Summary: In this paper, we prove the convergence of the multilevel iterative method for solving linear equations that arise from elliptic partial differential equations. Our theory is presented entirely in terms of the generalized condition number κ of the matrix A and the smoothing matrix B. This leads to a completely algebraic analysis of the method as an iterative technique for solving linear equations; the properties of the elliptic equation and discretization procedure enter only when we seek to estimate κ, just as in the case of most standard iterative methods. Here we consider the fundamental two level iteration, and the V and W cycles of the j-level iteration ($j > 2$). We prove that the V and W cyles converge even when only one smoothing iteration is used. We present several examples of the computation of κ using both Fourier analysis and standard finite element techniques. We compare the predictions of our theorems with the actual rate of convergence. Our analysis also shows that adaptive iterative methods, both fixed (Chebyshev) and adaptive (conjugate gradients and conjugate residuals), are effective as smoothing procedures.

Sharp Estimates for Multigrid Rates of Convergence with General Smoothing and Acceleration

Randolph E. Bank[1] and Craig C. Douglas[2]

Research Report YALEU/DCS/RR-277

[1] Department of Mathematics, University of California at San Diego
[2] Department of Computer Science, Yale University
This research was supported in part by ONR Grant N00014-82-K-0197[1], ONR Grant N00014-82-K-0184[2], FSU-ONR Grant F1.N00014-80-C-0076[2], and NSF MCS-8106181[2].

1 Introduction

In this paper, we prove the convergence of the multilevel iterative method for solving linear equations that arise from elliptic partial differential equations. While many convergence proofs already exist (e.g., Astrakhantsev [2], Bakhvalov [4], Bank and Dupont [5], Braess and Hackbusch [7], Douglas [10], Federenko [11], Hackbusch [13, 14, 15, 16], Nicolaides [18], Van Rosendale [20], Verfuth [21], Wesseling [22], and Yserentant [23]), our assumptions and proof techniques are different and (we believe) enlightening. While we are principally interested in partial differential equations, our theory can be applied to symmetric, positive definite linear systems of equations which do not arise from partial differential equations. A different perspective on algebraic aspects of the two level scheme is given by Greenbaum [12].

We do not make the usual assumptions about elliptic regularity of the continuous problem or approximation properties of the finite dimensional spaces M_j corresponding to the discretization procedure. Indeed, our theory is presented entirely in terms of the generalized condition number κ of the matrix A and the smoothing matrix B. This is analagous to the generalized condition number now used routinely in the analysis of preconditioned conjugate gradient iterations (where B is commonly called a preconditioner). This leads to a completely algebraic analysis of the method as an iterative technique for solving linear equations; the properties of the elliptic equation and discretization procedure enter only when we seek to estimate κ, just as in the case of most standard iterative methods.

For the multilevel method, however, we compute κ with respect to a certain subspace (the orthogonal complement of the coarser space M_{j-1}) rather than the entire space M_j. When κ can be bounded independent of the level j, as is often the case, the multilevel method has a rate of convergence bounded independent of j.

It has often been remarked that there is a substantial gap between theory and practice in multilevel methods [8] because the constants in the proof are much too pessimistic. As we show by example in Section 3, sharp estimates for κ lead to sharp or

reasonably sharp estimates of the rate of convergence, just as in the case of standard iterative methods. It is interesting to note that for small values of the number of smoothing iterations m, the rate of convergence predicted by our analysis in Theorem 4 is bounded by $((\kappa-1)/\kappa)^m$. The number $(\kappa-1)/\kappa$ is similar to the "smoothing rate" often used to predict the rate of convergence of the multilevel method [9]. Smoothing rates are computed by (generally nonrigorous) techniques, sometimes involving Fourier analysis, in which the effect of the smoothing procedure is estimated on a certain subspace. Our analysis suggests that when this subspace is close to the right one, then such techniques can lead to useful estimates of the convergence rates, at least for small values of m.

Finally, our analysis shows that adaptive iterative methods (e.g., preconditioned conjugate gradients) are effective as smoothing procedures. Unlike the analysis of Axelsson and Gustafsson [3] and Kettler and Meijerink [17], we use conjugate gradients as a smoother to multigrid. In their analysis, multigrid is used as a preconditioner to conjugate gradients. The use of acceleration techniques in the smoothing iteration essentially squares the rate of convergence (for large m), just as in the case of acceleration of regular iterative methods. The added cost of conjugate gradient acceleration may not be cost effective for problems with smooth solutions on a sequence of uniform grids, since only a modest reduction in the error components of the error on any smoothing step is required. It is usually not hard to devise smoothing procedures to achieve that in this case. However, if the coefficients of the partial differential equation and the solution are rough and the grids are irregular and not uniformly refined, finding a simple but effective smoothing procedure may be considerably more difficult. Here, conjugate gradient accleration can pay off handsomely since its adaptive nature allows it to compensate to some extent for the shortcomings of the basic smoothing procedure.

The remainder of this paper is organized as follows: in Section 2 we define and analyze the multilevel iteration. Here we consider the fundamental two level iteration, and the V and W cycles of the j-level iteration ($j > 2$). Our analysis of the V cycle was inspired by the recent paper of Braess and Hackbusch [7]. In Section 3 we present several

examples of the computation of κ using both Fourier analysis and standard finite element techniques. We compare the predictions of our theorems with the actual rate of convergence. In Section 4 we analyze the use of acceleration procedures, both fixed (Chebyshev) and adaptive (conjugate gradients and conjugate residuals) in conjunction with the smoothing procedure.

2 Analysis

2.1 The Problem and the Multigrid Algorithm

Let $\mathcal{H}$ be a Hilbert space and $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \ldots \subseteq \mathcal{M}_j \subseteq \mathcal{H}$ be a sequence of finite dimensional subspaces. We denote the dimension of $\mathcal{M}_j$ by N_j. Let $a(\cdot,\cdot)$: $\mathcal{H}\otimes\mathcal{H} \rightarrow R$ be a continuous, symmetric, positive definite bilinear form and let $g(\cdot)$: $\mathcal{H} \rightarrow \mathcal{R}$ be a continuous, bounded, linear functional. The problem to be solved is

$$a(z,v) = g(v) \quad \text{for all } v \in \mathcal{M}_j. \tag{1}$$

In a typical situation, the bilinear form $a(\cdot,\cdot)$ arises from the weak formulation of an elliptic boundary value problem, $\mathcal{H}$ is an appropriate Sobolev space, and the $\mathcal{M}_j$ are finite element subspaces in which we seek approximate solutions. Usually the N_j are required to grow geometrically in order to prove an optimal order work estimate. The right hand side of (1) can be either the right hand side of the discrete elliptic problem or a residual from the (j+1)-level iteration.

We will solve (1) by a standard multigrid algorithm, MG(j,z,g,m,p,r), which has six arguments:

- j the current level number
- z the approximate solution
- g the approximate right hand side
- m the number of smoothing iterations
- p the number of correction iterations used
- r the number of iterations of MG on level j

The precise definition of Algorithm MG is given by

Algorithm 1: MG(j, z, g, m, p, r): r iterations of the j-level scheme is defined by

(1) If j = 1 then solve directly.

(2) If j > 1 then

(a) smooth m times on z

(b) do r times:

(i) compute a residual correction problem c and set e = 0

(ii) MG(j−1, e, c, m, p, p)

(iii) z = z + e

(iv) smooth m times on z

Later in this section we will prove convergence results for this algorithm for choices of the parameters used in practice.

