

**A Complex Analysis Based Derivation of Multigrid
Smoothing Factors for Lexicographic Gauss-Seidel**

by

Laird Robert Hocking

B.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES
(Computer Science)

The University Of British Columbia
(Vancouver)

August 2011

© Laird Robert Hocking, 2011

Abstract

This thesis aims to present a unified framework for deriving analytical formulas for multigrid smoothing factors in arbitrary dimensions, under certain simplifying assumptions. To derive these expressions we rely on complex analysis and geometric considerations, using the maximum modulus principle and Möbius transformations. We restrict our attention to pointwise and block lexicographic Gauss-Seidel smoothers on a d -dimensional uniform mesh, where the computational molecule of the associated discrete operator forms a $2d + 1$ point star. In the pointwise case the effect of a relaxation parameter, as well as different choices of mesh ratio, are analyzed. The results apply to any number of spatial dimensions, and are applicable to high-dimensional versions of a few common model problems with constant coefficients, including the Poisson and anisotropic diffusion equations as well as the convection-diffusion equation. We show that in most cases our formulas, exact under the simplifying assumptions of Local Fourier Analysis, form tight upper bounds for the asymptotic convergence of geometric multigrid in practice. We also show that there are asymmetric cases where lexicographic Gauss-Seidel smoothing outperforms red-black Gauss-Seidel smoothing; this occurs for certain model convection-diffusion equations with high mesh Reynolds numbers.

Preface

This thesis describes the results of a refereed journal paper,

L. Robert Hocking and Chen Greif. Closed-Form Multigrid Smoothing Factors for Lexicographic Gauss-Seidel. *IMA Journal of Numerical Analysis*, accepted for publication, August 2011.

A body of unpublished work is described as well.

The results of the paper are described in Chapters 3 and 5, while Chapters 2 and 4 represent new additions.

I have taken a central role in all aspects of the paper. I am the first author, with collaboration by my supervisor Chen Greif.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents	iv
List of Tables	vi
List of Figures	viii
Acknowledgments	xi
1 Introduction	1
1.1 Thesis Contributions	3
1.2 Outline	4
2 Overview of Multigrid	6
2.1 Quantitative Convergence Analysis	11
2.2 A Closer Look at the Interplay Between Errors on the Fine and Coarse Grids	15
3 Smoothing Analysis	20
3.1 Pointwise Relaxation	22
3.2 SOR Relaxation	28
3.3 Block Relaxation	29
4 Alternative Mesh-Ratios	33

5	Examples	39
5.1	The d -Dimensional Poisson Problem	40
5.2	The Anisotropic Steady-State Diffusion Problem	41
5.3	Alternative Coarsening for the 2D Poisson Problem	42
5.4	The Convection-Diffusion Problem	44
5.5	The Time-dependent Diffusion Problem	47
6	Conclusions	49
	Bibliography	51

List of Tables

Table 4.1	Relative work of optimal coarsening and corresponding values of β_{opt} , for various d -dimensional Poisson problems.	36
Table 5.1	Smoothing factors of GS-LEX with point, line, and plane relaxation for the Poisson problem in 1D, 2D, and 3D. Moving from point to line or from line to plane relaxation is equivalent to reducing the dimension by one.	40
Table 5.2	Comparison of predicted smoothing factors of GS-LEX with measured convergence rates of multigrid $V(1, 1)$ -cycles, for the PDE $-\varepsilon u_{xx} - u_{yy} = f$. The finest grid is 1023×1023 , while the coarsest is 1×1 . Various values of ε are considered.	42
Table 5.3	Predicted smoothing factors smoothing vs. measured convergence rates of multigrid with $H = kh$ coarsening and GS-LEX smoothing, applied to the 2D Poisson problem discretized using centered differences, for various values of k . $V(1, 0)$ -cycles are used, with a finest grid of size $k^{N_k} - 1$ in each direction, and a coarsest grid consisting of a single point. $N_k = \lfloor \log_k(999) \rfloor$ is chosen to make the total number of grid points approximately one million. Also included is the total work as measured by the running time of the program divided by the size of the problem - this is given both in microseconds per grid point and as a ratio relative to the reference case $H = 2h$. This latter form is compared against the theoretical relative work (4.2).	43

Table 5.4	Smoothing factors of pointwise GS-LEX applied to two common discretizations of the convection-diffusion equation with all mesh Reynolds numbers equal to γ , for $d = 1, 2, 3$	45
Table 5.5	Comparison of the smoothing factor μ of GS-LEX with the measured asymptotic convergence rate of multigrid applied to the PDE $-u_{xx} - u_{yy} + \sigma u_x + \sigma u_y = f$, discretized uniformly. The problem is discretized using upwinding on a 1023×1023 grid. The Galerkin coarse grid operator is used. $W(1,0)$ -cycles are used, and various values of $\gamma = \sigma h/2$, where h is the grid spacing on the finest mesh, are considered. ‘NC’ stands for no convergence. ‘ nL ’ (with $n = 2, 3, 4$) signifies the number of levels.	45
Table 5.6	Comparison of the smoothing factor μ of GS-LEX with measured convergence rates of multigrid $V(1,0)$ and $W(1,0)$ -cycles, for the PDE $-u_{xx} - u_{yy} + \sigma(u_x + u_y) = f$, discretized uniformly. The finest grid is 1023×1023 , while the coarsest is 1×1 . The coarse grid operator is obtained by direct discretization of the PDE on the coarse mesh. Various values of $\gamma = \sigma h/2$ are considered, where h is the grid spacing of the finest mesh. ‘ $2L$ ’ stands for a two-level scheme, whereas $V(1,0)$ and $W(1,0)$ signify V and W -cycles, respectively, with one pre-smoothing sweep and no post-smoothing.	46
Table 5.7	Comparison of the smoothing factor μ of GS-LEX with the measured asymptotic convergence rate of two-level and multi-level multigrid applied to the PDE $-u_{xx} - u_{yy} + \sigma(u_x + u_y) = f$, discretized uniformly. The problem is discretized using centered differences on a 1023×1023 grid. The coarse grid operator is obtained by direct discretization of the PDE on the coarse mesh. $W(1,0)$ -cycles are used, and various values of $\gamma = \sigma h/2$, where h is the grid spacing on the finest mesh, are considered. ‘NC’ stands for no convergence.	47

List of Figures

Figure 1.1	The computational molecule associated with the operator \mathcal{L}^h ($d = 2$).	3
Figure 2.1	Top: a ‘smooth’ image is shrunk by a factor of 4 in each direction, and then blown back up. Bottom: the same is done for a ‘rough’ image. Reduction and subsequent enlargement were done using Photoshop’s bilinear image scaling option.	6
Figure 2.2	Radii Petrovich Fedorenko (1930-2009) is credited with inventing the multigrid algorithm.	7
Figure 2.3	A single iteration of two-level multigrid applied to the two-dimensional Poisson equation on the square $(0, 1)^2$, discretized on a 23×23 uniform grid using centered differences. The initial error, containing both high and low frequencies, is shown on the left. In the middle we see the error after the smoother S (in this case three iterations of lexicographic Gauss-Seidel) is applied - the high frequency components of the error have been significantly dampened. On the right we see the new error after the coarse grid correction K has been applied - all frequencies are now highly reduced.	8
Figure 2.4	The sequence of grids visited by one iteration of four level multigrid, for two different choices of γ . (a) $\gamma = 1$: a V-cycle (b) $\gamma = 2$: a W-cycle.	10

Figure 2.5	$\ e^{(k)}\ _2$ vs. k for two-level multigrid with lexicographic SOR smoothing applied to anisotropic diffusion problem $\frac{1}{100}u_{xx} + u_{yy} + u_{zz} = 0$, discretized using centered differences. (a) $\omega = 1$. (GS-LEX) (b) $\omega = 1.63$ (SOR-LEX). Note that for $\omega = 1$, the norm of the error goes down monotonically, whereas for $\omega = 1.63$ there is a initial increase before the asymptotic regime is entered. Convergence is faster overall for $\omega = 1.63$, despite the erratic behavior of the error initially.	13
Figure 2.6	The distinct Fourier modes $\varphi_1(k) = \sin(k\pi/20)$ and $\varphi_{19}(k) = \sin(19k\pi/20)$ on the fine grid $G_h = \{k/20\}_{k=0}^{20}$ restrict to the same mode $\varphi_1(K) = \sin(K\pi/10)$ on the coarse grid $G_H = \{K/10\}_{K=0}^{10}$	17
Figure 3.1	An illustration of Example 3.1.3: (a) If a double pendulum with arms of length $\frac{1}{6}$ and $\frac{1}{3}$ is allowed to swing freely, it remains inside a disk of radius 0.5. (b) If one arm is constrained to lie in the left half plane, the pendulum swings in the heart-shaped region shown. In both cases the pendulum is unable to reach a set of points surrounding the origin.	23
Figure 3.2	The heart-shaped set D is the union of two disks and a half disk.	25
Figure 3.3	Left: Smoothing factor of SOR-LEX with relaxation parameter ω , as a function of ω , for the standard seven-point centered differences discretization of the 3D Poisson problem. Right: Smoothing factors of GS-LEX and optimal SOR-LEX applied to the anisotropic diffusion problem $-\epsilon u_{xx} - u_{yy} - u_{zz} = f$, as a function of ϵ , assuming seven-point centered differences discretization.	29
Figure 4.1	(a) An example pacman set with $c = 0.5$ and $\beta = 1/4$. (b) The corresponding set D_β when $c_m = c_r = 0.25$. This example corresponds the standard centered differences discretization of the two dimensional Poisson problem, with the coarsening strategy $H = 4h$	34

Figure 4.2	The relative work $W_\beta/W_{0.5}$ vs β for the d -dimensional Poisson problem with $d = 2, 3, 4$	36
Figure 4.3	A fine and coarse grid with mesh ratio $h/H = 2/3$. The stencil of the projection operator varies from point to point, depending on whether the coarse grid lines up with fine grid.	37
Figure 4.4	The relative work $W_\beta/W_{0.5}$ for the d -dimensional Poisson problem converges to the limit curve shown as $d \rightarrow \infty$	38
Figure 5.1	Theoretical smoothing factors of GS-LEX vs. experimental asymptotic convergence rates of two-level multigrid with GS-LEX and GS-RB smoothers, applied to the steady-state diffusion equation $-u_{xx} - u_{yy} - \epsilon u_{zz} = f$. Pointwise relaxation appears on the left, while x -oriented line relaxation is depicted on the right. The discretization is on a $23 \times 23 \times 23$ grid with centered differences.	41

Acknowledgments

I would like to thank my supervisor Chen Greif for his ongoing guidance, support, and friendship.

I would also like to acknowledge the group of rotations and scalings in two dimensional Euclidean space, for possessing a structure such that $\sqrt{-1}$ and complex analysis may exist¹, thereby making the present work possible.

¹It is perhaps unfair to give sole credit to the aforementioned group of matrices. I should also give thanks to the structure of polynomials, in particular with respect to a certain equivalence relation. Perhaps more fundamentally I should thank \mathbb{R}^2 , for admitting a field, or maybe even reason itself, for making this inevitable.

Chapter 1

Introduction

Multigrid algorithms, a class of iterative methods for the numerical solution of elliptic partial differential equations, are based on the alternating application of two numerical procedures having complementary effects on the error. The first, called a *smoother*, is typically a simple relaxation scheme such as damped Jacobi or Gauss-Seidel. While ineffective at making the error *small*, smoothers are effective at making it *smooth* (in a geometric sense, relative to the underlying mesh). The second, called *coarse grid correction*, projects the error down to a coarser grid (with little loss in information, because the error is smooth), solves for it there, and interpolates the result onto the original grid. The result is an approximate error which may be used to update the current guess in such a way that the new error is *small*, but not necessarily *smooth*. The smoother may then be applied again, and the process continued. A detailed overview of the method is provided in Chapter 2.

For certain elliptic problems (such as the Poisson problem) multigrid is the most efficient known method to date, the total work being linear in the number of unknowns with small hidden constants [19]. Unfortunately, the performance of the method is quite sensitive to the nature of the problem at hand. For example, if multigrid with lexicographic Gauss-Seidel (GS-LEX) smoothing is applied to the anisotropic diffusion problem

$$\varepsilon u_{xx} + u_{yy} = f$$

discretized using centered differences, convergence becomes arbitrarily slow as $\varepsilon \rightarrow 0$ (see Chapter 5 Section 5.2). In this case, the issue is the smoother - although effective for $\varepsilon \approx 1$, GS-LEX fails to provide adequate smoothing for $\varepsilon \ll 1$. The problem may be resolved by, for example, replacing the *point* GS-LEX smoother with a *block* GS-LEX procedure that relaxes unknowns in the y -direction simultaneously. In general, one must tailor the smoother to the problem at hand, and finding an effective smoother is a major component in the design of an efficient multigrid algorithm.

An important tool in this respect is *Local Fourier Analysis* (LFA), which provides a quantitative estimate (called the smoothing factor and denoted by μ) of the asymptotic error reduction factor per iteration of multigrid with a given smoother applied to a given problem [3]. LFA isolates the effect of the smoother by making idealizing assumptions regarding the coarse grid correction and doing away with the issue of boundary conditions by assuming an infinite grid - it therefore represents the efficiency a multigrid method can attain if the coarse grid correct works perfectly¹ and if boundary conditions are handled suitably (in some cases by additional relaxations near the boundary) [13, p. 100].

Although smoothing factors may be computed numerically by straightforward techniques, closed form formulas are desirable if possible. This is particularly true when one wishes to understand how the smoothing factor varies as a function of one or more parameters in the PDE of interest (for example the mesh Reynolds numbers of a discrete convection-diffusion problem).

An example of the success of the analytical approach in this regard is the 1995 paper of Irad Yavneh [17], where closed form expressions for pointwise and block Gauss Seidel with *red-black* ordering (GS-RB) are derived for the family of problems

$$c_1 u_{x_1 x_1} + c_2 u_{x_2 x_2} + \dots + c_d u_{x_d x_d} = f,$$

discretized using centered differences, and normalized so that $\sum_{k=1}^d c_k = 1$. Here, among other observations, it is found that in the pointwise case μ depends only on $c_{\min} \equiv \min(c_1, c_2, \dots, c_d)$.²

¹In Chapter 5 Section 5.4 we will see an example where the assumptions of LFA regarding the coarse grid correction are *not* realistic.

²In Chapter 3, Theorem 3.1.5 I will show that the same is true of GS-LEX.

In contrast to GS-RB, the literature is relatively sparse with regards to closed form formulas for smoothing factors of GS-LEX, and seems to be limited to the case where $d \leq 2$. In particular, while the smoothing factor of GS-LEX applied to the standard 5-point discretization of the 2D Poisson problem has been known to be $\mu_{2D} = 0.5$ since 1977 [3], an analogous expression for the 3D Poisson problem does not seem to have appeared in the literature (however, the numerical result $\mu_{3D} \approx 0.567$ has appeared for example in [13, p. 74]).

