Yale University
Department of Computer Science

A Generalized Multigrid Theory in the Style of Standard Iterative Methods

Craig C. Douglas

YALEU/DCS/TR-976
November 10, 1993

This work was supported in part by the Office of Naval Research (grant N00014-91-J-1576), Yale University, and the Research Division of International Business Machines.

A Generalized Multigrid Theory in the Style of Standard Iterative Methods

Craig C. Douglas*

July 5, 1993

Abstract

A basic error bound for multigrid methods is given in terms of residuals on neighboring levels. The terms in this bound derive from the iterative methods used as solvers on each level and the operators used to go from a level to the next coarser level. This bound is correct whether the underlying operator is symmetric or nonsymmetric, definite or indefinite, and singular or nonsingular. We allow any iterative method as a smoother (or rougher) in the multigrid cycle.

One of the advantages of this theory is that all of the parameters are available during execution of a computer program. Hence, adaptively changing levels can be achieved with certainty of success. This is particularly important for solving problems in which there is no known useful convergence analysis. Two problems arising in modeling combustion problems (flame sheets and laminar diffusion flames with full chemistry) are discussed.

While this theory is quite general, it is not always the correct approach when analyzing the convergence rate for a given problem. A discussion of when this theory is useful and when it is hopelessly nonsharp is provided.

1 Introduction

In this paper, linear problems

$$Au + f = 0, \quad u, f \in \mathcal{M},\ A \in \mathcal{L}(\mathcal{M}) \tag{1}$$

are solved using a nested space multigrid iterative method. The operator (matrix) A is typically the discretized (by finite elements, differences, or volumes) version of a partial differential equation.

*Department of Computer Science, Yale University, P. O. Box 2158, New Haven, CT 06520-2158. and Mathematical Sciences Department, IBM Research Division, Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598-0218 E-mail: *na.cdouglas@na-net.ornl.gov.*

Many multigrid papers begin by narrowing their scope just to problems which are symmetric and positive definite, symmetric and indefinite, or nonsymmetric and indefinite. In each case, these papers assume the problem is nonsingular, a set of smoothers is defined, and one or more specific multigrid algorithms are defined (e.g., a V, W, or F cycle). Finally, analysis is provided, usually in only one particular norm. For excellent traditional multigrid theoretical treatments of problems, see [1], [8], [17], and [21].

The analysis in this paper is correct whether the underlying operator is symmetric or nonsymmetric, definite or indefinite, and singular or nonsingular. Any iterative method is allowed as a smoother or rougher in the multigrid cycle. Any multigrid cycle is allowed, including adaptively chosen ones. Finally, the analysis is not dependent on any specific norm. In fact, different norms can be used on different levels (though doing this can produce misleading convergence rates).

The purpose of this paper is to provide a discussion on when to use the theoretical tool in [12] for analyzing nested space multilevel algorithms that are applied to any problem with any set of properties. The approach is simple enough to implement in computer programs without adding an excessive amount of overhead. There are similar procedures, known as aggregation-disaggregation methods (see [6]) when A is not derived from partial differential equations; the theory in this paper applies directly to these methods.

The basic correction multigrid algorithm is defined in the traditional recursive style in §2. This is then rephrased into a nonstandard form in §3. This leads to the two flavors of analysis in §4, one quite simple (and rarely sharp) and the other somewhat more complicated (and sharper). Examples and the practicality of this analysis are given in §5.

The theory in §4 depends on three sets of parameters which are available either dynamically or in advance. The basic convergence (divergence) result is not stated in a "nice" closed form, as is usual in multigrid papers, but in terms of the convergence rate of the next coarser level's rate.

2 A standard multilevel formulation

Suppose that there is a set of solution spaces $\{\mathcal{M}_k\}_{k=1}^{j}$, which approximate $\mathcal{M}=\mathcal{M}_j$ in some sense, and that $\dim(\mathcal{M}_k) \leq \dim(\mathcal{M}_{k+1})$. In the partial differential equation case, the $\mathcal{M}_k$ correspond to discrete problems on given grids (which are not necessarily nested). Then the multigrid approximation to (1) requires solving a sequence of problems of the form

$$A_k u_k + f_k = 0, \quad u_k, f_k \in \mathcal{M}_k, \; A_k \in \mathcal{L}(\mathcal{M}_k). \tag{2}$$

That there exist mappings between the neighboring spaces is assumed:

$$R_k : \mathcal{M}_k \to \mathcal{M}_{k-1} \quad \text{and} \quad P_{k-1} : \mathcal{M}_{k-1} \to \mathcal{M}_k$$

as well as mappings

$$Q_k : \mathcal{M}_k \to \mathcal{M}_{k-1} \quad \text{such that} \quad A_{k-1} = Q_k A_k P_{k-1}.$$

For partial differential equations, there are natural definitions of Q_k depending on the discretization method and the grids. See [12] for a more complete discussion of natural choices for Q_k.