Algorithm MG can be imbedded in another multigrid algorithm (sometimes referred to as nested iteration or full multigrid) in which the sequence of discrete problems (1) associated with M_1, M_2, ... is solved. The solution of the j-th problem serves as the initial guess for the (j+1)-st. Since analysis of several such schemes for both linear and nonlinear problems is available elsewhere (e.g., Bank and Dupont [5], Bank and Rose [6], Douglas [10], Hackbusch [13, 15, 16]), we do not repeat that analysis here. Instead, we obtain reasonably sharp estimates for the spectral radius of the iteration matrix associated with Algorithm MG.

There are two main components of Algorithm MG when j > 1: smoothing and coarse grid correction. We will present our discussion of these topics using both inner product and matrix notation. The former is more convenient for our proofs while the latter is useful for clarifying details of the implementation.

Let each space M_j have a computational basis ϕ_k (when required, we will add a superscript j to denote the level, e.g., $\phi_k^{(j)}$). Define

$$A_{ik} = a(\phi_i,\phi_k) \quad \text{and} \quad G_i = g(\phi_i).$$

The stiffness matrix A is symmetric, positive definite, and usually sparse. Equation (1) can be rewritten as

$$AZ = G, \qquad \text{where } Z = \sum_{i=1}^{N} Z_k \phi_k.$$

As with the computational basis, we will add a superscript j to A, Z, and G if it is necessary.

2.2 Smoothing Process

We first consider the smoothing process. Let $b_j(\cdot,\cdot)$: $M_j \otimes M_j \rightarrow R$ be a continuous, symmetric, positive definite bilinear form (called a *smoother*) associated with M_j. For simplicity we assume

$$\sup_{v \in M_j} a(v,v) \,/\, b_j(v,v) \leq 1. \tag{2}$$

This assumption can always be satisfied in practice by scaling any proposed smoother by an appropriate damping constant. The smoother has the property that solving the linear system BX = Y (where B, X, and Y are analagous to A, Z, and G) should be easy in comparison to solving AZ = G. In particular, the cost should be proportional to N_j in order to obtain an optimal order work estimate [4, 5, 7, 10, 11, 14, 18].

Consider the generalized eigenvalue problem

$$a(\psi_k, v) = \lambda_k b_j(\psi_k, v) \qquad \text{for all } v \in M_j.$$

Without loss of generality, we order the eigenvalues such that

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N_j} \leq 1,$$

and normalize $\{\psi_k\}$ such that

$$b_j(\psi_k, \psi_i) = \delta_{ki} \qquad \text{and} \qquad a(\psi_k, \psi_i) = \lambda_k \delta_{ki}.$$

From the viewpoint of the smoothing process, eigenfunctions corresponding to larger eigenvalues are called "rough" while those corresponding to smaller eigenvalues are called "smooth".

Since the $\{\psi_k\}$ form an orthonormal basis for M_j, we can define discrete norms $|||v|||_\theta \equiv |||v|||_{\theta,j}$ for $v \in M_j$, $\theta \in R$ by

$$|||v|||_\theta^2 = \sum_{i=1}^{N} c_i^2 \lambda_i^\theta, \qquad \text{where } v = \sum_{i=1}^{N} c_i \psi_i.$$

Thus, $|||v|||^2 = |||v|||_1^2 = a(v,v)$ and $|||v|||_0^2 = b_j(v,v)$. The 1-norm (or *energy* norm) is

well defined on the spaces M_j and H. The convergence of Algorithm MG will be analyzed in this norm.

If $N = N_j$ and $v = \sum_{i=1}^{N} V_i \phi_i \in M_j$, then in matrix notation

$$|||v|||_\theta^2 = V^T A^{1/2}(A^{-1/2}BA^{-1/2})^{1-\theta}A^{1/2}V = V^T B^{1/2}(B^{-1/2}AB^{-1/2})^{\theta}B^{1/2}V. \tag{3}$$

The basic smoothing step S: $w \in M_j \rightarrow \bar{w} \in M_j$ is defined by

$$b_j(\bar{w}-w,v) = g(v)-a(w,v) \quad \text{for all } v \in M_j \quad \text{or} \tag{4}$$

$$B(\bar{W}-W) = G-AW \quad \text{in matrix terminology.}$$

Letting $e = z-w = \sum_{i=1}^{N} c_i \psi_i$, we see from (1) and (4) that

$$\bar{e} = z-\bar{w} = S(e) = \sum_{i=1}^{N} c_i \psi_i (1-\lambda_i).$$

We can think of $\bar{e}$ as "smoother" than e in the sense that the rough components of the error (the ones corresponding to eigenvalues close to one) are damped quickly. Note that

$$|||S(e)|||_\theta \leq |||e|||_\theta \quad \text{for all } \theta \in \mathcal{R}.$$

In matrix notation we have for $e = \sum_{i=1}^{N} E_i \psi_i$,

$$A^{1/2}\bar{E} = SA^{1/2}E \quad \text{where} \quad S = I-A^{1/2}B^{-1}A^{1/2}.$$

S is the symmetric positive semi-definite error propagation matrix for the smoothing process.

2.3 Coarse Grid Correction

The other feature of Algorithm MG is the coarse grid correction. Given some approximation $\hat{z}$ to z in (1), we pose the following coarse grid problem: find $\delta \in M_{j-1}$ such that

$$a(\delta,v) = g(v)-a(\hat{z},v) \quad \text{for all } v \in M_{j-1}. \tag{5}$$

Problem (5) is solved exactly if $j = 2$. Otherwise, we apply Algorithm MG to this new (j-1)-level residual problem. Setting $p = 1$ (*V Cycle*) or $p = 2$ (*W Cycle*) in Algorithm MG is usually sufficient to obtain both good error reduction and optimal order work estimates.

If $e = z-\hat{z}$, then (5) can be written as

$$a(e-\delta,v) = 0 \quad \text{for all } v\in M_{j-1}. \tag{6}$$

δ is the orthogonal projection of the error from M_j to M_{j-1}. We define the coarse grid correction by

$$C(e) = e-\delta = z-(\hat{z} + \delta). \tag{7}$$

From the viewpoint of the coarse grid correction, elements of M_{j-1} are "smooth" and the elements of $M_{j-1}^{\perp}$ are "rough".

Let $\phi_i^{(j-1)}$ be the computational basis for M_{j-1} and let

$$\phi_k^{(j-1)} = \sum_{i=1}^{N} R_{ik}\phi_i^{(j)}, \quad 1 \le k \le N = N_{j-1},$$

define the $N_{j-1}xN_{j-1}$ matrix R. In matrix notation, (5) becomes

$$(RAR^T)\tilde{\Delta} = R(G-AW),$$

where $\tilde{\Delta}$ corresponds to the computational basis in M_{j-1}. Setting $\Delta = R^T\tilde{\Delta}$ changes the basis to that of M_j. Letting $\bar{e} = C(e)$, the coarse grid correction has the matrix representation

$$A^{1/2}\bar{E} = \{I - A^{1/2}R^T(RAR^T)^{-1}RA^{1/2}\}A^{1/2}E = CA^{1/2}E, \tag{8}$$

where C is the symmetric positive semi-definite projector for the coarse grid correction. If (5) is not solved exactly, then (8) is replaced by

$$A^{1/2}\bar{E} = \{I - A^{1/2}R^T(RAR^T)^{-1/2}(I-\tilde{Q}_{j-1}(p))(RAR^T)^{-1/2}RA^{1/2}\}A^{1/2}E$$
$$= \tilde{C}A^{1/2}E,$$

where $\tilde{Q}_{j-1}(p)$ is the symmetric positive semi-definite error propagation matrix for p iterations of Algorithm MG applied to the (j-1)-level problem.