It is the goal of this thesis to fill a hole in the literature by providing closed form expressions for the smoothing factors of pointwise and block GS-LEX, valid in arbitrary dimensions, for a broad class of difference operators. I will show (Chapter 5) how the smoothing factors of a variety of d -dimensional problems, including Poisson's equation, steady state diffusion, and a special case of convection-diffusion, can all be obtained as special cases of a single expression. In particular, the family of problems considered for GS-RB in [17] is such a special case.

My analysis relies on results from the theory of analytic functions of a complex variable, in particular the maximum modulus principle and properties of Möbius Transformations [10].

1.1 Thesis Contributions

The purpose of this thesis is to provide smoothing factors for a family of discrete elliptic operators that includes a number of common model problems.

To that end, suppose $\mathcal{L}u = f$ is a given elliptic differential equation (assumed linear), and let

$$\sum_{\vec{J}} \mathcal{L}_{\vec{I},\vec{J}}^h u_{\vec{J}}^h = f_{\vec{I}}^h \quad (1.1)$$

denote the linear system arising from its discretization on a uniform rectangular mesh, extending to infinity in all directions, so that \vec{I} and \vec{J} vary over \mathbb{Z}^d . We will assume the computational molecule forms a $2d + 1$ point star, or equivalently that $\mathcal{L}_{\vec{I},\vec{J}}^h$ is of the form

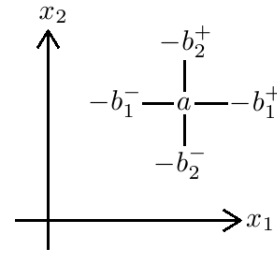


Figure 1.1: The computational molecule associated with the operator \mathcal{L}^h ($d = 2$).

$$\mathcal{L}_{\vec{I}, \vec{J}}^h = \begin{cases} a & \text{if } \vec{I} - \vec{J} = \vec{0} \\ -b_k^+ & \text{if } \vec{I} - \vec{J} = e_k \\ -b_k^- & \text{if } \vec{I} - \vec{J} = -e_k \\ 0 & \text{otherwise} \end{cases} \quad (1.2)$$

(here e_k denotes the k th basis vector of \mathbb{R}^d).

The connection between $\mathcal{L}_{\vec{I}, \vec{J}}^h$ and the computational molecule is illustrated in Figure 1.1.

I assume throughout that $\mathcal{L}_{\vec{I}, \vec{J}}^h$ satisfies

$$a \geq \sum_{k=1}^d (b_k^+ + b_k^-) \quad \text{and} \quad b_k^+, b_k^- > 0 \quad \text{for all } k \in \{1, \dots, d\}, \quad (1.3)$$

so that $[\mathcal{L}_{\vec{I}, \vec{J}}^h]$ is a diagonally dominant M-matrix [15, p. 85].

In this thesis we will derive analytical smoothing factors for \mathcal{L}^h of the form (1.2) obeying the constraints (1.3), as well as one additional technical assumption (3.2) in Chapter 3. My analysis relies on results from the theory of analytic functions of a complex variable, in particular the maximum modulus principle and properties of Möbius Transformations [10].

1.2 Outline

- Chapter 2 provides an overview of multigrid. Special attention is given to the theoretical justification for idealization of the coarse grid correction in Local Fourier Analysis, with a novel interpretation in terms of the Nyquist-Shannon theorem [12] from sampling theory. The more conventional justification is also given, although in somewhat unorthodox terms using equivalence classes.
- Chapter 3 is the main analytical portion of the thesis, in which the main results are derived. In particular, Theorem 3.1.5 provides an expression for the smoothing factor of point GS-LEX applied to operators of the form (1.2)

obeying the constraints (1.3) as well as the additional constraint (3.2). These results are then extended in Section 3.2 to incorporate the effect of a relaxation parameter. In Theorem 3.3.1, the results are extended in a different direction to *block* GS-LEX relaxations in which a subset $\mathcal{I}^b \subset \{1, \dots, d\}$ of indices are relaxed simultaneously. However, in this case the additional complexity is such that our results only apply to the symmetric case where $b_k^+ = b_k^-$ for all k , and a relaxation parameter is not included.

- Chapter 4, explores the dependence of the smoothing factor on the ratio of the mesh spacing on the fine and coarse grids. An estimate of the total work is included, and the question of the optimal ratio minimizing this work is considered.
- In Chapter 5, a series of example problems are considered, and the results of Chapter 3 is used to derive smoothing factors in each case. These smoothing factors are then compared against the actual performance of multigrid in numerical experiments. Differences between theoretical and experimental results are explored where they exist.
- Finally, Chapter 6 draws some conclusions.

Chapter 2

Overview of Multigrid

If one shrinks a sufficiently smooth image down to a reduced size and then blows it back up, it is essentially unaffected. However, the same is not true of a ‘rough’ image, in which the color changes rapidly from pixel to pixel - see Figure 2.1.

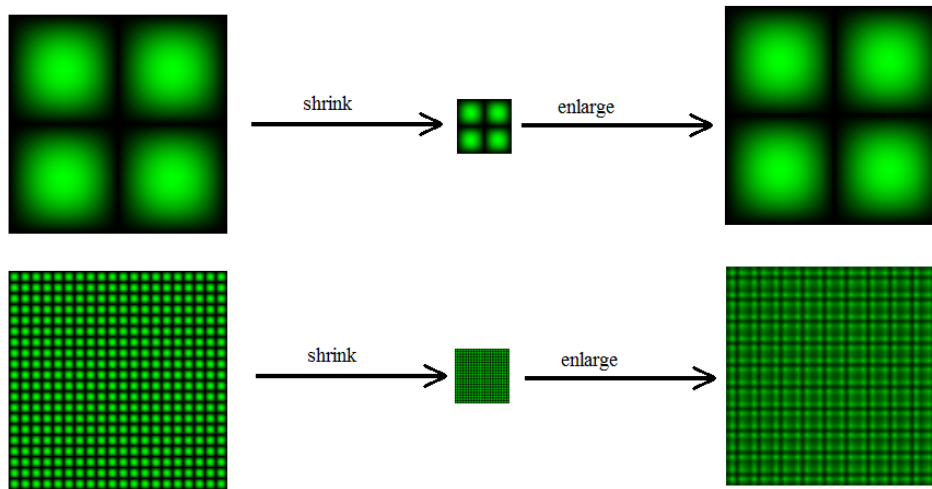


Figure 2.1: Top: a ‘smooth’ image is shrunk by a factor of 4 in each direction, and then blown back up. Bottom: the same is done for a ‘rough’ image. Reduction and subsequent enlargement were done using Photoshop’s bilinear image scaling option.

A similar situation arises in the iterative solution of difference equations arising from the discretization of elliptic partial differential equations (PDEs). If the error between the current approximation and the exact solution is sufficiently smooth, one may project the problem onto a coarser mesh (multigrid researchers call this restriction), solve for the error there, and interpolate it back to the original mesh (prolongation), obtaining a good approximation to the error on the fine mesh. This process is called *coarse grid correction*.

Let $\mathcal{L}u = f$ denote the PDE of interest (assumed linear) and let

$$\mathcal{L}^h u^h = f^h \tag{2.1}$$

denote the linear system arising from its discretization; in this thesis it is always assumed that the discretization takes place on a uniform rectangular grid. If \hat{u}^h denotes the current approximate solution, then the error $e^h = u^h - \hat{u}^h$ can be computed by solving the residual equation

$$\mathcal{L}^h e^h = r^h, \tag{2.2}$$

where $r^h \equiv f^h - \mathcal{L}^h \hat{u}^h$.

The coarse grid correction calculates an approximate error \tilde{e}^h by computing r^h on the fine grid, restricting it to a suitable coarse grid, solving the residual equation (2.2) on the coarse grid using the restricted residual as the right hand side, and prolonging the result back up to the fine grid. The current guess \hat{u}^h is then updated via $\hat{u}^h \rightarrow \hat{u}^h + \tilde{e}^h$. Note that if \tilde{e}^h was the true error, this update would yield the exact answer u^h .

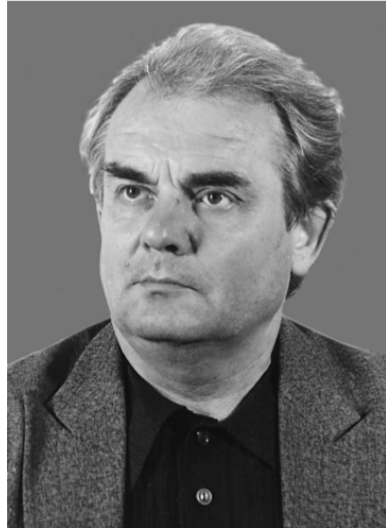


Figure 2.2: Radii Petrovich Fedorenko (1930-2009) is credited with inventing the multigrid algorithm.

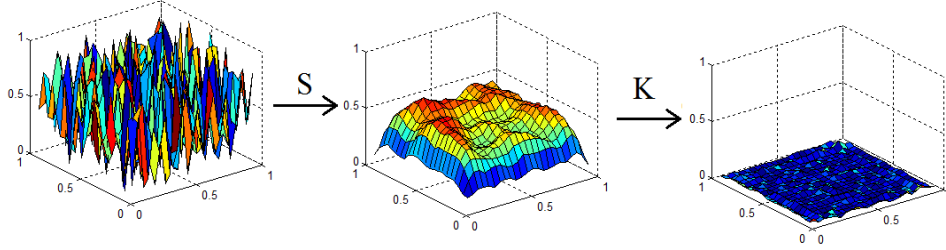


Figure 2.3: A single iteration of two-level multigrid applied to the two-dimensional Poisson equation on the square $(0, 1)^2$, discretized on a 23×23 uniform grid using centered differences. The initial error, containing both high and low frequencies, is shown on the left. In the middle we see the error after the smoother S (in this case three iterations of lexicographic Gauss-Seidel) is applied - the high frequency components of the error have been significantly dampened. On the right we see the new error after the coarse grid correction K has been applied - all frequencies are now highly reduced.

This process may be represented symbolically as

$$\hat{u}^h \rightarrow \hat{u}^h + \underbrace{I_H^h (\mathcal{L}^H)^{-1} I_h^H (f^h - \mathcal{L}^h \hat{u}^h)}_{\tilde{e}^h}, \quad (2.3)$$

where I_h^H and I_H^h denote restriction and prolongation operators respectively between the fine grid (mesh spacing h) and the coarse grid (mesh spacing H), and \mathcal{L}^H denotes an appropriate representation of the operator \mathcal{L} on the coarse grid. The quantity h/H is called the *mesh ratio*. Unless specified otherwise, in this thesis it is assumed that $H = 2h$ - see Chapter 4 for a discussion of alternatives.

We cannot expect (2.3) alone to be effective, since it requires the error to be smooth. However, if (2.3) is combined with a procedure for efficiently *smoothing* the error, a highly effective iterative method is born: one smooths the error, applies the coarse grid correction, smooths the new error, applies the coarse grid correction again, and so on (an illustration is provided in Figure 2.3).

This idea is the basis of *multigrid* [11, 13, 16]. Credit for the inventing the algorithm is given to the late Soviet mathematician Rüdiger Petrovich Fedorenko

(Figure 2.2), who wrote the first paper on the subject, “A Relaxation Method for Solving Elliptic Difference Equations”, in 1962 [8]. Achi Brandt popularized the method through his work in the 1970s, most notably his seminal 1977 paper “Multi-Level Adaptive Solutions to Boundary Value Problems” [3]; see also [1], [13, p. 23] for a historical account.

We represent one iteration of this joint process symbolically as

$$\hat{u}^h \rightarrow S^{v_2} (S^{v_1} \hat{u}^h + I_H^h (\mathcal{L}^H)^{-1} I_h^H (f^h - \mathcal{L}^h S^{v_1} \hat{u}^h)), \quad (2.4)$$

where S is the smoothing operator of choice and v_1, v_2 are the number of smoothing steps before and after the coarse grid correction. Typical values are $v_1 = v_2 = 1$ or $v_1 = 1, v_2 = 0$. The smoother S is typically a simple relaxation scheme such as damped Jacobi or Gauss-Seidel, which, while not necessarily effective at reducing the magnitude of the error, has a strong smoothing effect on it.

Equation (2.4) represents what is called *two-level multigrid*, since it involves only two grids. It is not particularly efficient - discounting smoothing sweeps, two-level multigrid turns the solution of a problem with mesh spacing h into the solution of a small number of problems with mesh spacing H , leading to a speed up of at most a constant factor. Depending on what method is used to solve the coarse problem, there may be no benefit at all. For example, if the coarse grid problem is solved naively with Gaussian elimination, this is roughly 2^{3d} (d denotes the dimension of the problem) times cheaper on the coarse grid than it is on the fine grid. Clearly, convergence must be achieved in fewer than 2^{3d} iterations for there to be any gains in this case. Therefore, while two-level multigrid is of theoretical interest because of its relative simplicity, it is of little practical use.

Full multigrid improves upon its two-level counterpart by (1) acknowledging that it is only necessary to solve the coarse grid problem approximately and (2) doing so recursively with multigrid. The operator $(\mathcal{L}^H)^{-1}$ (whose role is to solve for the error on the coarse grid) is replaced with a small number γ (called the cycle index) of iterations of multigrid. Full multigrid thus involves a whole hierarchy of grids - the coarsest of which is sufficiently small that the error may be cheaply computed by direct methods. See Figure 2.4 for a sketch of the sequence of grids visited in a single iteration of full multigrid (with four grids) for $\gamma = 1$ (called a

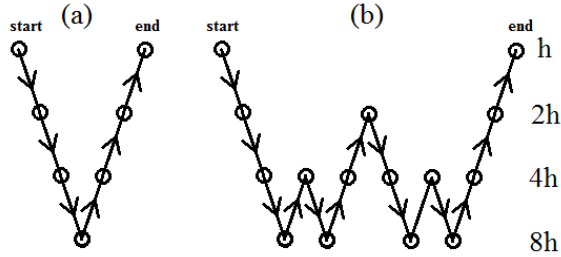


Figure 2.4: The sequence of grids visited by one iteration of four level multigrid, for two different choices of γ . (a) $\gamma = 1$: a V-cycle (b) $\gamma = 2$: a W-cycle.

V-cycle) and $\gamma = 2$ (called a W-cycle).

