & & \vdots \\
\vdots & & \ddots & \ddots & & \vdots \\
\vdots & & & \ddots & \ddots & 0 \\
0 & \cdots & & 0 & -1 & 4
\end{bmatrix}
$$

The reader may verify that Λ^{-1} has the form

$$
\text{(B-3)} \qquad
\begin{array}{cccccc}
1/4 & 0 & 0 & 0 & \cdots & 0 \\
1/16 & 1/4 & 0 & 0 & & \cdot \\
1/64 & 1/16 & 1/4 & 0 & & \cdot \\
\cdot & & & \cdot & & \cdot \\
\cdot & & & & \cdot & 0 \\
\cdot & & & & \cdot & \\
 & \cdots & & & 1/16 & 1/4
\end{array}
$$

where in the lower triangle the (i,j) entry is 4 to the power $-1 - i + j$.

L^{-1} has the form

$$
\text{(B-4)} \qquad
\begin{array}{cccccc}
\Lambda^{-1} & 0 & 0 & 0 & \cdots & 0 \\
\Lambda^{-2} & \Lambda^{-1} & 0 & 0 & & \cdot \\
\Lambda^{-3} & \Lambda^{-2} & \Lambda^{-1} & 0 & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & & \cdot & \cdot & 0 \\
\cdot & & \cdots & & \Lambda^{-2} & \Lambda^{-1}
\end{array}
$$

where the block in position (i,j) is Λ to the power $-1-i+j$.

If we compare the (i,j) entry in $\Lambda\Lambda^{-k}$ with the (i,j) entry in Λ^{-k+1}, we obtain

$$\text{(B-5)} \quad 4\Lambda^{-k}(i,j) = \Lambda^{-k}(i-1,j) + \Lambda^{-k+1}(i,j).$$

Clearly the elements along the main diagonal of Λ^{-k} are

$$\text{(B-6)} \quad \Lambda^{-k}(i,i) = \frac{1}{4^k}.$$

Using (B-6) and (B-5) the reader may verify that

$$(B\text{-}7)\quad \Lambda^{-k}(i,j) = \frac{(k+i-j-1)!}{(k-1)!\ (i-j)!}\Big/4^{k+i-j}, \text{ if } i\geq j, \text{ and}$$

$$= 0, \qquad \text{if } i<j.$$

Note that the entries in the rows are equal to the entries in the columns under the correspondence

$$(B\text{-}8)\quad \Lambda^{-k}(i,j) = \Lambda^{-k}(n-j+1,n-i+1).$$

Inspection of the form of (B-4) shows that then

$$(B\text{-}9)\quad L^{-1}(i,j) = L^{-1}(n^2-j+1,n^2-i+1).$$

2. Norms

The square of the 2-norm induced on L by the 2-norm on grid vectors x is

$$(B\text{-}10)\quad |L|_2^2 = \max\{(Lx,Lx)\mid\ |x|_2=1\}.$$

Thus $|L|_2 = \sqrt{\kappa}$ where κ is the largest eigenvalue of L^TL. Let z be an eigenvector of L^TL with eigenvalue κ. Then

$$(B\text{-}11)\quad |\kappa| = |L^TLz|_1 / |z|_1 \leq |L^T|_1 |L|_1 = |L|_\infty |L|_1,$$

where $|L|_1$ denotes the induced 1-norm which is the maximum of the sums of the entries in the columns of L, and $|L|_\infty$ denotes

the induced supremum norm which is the maximum of the sums of the entries in the rows of L. From (B-1) and (B-2) both are 6. Thus $\kappa \leq 36$ and

(B-12) $|L|_2 \leq 6.$

By a similar argument

(B-13) $|L^{-1}|_2^2 \leq |L^{-1}|_\infty |L^{-1}|_1.$

Because of the correspondence (B-9),

$|L^{-1}|_\infty = |L^{-1}|_1,$

and we may write

(B-14) $|L^{-1}|_2 \leq |L^{-1}|_1.$

The reader may verify by inspection of (B-4) and (B-7) that the sum of the entries in the first column of L^{-1} is larger than the sum of the entries in any other column. Thus

(B-15) $$|L^{-1}|_2 \leq |L^{-1}|_1$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{(k+i-2)!}{(k-1)!(i-1)!} \Big/ 4^{k+i-1}$$

$$< \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} \frac{(k+i-2)!}{(k-1)!(i-1)!} \Big/ 4^{k+i-1}$$

$$= \sum_{m=1}^{\infty} \frac{1}{4^{m+1}} \left(\sum_{j=0}^{m} \frac{m!}{(m-j)!j!} \right)$$

$$= \sum_{m=0}^{\infty} \frac{1}{4^{m+1}} \cdot (2^m)$$

$$= \sum_{m=0}^{\infty} \frac{1}{2^{m+2}}$$

$$= 1/2.$$