2.4 Matrix Formulation of Algorithm MG

Suppose $r \ge 1$ iterations of MG(j,z,g,m,p,r) are used. If E_0 is the initial error and E_f is the final error, then

$$A^{1/2}E_f = \begin{cases} S^m(CS^m)^rA^{1/2}E_0 & \text{if exact corrections are used} \\ S^m(\tilde{C}S^m)^rA^{1/2}E_0 & \text{otherwise.} \end{cases}$$

Setting

$H_j = H = S^{m/2}CS^{m/2}$ and $Q_j(r) = Q(r) = S^{m/2}H^rS^{m/2}$

we have $A^{1/2}E_f = Q(r)A^{1/2}E_0$ for exact coarse grid corrections and $A^{1/2}E_f = \widetilde{Q}(r)A^{1/2}E_0$ otherwise. $\widetilde{Q}(r)$ and $\widetilde{H}$ are defined analagously to Q(r) and H.

Although to this point we have only considered partial differential equations, we can apply Algorithm MG to linear systems of equations of the form

$$AZ = G, \qquad Z,G \in \mathcal{R}^N,$$

where A is an arbitrary symmetric, positive definite matrix. We define

$$M_j = \mathcal{R}^N, \qquad N_j = N, \quad \text{and}$$

$$A_j = A.$$

Suppose we define an arbitrary full rank matrix $R_j: M_j \rightarrow M_{j-1} = \mathcal{R}^{N'}$, $N' \leq N$, such that $R_j^T R_j X = X$ for all $X \in Sm(j-1)$. R_j need not be an "interpolation" matrix since the concept of a grid may be meaningless. We define a coarser system by

$$M_{j-1} = \mathcal{R}^{N'}, \quad N_j = N', \quad \text{and}$$

$$A_{j-1} = R_j A_j R_j^T.$$

We define $\{M_k\}$ and $\{A_k\}$, $1 \leq k \leq j-1$, inductively in the obvious fashion. The inner product is defined by

$$a_j(x,y) = x^T A_j y, \qquad \text{for all } x,y \in M_j.$$

A smoother $b_j(\cdot,\cdot)$ is defined from a B_j analogously. Norms $|||\cdot|||_{\theta,j}$ are defined as in (3).

The choice of R_j is quite broad. However, R_j could be based on the operator A_j. Using this technique, many finite difference multigrid schemes (e.g., [1]) can be analyzed using the techniques in Sections 2.5 and 4.

2.5 Convergence of Algorithm MG

The central question now becomes under what conditions can we expect to bound the spectral radius $\rho(H_j) < 1$, independent of j? The answer clearly lies in the relation between smoothing iterations and coarse grid corrections. In particular, the smoothing iteration must effectively damp out the components of the error that cannot be approximated in M_{j-1}, i.e., those elements in $M_{j-1}^{\perp}$. In other words, the notions of

"smooth" and "rough" from the standpoints of smoothing and coarse grid corrections must coincide to a great extent.

In our analysis, the function $f(\alpha,\beta) = \alpha^{\alpha}\beta^{\beta}(\alpha+\beta)^{-(\alpha+\beta)}$, $\alpha,\beta > 0$, will play an important role. In Lemma 2, we summarize some of its properties:

Lemma 2: Let $p,\alpha,\beta > 0$. Then

(a) $\sup_{x\in[0,1]} x^{\alpha}(1-x)^{\beta} = f(\alpha,\beta)$

(b) $[\, f(\alpha,\beta)\,]^{p} = f(p\alpha,p\beta)$

We also need the following norm bound:

Lemma 3: Assume there exist constants $\kappa \geq 1$ and $\alpha > 0$ (both independent of j) such that for $u\in M_j$,

$$|||u|||_{1-\alpha} \leq \kappa^{\alpha/2}|||u|||. \tag{9}$$

Then

$$|||u|||_{1-\beta} \leq \kappa^{\beta/2}|||u|||, \qquad 0 \leq \beta \leq \alpha.$$

Proof: The proof follows from (9) and the Holder inequality

$$|||u|||_{1-\beta} \leq |||u|||_{1-\alpha}^{\beta/\alpha}|||u|||^{1-\beta/\alpha}.$$

QED

We now consider in detail the convergence of one iteration of Algorithm MG for the case of two levels (j = 2). Here, coarse grid correction problems are solved directly.

Theorem 4: Assume there exist constants $\kappa \geq 1$ and $\alpha > 0$ (both independent of j) such that for $u\in M_{j-1}^{\perp}\cap M_j$,

$$|||u|||_{1-\alpha} \leq \kappa^{\alpha/2}|||u|||.$$

Then $|||S^{m/2}(C(S^{m/2}(v)))||| \leq \gamma|||v|||$, where

$$\gamma = \begin{cases} [(\kappa-1)/\kappa]^{m} & \text{if } m \leq (\kappa-1)\alpha \\ \kappa^{\alpha} f(m,\alpha) & \text{if } m > (\kappa-1)\alpha \end{cases}$$

Note that γ monotonically decreases to zero as a function of m.

Proof: Let $N = N_j$ and $v = \sum_{i=1}^{N} c_i\psi_i$. Then $\bar{v} = S^{m/2}(v) = \sum_{i=1}^{N} c_i\psi_i(1-\lambda_i)^{m/2}$. Taking $\chi = C(\bar{v})-\bar{v}\in M_{j-1}$, we have

$$|||C(\bar{v})|||^2 = a(C(\bar{v}),C(\bar{v}))$$

$$= a(C(\bar{v}),C(\bar{v})-\chi)$$

$$= a(C(\bar{v}),\bar{v})$$

$$\leq |||C(\bar{v})|||_{1-\alpha}|||\bar{v}|||_{1+\alpha}$$

$$\leq \kappa^{\alpha/2}|||C(\bar{v})|||\cdot|||\bar{v}|||_{1+\alpha} \tag{10}$$

Now

$$|||\bar{v}|||^2_{1+\alpha} = \sum_{i=1}^{N} c_i^2(1-\lambda_i)^m\lambda_i^{1+\alpha} \leq f(m,\alpha)|||v|||^2$$

Taking square roots and substituting into (10) gives us

$$|||C(\bar{v})||| \leq \kappa^{\alpha/2}f(m/2,\alpha/2)|||v|||$$

Using an eigenvector expansion of $C(\bar{v})$ shows

$$|||S^{m/2}(C(\bar{v}))||| \leq f(m/2,\alpha/2)|||C(\bar{v})|||_{1-\alpha}$$

$$\leq \kappa^{\alpha/2}f(m/2,\alpha/2)|||C(\bar{v})|||$$

$$\leq \kappa^{\alpha}f(m,\alpha)|||v||| \tag{11}$$

If we substitute β instead of α, $0 < \beta \leq \alpha$, into (11) (using Lemma 3) and minimize (11) with respect to β, we have

$$D_\beta[\kappa^\beta f(m,\beta)] = \log[(\kappa\beta)/(m+\beta)]\cdot\kappa^\beta f(m,\beta)$$

$$= 0 \Leftrightarrow \kappa\beta/(m+\beta) = 1$$

$$\Rightarrow \beta = m/(\kappa-1)$$

There are two cases: the first is when $m/(\kappa-1) > \alpha$. Here the minimum occurs at $\beta = \alpha$. Hence, for $m > (\kappa-1)\alpha$,

$$\gamma = \kappa^\alpha f(m,\alpha).$$

The second case is when $m/(\kappa-1) \leq \alpha$. Here $\beta = m/(\kappa-1)$ and

$$\gamma = [(\kappa-1)/\kappa]^m.$$

QED

We now consider the convergence of one iteration of Algorithm MG when the coarse grid corrections are not exact. Theorem 5 analyzes the *W Cycle* ($j > 2$, $p = 2$ in Algorithm MG). The proof can be extended to any $p > 2$ trivially.