Unlike two-level multigrid, it can be shown that under certain conditions the work required for a single iteration of full multigrid (henceforth referred to as multigrid) is proportional to the number of grid points N . In particular, let us assume that the work required for one iteration of the smoother and intergrid operators is linear in N (as is the case for standard intergrid operators and pointwise smoothers) and that the cycle index obeys $\gamma < 2^d$, where d denotes the dimension of the problem. If we let L denote the number of levels and assume $H = 2h$, then we find that the total work W obeys

$$\begin{aligned}
 W &= CN + \gamma CN/2^d + \gamma^2 CN/2^{2d} + \dots + \gamma^{L-1} CN/2^{(L-1)d} \\
 &\leq CN + \gamma CN/2^d + \gamma^2 CN/2^{2d} + \dots \\
 &= \frac{CN}{1 - 2^{-d}\gamma}.
 \end{aligned}$$

The cost of a single iteration is thus linear in N , just like alternate iterative methods such as Gauss-Seidel, SOR, or the Conjugate Gradient method, provided the aforementioned conditions are met.

For most iterative methods, however, the *total work*, computed as the work per iteration times the number of iterations required for convergence, is *not* linear in N . This is because the number of iterations required for convergence is typically not a constant, but grows with N . Consider, for example, the three dimensional Poisson problem discretized on a uniform rectangular grid with mesh spacing h .

In this case N is proportional to h^{-3} , Gauss-Seidel and optimal SOR reduce the error by $1 - O(h^2)$ and $1 - O(h)$ per iteration [2, p. 181], and it follows, using the Taylor expansion $\log(1 + \varepsilon) = \varepsilon + O(\varepsilon^2)$, that the number of iterations required for convergence grows like $N^{2/3}$ and $N^{1/3}$ respectively. The total work is then proportional to $N^{5/3}$ and $N^{4/3}$ respectively. By contrast, if multigrid with lexicographic Gauss-Seidel (GS-LEX) as a smoother is used (on the aforementioned 3D Poisson problem), it can be shown that the number of iterations required for convergence does not depend on h [9], and therefore the total work is linear in N . Conversely, one can easily show that *no* algorithm can do better than $O(N)^1$, so multigrid is optimal in terms of its scalability.

However, optimal scaling is not enough to demonstrate the suitability of multigrid as a numerical method for real world problems. For example, a multigrid algorithm that solves the 3D Poisson problem in a constant $10^{10^{10}}$ iterations is optimal in terms of its scaling, but will be beaten by simpler methods such as Gauss-Seidel unless the size of the problem is astronomical.

With this in mind, we turn our attention to the quantitative convergence analysis of multigrid.

2.1 Quantitative Convergence Analysis

The error in a multigrid algorithm with $\ell + 1$ grids (labeled 0 through ℓ from coarsest to finest) obeys the recurrence relation $e^{(k)} = M_\ell e^{(k-1)}$, where $e^{(k)}$ denotes the error after k iterations and M_ℓ is the iteration matrix of ℓ -level multigrid, defined recursively by

$$M_\ell = S^{v_2}(I_\ell - I_{\ell-1}^\ell(I_{\ell-1} - (M_{\ell-1})^\gamma)(\mathcal{L}_{\ell-1})^{-1}I_\ell^{\ell-1}\mathcal{L}_\ell)S^{v_1} \equiv S^{v_2}K_\ell S^{v_1}, \quad (2.5)$$

where I_ℓ is the identity operator on the ℓ th grid, $I_{\ell-1}^\ell$ and $I_\ell^{\ell-1}$ are interpolation and projection operators respectively between grids ℓ and $\ell - 1$, \mathcal{L}_ℓ is the discrete representation of \mathcal{L} on grid ℓ , and $M_0 = 0$ [13, p. 48]. The operator K_ℓ is called the coarse grid operator.

To gain quantitative predictions of convergence rates we are interested in the

¹Any algorithm must specify the value of the solution at each of the N grid points. Even if this can be done in $O(1)$ operations per grid point, this is still $O(N)$ overall work.

spectral radius of the iteration matrix, defined by

$$\rho(M_\ell) = \max_i |\lambda_i|,$$

where $\{\lambda_i\}$ denotes the spectrum of M_ℓ . The spectral radius provides information on the asymptotic convergence rate, via the identity

$$\rho(M_\ell) = \lim_{k \rightarrow \infty} \frac{\|e^{(k)}\|}{\|e^{(k-1)}\|}.$$

Although we are interested in $\rho(M_\ell)$, it is worth noting the limitations on what it can tell us. The spectral radius only provides us with asymptotic information, relevant after a sufficiently large number of iterations have occurred. However, it can in general say nothing regarding what happens before the asymptotic behavior takes over, or how long this will take.

Let us digress briefly from multigrid in particular to iterative methods in general, and suppose

$$Ax = b$$

is solved iteratively by a process in which the error obeys

$$e^{(k)} = Me^{(k-1)},$$

for a suitable matrix M .

Even if $\rho(M) < 1$, the norm of the error may actually *increase* in the initial phases of the process (see Figure 2.5(b)). In extreme cases, overflow may occur before asymptotic behavior takes over.

A quantity which has more to say regarding early behavior (and is unfortunately often much harder to compute) is the *norm* of the iteration matrix, defined by

$$\|M\| \equiv \max_{x \neq 0} \frac{\|Mx\|}{\|x\|},$$

where $\|\cdot\|$ is a suitable vector norm. This quantity provides a bound on the norm

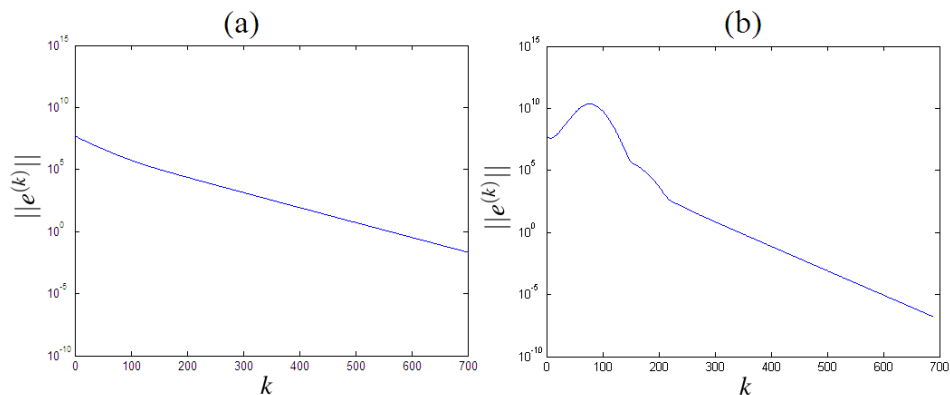


Figure 2.5: $\|e^{(k)}\|_2$ vs. k for two-level multigrid with lexicographic SOR smoothing applied to anisotropic diffusion problem $\frac{1}{100}u_{xx} + u_{yy} + u_{zz} = 0$, discretized using centered differences. (a) $\omega = 1$. (GS-LEX) (b) $\omega = 1.63$ (SOR-LEX). Note that for $\omega = 1$, the norm of the error goes down monotonically, whereas for $\omega = 1.63$ there is a initial increase before the asymptotic regime is entered. Convergence is faster overall for $\omega = 1.63$, despite the erratic behavior of the error initially.

of the error after k iterations via the identity:

$$\|e^{(k)}\| \leq \|M^k\| \|e^{(0)}\| \leq \|M\|^k \|e^{(0)}\|.$$

In particular, if $\|M\| < 1$, then monotonic reduction in the norm of the error is guaranteed. On the other hand, if $\|M\| > 1$, the quantity $\|M^k\|$ can sometimes provide useful information on the duration of the transient phase before asymptotic convergence takes over - see for example [7], where this is done in the context of Gauss-Seidel with various ordering strategies applied to the one-dimensional convection-diffusion equation.

It is worth noting the inequality $\rho(M) \leq \|M\|$, which becomes an equality when the matrix M is *normal*, that is $M^*M = MM^*$ (here M^* denotes the *conjugate transpose* of M). In this special case $\rho(M)$ gives a global upper bound on convergence rates. An example of this occurs when the matrix A is symmetric (for example, the discrete Poisson problem), and damped Jacobi is used as an iterative solver. In this case the iteration matrix is also symmetric, and therefore normal.

However, iteration matrices of GS-LEX are generally not normal, and neither are the iteration matrices of multigrid with GS-LEX smoothing. As a result, the convergence of multigrid is in general not guaranteed to be monotonic, and Figure 2.5 shows that it isn't in general.

Having seen that care must be taken interpreting the information provided by the spectral radius, we now move on to the method of its computation.

The definition of iteration matrix given in (2.5) is typically too complex to derive analytic convergence rates from. Bounds may be obtained (see for example [13, p. 77]), but they typically lead to estimates of the work which are several orders of magnitude too large [3].

A practical alternative is Local Fourier Analysis, first described by Brandt [3]. Although non-rigorous, this approach can yield estimates of $\rho(M_\ell)$ that are much closer to the actual convergence rate of multigrid than the bounds provided by the rigorous approach [3].

As stated in the Introduction, the main idea behind Local Fourier Analysis is to idealize the effect of the coarse grid correction. Representing the error as a sum of Fourier modes, one partitions the modes into 'low frequency's (also called 'smooth modes') and 'high frequency's ('rough modes'), and assumes the coarse grid correction leaves the high frequencies alone while wiping the low frequencies out completely. The smoothing factor μ is then defined in terms of this idealized coarse grid operator \hat{K} by the formula

$$\mu \equiv \sqrt[v]{\rho(S^{v_2} \hat{K}_h S^{v_1})},$$

where $v = v_1 + v_2$.

μ represents the *average asymptotic error reduction factor per smoothing sweep*. If bounds independent of h are desired, an alternative definition is

$$\mu \equiv \sup_h \sqrt[v]{\rho(S^{v_2} \hat{K}_h S^{v_1})}.$$

Smoothing factors are often analytically computable in situations where $\rho(M_\ell)$ is not. In particular, since \hat{K}_ℓ is by definition diagonal when written with respect to the Fourier basis, a highly tractable situation arises when the same is true of

the smoother S . In this case the smoothing factor reduces to the largest magnitude eigenvalue of the high frequency eigenfunctions of S .

It is intuitively clear that a Fourier mode whose wavelength is as long as the grid should be classified as low frequency and conversely, a mode with a wavelength on the order of the mesh spacing h should be called high frequency. It is also clear that the former can be well represented on the coarse grid while the latter cannot (see Figure 2.1) - therefore, it is reasonable to idealize the coarse grid correction as totally effective and totally ineffective respectively in these cases. What isn't so clear is where the line should be drawn between high and low frequencies, and whether it is realistic to draw such a line at all.

2.2 A Closer Look at the Interplay Between Errors on the Fine and Coarse Grids

I conclude this chapter with a theoretical justification for the partitioning of modes into high and low frequencies, and their subsequent treatment by the idealized coarse grid correction. First, an (essentially) standard argument is provided based on the observation that distinct Fourier modes *alias* when restricted to the coarse mesh. It is followed by a novel argument based on the Nyquist-Shannon Theorem from sampling theory.

For simplicity, let us assume our error lives on a one dimensional grid $G_h = \{h, 2h, \dots, 1 - h\} \subset (0, 1)$. On the continuous interval $(0, 1)$, any $f \in L^2((0, 1))$ may be represented as an infinite sum of Fourier modes belonging to the set

$$\Psi = \{\varphi_n \equiv \sin(n\pi x) : n \in \mathbb{N}\}.$$

Although distinct on $(0, 1)$, modes φ_n and φ_m look the same² when restricted to G_h provided $n \pm m$ is a integer multiple of $2h^{-1}$. As a result, one may show that the infinity of possible modes on $(0, 1)$ collapses down to just N on G_h ³, where $N \equiv |G_h|$ is assumed odd. These modes span the function space $\{f : G_h \rightarrow \mathbb{R}\}$, viewed as a vector space of dimension N .

²i.e differ at most in sign.

³Technically there are $N + 1$ possible modes, but one is the trivial mode which is zero everywhere, so I discount it.

We formalize this observation by introducing an equivalence relation \sim_h on Ψ , where

Definition 2.2.1 *We say two modes $\varphi_n, \varphi_m \in \Psi$ are h -equivalent if they cannot be distinguished on G_h . In this case we write*

$$\varphi_n \sim_h \varphi_m.$$

It is easy to verify that \sim_h is an equivalence relation, and that

$$\varphi_n \sim_h \varphi_m \Leftrightarrow \varphi_n|_{G_h} = \pm \varphi_m|_{G_h} \Leftrightarrow n \equiv \pm m \pmod{2h^{-1}}.$$

In light of definition 2.2.1 we may identify the Fourier modes on G_h with equivalence classes of modes on $(0, 1)$. That is, we identify the Fourier mode $\varphi_n|_{G_h}$ on G_h with the equivalence class $\{\varphi_n : \sim_h\}$ of modes on $(0, 1)$ modulo the equivalence relation \sim_h . The set of distinct Fourier modes on G_h , denoted by Ψ_h , may then be represented as

$$\Psi_h = \{\{\varphi_n : \sim_h\} : n = 1, \dots, N\}.$$

Let us now introduce a second grid $G_H = \{H, 2H, \dots, 1 - H\}$, where for simplicity $H = 2h$ (note that this ensures $G_H \subset G_h$). It is natural to ask how the equivalence classes Ψ_H on G_H are related to those on G_h - in particular, one might wonder if there are modes which can be distinguished on G_h but not on G_H . To answer this, one may check that $\varphi_n \sim_H \varphi_m$ and $\varphi_n \sim_h \varphi_m$ if and only if $n + m$ is a multiple of $h^{-1} = N + 1$, from which it follows that

$$\{\varphi_n : \sim_H\} = \{\varphi_n : \sim_h\} \cup \{\varphi_{N+1-n} : \sim_h\}.$$

In other words, when the mesh spacing is doubled the equivalence classes of Fourier modes merge pairwise; conversely, when the mesh spacing is halved each equivalence class splits in two.

Let us now imagine a restriction operator $I_h^H : \{f : G_h \rightarrow \mathbb{R}\} \rightarrow \{f : G_H \rightarrow \mathbb{R}\}$ defined by

$$I_h^H(f) = f|_{G_H}.$$

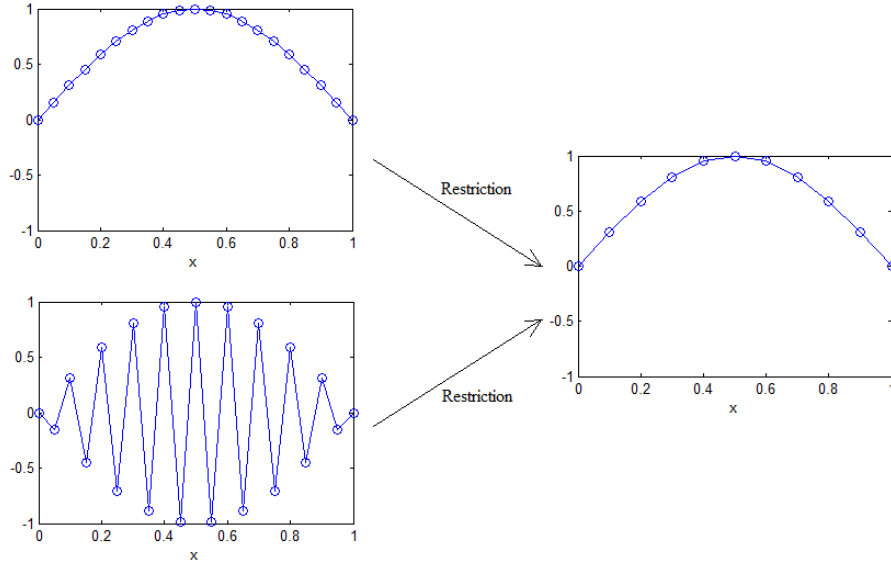


Figure 2.6: The distinct Fourier modes $\varphi_1(k) = \sin(k\pi/20)$ and $\varphi_{19}(k) = \sin(19k\pi/20)$ on the fine grid $G_h = \{k/20\}_{k=0}^{20}$ restrict to the same mode $\varphi_1(K) = \sin(K\pi/10)$ on the coarse grid $G_H = \{K/10\}_{K=0}^{10}$.

Then I_h^H is an operation that maps Fourier modes on G_h to Fourier modes on G_H in a 2 – 1 fashion since

$$I_h^H(\{\varphi_n : \sim_h\}) = I_h^H(\{\varphi_{N+1-n} : \sim_h\}) = \{\varphi_n : \sim_H\},$$

see Figure 2.6.

This fact creates a problem in the phase of the coarse grid correction in which the error on the coarse grid is sent to the fine grid. For if an error of $\{\varphi_n : \sim_H\}$ has just been solved for on G_H , we can only conclude that the error on G_h is some linear combination of $\{\varphi_n : \sim_h\}$ and $\{\varphi_{N+1-n} : \sim_H\}$ - in other words, we do not have enough information to reconstruct the error on G_h .

However, if we somehow know that the error on G_H contains no modes $\{\varphi_n : \sim_H\}$ with $n \geq N/2$ (for example if they had all been removed by a process of smoothing), we can say with confidence that an error of $\{\varphi_n : \sim_H\}$ on G_H implies an error of $\{\varphi_n : \sim_h\}$ on G_h . Under these conditions we can construct a prolongation

operator $I_H^h : \{f : G_H \rightarrow \mathbb{R}\} \rightarrow \{f : G_h \rightarrow \mathbb{R}\}$ by defining

$$I_H^h(\{\varphi_n : \sim_H\}) = \{\varphi_n : \sim_h\}, \quad (2.6)$$

and transfer errors between G_H and G_h without loss of information.

Let us define a Fourier mode to be ‘high frequency’ if $n \geq N/2$, and low frequency otherwise. It follows from (2.6) that the range of the prolongation operator I_H^h is contained in the span of the low frequency Fourier modes, and therefore all high frequency error components are unaffected by the coarse grid correction. On the other hand, restriction combines the low frequency mode $\{\varphi_n : \sim_h\}$ and its high frequency ‘partner’ $\{\varphi_{N+1-n} : \sim_h\}$ into a single mode on the coarse grid. The coarse grid correction will thus overcorrect $\{\varphi_n : \sim_h\}$ by an amount equal to the amplitude of $\{\varphi_{N+1-n} : \sim_h\}$. However, if we assume the amplitude of the high frequency mode $\{\varphi_{N+1-n} : \sim_h\}$ is sufficiently small, then this overcorrection is negligible and we may assume that the coarse grid correction simply eliminates all smooth modes.

For another way of thinking about things, we turn to sampling theory, where the coarse grid operator can be thought of as a process that attempts to construct the error on G_h from a sampling on the subgrid G_H . By the Nyquist-Shannon theorem, this is feasible if and only if the sampling frequency is at least twice the highest frequency present in the error (expanded in a Fourier series) [12].

Let us write the error on the grid G_h as a sum of complex Fourier modes of the form $e^{i\vec{\theta} \cdot \vec{x}/h}$,⁴ where $\vec{x} = \vec{I}h \in G_h$, and $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_d) \in [-\pi, \pi]^d$. Then the sampling frequency on G_H is $\frac{1}{H}$ in each direction, the frequency of mode $e^{i\theta x/h}$ in direction e_k is given by $f_k = \theta_k/2\pi h$, and we conclude that the coarse grid correction will be successful only if all modes in the error obey

$$\theta_k \leq \frac{h}{H}\pi \quad \text{for all } k = 1, \dots, n. \quad (2.7)$$

Let us assume that the components of the error violating (2.7) (called ‘rough’ modes) have been dampened in the smoothing stage to the point that they are small in comparison with the components that obey (2.7) (called ‘smooth’ modes). Then we idealize the coarse grid operator by assuming it succeeds at reconstructing the

⁴This is more convenient than the real Fourier modes used previously, and is preferred in LFA.

error in the fullest sense possible given (2.7); the ‘smooth’ components of the error are reconstructed perfectly while the ‘rough’ components are ignored. As a result, after the reconstructed error \tilde{e}^h is added to the current guess, the ‘smooth’ part of the error is gone (because it was perfectly constructed) and the ‘rough’ part is as it was at the end of the smoothing step.

Chapter 3

Smoothing Analysis

Let \mathcal{L}^h denote the elliptic difference operator defined in (1.2) and illustrated in Figure 1.1. Recall that \mathcal{L}^h is a d -dimensional operator with a stencil that forms a $2d + 1$ -point star. The central element of \mathcal{L}^h is denoted by a , while the element one unit in the $\pm e_k$ direction from a is denoted by $-b_k^\pm$.

I will consider an ordering in which grid points are ordered according to the following rule: along dimension k , unknowns are ordered from $-e_k$ to $+e_k$ if $b_k^+ \leq b_k^-$, and from $+e_k$ to $-e_k$ otherwise. For convenience, let us define:

$$\begin{aligned} c_k &\equiv \max(b_k^+, b_k^-)/a \\ d_k &\equiv \min(b_k^+, b_k^-)/a. \end{aligned}$$

It is straightforward to show (following the technique of [13, p. 103] for example) that the smoothing factor of pointwise lexicographic Gauss-Seidel applied to \mathcal{L}^h is given by

$$\mu^{pt} = \max_{\vec{\theta} \in \Theta^d} \left| \frac{\sum_{k=1}^d d_k e^{i\theta_k}}{1 - \sum_{k=1}^d c_k e^{-i\theta_k}} \right|, \quad (3.1)$$

where $\Theta^d = [-\pi, \pi]^d \setminus (-\pi/2, \pi/2)^d$ is the set of rough modes in d dimensions and $\iota \equiv \sqrt{-1}$. The argument of the max in (3.1) represents the eigenvalue of the point GS-LEX iteration matrix corresponding to the eigenfunction $\varphi_{\vec{\theta}} : G_h \rightarrow \mathbb{C}$ defined by $\varphi_{\vec{\theta}}(\vec{I}h) = e^{i\vec{\theta} \cdot \vec{I}h}$. Although not an eigenfunction of GS-LEX on any finite grid, $\varphi_{\vec{\theta}}$ becomes an eigenfunction when the grid G_h is assumed infinite, as in LFA.

We restrict our analysis to the case

$$d_k = \alpha c_k, \quad (3.2)$$

where $\alpha \in (0, 1]$ is a constant independent of k . This restriction includes all cases where $[\mathcal{L}_{\vec{I}, \vec{J}}^h]$ is symmetric, and other cases, for example the case of the convection-diffusion equation with all mesh Reynolds numbers equal.

It is well-known that pointwise smoothing is ineffective for highly anisotropic problems [13, p. 131]; one way to resolve this is the use of block smoothers [13, p. 134]. Such smoothers are defined by a partitioning of $\{1, \dots, d\}$ into disjoint sets

$$\mathcal{S}_b \subset \{1, \dots, d\} \quad \text{and} \quad \mathcal{S}_p = \{1, \dots, d\} \setminus \mathcal{S}_b,$$

where coordinates belonging to \mathcal{S}_b are relaxed simultaneously. To avoid degenerate cases, we assume both \mathcal{S}_p and \mathcal{S}_b are non-empty. In this case the smoothing factor is given by

$$\mu^{block} = \max_{\vec{\theta} \in \Theta^d} \left| \frac{\sum_{k \in \mathcal{S}_p} d_k e^{i\theta_k}}{1 - \sum_{k \in \mathcal{S}_p} c_k e^{-i\theta_k} - \sum_{k \in \mathcal{S}_b} (c_k e^{-i\theta_k} + d_k e^{i\theta_k})} \right|. \quad (3.3)$$

It is worth noting that the computational cost per iteration of a block smoother is higher than that of a point smoother. However, for highly anisotropic cases this overhead is typically small compared to the gains in convergence rates.

Our analysis shows that μ^{pt} depends only on α and any two of the quantities

$$c \equiv c_1 + c_2 + \dots + c_d \quad c_m \equiv \min(c_1, \dots, c_d) \quad c_r \equiv c - c_m, \quad (3.4)$$

regardless of the number of dimensions, d .

Similarly, in the case that \mathcal{L}^h is symmetric, our analysis shows that the block smoothing factor μ^{block} depends on any two of the quantities

$$C^p = \sum_{k \in \mathcal{S}_p} c_k, \quad C_m^p = \min_{k \in \mathcal{S}_p} c_k, \quad C_r^p = C^p - C_m^p \quad (3.5)$$

as well as any two of C^b , C_m^b and C_r^b (which are defined analogously).

Expressions for μ^{pt} and μ^{block} are given in Theorems 3.1.5 and 3.3.1 respectively.

3.1 Pointwise Relaxation

Starting from (3.1), we complex-conjugate the denominator and apply (3.2) to obtain

$$\mu^{pt} = \max_{\vec{\theta} \in \Theta^d} \left| \frac{\sum_{k=1}^d d_k e^{i\theta_k}}{1 - \sum_{k=1}^d c_k e^{-i\theta_k}} \right| = \max_{\vec{\theta} \in \Theta^d} \left| \frac{\alpha \sum_{k=1}^d c_k e^{i\theta_k}}{1 - \sum_{k=1}^d c_k e^{i\theta_k}} \right| \equiv \max_{\vec{\theta} \in \Theta^d} \left| \frac{\alpha g(\vec{\theta})}{1 - g(\vec{\theta})} \right|,$$

where $g(\vec{\theta}) \equiv \sum_{k=1}^d c_k e^{i\theta_k}$.

The key to our analysis lies in the observation that this may be rewritten as

$$\mu^{pt} = \max_{z \in g(\Theta^d)} |f(z)|,$$

where

$$f(z) \equiv \frac{\alpha z}{1 - z}$$

is analytic in the punctured plane $\mathbb{C} \setminus \{1\}$.

If the set $g(\Theta^d)$ is sufficiently well-behaved (a connected open set or the closure of one) and does not contain the point $z = 1$, the maximum principle [10, p. 88] applies and we can write

$$\mu^{pt} = \max_{z \in \partial g(\Theta^d)} |f(z)|,$$

where $\partial g(\Theta^d)$ denotes the boundary of $g(\Theta^d)$.

However, computing $g(\Theta^d)$ explicitly is challenging because for certain values of $\{c_k\}$ it contains a hole in the vicinity of the origin - see Example 3.1.3. Keeping track of this hole is difficult - we avoid the issue by proving we can replace $g(\Theta^d)$ with a simply-connected set D equivalent to it in the following sense:

Definition 3.1.1 *Let $A, B \subset \mathbb{C}$. We say $A \sim B$ if*

$$\sup_A |f| = \sup_B |f|$$

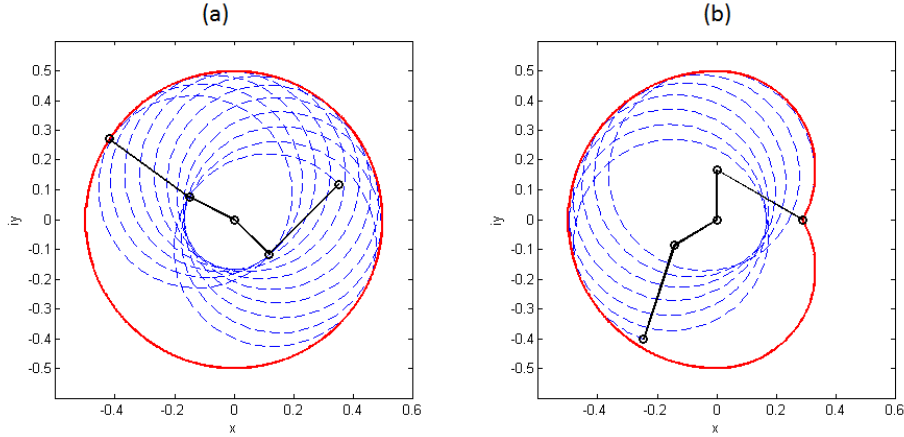


Figure 3.1: An illustration of Example 3.1.3: (a) If a double pendulum with arms of length $\frac{1}{6}$ and $\frac{1}{3}$ is allowed to swing freely, it remains inside a disk of radius 0.5. (b) If one arm is constrained to lie in the left half plane, the pendulum swings in the heart-shaped region shown. In both cases the pendulum is unable to reach a set of points surrounding the origin.

for all functions f complex differentiable on $A \cup B$.

The following observation establishes an important sufficient condition for two sets to be equivalent.

Observation 3.1.2 Let $A, B \subset \mathbb{C}$, and suppose B is a bounded, connected open set (domain). If $\partial B \subseteq A \subseteq \bar{B}$, then $A \sim B \sim \bar{B}$.

To see this, let f be any function complex differentiable on $A \cup B = \bar{B}$ – since B is a domain, f is *analytic* on B . We clearly have $\sup_{\partial B} |f| \leq \sup_A |f| \leq \sup_B |f|$, but the maximum modulus principle implies the equality of the leftmost and rightmost terms.

In light of Observation 3.1.2, our goal is to find a simply-connected domain D such that (i) $D = \bar{E}$, where E is a domain (ii) $\partial D \subseteq g(\Theta^d) \subseteq D$. This is done in Lemma 3.1.4, but first we build some intuition with a simple example.

Example 3.1.3 Suppose $d = 2$ and \mathcal{L}^h arises from a centered differences discretization of the Laplacian-like operator $\mathcal{L} = 0.5\partial_{xx} + \partial_{yy}$, giving $a = 3$, $b_1^+ =$

$b_1^- = 0.5$ and $b_2^+ = b_2^- = 1$. Then $c_1 = 1/6$, $c_2 = 1/3$, and

$$g(\theta_1, \theta_2) = \frac{1}{6}e^{i\theta_1} + \frac{1}{3}e^{i\theta_2}.$$

Geometrically, $g(\theta_1, \theta_2)$ may be viewed as a double pendulum with arms of length $1/6$ and $1/3$, making angles θ_k ($k=1,2$) with the x -axis. As (θ_1, θ_2) varies over $[-\pi, \pi]^2$, this pendulum (and therefore the range of g) remains confined to the disk of radius 0.5 shown in Figure 3.1(a). Furthermore, the boundary of the disk is swept out as the double pendulum completes a full revolution with a constant angle of 180° between the arms. By Observation 3.1.2, $g([-\pi, \pi]^2)$ is equivalent to this disk in the sense of definition 3.1.1.