Theorem 5: Suppose

$$|||S^{m/2}(C(S^{m/2}(v)))||| \leq \gamma|||v||| \quad \text{for all } v\in M_j\cap M_{j-1}^{\perp}$$

and for some $0 \leq \gamma < .5$ (γ is defined in Theorem 4). Further, assume for all $v\in M_j$ that

$$|||\widetilde{C}(v)-C(v)||| \leq \widetilde{\gamma}^2|||C(v)-v|||,$$

where $\widetilde{\gamma} = \gamma/(1-\gamma)$. Then

$$|||S^{m/2}(\widetilde{C}(S^{m/2}(v)))||| \leq \widetilde{\gamma}|||v||| \quad \text{for all } v\in M_j.$$

Proof: From (6) and (7) we know that $a(C(v),\chi) = 0$ for all $\chi\in M_{j-1}$. In particular, both $\widetilde{C}(v)-C(v)\in M_{j-1}$ and $C(v)-v\in M_{j-1}$. Thus,

$$\begin{aligned} |||C(v)-\widetilde{C}(v)+\widetilde{\gamma}^2C(v)|||^2 &= |||C(v)-\widetilde{C}(v)|||^2 + \widetilde{\gamma}^4|||C(v)|||^2 \\ &\leq \widetilde{\gamma}^4\{|||C(v)-(v)|||^2 + |||C(v)|||^2\} \\ &= \widetilde{\gamma}^4|||v|||^2 \end{aligned} \tag{12}$$

Finally,

$$\begin{aligned} |||S^{m/2}(\widetilde{C}(S^{m/2}(v)))||| &\leq (1-\widetilde{\gamma}^2)|||S^{m/2}C(S^{m/2}(v)))||| + \\ &\qquad |||S^{m/2}((\widetilde{C}-C+\widetilde{\gamma}^2C)(S^{m/2}(v)))||| \\ &\leq [(1-\widetilde{\gamma}^2)\gamma+\widetilde{\gamma}^2]\cdot|||v||| \\ &= \widetilde{\gamma}|||v||| \end{aligned}$$

QED

We now consider the convergence of one iteration of the *V Cycle* ($j > 2$, $p = 1$ in Algorithm MG). Our proof corresponds closely to that of Braess and Hackbush [7], although our assumptions are weaker (just those of Theorem 4) and general smoothers are allowed. The proof uses two simple observations about the smoothing procedure and the discrete norms $|||\cdot|||_\theta$.

Lemma 6: Let $v\in M_j$ and $0 < \alpha \leq 1$. Then

$$|||S^{m/2}(v)||| \;/\; |||v||| \;\leq\; \{\; |||S^{(m+1)/2}(v)|||^2 \;/\; |||S^{m/2}(v)|||^2 \;\}^{m/2} \tag{13}$$

$$|||v|||_{1+\alpha} \;/\; |||v||| \;\leq\; \{\; 1 - |||S^{1/2}(v)|||^2 \;/\; |||v|||^2 \;\}^{\alpha/2} \tag{14}$$

Proof: The proof of both (13) and (14) follow from the definitions of the respective norms, the Holder inequalties

$$|||S^{1/2}(v)||| \le |||S^{(m+1)/2}(v)|||^{m/(m+1)}\cdot|||v|||^{1/(m+1)}$$

$$|||v|||_{1+\alpha} \le |||v|||^{1-\alpha}\cdot|||v|||_2^{\alpha},$$

and by noting that

$$|||v|||_2^2 = |||v|||^2 - |||S^{1/2}(v)|||^2,$$

QED

Theorem 7: Assume there exists a constant $\kappa \ge 1$ (independent of j) such that for $u \in M_{j-1}^{\perp} \cap M_j$,

$$|||u|||_0 \le \kappa^{1/2}|||u|||.$$

Further, assume for all $v \in M_j$ that

$$|||\tilde{C}(v)-C(v)||| \le \hat{\gamma}|||C(v)-v|||,$$

where $\hat{\gamma} = \kappa/(\kappa+m)$. Then

$$|||S^{m/2}(\tilde{C}(S^{m/2}(v)))||| \le \hat{\gamma}|||v||| \quad \text{for all } v \in M_j.$$

Proof: For any $v \in M_j$, let

$$\chi = \chi(v) = 1 - |||S^{(m+1)/2}(v)|||^2/|||S^m(v)|||^2$$

and set $\bar{v} = S^{m/2}(v)$. Using (10), (12), (13), and (14) we have

$$\begin{aligned}
|||\tilde{C}(\bar{v})||| &\le (1-\hat{\gamma})|||\tilde{C}(\bar{v})||| + |||(\tilde{C}-C+\hat{\gamma}C)(\bar{v})||| \\
&\le (1-\hat{\gamma})\kappa^{1/2}|||(\bar{v})|||_2 + \hat{\gamma}|||\bar{v}||| \\
&\le \{(1-\hat{\gamma})\kappa^{1/2}\chi^{1/2}(v) + \hat{\gamma}\}|||\bar{v}||| \\
&\le \{(1-\hat{\gamma})\kappa^{1/2}\chi^{1/2}(v) + \hat{\gamma}\}(1-\chi(v))^{m/2}|||v|||
\end{aligned}$$

Let $w \in M_j$ and $\bar{w} = S^{m/2}(w)$. Since $C = C^2$, we have

$$\begin{aligned}
a(S^{m/2}(\tilde{C}(S^{m/2}(v))),w) &= a(\tilde{C}(\bar{v}),\bar{w}) \\
&= (1-\hat{\gamma})a(C(\bar{v}),C(\bar{w})) + a((\tilde{C}-C+\hat{\gamma}C)(\bar{v}),\bar{w}) \\
&\le \{(1-\hat{\gamma})\kappa\chi^{1/2}(v)\chi^{1/2}(w) + \hat{\gamma}\}|||\bar{v}|||\cdot|||\bar{w}||| \\
&\le \{(1-\hat{\gamma})\kappa\chi^{1/2}(v)\chi^{1/2}(w) + \hat{\gamma}\}\cdot \\
&\qquad \{(1-\chi(v))(1-\chi(w))\}^{m/2}|||v|||\cdot|||w|||
\end{aligned}$$