However, if (θ_1, θ_2) is constrained to lie in $\Theta^2 \equiv [-\pi, \pi]^2 \setminus [-\frac{\pi}{2}, \frac{\pi}{2}]^2$, then $|\theta_k| \geq \pi/2$ for at least one $k \in \{1, 2\}$. In other words, at least one arm is constrained to lie in the left half plane. With this constraint, $g(\theta_1, \theta_2)$ now lives in the set shown in Figure 3.1(b) - the union of two disks and a half disk, as demonstrated in Figure 3.2. It is simple to show the boundary of this set is once again in the range of g , so this heart-shaped set provides us with the D we are looking for.

Note that while $g(\Theta^2) \sim D$ it is not true that $g(\Theta^2) = D$, since $0 \in D$ but $|g| \geq \frac{1}{3} - \frac{1}{6} > 0$.

Lemma 3.1.4 Let $B(r, z)$ denote the closed ball of radius $r \geq 0$ centered at $z \in \mathbb{C}$. Let \mathbb{C}^- denote the set of all complex numbers with negative real part, and let \sim be the relation in definition 3.1.1. Then

$$\begin{aligned} g(\Theta^d) &\sim [B(c, 0) \cap \mathbb{C}^-] \cup B(c_r, ic_m) \cup B(c_r, -ic_m) \equiv D, \\ g([-\pi, \pi]^d) &\sim B(c, 0), \end{aligned}$$

where c , c_m , and c_r are defined in (3.4).

Proof. By Observation 3.1.2 we clearly have $g([-\pi, \pi]^d) \sim B(c, 0)$, since $|g(\theta_1, \dots, \theta_d)| \leq c$ and $g(\theta, \dots, \theta) = ce^{i\theta}$. On the other hand, if $\vec{\theta} \in \Theta^d$ then there is a j such that $|\theta_j| \geq \pi/2$. Assuming we have made some fixed choice of j and set θ_j to a fixed value α , the remaining variables $\{\theta_k\}_{k \neq j}$ range over $[-\pi, \pi]^{d-1}$. It

follows that g is confined to $B(c - c_j, c_j e^{i\alpha})$, and therefore

$$g(\Theta^d) \subseteq \bigcup_{|\alpha| \geq \frac{\pi}{2}} \bigcup_{j=1}^d B(c - c_j, c_j e^{i\alpha}).$$

But $B(c - c_j, c_j e^{i\alpha}) \subseteq B(c_r, c_m e^{i\alpha})$ for all j since $|z - c_j e^{i\alpha}| \leq c - c_j$ implies

$$|z - c_m e^{i\alpha}| \leq |z - c_j e^{i\alpha}| + |c_j e^{i\alpha} - c_m e^{i\alpha}| \leq (c - c_j) + (c_j - c_m) = c_r.$$

Therefore

$$g(\Theta^d) \subseteq \bigcup_{|\alpha| \geq \frac{\pi}{2}} B(c_r, c_m e^{i\alpha}).$$

It can be shown that any $z \in B(c_r, c_m e^{i\alpha})$ obeying $\operatorname{Re}(z) \geq 0$ lies in $B(c_r, ic_m)$ if $\operatorname{Im}(z) \geq 0$ and $B(c_r, -ic_m)$ if $\operatorname{Im}(z) \leq 0$. Similarly, since $|z| \leq c$ we have $z \in B(c, 0) \cap \mathbb{C}^-$ if $\operatorname{Re}(z) < 0$. It follows that

$$g(\Theta^d) \subseteq D.$$

It is trivial to show $\partial D \subseteq g(\Theta^d)$, which completes the proof. \square

Lemma 3.1.4 and Definition 3.1.1 allow us to conclude

$$\mu^{pt} = \max_{\partial D} |f| \quad (3.6)$$

provided f is complex differentiable in D , that is, provided we can show $1 \notin D$. To that end, we note that the diagonal dominance of \mathcal{L}^h (1.3) implies

$$\begin{aligned} c &= \frac{\sum_{k=1}^d \max(b_k^+, b_k^-)}{a} \\ &\leq \frac{\sum_{k=1}^d \max(b_k^+, b_k^-)}{\sum_{k=1}^d (b_k^+ + b_k^-)} < 1. \end{aligned}$$

The desired result follows from the obser-

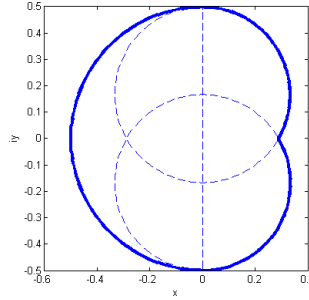


Figure 3.2: The heart-shaped set D is the union of two disks and a half disk.

vation

$$D \cap \mathbb{R} = [-c, \sqrt{c_r^2 - c_m^2}] \subseteq [-c, c].$$

Theorem 3.1.5 *The smoothing factor of pointwise GS-LEX applied to discrete operators of the form (1.2) obeying the constraints (1.3) and (3.2) is given by*

$$\mu^{pt}(c_m, c_r, \alpha) = \frac{\alpha c_r + \alpha \sqrt{c_m^2 + (c_m^2 - c_r^2)^2}}{1 + c_m^2 - c_r^2}, \quad (3.7)$$

where c_m and c_r are defined in (3.4) and α is defined in (3.2).

Proof. The set ∂D is the union of the three semicircular arcs

$$\begin{aligned} S_1 &= \left\{ ce^{i\theta} : \theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2} \right] \right\}, & S_2 &= \left\{ ic_m + c_r e^{i\theta} : \theta \in \left[-\sin^{-1} \left(\frac{c_m}{c_r} \right), \frac{\pi}{2} \right] \right\}, \\ S_3 &= \left\{ z \in \mathbb{C} : \bar{z} \in S_2 \right\}. \end{aligned}$$

By (3.6), it suffices to compute the maximum of the maxima of $|f|$ over S_1 , S_2 , and S_3 . However, since $|f(\bar{z})| = |f(z)|$, the maxima over S_2 and S_3 are the same and hence we omit S_3 from consideration. Furthermore, it is trivial to show that the maximum of $|f|$ over S_1 is achieved at ic , the point of intersection of S_1 and S_2 . Therefore, the maximum of $|f|$ over S_2 is at least as large as the maximum over S_1 - hence we may omit S_1 as well.

It follows that the maximum of $|f|$ over D is attained on the semicircle S_2 . However, it is also true that the maximum is attained on the full circle $|z - ic_m| = c_r$, as this circle contains S_2 and is contained in D . It is more convenient to work with the full circle, so we conclude

$$\mu^{pt} = \max_{|z - ic_m| = c_r} |f(z)|. \quad (3.8)$$

Now, $f(z) = \frac{\alpha z}{1-z}$ is a *Möbius transform*, and hence the image of a circle is either a circle or a line [10, p. 65]. Since the circle $|z - ic_m| = c_r$ does not contain any poles of f , it follows that $f(\{|z - ic_m| = c_r\})$ is a circle.

If we let $z_c \in \mathbb{C}$ and $r \geq 0$ denote the center and radius respectively of this

circle, then from (3.8) we have

$$\mu^{pt} = \max_{z \in f(\{|z - ic_m| = c_r\})} |z| = |z_c| + r. \quad (3.9)$$

To find the parameters z_c and r , we first write f as a sequence of elementary Möbius transformations: $f = f_4 \circ f_3 \circ f_2 \circ f_1$, where $f_1(z) = z - 1$, $f_2(z) = z^{-1}$, $f_3(z) = -\alpha z$, and $f_4(z) = z - \alpha$. Then, starting with the circle $|z - ic_m| = c_r$, we track the changes in its center and radius as we compute its image under f_1 , then compute the image of the result under f_2 , and so on.

Since f_1 , f_3 , and f_4 are all either translations or dilations, the steps involving them are straightforward. For f_2 , we make the observation that the image of a circle with radius R and center z_0 under inversion is the circle with radius R' and center z'_0 given by

$$z'_0 = \frac{\bar{z}_0}{|z_0|^2 - R^2} \text{ and } R' = \frac{R}{\left| |z_0|^2 - R^2 \right|};$$

this fact can easily be derived from the argument found in [10, p. 66].

In the end we find

$$r = \frac{\alpha c_r}{1 + c_m^2 - c_r^2}$$

and

$$z_c = \alpha \frac{c_r^2 - c_m^2}{1 + c_m^2 - c_r^2} + i\alpha \frac{c_m}{1 + c_m^2 - c_r^2},$$

which completes the proof. \square

Notice that the smoothing factor μ^{pt} is proportional to α . Since the latter is small when the operator \mathcal{L}^h is strongly asymmetric, this suggests that GS-LEX may be particularly effective in this regime. In particular, as $\alpha \rightarrow 0$ the part of \mathcal{L}^h above its diagonal is converging to the zero matrix, and for $\alpha = 0$ we have that \mathcal{L}^h is lower triangular and GS-LEX becomes a (direct) solver. See also Chapter 5 for further illustrations of this behavior.

3.2 SOR Relaxation

It is well-known that the smoothing factors of GS-RB, Jacobi-RB¹, and Jacobi can be greatly improved by incorporating a relaxation parameter [13, 18, 20]. For example, for the standard seven-point centered difference discretization of the three-dimensional Poisson problem, overrelaxation with $\omega = 1.15$ improves the smoothing factor of GS-RB (equivalent to Jacobi-RB in this specific case) from $\mu \approx 0.44$ to $\mu \approx 0.23$ [18]. For the same problem, the smoothing factor of Jacobi is improved from $\mu = 1$ (no convergence) to $\mu = 5/7$ by underrelaxation with $\omega = 6/7$ [13, p. 73].

We can analyze the effect of a relaxation parameter ω in GS-LEX smoothing by repeating the steps of Theorem 3.1.5 with the Möbius transformation

$$f_\omega(z) = \frac{\alpha z + \frac{1}{\omega} - 1}{\frac{1}{\omega} - z}$$

used in place of f . Note that we must assume $\omega < \frac{1}{\sqrt{c_r^2 - c_m^2}}$ in order to ensure that f_ω has no poles in D .

Making use of the factorization $f_\omega = f_4 \circ f_3 \circ f_2 \circ f_1$, where $f_1(z) = z - \frac{1}{\omega}$, $f_2(z) = z^{-1}$, $f_3(z) = (1 - \frac{1+\alpha}{\omega})z$, $f_4(z) = z - \alpha$, we obtain

$$\mu_{SOR}^{pt}(\omega) = \frac{c_r \left| 1 - \frac{1+\alpha}{\omega} \right| + \sqrt{c_m^2 \left(1 - \frac{1+\alpha}{\omega} \right)^2 + \left(\frac{1}{\omega} \left(1 - \frac{1}{\omega} \right) + \alpha(c_m^2 - c_r^2) \right)^2}}{\frac{1}{\omega^2} + c_m^2 - c_r^2}. \quad (3.10)$$

Figure 3.3(a) shows a plot of μ_{SOR}^{pt} as a function of $\omega \in [0, 2]$ for the 3D Poisson problem mentioned above ($c_m = 1/6$, $c_r = 1/3$, $\alpha = 1$). The optimal smoothing factor $\mu \approx 0.551$ is attained at $\omega \approx 1.1$, a modest improvement over the smoothing factor $\mu \approx 0.567$ obtained when $\omega = 1$. This finding supports the conclusion [13, p. 105] that for GS-LEX, the inclusion of a relaxation parameter is not necessarily worth the extra work per iteration (two operations per point per relaxation sweep).

In Figure 3.3(b), the anisotropic problem $-\varepsilon u_{xx} - u_{yy} - u_{zz} = f$ is discretized using standard seven-point centered differences, and the theoretical smoothing fac-

¹Jacobi-RB consists of a Jacobi sweep over the red points, followed by a Jacobi sweep over the black points using the updated values at red points; see for example [13, p 173].

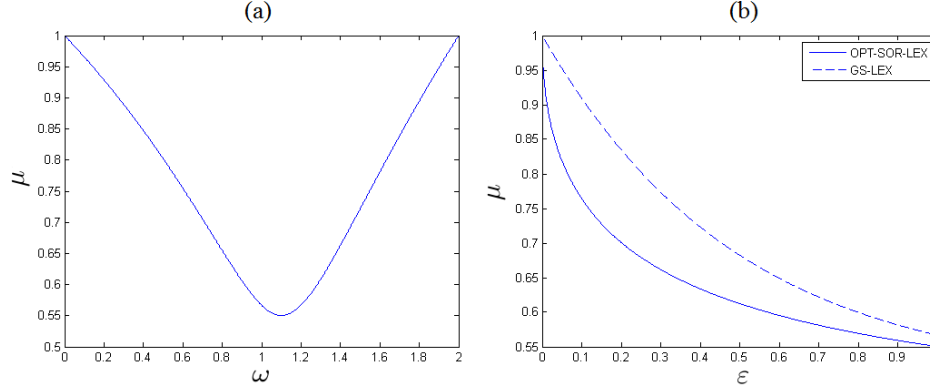


Figure 3.3: Left: Smoothing factor of SOR-LEX with relaxation parameter ω , as a function of ω , for the standard seven-point centered differences discretization of the 3D Poisson problem. Right: Smoothing factors of GS-LEX and optimal SOR-LEX applied to the anisotropic diffusion problem $-\varepsilon u_{xx} - u_{yy} - u_{zz} = f$, as a function of ε , assuming seven-point centered differences discretization.

tors of both GS-LEX and optimal SOR-LEX (obtained by numerical minimization of (3.10)) are plotted against ε .

3.3 Block Relaxation

In this section we analyze the smoothing properties of block GS-LEX relaxation. To keep the algebra manageable, we restrict the scope of our analysis to the case that $[\mathcal{L}_{I,J}^h]$ is symmetric ($\alpha = 1$). Our main result is Theorem 3.3.1, which establishes a connection between block and pointwise smoothing factors.

Since $\alpha = 1$, we have $c_k = d_k$ for all k - substituting this into equation (3.3), complex-conjugating the denominator and then simplifying, we obtain

$$\begin{aligned} \mu^{block} &= \max_{\vec{\theta} \in \Theta^d} \left| \frac{\sum_{k \in \mathcal{J}_p} c_k e^{i\theta_k}}{1 - 2 \sum_{k \in \mathcal{J}_b} c_k \cos \theta_k - \sum_{k \in \mathcal{J}_p} c_k e^{i\theta_k}} \right| \\ &= \max_{(z,x) \in G(\Theta^d)} \left| \frac{z}{1 - x - z} \right|, \end{aligned}$$

where $G : \Theta^d \rightarrow \mathbb{C} \times \mathbb{R}$ is given by $G(\vec{\theta}) = (G_1(\vec{\theta}), G_2(\vec{\theta}))$ with

$$G_1(\vec{\theta}) = \sum_{k \in \mathcal{J}_p} c_k e^{i\theta_k} \text{ and } G_2(\vec{\theta}) = 2 \sum_{k \in \mathcal{J}_b} c_k \cos \theta_k.$$

Defining $z' = z/(1-x)$, this becomes

$$\mu^{block} = \max_{z' \in \Lambda} \left| \frac{z'}{1-z'} \right| = \max_{z' \in \Lambda} |f(z')|, \quad (3.11)$$

where f is defined in section 3.1 and

$$\Lambda \equiv \{z/(1-x) : (z, x) \in G(\Theta^d)\}.$$

We will show that Λ is \sim (in the sense of Definition 3.1.1) to the union of a scaled copy of the set D from Lemma 3.1.4 and a particular ball centered at the origin.

This fact will allow us to express μ^{block} in terms of μ^{pt} .

If $\vec{\theta} \in \Theta^d$, then there is a j such that $|\theta_j| \geq \pi/2$. If $j \in \mathcal{J}_p$, it follows that

$$\{\theta_k\}_{k \in \mathcal{J}_p} \in \Theta^{|\mathcal{J}_p|} \text{ and } \{\theta_k\}_{k \in \mathcal{J}_b} \in [-\pi, \pi]^{|\mathcal{J}_b|},$$

while if $j \in \mathcal{J}_b$, then

$$\{\theta_k\}_{k \in \mathcal{J}_p} \in [-\pi, \pi]^{|\mathcal{J}_p|} \text{ and } \{\theta_k\}_{k \in \mathcal{J}_b} \in \Theta^{|\mathcal{J}_b|}.$$

The function G_1 has the same form as g of Section 3.1 with the variables $\{\theta_k\}_{k=1}^d$ replaced by $\{\theta_k\}_{k \in \mathcal{J}_p}$. Thus, Lemma 3.1.4 applies with C^p , C_m^p and C_r^p playing the roles of c , c_m , c_r :

$$G_1(\Theta^d) \sim \begin{cases} D(C_m^p, C_r^p) & \text{if } k \in \mathcal{J}_p, \\ B(C^p, 0) & \text{otherwise.} \end{cases}$$

At the same time we have

$$G_2(\Theta^d) = \begin{cases} [-2C^b, 2C^b] & \text{if } k \in \mathcal{J}_p, \\ [-2C^b, 2C_r^b] & \text{otherwise.} \end{cases}$$

Assuming $j \in \mathcal{J}_p$, since G_1 and G_2 depend on disjoint variables we have

$$\Lambda \Big|_{j \in \mathcal{J}_p} = \bigcup_{x \in G_2(\Theta^d) \Big|_{j \in \mathcal{J}_p}} \frac{1}{1-x} G_1(\Theta^d) \Big|_{j \in \mathcal{J}_p} \sim \bigcup_{x \in [-2C^b, 2C^b]} \frac{1}{1-x} D(C_m^p, C_r^p).$$

Note that if $2C^b < 1$, then for all $x \in [-2C^b, 2C^b]$ we have

$$\frac{1}{1-x} D(C_m^p, C_r^p) \subseteq \frac{1}{1-2C^b} D(C_m^p, C_r^p).$$

It follows from the diagonal dominance of \mathcal{L}^h and the condition $\alpha = 1$ that $c \leq 0.5$. Since $C^b < c$, we have $2C^b < 1$ as desired, and therefore

$$\Lambda \Big|_{j \in \mathcal{J}_p} \sim \frac{1}{1-2C^b} D(C_m^p, C_r^p) = D\left(\frac{C_m^p}{1-2C^b}, \frac{C_r^p}{1-2C^b}\right). \quad (3.12)$$

A similar analysis with $j \in \mathcal{J}_b$ shows

$$\Lambda \Big|_{j \in \mathcal{J}_b} \sim B\left(\frac{C^p}{1-2C_r^b}, 0\right).$$

Theorem 3.3.1 *The smoothing factor of block GS-LEX applied to symmetric discrete operators of the form (1.2) obeying the constraints (1.3) and (3.2) is given by*

$$\mu^{block} = \max\left(\mu^{pt}(\tilde{c}_m, \tilde{c}_r, \alpha = 1), \frac{C^p}{1-C^p-2C_r^b}\right), \quad (3.13)$$

where

$$\tilde{c}_m = \frac{C_m^p}{1-2C^b}, \tilde{c}_r = \frac{C_r^p}{1-2C^b},$$

and $C^p, C_m^p, C_r^p, C^b, C_r^b$ are defined in (3.5).

Proof. Clearly the maximum of $|f|$ over Λ is the maximum of the separate maxima over $\Lambda \Big|_{j \in \mathcal{J}_p}$ and $\Lambda \Big|_{j \in \mathcal{J}_b}$. By (3.12) and Theorem 3.1.5 we have

$$\max_{z \in \Lambda \Big|_{j \in \mathcal{J}_p}} |f(z)| = \mu^{pt}(\tilde{c}_m, \tilde{c}_r, 1),$$

provided f is analytic in $D(\tilde{c}_m, \tilde{c}_r)$. From the inequalities $C^p, C^b < c \leq 1/2$ it is

easy to show $\tilde{c} \equiv \tilde{c}_m + \tilde{c}_r \leq 1$, and the desired result follows in the same way as in Section 3.1.

Similarly, since $C_r^b, C^p < 1/2$ we have

$$1 \notin B\left(\frac{C^p}{1-2C_r^b}, 0\right) \sim \Lambda\Big|_{j \in \mathcal{J}_b}.$$

It follows from the maximum modulus principle that $f(z)$ attains its maximum modulus over $\Lambda\Big|_{j \in \mathcal{J}_b}$ at $z = C^p/(1-2C_r^b)$, and therefore

$$\max_{z \in \Lambda\Big|_{j \in \mathcal{J}_b}} |f(z)| = \frac{C^p}{1-C^p-2C_r^b},$$

which completes the proof. □

Theorem 3.3.1 shows that applying a block smoother can be equivalent to applying a point-smoother to a related problem.

Chapter 4

Alternative Mesh-Ratios

In his comprehensive 1977 paper [3], Brandt had the following to say regarding the choice of mesh ratio h/H .

“...it is evident that the mesh-size ratio ($H = 2h$) is close to optimal...this ratio is more convenient and more economical in the interpolation process than any other efficient ratio. In practice, therefore, the ratio ($H = 2h$) should always be used, giving also a very desirable standardization.”

Indeed, the choice $H = 2h$ (called double coarsening) is very common amongst multigrid practitioners. However, under certain conditions alternatives may be more appropriate. For example, it has been argued that for cell-based discretization schemes *triple coarsening* ($H = 3h$) is a better choice because it better preserves cell structure [6]. See also [20] for an example involving the choice $H = 4h$.

Therefore, to keep my results as general as possible, this chapter considers general mesh ratios $h/H = \beta \in (0, 1)$. The smoothing factor of point GS-LEX is derived as a function of β , along with an estimate of the total work of multigrid. One finds that the standard choice $\beta = 0.5$ is nearly optimal when the dimension is small. In Theorem 4.0.2 I will show that a judicious choice of β can reduce the overall work by up to 50% in high dimensions, relative to the standard choice $H = 2h$; I call this *optimal coarsening*.

High dimensional Poisson problems are of interest in applications such as mathematical finance; their efficient numerical solution by multigrid has been studied in [20]. It is possible that optimization of the mesh ratio could lead to reductions

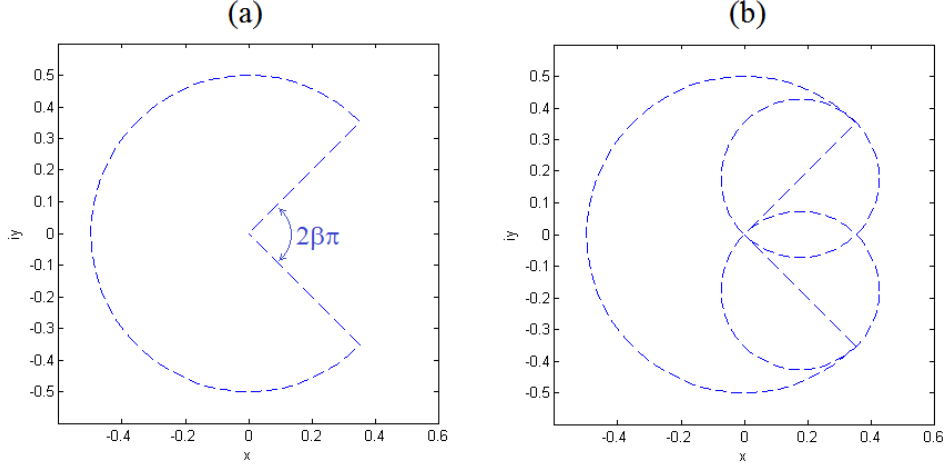


Figure 4.1: (a) An example pacman set with $c = 0.5$ and $\beta = 1/4$. (b) The corresponding set D_β when $c_m = c_r = 0.25$. This example corresponds the standard centered differences discretization of the two dimensional Poisson problem, with the coarsening strategy $H = 4h$.

in the overall work for problems of this type.

However, since the optimal β (denoted by β_{opt}) is typically such that $h < H_{opt} \equiv \frac{h}{\beta_{opt}} < 2h$, a potentially significant price is that the coarse grid nodes do not line up with those of the fine grid, leading to complications in the intergrid transfer operators (see Figure 4.3).

The modified smoothing factor μ^{pt} may be computed as a function of β from (3.1), replacing the rough modes Θ^d with their β -dependent generalization derived from (2.7), namely

$$\Theta_\beta^d = [-\pi, \pi]^d \setminus (-\beta\pi, \beta\pi)^d.$$

It is not hard to generalize the result of Theorem 3.1.5 to include $\beta \neq 1/2$. The set D is replaced with a set D_β composed of two disks, centered at $c_m e^{\pm i\beta\pi}$, and a ‘pacman’ shape with mouth subtending an angle of $2\beta\pi$ (see Figure 4.1)¹. In symbols, we have

$$D_\beta = P(c, \beta) \cup B(c_r, c_m e^{i\beta\pi}) \cup B(c_r, c_m e^{-i\beta\pi}),$$

¹To see this, simply follow the proof of Lemma 3.1.4, replacing occurrences of $\pi/2$ with $\beta\pi$.

where $P(c, \beta)$ is the pacman set

$$P(c, \beta) \equiv \{re^{i\theta} : r \leq c \text{ and } \theta \in [-\pi, \pi] \setminus [-\beta\pi, \beta\pi]\}.$$

Mimicking the approach taken in Chapter 3, the smoothing factor is found by taking the image of the circle $|z - c_m e^{i\beta\pi}| = c_r$ under the Möbius transform $f(z)$. In the end one finds

$$\mu^{pt}(\beta) = \frac{\alpha c_r + \alpha \sqrt{c_m^2 + 2c_m(c_r^2 - c_m^2)\cos(\beta\pi) + (c_r^2 - c_m^2)^2}}{1 + c_m^2 - c_r^2 - 2c_m \cos(\beta\pi)}. \quad (4.1)$$

The qualitative behavior of μ^{pt} as a function of β may not be obvious from (4.1). However, noting that $\beta' > \beta$ implies $\Theta_{\beta'}^d \subset \Theta_{\beta}^d$, it follows that μ^{pt} is a decreasing function of β . Therefore, the larger β is, the fewer iterations are required for convergence.

On the other hand, the larger β is, the more grids are required between the coarsest and finest. As an example, suppose we are in two dimensions and the finest grid has 100^2 points. With the mesh ratio $H = 2h$, five grids are required to obtain a coarse grid with fewer than 5^2 unknowns; for $H = 1.5h$, however, eight grids are required. Thus, the work per iteration increases with β .

A natural question now arises: which $\beta \in (0, 1)$ minimizes the total work, computed as the product of the work per iteration and the number of iterations required for convergence?

Let N denote the number of grid points on the finest grid. Let us assume for simplicity that V -cycles are employed, and consider the limit where the number of grids tends to infinity. Then the work per iteration is

$$w_i = C(N + \beta^d N + \beta^{2d} N + \dots) = \frac{CN}{1 - \beta^d},$$

where C is a constant of proportionality determined by the smoother and intergrid operators. At the same time, the number of iterations required to bring the error down to a predetermined tolerance is proportional to $\log(\mu^{pt})^{-1}$.

It is natural to express the total work as a fraction of the work for the standard

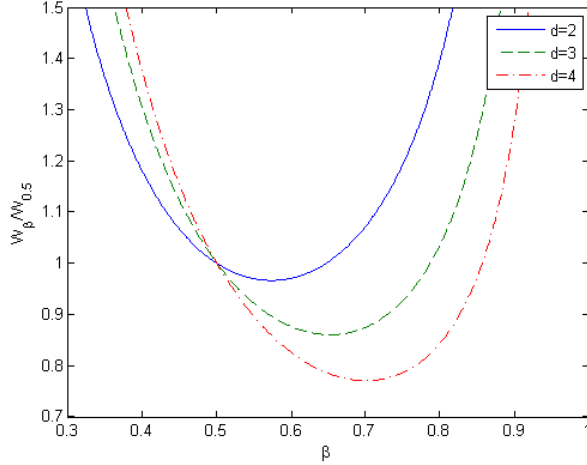


Figure 4.2: The relative work $W_\beta/W_{0.5}$ vs β for the d -dimensional Poisson problem with $d = 2, 3, 4$.

	$d = 2$	$d = 3$	$d = 4$	$d = 8$	$d = 16$	$d = 32$	$d = 64$
β_{opt}	0.57	0.65	0.7	0.76	0.82	0.88	0.92
$W_{opt}/W_{0.5}$	0.97	0.86	0.77	0.63	0.56	0.53	0.51

Table 4.1: Relative work of optimal coarsening and corresponding values of β_{opt} , for various d -dimensional Poisson problems.

choice $\beta = 0.5$. One obtains

$$\frac{W_\beta}{W_{0.5}} = \frac{(1 - 0.5^d) \log(\mu^{pt}(0.5))}{(1 - \beta^d) \log(\mu^{pt}(\beta))}. \quad (4.2)$$

Figure 4.2 shows a plot of $\frac{W_\beta}{W_{0.5}}$ for a few d -dimensional Poisson problems. Table 4.1 displays the optimal β and corresponding savings for various values of d .

From Figure 4.2 and Table 4.1, two things are apparent. Firstly, the choice $\beta < 0.5$ (that is $H > 2h$) seems to lead to an increase in the overall work. Thus, mesh ratios such as $H = 3h$ or $H = 4h$ will tend to increase the overall work (but may be desirable for other reasons). Second, while values of $\beta > 0.5$ ($h < H = \frac{h}{\beta} < 2h$) can lead to decreased work, the gains appear to be small in low dimensions - in particular, for the two-dimensional Poisson problem optimal coarsening yields

a theoretical reduction of only 3% in the overall work.