Hence,

$$|||S^{m/2}(\tilde{C}(S^{m/2}(v)))||| \leq \sup_{x\in[0,1]} ((1-\hat{\gamma})\kappa x + \hat{\gamma})(1-x)^m |||v||| \qquad (15)$$

$$\leq [\kappa/(\kappa+m)] \sup_{x\in[0,1]} (mx+1)(1-x)^m |||v|||$$

$$\leq [\kappa/(\kappa+m)]\ |||v|||.$$

QED

For large values of m, this estimate for $\hat{\gamma}$ does not approach the estimate γ for exact coarse grid corrections (as is the case for the W Cycle). For $\alpha = 1$, $m \geq \kappa-1$, we have

$$\hat{\gamma} = \kappa\ /\ (\kappa+m) = \gamma\cdot\{\gamma + (m/(m+1))^{m+1}\}^{-1} \rightarrow e\gamma,$$

γ as in Theorem 4. For large values of κ, this estimate requires about e times as many smoothing iterations to obtain a comparable bound on the convergence rate. Another drawback is that the proof does not work for $0 < \alpha < 1$. Instead of (15), we get

$$|||S^{m/2}(\tilde{C}(S^{m/2}(v)))||| \leq \sup_{x\in[0,1]} [(1-\hat{\gamma})\kappa^\alpha x^\alpha + \hat{\gamma}](1-x)^m |||v|||$$

$$= F(\hat{\gamma},\alpha,m)|||v|||.$$

The fixed point of $F(\hat{\gamma},\alpha,m) = \hat{\gamma}$ is $\hat{\gamma} = 1$.

One benefit of the V Cycle is that it can be used to improve the estimate of Theorem 5 for the W Cycle when $\alpha = 1$.

Corollary 8: Let the assumptions of both Theorem 5 and 7 hold for $\alpha = 1$. Then

$$|||S^{m/2}(\tilde{C}(S^{m/2}(v)))||| \leq \bar{\gamma}|||v||| \qquad \text{for all } v\in M_j,$$

where $\bar{\gamma} = \min\{\tilde{\gamma},\hat{\gamma}\}$ and $\tilde{\gamma}$ is defined in Theorem 5 and $\hat{\gamma}$ in Theorem 7.

3 Examples

In this section, we give estimates for the constant κ of the theorems of the previous section for three model problems. The first two are constant coefficient problems. Due to their simplicity, we can compute κ exactly for these problems. The third is a linear second order variable coefficient self adjoint problem in a general closed domain in $\mathcal{R}^2$. The estimate of κ given for this problem is not as sharp as for those of the first two

problems.

The first two examples are Poisson's equation in one and two dimensions:

$$-u'' = f \text{ in } \Omega = (0,1) \ ; \ u(0) = u(1) = 0. \tag{16}$$

$$-\Delta u = f \text{ in } \Omega = (0,1)x(0,1) \ ; \ u = 0 \text{ on } \partial\Omega. \tag{17}$$

Setting $\mathcal{H} = \mathcal{H}_0^1(\Omega)$, where $\mathcal{H}_0^1$ is the usual Sobolev space whose functions satisfy the Dirichlet boundary conditions, the weak forms of (16) and (17) are: find $u \in \mathcal{H}$ such that for all $v \in \mathcal{H}$,

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, dx \ = \ (f,v) = \int_\Omega fv \, dx.$$

First, consider the one dimensional Poisson equation (16). Let $N_j = (N_1+1)\cdot 2^{j-1}$ for some $N_1 \geq 1$ and set $h_j = (N_j+1)^{-1}$. Let $\mathcal{M}_j \subseteq \mathcal{H}$ denote the N_j-dimensional space of C^0 piecewise linear polynomials with respect to a uniform mesh of size h_j on Ω. The computational basis are the usual nodal basis. The matrices A_j are tridiagonal:

$$A_j = h_j^{-1} [-1, 2, -1].$$

As a smoother, we use the damped Jacobi scheme

$$B_j = 4h_j^{-1} I.$$

We can obtain the same matrix problem using central finite differences on the same mesh [19].

While we do not advocate using a multigrid algorithm to solve this problem, it is of theoretical interest because it can be completely analyzed. The eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\psi_i\}$ are

$$\lambda_i = 2h_j^{-1}(1-\cos(i\pi h_j)) \quad \text{and} \quad (\psi_i)_k = (2h_j)^{1/2}\sin(k\pi i h_j).$$

The two dimensional subspaces $\text{Span}<\psi_i, \psi_{N_j+1-i}>$ remain invariant under the coarse grid correction. Thus, with respect to the eigenvector basis, the process reduces to the study of 2x2 error propagation matrices. See Douglas [10] for more details.

Let $x_i = (1-\cos(i\pi h_j))/2$. With respect to the eigenvector basis the transformed matrices $\bar{A}$, $\bar{B}$, ... have the form

$$\bar{A}_i = 4h_j^{-1}\begin{bmatrix} x_i & 0 \\ 0 & 1-x_i \end{bmatrix}$$

$$\bar{B}_i = 4h_j^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\bar{R}_i = [\, 1-x_i \quad x_i \,]$$

$$\bar{S}_i = \begin{bmatrix} 1-x_i & 0 \\ 0 & x_i \end{bmatrix}$$

$$\bar{C}_i = \begin{bmatrix} x_i & -(x_i(1-x_i))^{1/2} \\ -(x_i(1-x_i))^{1/2} & 1-x_i \end{bmatrix}$$

$$\bar{S}_i^{m/2}\bar{C}_i\bar{S}_i^{m/2} = \begin{bmatrix} x_i(1-x_i)^m & -(x_i(1-x_i))^{m+1/2} \\ -(x_i(1-x_i))^{m+1/2} & x_i^m(1-x_i) \end{bmatrix}$$

The spectral radius $\rho(\bar{S}_i^{m/2}\bar{C}_i\bar{S}_i^{m/2})$ can be computed exactly:

$$\rho(\bar{S}_i^{m/2}\bar{C}_i\bar{S}_i^{m/2}) = x_i(1-x_i)^m + x_i^m(1-x_i).$$

Hence,

$$\rho(S^{m/2}CS^{m/2}) = \max_i \; x_i(1-x_i)^m + x_i^m(1-x_i).$$

$$\leq \sup_{x\in[0,1]} x(1-x)^m + x^m(1-x).$$

The part of the two dimensional invariant subspace corresponding to M_{j-1} is Span$<[1-x_i, x_i]^T>$ and the part corresponding to $M_{j-1}^{\perp}$ is Span$<[1, -1]^T>$. In this case, we can take $\alpha = 1$ and

$$\kappa = \sup_{v\in M_{j-1}^{\perp}} |||v|||_0^2 \,/\, |||v|||^2 = 2.$$

The corresponding estimate for $\rho(S^{m/2}CS^{m/2})$ obtained from Theorem 4 is

$$\rho(S^{m/2}CS^{m/2}) \leq \begin{cases} .5 & \text{if } m = 1 \\ 2\sup\limits_{x\in[0,1]} x(1-x)^m & \text{if } m > 1. \end{cases}$$

For m = 1 the estimate is sharp, while for large m it is an overestimate by a factor of 2.