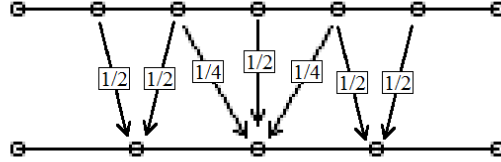


Figure 4.3: A fine and coarse grid with mesh ratio $h/H = 2/3$. The stencil of the projection operator varies from point to point, depending on whether the coarse grid lines up with fine grid.

For three-dimensional Poisson, however, the gain is almost 15%, and may be worthwhile. To give an idea of how this method might be implemented in practice, Figure 4.3 provides a sketch of a one-dimensional fine and coarse grid with $h/H = 2/3$ (sufficiently close to β_{opt} to give good results, but more convenient in terms of alignment of gridpoints), and gives a possible restriction operator.

It is clear from Table 4.1 that the advantages of optimal coarsening become more significant as the dimension grows. In fact, we have the following Theorem:

Theorem 4.0.2 Consider the standard centered differences discretization of the Poisson problem in d dimensions. As $d \rightarrow \infty$,

$$\beta_{opt} \rightarrow 1 \quad \text{and} \quad \frac{W_{opt}}{W_{0.5}} \rightarrow 0.5.$$

Proof. As $d \rightarrow \infty$, $1 - 0.5^d \rightarrow 1$, while $1 - \beta^d$ converges to the step function which is 1 for $\beta \in [0, 1)$ and 0 for $\beta = 1$. As a result, (4.2) becomes

$$\lim_{d \rightarrow \infty} \frac{W_\beta}{W_{0.5}} = \begin{cases} \frac{\log(\mu^{pt}(0.5))}{\log(\mu^{pt}(\beta))} & \text{if } \beta < 1 \\ \infty & \text{if } \beta = 1 \end{cases}$$

and one finds

$$\lim_{d \rightarrow \infty} \frac{W_{opt}}{W_{0.5}} = \frac{\log(\mu^{pt}(0.5))}{\log(\mu^{pt}(1))},$$

while β_{opt} tends to 1 (see Figure 4.4).

One may expand $\mu^{pt}(\beta)$ as a Taylor series in $1/d$, obtaining

$$\mu^{pt} = 1 - 2(1 - \cos(\beta\pi))\frac{1}{d} + O\left(\frac{1}{d^2}\right).$$

Then, using the identity $\log(1+x) = x + O(x^2)$ (valid for $x \ll 1$), one obtains

$$\lim_{d \rightarrow \infty} \frac{W_{opt}}{W_{0.5}} = 0.5$$

for the d -dimensional poisson problem. □

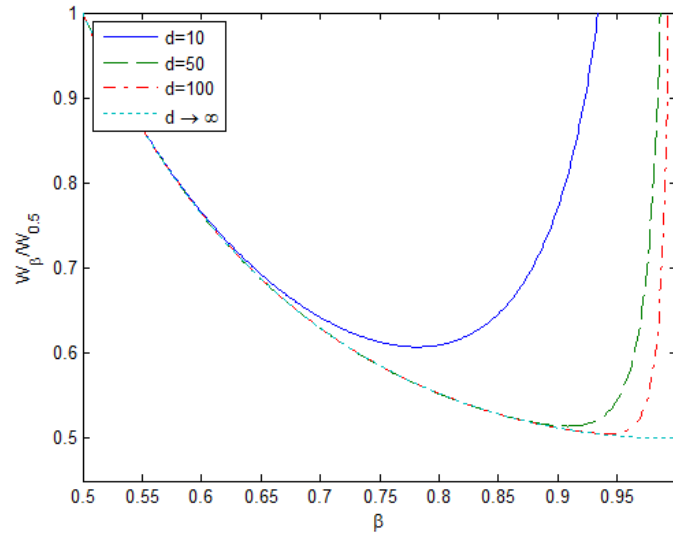


Figure 4.4: The relative work $W_\beta/W_{0.5}$ for the d -dimensional Poisson problem converges to the limit curve shown as $d \rightarrow \infty$.

Chapter 5

Examples

In this section we apply Theorems 3.1.5 and 3.3.1 to a few examples, obtaining formulas for the smoothing factors of standard discretizations of several common PDEs; these are then compared with the measured asymptotic convergence rate of multigrid with GS-LEX smoothing. We also include measured convergence rates with GS-RB smoothing, and show that while the latter has better smoothing properties for the symmetric case [17], lexicographic smoothing is more effective in the strongly asymmetric setting of the convection-dominated convection-diffusion equation. All experiments are done on a rectangular domain with homogeneous Dirichlet boundary conditions. Unless stated otherwise, we use the Galerkin coarse grid operator $L^H = I_h^{2h} L^h I_{2h}^h$ in our experiments, where I_{2h}^h is the prolongation operator with linear interpolation, and $I_h^{2h} = 2^d (I_{2h}^h)^T$ is the restriction operator with full weighting. The asymptotic convergence rate ρ is estimated from the sequence of residuals $\{r^{(k)}\}_{k=0}^m$ using the identity

$$\rho = \lim_{m \rightarrow \infty} \sqrt[m]{\frac{\|r^{(m)}\|_2}{\|r^{(0)}\|_2}}$$

and approximating the limit with a sufficiently large value of m (see [13, p. 54] for justification). All experiments were done in MATLAB 7.9.0 on a 1.80 GHz Intel(R) Core(TM)2 Duo CPU.

relaxation/dimension	$d = 1$	$d = 2$	$d = 3$
point relaxation	0.447	0.5	0.567
line relaxation		0.447	0.5
plane relaxation			0.447

Table 5.1: Smoothing factors of GS-LEX with point, line, and plane relaxation for the Poisson problem in 1D, 2D, and 3D. Moving from point to line or from line to plane relaxation is equivalent to reducing the dimension by one.

5.1 The d -Dimensional Poisson Problem

Suppose the d dimensional Poisson problem

$$-\Delta u = f$$

is discretized using centered differences on a uniform grid. Theorem 3.1.5 gives a general formula for the smoothing factor of pointwise GS-LEX, namely

$$\mu_{Poisson}^{pt}(d) = \frac{2(d-1) + \sqrt{d^2 - 4d + 8}}{3d + 2}.$$

It can be verified that this reduces to the known values $1/\sqrt{5}$ and $1/2$ for $d = 1$ and $d = 2$ respectively [3], but also yields the exact expression

$$\mu_{Poisson}^{pt}(3) = \frac{4 + \sqrt{5}}{11} \approx 0.5669,$$

which as far as we know has not appeared in the literature in closed form.

Similarly, Theorem 3.3.1 provides a general formula for block GS-LEX applied to the d -dimensional Poisson problem. If each block is a k -dimensional subproblem, then we have

$$\mu_{Poisson}^{block}(d, k) = \begin{cases} \mu_{Poisson}^{pt}(d-k) & \text{if } d-k < 3, \\ \frac{d-k}{2+d-k} & \text{otherwise;} \end{cases}$$

see Table 5.1.

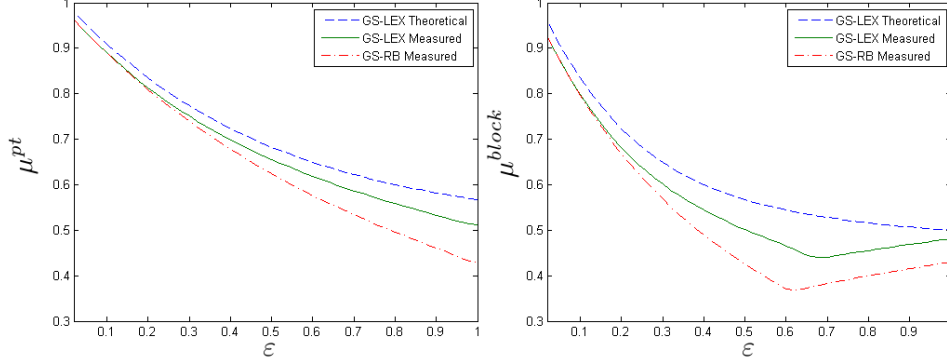


Figure 5.1: Theoretical smoothing factors of GS-LEX vs. experimental asymptotic convergence rates of two-level multigrid with GS-LEX and GS-RB smoothers, applied to the steady-state diffusion equation $-u_{xx} - u_{yy} - \varepsilon u_{zz} = f$. Pointwise relaxation appears on the left, while x -oriented line relaxation is depicted on the right. The discretization is on a $23 \times 23 \times 23$ grid with centered differences.

5.2 The Anisotropic Steady-State Diffusion Problem

Suppose the anisotropic steady-state diffusion problem

$$-\rho_1 u_{x_1 x_1} - \rho_2 u_{x_2 x_2} - \dots - \rho_d u_{x_d x_d} = f \quad (5.1)$$

is discretized over \mathbb{R}^d with standard second-order centered differences. We then obtain a linear system with \mathcal{L}^h of the form (1.2) where $b_k^+ = b_k^- = \rho_k$, $a = 2 \sum_{k=1}^d \rho_k$.

In particular, if $d = 3$ and $\rho_1 = \rho_2 = 1$ while $\rho_3 = \varepsilon \in (0, 1]$, then

$$\mu^{pt} = \frac{4 + \sqrt{5\varepsilon^2 - 4\varepsilon + 4}}{6 + 5\varepsilon}.$$

Note that $\mu^{pt} \rightarrow 1$ as $\varepsilon \rightarrow 0$. If line relaxation is performed in which unknowns along the x or y directions are relaxed simultaneously, one obtains

$$\mu^{line} = \frac{2 + \sqrt{5\varepsilon^2 - 2\varepsilon + 1}}{3 + 5\varepsilon}.$$

ε	1.00	0.50	0.10	0.05
μ^2	0.25	0.32	0.70	0.83
$V(1, 1)$	0.12	0.27	0.68	0.81

Table 5.2: Comparison of predicted smoothing factors of GS-LEX with measured convergence rates of multigrid $V(1, 1)$ -cycles, for the PDE $-\varepsilon u_{xx} - u_{yy} = f$. The finest grid is 1023×1023 , while the coarsest is 1×1 . Various values of ε are considered.

It is interesting to note that μ^{line} above is identical to μ^{pt} in the lower dimensional case $d = 2$ and $\rho_1 = 1$, $\rho_2 = \varepsilon$. Figure 5.1 provides a comparison between the formulas for μ^{pt} and μ^{line} and measured asymptotic converge rates of two-level multigrid on a $23 \times 23 \times 23$ grid, for both GS-LEX and GS-RB. The graphs illustrate that GS-RB is superior to GS-LEX in this symmetric case. We also consider multigrid V -cycles with one pre and one post GS-LEX smoothing step, applied to model problem (5.1) with $d = 2$, $\rho_1 = 1$, $\rho_2 = \varepsilon$ on a 1023×1023 grid. These are compared against theoretical smoothing factors in Table 5.2.

5.3 Alternative Coarsening for the 2D Poisson Problem

Suppose we wish to solve the two-dimensional Poisson problem

$$u_{xx} + u_{yy} = f,$$

(discretized using centered differences) by way of multigrid with a coarsening strategy $H = kh$, where $k = 2, 3, 4, \dots$

It follows from Equation (4.1) that the smoothing factor of point GS-LEX in this case is given by

$$\mu^{pt} = \frac{1}{2 - \cos\left(\frac{\pi}{k}\right)}.$$

Increasing k decreases the work per iteration by making the coarse grids diminish in size more rapidly, however, since $\mu^{pt} \rightarrow 1$ as $k \rightarrow \infty$, the number of iterations is at the same time increasing without bound.

The analysis of Chapter 4 predicts that the standard choice $k = 2$ balances these

	$H = 2h$	$H = 3h$	$H = 4h$	$H = 5h$	$H = 6h$
Smoothing factor	0.5	0.66	0.77	0.84	0.88
$V(1,0)$ -cycle convergence rate	0.43	0.64	0.76	0.84	0.88
Clock time per grid point (microseconds)	74	125	188	295	424
Relative work (measured)	1	1.7	2.5	3.9	5.7
Relative work (theoretical)	1	1.4	2.2	3.1	4.3

Table 5.3: Predicted smoothing factors smoothing vs. measured convergence rates of multigrid with $H = kh$ coarsening and GS-LEX smoothing, applied to the 2D Poisson problem discretized using centered differences, for various values of k . $V(1,0)$ -cycles are used, with a finest grid of size $k^{N_k} - 1$ in each direction, and a coarsest grid consisting of a single point. $N_k = \lfloor \log_k(999) \rfloor$ is chosen to make the total number of grid points approximately one million. Also included is the total work as measured by the running time of the program divided by the size of the problem - this is given both in microseconds per grid point and as a ratio relative to the reference case $H = 2h$. This latter form is compared against the theoretical relative work (4.2).

opposing forces in such a way as to minimize their product¹.

To test this prediction, numerical experiments with $H = kh$ for $k = 2, 3, 4, 5, 6$ have been performed. The overall work is measured in terms of clock time on the computer running the experiment, however, this value is not used directly. Standard multigrid prolongation and restriction operators require that the resolution of the finest grid be $(k^n - 1) \times (k^n - 1)$ for some $n \in \mathbb{N}$. As a result it is not possible run all experiments on the same grid; for this reason the total work is calculated by dividing the running time of the program by the number of grid points on the finest grid. For each k , a grid of size $(k^{N_k} - 1) \times (k^{N_k} - 1)$ is used, where N_k is chosen to be the largest integer such that the number of points on the finest grid is less than 10^6 . Thus $N_2 = 9$, $N_3 = 6$, $N_4 = N_5 = 4$, $N_6 = 3$.

The results are shown in Table 5.3. The first two rows compare predicted smoothing factors with measured $V(1,0)$ -cycle convergence rates, showing close agreement. The next three rows compare the total work of $V(1,0)$ -cycles with the theoretical work (4.2).

¹To be precise, $H = 2h$ is optimal amongst the choices $H = kh$ where $k \in \mathbb{N}$. In fact, $H = 1.75h$ is in theory slightly more efficient.

It is found that $k = 2$ is the most efficient of these possibilities. However, the agreement between the measured and theoretical work is not as close as the agreement between smoothing factors and measured convergence rates. Experimental results show that $k > 2$ is somewhat less efficient than expected. One possible explanation for this discrepancy lies in the observation that MATLAB (the platform on which the experiments were run) has highly optimized built in functions and vectorization, making estimates of the form (4.2) less meaningful.

5.4 The Convection-Diffusion Problem

Suppose the constant coefficient convection-diffusion problem

$$-\Delta u + \vec{w} \cdot \nabla u = f$$

is discretized on a uniform rectangular grid with mesh spacing h . We define the mesh Reynolds numbers by $\gamma_k = w_k h / 2$, and assume they are all equal, that is

$$\gamma_1 = \gamma_2 = \dots = \gamma_d = \gamma.$$

If we discretize this PDE using centered differences, we obtain $a = 2d$, $b_k^+ = 1 - \gamma$, $b_k^- = 1 + \gamma$ for all k . Provided $|\gamma| < 1$, (1.3) is satisfied and Theorem 3.1.5 gives

$$\mu^{pt} = (1 - |\gamma|) \left[\frac{(d-1) + \sqrt{1 + (1 + |\gamma|)^2 (1 - \frac{d}{2})^2}}{2d + (1 + |\gamma|)^2 (1 - \frac{d}{2})} \right].$$

Note that the condition $|\gamma| < 1$ means we are not in the convection-dominated regime. On the other hand, if we use first order upwinding, then (assuming for simplicity that $\gamma \geq 0$) we obtain $a = 2d(1 + \gamma)$, $b_k^+ = 1$, $b_k^- = 1 + 2\gamma$ for all k . The constraints (1.3) are satisfied for all $\gamma \geq 0$ (in other words, the convection-dominated case is included) and

$$\mu^{pt} = \frac{1}{1 + \gamma} \left[\frac{(d-1) + \sqrt{1 + \left(1 + \frac{\gamma}{1+\gamma}\right)^2 (1 - \frac{d}{2})^2}}{2d + \left(1 + \frac{\gamma}{1+\gamma}\right)^2 (1 - \frac{d}{2})} \right].$$

	centered differences, $\gamma \in (-1, 1)$	upwinding, $\gamma \geq 0$
$d = 1$	$\frac{1- \gamma }{\sqrt{4+(1+ \gamma)^2}}$	$\frac{1}{\sqrt{8\gamma^2+12\gamma+5}}$
$d = 2$	$\frac{1- \gamma }{2}$	$\frac{1}{2(1+\gamma)}$
$d = 3$	$\frac{(1- \gamma)\left(2+\sqrt{1+\left(\frac{1+ \gamma }{2}\right)^2}\right)}{6-\frac{1}{2}(1+ \gamma)^2}$	$\frac{2+\sqrt{1+\frac{1}{4}\left(1+\frac{\gamma}{1+\gamma}\right)^2}}{(1+\gamma)\left(6-\frac{1}{2}\left(1+\frac{\gamma}{1+\gamma}\right)^2\right)}$

Table 5.4: Smoothing factors of pointwise GS-LEX applied to two common discretizations of the convection-diffusion equation with all mesh Reynolds numbers equal to γ , for $d = 1, 2, 3$.

Table 5.4 lists the above smoothing factors for the cases of interest $d = 1, 2, 3$.

For asymmetric problems such as convection-diffusion, Galerkin coarse grid operators obtained by linear interpolation and full weighting may cause unstable coarse grid discretizations (even in diffusion dominated settings), leading to multigrid convergence problems, especially as the number of levels increases [14]. Indeed, in Table 5.5 we observe divergence for multigrid with three or more levels for upwinding in the convection-dominated regime; for two levels, however, we see a good agreement between measured convergence rates and the predicted smoothing factor. We note that there are ways of dealing with these instabilities within the Galerkin framework; see for example [5].

γ	0.00	0.33	0.66	1.00	3.00	5.00
μ	0.50	0.38	0.30	0.25	0.13	0.08
$2L$	0.39	0.36	0.30	0.25	0.13	0.08
$3L$	0.41	0.36	0.31	NC	NC	NC
$4L$	0.41	NC	NC	NC	NC	NC

Table 5.5: Comparison of the smoothing factor μ of GS-LEX with the measured asymptotic convergence rate of multigrid applied to the PDE $-u_{xx} - u_{yy} + \sigma u_x + \sigma u_y = f$, discretized uniformly. The problem is discretized using upwinding on a 1023×1023 grid. The Galerkin coarse grid operator is used. $W(1,0)$ -cycles are used, and various values of $\gamma = \sigma h/2$, where h is the grid spacing on the finest mesh, are considered. ‘NC’ stands for no convergence. ‘ nL ’ (with $n = 2, 3, 4$) signifies the number of levels.

The coarse grid operator \mathcal{L}^H may also be defined by direct discretization of the underlying PDE on the coarse mesh. For upwinding with constant coefficients, this yields a stable discretization on all grid levels, and we can expect multigrid methods to converge. However, it has been shown that in this setting, for strong convection the determining factor of the multigrid convergence rate may not be the smoother, but rather the coarse grid correction [4]. In particular, in the high convection limit, there are certain smooth modes which are reduced by a constant factor that grows as the number of levels increases. In the convection-dominated regime where the smoothing factor μ becomes smaller than this factor, it is these smooth modes rather than the rough modes with reduction factor μ , that dominate the error asymptotically. Therefore, we expect that the asymptotic convergence of multigrid will follow the trend indicated by our smoothing analysis in the diffusion-dominated regime, while tending to a constant in the convection-dominated regime; see Table 5.6. We also include GS-RB W -cycles in the table, to illustrate the superiority of GS-LEX in this case.

γ	0.0	1.0	2.0	3.0	4.0	5.0
μ	0.50	0.25	0.17	0.13	0.10	0.08
$2L$	0.37	0.25	0.24	0.27	0.29	0.30
$V(1,0)$	0.42	0.56	0.54	0.52	0.50	0.48
$W(1,0)_{LEX}$	0.37	0.29	0.34	0.35	0.36	0.36
$W(1,0)_{RB}$	0.23	0.37	0.46	0.59	0.71	0.79

Table 5.6: Comparison of the smoothing factor μ of GS-LEX with measured convergence rates of multigrid $V(1,0)$ and $W(1,0)$ -cycles, for the PDE $-u_{xx} - u_{yy} + \sigma(u_x + u_y) = f$, discretized uniformly. The finest grid is 1023×1023 , while the coarsest is 1×1 . The coarse grid operator is obtained by direct discretization of the PDE on the coarse mesh. Various values of $\gamma = \sigma h/2$ are considered, where h is the grid spacing of the finest mesh. ‘ $2L$ ’ stands for a two-level scheme, whereas $V(1,0)$ and $W(1,0)$ signify V and W -cycles, respectively, with one pre-smoothing sweep and no post-smoothing.

For the centered difference discretization of the convection-diffusion problem, our smoothing analysis applies to the case where the mesh Reynolds number γ is less than one in magnitude. But in a multigrid setup, we have different mesh

γ	0.00	0.20	0.40	0.60	0.80	0.99
μ	0.50	0.40	0.30	0.20	0.10	0.005
$2L$	0.37	0.36	0.30	0.21	0.10	0.005
$3L$	0.37	0.36	0.30	NC	NC	NC
$4L$	0.37	0.36	NC	NC	NC	NC
$5L$	0.37	NC	NC	NC	NC	NC

Table 5.7: Comparison of the smoothing factor μ of GS-LEX with the measured asymptotic convergence rate of two-level and multi-level multigrid applied to the PDE $-u_{xx} - u_{yy} + \sigma(u_x + u_y) = f$, discretized uniformly. The problem is discretized using centered differences on a 1023×1023 grid. The coarse grid operator is obtained by direct discretization of the PDE on the coarse mesh. $W(1,0)$ -cycles are used, and various values of $\gamma = \sigma h/2$, where h is the grid spacing on the finest mesh, are considered. ‘NC’ stands for no convergence.

Reynolds numbers on different grids - for our analysis to apply, they must all obey this constraint. Therefore, if L denotes the number of levels in our algorithm and γ denotes the mesh Reynolds number on the finest grid, what we really require is $|\gamma| < 2^{2-L}$.

In Table 5.7, we compare our predictions with measured convergence rates for various values of L ; $W(1,0)$ -cycles are applied. In practice, we can get away with a mesh Reynolds number moderately larger than 1 on the coarsest grid and still maintain expected convergence rates.

5.5 The Time-dependent Diffusion Problem

Suppose the time-dependent diffusion problem

$$u_t = \rho_1 u_{x_1 x_1} + \rho_2 u_{x_2 x_2} + \dots + \rho_d u_{x_d x_d}$$

is discretized on a uniform mesh using backward Euler in time and centered differences in space, with multigrid used to solve the resulting linear system at each time step. If τ denotes the timestep size, then μ^{pt} follows from Theorem 3.1.5 by setting

$$c_r = \frac{\rho - \rho_m}{2\rho + h^2/\tau}, \quad c_m = \frac{\rho_m}{2\rho + h^2/\tau}, \quad \alpha = 1.$$

where $\rho = \sum_{k=1}^d \rho_k$ and $\rho_m = \min(\{\rho_k\})$.

In particular, in the isotropic case $\rho_1 = \rho_2 = \dots = \rho_d = 1$, we obtain

$$\mu^{pt} = \frac{(d-1)(2d + h^2/\tau) + \sqrt{(2d + h^2/\tau)^2 + d^2(2-d)^2}}{(2d + h^2/\tau)^2 + d(2-d)}.$$

When $d = 2$ this reduces to the particularly simple expression

$$\mu^{pt} = \frac{1}{2 + \frac{h^2}{2\tau}}.$$

Note the (expected) behavior of the smoothing factor as a function of τ : when h is fixed and τ gets smaller, the smoothing factor becomes smaller too.

Chapter 6

Conclusions

Using results from complex analysis, primarily the maximum modulus principle and properties of Möbius transformations, we have derived closed-form expressions for the smoothing factors of lexicographic pointwise and block Gauss-Seidel (Theorems 3.1.5 and 3.3.1). An extension of Theorem 3.1.5 that incorporates a relaxation parameter is provided in (3.10). In the pointwise case our results are valid for general operators of the form (1.2) satisfying the constraints (1.3) and (3.2), whereas in the block case we require the additional assumption of symmetry. In some cases, block smoothing on a high dimensional problem has the same smoothing factor as pointwise smoothing on a related lower dimensional problem.

Our analysis provides smoothing factors for, among other equations, the following:

- Pointwise GS-LEX applied to the d -dimensional *anisotropic steady-state diffusion* equation $-\sum_{k=1}^d \rho_k u_{x_k x_k} = f$, discretized with centered differences.
- Pointwise and block GS-LEX applied to the d -dimensional *Poisson* equation $-\Delta u = f$, discretized with centered differences.
- Pointwise GS-LEX applied to the d -dimensional constant coefficient *convection-diffusion* equation $-\Delta u + \vec{w} \cdot \nabla u = f$, discretized with centered differences or upwinding, and all mesh Reynolds numbers equal.
- Pointwise GS-LEX applied to the linear systems arising in each timestep of

the solution of the d -dimensional *time-dependent diffusion* equation $u_t = \sum_{k=1}^d \rho_k u_{x_k x_k}$, discretized with centered differences in space and backward Euler in time.

For pointwise GS-LEX, we have generalized our results to allow mesh ratios besides $H = 2h$. We have shown that for high dimensional Poisson problems, a judicious choice of $2h > H > h$ can in theory lead to a reduction of up to 50% in the overall work.

We have also observed that lexicographic Gauss-Seidel smoothing is effective for equations with strong asymmetry. In particular, for the constant coefficient convection-diffusion equation with equal mesh Reynolds numbers we have shown for upwind discretizations that GS-LEX has a smaller smoothing factor than GS-RB in the convection-dominated regime.

Bibliography

- [1] E. L. Akim, O. M. Belotserkovskii, K. V. Brushlinskii, S. K. Godunov, V. F. Dyachenko, V. T. Zhukov, M. K. Kerimov, A. I. Lobanov, M. V. Maslennikov, I. B. Petrov, Y. P. Popov, G. P. Prokopov, V. S. Ryabenskii, L. G. Strakhovskaya, A. S. Kholodov, B. N. Chetverushkin, and T. M. Eneev. In Memory of Radii Petrovich Fedorenko. *Computational Mathematics and Mathematical Physics*, 50:1459–1463, 2010.
- [2] U. M. Ascher and C. Greif. *A First Course in Numerical Methods*. Computational Science and Engineering. SIAM, Philadelphia, 2011.
- [3] A. Brandt. Multi-Level Adaptive Solutions to Boundary-Value Problems. *Mathematics of Computation*, 31(138):333–390, 1977.
- [4] A. Brandt and I. Yavneh. Accelerated Multigrid Convergence and High-Reynolds Recirculating Flows. *SIAM J. Sci. Comput.*, 14(3):607–626, 1993.
- [5] P. de Zeeuw. Matrix-dependent Prolongations and Restrictions in a Blackbox Multigrid Solver. *Journal of Computational and Applied Mathematics*, 33(1):1–27, 1990.
- [6] J. Dendy. Black Box Multigrid with Coarsening by a Factor of Three. *Numerical Linear Algebra with Applications*, 17:577–598, 2010.
- [7] H. C. Elman and M. P. Chernesky. Ordering Effects on Relaxation Methods Applied to the Discrete One- Dimensional Convection-Diffusion Equation. *SIAM Journal on Numerical Analysis*, 30(5):1268–1290, 1993.
- [8] R. P. Fedorenko. A relaxation method for solving elliptic difference equations. *USSR Computational Mathematics and Mathematical Physics*, 1(4):1092–1096, 1962.

- [9] R. P. Fedorenko. The speed of convergence of one iterative process. *USSR Computational Mathematics and Mathematical Physics*, 4(3):227–235, 1964.
- [10] T. W. Gamelin. *Complex Analysis*. Springer, 2001.
- [11] W. Hackbusch. *Multigrid methods and applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.
- [12] H. Nyquist. Certain Topics in Telegraph Transmission Theory. *Trans. Amer. Inst. Electr. Eng.*, pages 617–644, 1928.
- [13] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid*. Academic Press Inc., San Diego, CA, 2001.
- [14] E. J. van Asselt and P. de Zeeuw. The Convergence Rate of Multi-Level Algorithms Applied to the Convection-Diffusion Equation. *SIAM Journal on Scientific and Statistical Computing*, 6(2):492–503, 1985.
- [15] R. S. Varga. *Matrix Iterative Analysis*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1962.
- [16] P. Wesseling. *An Introduction to Multigrid Methods*. Pure and Applied Mathematics (New York). John Wiley & Sons Ltd., Chichester, 1992.
- [17] I. Yavneh. Multigrid Smoothing Factors of Red-Black Gauss-Seidel Applied to a Class of Elliptic Operators. *SIAM J. Numer. Anal.*, 32(6):1126–1138, 1995.
- [18] I. Yavneh. On Red-Black SOR Smoothing in Multigrid. *SIAM J. Sci. Comput.*, 17(1):180–192, 1996.
- [19] I. Yavneh. Why Multigrid Methods are so Efficient. *Computing in Science and Engineering*, 8(6):12–22, 2006.
- [20] H. B. Zubair, C. Oosterlee, and R. Wienands. Multigrid for High-Dimensional Elliptic Partial Differential Equations on Non-equidistant Grids. *SIAM J. Sci. Comput.*, 29(4):1613–1636, 2007.