The crossover point for the two forms of γ in Theorem 4 is predicted to be at $m \geq \kappa - 1 = 1$. However, direct computation shows $\rho(S^{m/2} C S^{m/2})$ behaves like 2^{-m} for $1 \leq m \leq 3$. For $1 \leq m \leq 5$ we have

	Two Levels,			W Cycle	V Cycle
m	ρ	γ	γ/ρ	$\bar{\gamma}$	$\hat{\gamma}$
1	.50000	.50000	1.0000	.66667	.66667
2	.25000	.29630	1.1852	.42105	.50000
3	.12500	.21094	1.6875	.26733	.40000
4	.083333	.16384	1.9661	.19594	.33333
5	.067088	.13414	1.9995	.15480	.28571

(γ is from Theorem 4, $\bar{\gamma}$ is from Corollary 8, and $\hat{\gamma}$ is from Theorem 7)

Now consider the two dimensional Poisson equation (17). This is a problem which we do advocate using a multigrid algorithm to solve. We discretize this problem by either central finite differences on a uniform mesh or by finite elements on a uniform triangulation using C^0 piecewise linear polynomials and the usual nodal basis functions. In either case, the matrices A_j are block tridiagonal [19]:

$$A_j = [-I, T, -I],$$

where $T = [-1, 4, -1]$ is tridiagonal. As a smoother, we use the damped Jacobi scheme

$$B_j = 4I.$$

Analagous to the one dimensional case, the problem can be reduced to the study of 4x4 matrices.

$$\bar{A}_{ij} = 4 \begin{bmatrix} 2-x_i-x_j & 0 & 0 & 0 \\ 0 & 1-x_i+x_j & 0 & 0 \\ 0 & 0 & 1+x_i-x_j & 0 \\ 0 & 0 & 0 & x_i+x_j \end{bmatrix}$$

$$\bar{R}_{ij} = \begin{bmatrix} (1-x_i)(1-x_j) & -(1-x_i)x_j & -x_i(1-x_j) & x_i x_j \end{bmatrix}$$

$\bar{S}_i$, $\bar{C}_i$, and $\bar{S}_i^{m/2}\bar{C}_i\bar{S}_i^{m/2}$ can be computed from these matrices above. Once again we can take $\alpha = 1$ and compute

$$\kappa = \sup_{v \in M_{j-1}^{\perp}} |||v|||_0^2 \,/\, |||v|||^2 = 4.$$

The resulting estimate for $\rho(S^{m/2}CS^{m/2})$ obtained from Theorem 4 is

$$\rho(S^{m/2}CS^{m/2}) \leq \begin{cases} (.75)^m & \text{if } 1 \leq m \leq 3 \\ 4 \sup\limits_{x \in [0,1]} x(1-x)^m & \text{if } m \geq 4. \end{cases}$$

The crossover point for the two forms of γ in Theorem 4 is predicted to be at $m = 3$. However, direct computation shows $\rho(S^{m/2}CS^{m/2})$ behaves like $(.75)^m$ for $1 \leq m \leq 7$. For $1 \leq m \leq 5$ we have

	Two Levels			W Cycle	V Cycle
m	ρ	γ	γ/ρ	$\bar{\gamma}$	$\hat{\gamma}$
1	.75000	.75000	1.0000	.80000	.80000
2	.56250	.56250	1.0000	.66667	.66667
3	.42188	.42188	1.0000	.57143	.57143
4	.31641	.32768	1.0356	.48739	.50000
5	.23730	.26792	1.1290	.36597	.44444

(γ is from Theorem 4, $\bar{\gamma}$ is from Corollary 8, and $\hat{\gamma}$ is from Theorem 7)

We note that for $1 \leq m \leq 3$, the V Cycle Theorem 7 estimates a faster convergence rate than the W Cycle Theorem 5.

Theorem 4 can yield reasonably sharp estimates provided good estimates of κ are available. However, obtaining good estimates in specific cases appears to be a hard problem. The last example shows how the existence of κ can be proven without necessarily obtaining a meaningful estimate of its size.

Consider the two dimensional linear elliptic boundary value problem

$$-\nabla\cdot(a\nabla u) + bu = f \text{ in } \Omega\,;\; u_n = 0 \text{ on } \partial\Omega, \tag{18}$$

where Ω is a polygonal domain in R^2. We assume that $a \in C^1(\bar{\Omega})$, $b \in C(\bar{\Omega})$, and that there exist positive constants a_{min}, a_{max}, b_{min}, b_{max}, such that

$$0 < a_{min} \leq a(x) \leq a_{max} \quad \text{and} \quad 0 < b_{min} \leq b(x) \leq b_{max} \quad \text{for all } x \in \bar{\Omega}.$$

$\mathcal{H}^k(\Omega)$ are the usual Sobolev spaces equipped with norms $\|\cdot\|_k$ and $\mathcal{H} = \mathcal{H}^1(\Omega)$. We seek a weak form solution of (18): find $u \in \mathcal{H}(\Omega)$ such that for all $v \in \mathcal{H}(\Omega)$,

$$a(u,v) = \int_\Omega a\nabla u \cdot \nabla v + buv \, dx = (f,v) = \int_\Omega fv \, dx.$$

We discretize (18) using a finite element formulation. Let T_j, $j \geq 1$, be a nested sequence of triangulations of Ω. Take T_1 to be a fixed triangulation. We construct T_j, $j > 1$, inductively: divide every $T \in T_{j-1}$ into 4 congruent triangles by pairwise connecting the midpoints of the edges. Let M_j be the space of C^0 piecewise linear polynomials associated with T_j. Because of the refinement scheme used, $N_j \sim 4^{j-1}N_1$ and $h_j = 2^{1-j}h_1$. The computational basis for M_j are the usual nodal basis and the smoother will again be the damped Jacobi scheme. If D_j is the diagonal of the stiffness matrix A_j, then

$$B_j = \tau_j D_j, \quad \text{where} \quad \tau_j = \max_{X \in R^N} X^T A_j X \,/\, X^T D_j X.$$

τ_j is bounded by a constant which is independent of j.. In order to prove the existence of κ, and thus establish Theorem 4, we will use the following three facts:

1. There exists a constant $0 < \alpha \leq 1$, such that for all $f \in \mathcal{H}^{1-\alpha}$ there exists a unique solution $u \in \mathcal{H}^{1+\alpha}$ and

$$\|u\|_{1+\alpha} \leq C_2 \|f\|_{\alpha-1}, \tag{19}$$

where $C_2 = C(a,b,\Omega)$.

2. The spaces M_j satisfy the following standard approximation property: for $v \in \mathcal{H}^{1+\alpha}$, $0 \leq \alpha \leq 1$,

$$\inf_{\chi \in M_j} \|v-\chi\|_0 + h_j\|v-\chi\|_1 \leq C_3 h_j^{1+\alpha}\|v\|_{1+\alpha}, \tag{20}$$

where $C_3 = C(\Omega, T_1)$.

3. Since the mesh is shape regular and quasi uniform, the discrete norms

$|||\cdot|||_{\theta,j} = |||\cdot|||_\theta$ and $\|\cdot\|_\theta$, $\theta\in[0,1]$, are comparable in the following sense: for $v\in M_j$,

$$h_j^{1-\theta}C_4^{-1}|||v|||_\theta \le \|v\|_\theta \le h_j^{1-\theta}C_4|||v|||_\theta, \tag{21}$$

where $C_4 = C(a,b,\Omega,T_1)$. Inequality (21) (or its equivalent) is relatively easy to prove for Jacobi-like smoothers, which explains their popularity in theoretical analyses of multigrid methods.

To prove the existence of κ, we will bound $|||v|||_{1-\alpha}$, $v\in M_{j-1}^{\perp}$, for $\alpha\in(0,1]$ as given by (19). First note that by (21),

$$|||v|||_{1-\alpha} \le C_4 h_j^{-\alpha}\|v\|_{1-\alpha}. \tag{22}$$

To estimate $\|v\|_{1-\alpha}$, we use a standard duality argument. Let $z\in\mathcal{H}$ be the solution of

$$a(z,w) = (\mu,w) \quad \text{for all } w\in\mathcal{H},$$

where $\mu\in\mathcal{H}^{\alpha-1}(\Omega)$. Take $w=v\in M_{j-1}^{\perp}$ to see that for any $\chi\in M_{j-1}$,

$$\begin{aligned}(\mu,v) &= a(z-\chi,v)\\ &\le |||z-\chi|||\cdot|||v|||\\ &\le C_3C_4h_{j-1}^{\alpha}\|z\|_{1+\alpha}|||v|||\\ &\le C_2C_3C_4h_{j-1}^{\alpha}\|\mu\|_{\alpha-1}|||v|||\end{aligned}$$

from which it follows that

$$\|v\|_{1-\alpha} \le C_2C_3C_4h_{j-1}^{\alpha}|||v|||.$$

Combining with (22) and using $h_{j-1} = 2h_j$ gives us

$$|||v|||_{1-\alpha} \le (2^{\alpha}C_2C_3C_4^2)\cdot|||v||| \tag{23}$$

and the corresponding estimate

$$\kappa = 2^{\alpha}C_2C_3C_4^2. \tag{24}$$

An estimate like (23) is the heart of many multigrid convergence proofs. The bound is of the "right" form, but the constant is usually rather pessimistic.

4 Acceleration

In this section, we investigate the effect accelerating the smoothing scheme has on the overall convergence rate. In particular, we consider the Chebyshev, conjugate gradient, and conjugate residual schemes as smoothers and compare the convergence rates of Algorithm MG using these smoothers with the rates proven in Section 2.

We want to take some initial guess $z_0 \in M_j$ to a final guess $z_m = S_m(z_0) \in M_j$ by replacing m steps of (4) by

$$b(z_\mu - z_{\mu-1}, v) = \tau_\mu^{-1}(g(v) - a(z_{\mu-1}, v) \quad \text{for all } v \in M_j,\ \tau_\mu \in (0,1],\ 1 \le \mu \le m.$$

If $e_0 = \sum_{i=1}^{N} C_i \psi_i$, then $e_m = \sum_{i=1}^{N} c_i \psi_i p_m(\lambda_i)$, where

$$p_m(x) = \prod_{\mu=1}^{m} (1 - \tau_\mu^{-1} x).$$

The error propagation matrix,

$$S_m = \prod_{\mu=1}^{m} (I - \tau_\mu^{-1} A^{1/2} B^{-1} A^{1/2}),$$

may be indefinite, so it is convenient to modify our proof of Theorem 4.

Theorem 9: Suppose there exist $\kappa \ge 1$ and $\alpha > 0$, both independent of j, such that for $u \in M_{j-1}^{\perp} \cap M_j$,

$$|||u|||_{1-\alpha} \le \kappa^{\alpha/2} |||u|||.$$

Then

$$|||C^{1/2}(S_m(C^{1/2}(v)))||| \le \gamma |||v|||, \quad \text{for all } v \in M_j,$$

where

$$\gamma = \kappa^{\alpha} \hat{f}(m,\alpha) \quad \text{and} \quad \hat{f}(m,\alpha) = \sup_{x \in [0,1]} |x^{\alpha} p_m(x)|.$$

Proof: The coarse grid corrector C is a symmetric orthogonal projector, so $C^{1/2} = C$. Thus, it is sufficient to show

$$|||C(S_m(C(v)))||| \le \gamma |||v|||, \quad \text{for all } v \in M_j.$$

Let $\bar{v} = S_m(C(v))$. Then

$$|||C(\bar{v})|||^2 = a(C(\bar{v}),C(\bar{v})) \tag{25}$$

$$= a(C(\bar{v}),\bar{v})$$

$$\leq |||C(\bar{v})|||_{1-\alpha}\cdot|||\bar{v}|||_{1+\alpha}$$

$$\leq \kappa^{\alpha/2}|||C(\bar{v})|||\cdot|||\bar{v}|||_{1+\alpha}$$

and

$$|||\bar{v}|||_{1+\alpha} \leq \hat{f}(m,\alpha)|||C(\bar{v})|||_{1-\alpha}$$

$$\leq \kappa^{\alpha/2}\hat{f}(m,\alpha)|||v|||.$$

Combining the last two inequalities gives the desired result.

QED

This proof does not extend in a straightforward fashion to the W and V Cycles since $\tilde{C}$ is not necessarily a projector. If we had norm estimates for the nonsymmetric operators S^mC (S_mC) and CS^m (CS_m), such extensions would be straightforward. While it is true that

$$(S^mC)^rS^m = S^m(CS^m)^r$$

$$= S^{m/2}(S^{m/2}CS^{m/2})^rS^{m/2}$$

$$= S^mC^{1/2}(C^{1/2}S^mC^{1/2})^{r-1}C^{1/2}S^m$$

so that convergence rates obtained for any one of S^mC, CS^m, $S^{m/2}CS^{m/2}$, or $C^{1/2}S^mC^{1/2}$ are asymptotically valid for all, a judicious choice of which operator norm to estimate can greatly simplify the argument.

We now consider the problem of picking the r_μ to minimize $\hat{f}(m,\alpha)$. For the important case of $\alpha = 1$ the problem can be solved analytically in terms of Chebyshev polynomials. In particular, the solution is $q_{m+1}(x)$, a multiple of $T_{m+1}(\cos\theta) = \cos(m+1)\theta$ mapped to an interval $[-\delta,1]$. The conditions that determine δ and the normalization constant are $q_{m+1}(0) = 0$ and $q'_{m+1}(0) = 1$. The solution is

$$q_{m+1}(x) = \{\cos[(m+1)\theta]\sin\eta\} \;/\; \{(m+1)(1+\cos\eta)\}, \qquad \text{where } \eta = \pi/(2m+2)$$

$$\delta = (1-\cos\eta) \;/\; (1+\cos\eta)$$

$$\cos\theta = (2x+\delta-1)\cdot(\delta+1)^{-1}$$

Finally,

$$\tau_\mu = (\cos((2\mu+1)\eta)+\cos\eta) \,/\, (1+\cos\eta), \qquad 1 \le \mu \le m$$

$$\hat{f}(m,1) \le \hat{q}_{m+1} = \sup_{x\in[0,1]} |q_{m+1}(x)| = \sin\eta \,/\, [(m+1)(1+\cos\eta)].$$

For large m,

$$\hat{q}_{m+1} \sim \pi\cdot(4m^2)^{-1} \qquad \text{while} \qquad f(m,1) \sim (em)^{-1}.$$

Even for small m the advantage is appreciable, as is illustrated below:

m	$f(m,1) \,/\, \hat{q}_{m+1}$
1	1.207
2	1.659
3	2.121
4	2.586
5	3.053

In most situations the object of Algorithm MG is to reduce the error by a factor of 4–10. Usually m and the number of iterations r of MG is small, e.g., between 1 and 4 each. By accelerating the smoothing iteration it may be possible to reduce the number of smoothing iterations or correction cycles required to achieve the desired error reduction. Even if the number is reduced by 1 or 2, this may be a substantial portion of the total work.

A related issue is whether a fixed sequence $\{\tau_\mu\}$ should be used or an adaptive acceleration scheme like either conjugate gradients or conjugate residuals. One worry is that an adaptive scheme might waste time reducing "smooth" components of the error and be ill suited to the specialized requirements of a smoothing iteration. Quite the contrary, these schemes have been found to be quite robust and cost effective.

We end this section by estimating the convergence rate when either the conjugate residual or conjugate gradient algorithm is used as a smoother.

Theorem 10: Suppose there exist $\kappa \ge 1$, independent of j, such that for $u \in M_{j-1}^{\perp} \cap M_j$,

$$|||u|||_0 \le \kappa^{1/2}|||u|||.$$

Let S_m^{CR} be generated by the conjugate residual algorithm. Then

$$|||C^{1/2}(S_m^{CR}(C^{1/2}(v)))||| \le \gamma_{CR}|||v|||,$$

where

$$\gamma_{CR} \leq \kappa \inf_{p_m(x),p_m(0)=1} \hat{f}(m,1)|||v|||$$

and $\hat{f}(m,1)$ is defined as in Theorem 9.

Proof: The proof follows the one for Theorem 9. As in (25) with $\alpha = 1$,

$$|||C(\bar{v})||| \leq \kappa^{1/2}|||\bar{v}|||_2, \qquad \text{where } \bar{v} = S_m^{CR}(C^{1/2}(v)).$$

Since the $|||\cdot|||_2$ norm is minimized by the conjugate residual algorithm,

$$\begin{aligned} |||\bar{v}|||_2 &\leq \inf_{\tilde{S}_m} |||\tilde{S}_m(C(v))||| \qquad (26) \\ &\leq \inf_{p_m(x),p_m(0)=1} \hat{f}(m,1)|||C(v)|||_0 \\ &\leq \kappa^{1/2} \inf_{p_m(x),p_m(0)=1} \hat{f}(m,1)|||C(v)||| \\ &\leq \kappa^{1/2} \inf_{p_m(x),p_m(0)=1} \hat{f}(m,1)|||v||| \end{aligned}$$

Combining the last two inequalities gives the result for conjugate residuals.

QED

The corresponding theorem for conjugate gradients is probably not optimal, however. For conjugate gradients, the left hand side of (26) must be evaluated in the one norm instead of the two norm. Our analysis leads to a suboptimal estimate.

Theorem 11: Suppose there exist $\kappa \geq 1$, independent of j, such that for $u \in M_{j-1}^{\perp} \cap M_j$,

$$|||u|||_0 \leq \kappa^{1/2}|||u|||.$$

Let S_m^{CG} be generated by the conjugate gradient algorithm. Then

$$|||C^{1/2}(S_m^{CG}(C^{1/2}(v)))||| \leq \gamma_{CG}|||v|||,$$

where

$$\gamma_{CG} \leq \kappa^{1/2} \inf_{p_m(x),p_m(0)=1} \hat{f}(m,1/2)|||v|||$$

and $\hat{f}(m,1/2)$ is defined as in Theorem 9.

References

[1] R. E. Alcouffe, A. Brandt, J. E. Dendy, Jr., J. W. Painter, The multi-grid methods for the diffusion equation with strongly discontinuous coefficients, *SIAM J. Sci. Stat. Comp.*, 2 (1981), pp. 430-454.

[2] G. P. Astrakhantsev, An iterative method of solving elliptic net problems, *Z. Vycisl. Mat. i. Mat. Fiz.*, 11 (1971), pp. 439-448.

[3] O. Axelsson and I. Gustafsson, *A Combination of Preconditioning and Multigrid Methods*, Technical Report 8120, Universty of Nijmegen, 1981.

[4] N. S. Bakhvalov, On the convergence of a relaxation method under natural constraints on an elliptic operator, *Z. Vycisl. Mat. i. Mat. Fiz.*, 6 (1966), pp. 861-883.

[5] R. E. Bank and T. Dupont, An optimal order process for solving elliptic finite element equations, *Math. Comp.*, 36 (1981), pp. 35-51.

[6] R. E. Bank and D. J. Rose, *Analysis of a Multi-Level Iterative Method for Nonlinear Finite Element Equations*, Technical Report 202, Department of Computer Science, Yale University, 1981.

[7] D. Braess and W. Hackbusch, A New Convergence Proof for the Multigrid Method Including the V Cycle, 1982.

[8] A. Brandt, private communication.

[9] A. Brandt, Multi-level adaptive solution to boundary-value problems, *Math. Comp.*, 31 (1977), pp. 333-390.

[10] C. Douglas, *Multi-Grid Algorithms for Elliptic Boundary-Value Problems*, PhD Thesis, Department of Computer Science, Yale University, 1982. Also, Computer Science Department Technical Report 223.

[11] R. P. Federenko, The speed of convergence of one iteration process, *Z. Vycisl. Mat. i. Mat. Fiz.*, 4 (1964), pp. 559-563.

[12] A. Greenbaum, *Analysis of a Multigrid Method as an Iterative Technique for Solving Linear Systems*, Technical Report UCRL-87912, Lawrence Livermore National Laboratory, 1982.

[13] W. Hackbusch, Convergence of multi-grid iterations applied to difference equations, *Math. Comp.*, 34 (1980), pp. 425-440.

[14] W. Hackbusch, Multi-grid convergence theory, W. Hackbusch and U. Trottenberg, ed., *Multigrid Methods*, Springer-Verlag, New York, 1982, pp. 177-219.

[15] W. Hackbusch, *On the convergence of a multi-grid iteration applied to finite element equations*, Technical Report 77-8, Mathematisches Institut, Universitat zu Koln, 1977.

[16] W. Hackbusch, On the convergence of multi-grid iterations, *Beit. zur Num. Math.*, 9 (1981), pp. 213-239.

[17] R. Kettler and J. A. Meijerink, *A Multigrid Method and a Combined Multigrid-Conjugate Gradient Method for Elliptic Problems with Strongly Discontinuous Coeeficients in General Domains*, Technical Report 604, Shell Oil Company, 1981.

[18] R. A. Nicolaides, On the l^2 convergence of an algorithm for solving finite element equations, *Math. Comp.*, 31 (1977), pp. 892-906.

[19] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, New York, 1973.

[20] J. R. Van Rosendale, *Rapid Solution of Finite Element Equations on Locally Refined Grids by Multi-Level Methods*, PhD Thesis, University of Illinois, 1980.

[21] R. Verfurth, A Multilevel Algorithm for Mixed Problems, Mathematisches Institut, Ruhr-Universtat, Bochum, 1982.

[22] P. Wesseling, Theorectical and practical aspects of a multigrid method, *SIAM J. Sci. Stat. Comp.*, 4 (1982), pp. 387-407.

[23] H. Yserentant, *The Convergence of Multi-Level Methods for Solving Finite Element Equations in the Presence of Singularities*, Technical Report 14, Rheinische-Westfalische Technische Hochschule, 1982.