Since for most applications, $\dim(\mathcal{M}_k) < \dim(\mathcal{M}_{k+1})$, Q_k cannot be inverted. However, the theory in §4 uses Q_k^{-1}. Thus, the interpretation of Q_k^{-1} must be explained. For finite element methods commonly used in practice, $\mathcal{M}_k$ represents a refinement of $\mathcal{M}_{k-1}$ and

$$Q_k = \begin{cases} I & \text{on } \mathcal{M}_{k-1}, \\ \\ 0 & \text{on } \mathcal{M}_k - \mathcal{M}_{k-1}; \end{cases}$$

this is true for both the h-version and the p-version of the methods.

The same relation holds for refinements in the finite difference case. Hence, Q_k^{-1} can be taken to be injection of $\mathcal{M}_{k-1}$ into $\mathcal{M}_k$ in each of the cases described; otherwise Q_k^{-1} should be taken as a pseudoinverse. Note that a Moore-Penrose type pseudoinverse may not be the best choice; a Drazin type pseudoinverse may be better.

For $k \geq 1$, assume there are iterative methods, represented by M_k and N_k, and possibly dependent upon the data (e.g., conjugate gradients), which are used as smoothers (or roughers) on level k before and after, respectively, the residual correction step (on level 1, note that there is never a residual correction step nor, usually, a smoother N_1).

In the multigrid literature, the term smoother has become synonymous with the direct or iterative methods M_k and N_k. The term was used in [4] to describe the effect of one or more iterations of a relaxation method on each of the components of the error vector. For many relaxation methods (e.g., SSUR and Gauss-Seidel), the norm of each error component is reduced each iteration; hence, the term smoother. For many other iterative methods (e.g., SSOR or conjugate gradients), while the norm of the error vector is reduced each iteration, the norm of some of the components of the error may grow each iteration; hence, the term rougher. For some iterative methods (e.g., Bi-CGSTAB), the norm of the error vector does not necessarily decrease each iteration, much less smooth all of the error components. The term smoother in the traditional multigrid sense will be used, even though it is technically wrong.

Standard multigrid analysis assumes the smoothers have the form

$$B_k(w_k^{\ell+1} - w_k^{\ell}) = f_k + A_k w_k^{\ell}, \quad \ell = 0, 1, \cdots, \ell_k,$$

where B_k corresponds to some scaled iterative method on each level k (e.g., symmetric Gauss-Seidel or conjugate gradients). This frequently leads to an analysis

which assumes a fixed ℓ_k throughout the multigrid iterations. Neither assumption is required in §4.

There are two principal variants of multigrid algorithms. One variant is composed of correction schemes, which start on some level j and only use the coarser levels k, $k < j$, for solving residual correction problems. The other variant is composed of nested iteration schemes, which begin computation on level 1 and work their way to some level j, using each level k, $k < j$, both to generate an initial guess for level $k+1$ and for solving residual correction problems. Analysis of nested iteration algorithms in the context of this paper can be found in [12]; more traditional analyses can be found in [2], [7], [8], and [17].

In this paper, only correction schemes are considered. Define a k-level (*standard*) correction multigrid scheme by

ALGORITHM MG$(k, \{\mu_\ell\}_{\ell=1}^{k}, x_k, f_k)$
(1) If $k = 1$, then solve $A_1 x_1 = f_1$ exactly or by smoothing
(2) If $k > 1$, then repeat $i = 1, \cdots, \mu_k$:
($2a$) Smoothing: $x_k \leftarrow M_k^{(i)}(x_k, f_k)$
($2b$) Residual Correction:
$x_k \leftarrow x_k + P_{k-1}$ MG$(k-1, \{\mu_\ell\}_{\ell=1}^{k-1}, 0, R_k(A_k x_k + f_k))$
($2c$) Smoothing: $x_k \leftarrow N_k^{(i)}(x_k, f_k)$
(3) Return x_k

This definition requires that $\mu_1 = 1$. Steps ($2a$) and ($2b$) are sometimes referred to as *pre-smoothing* and *post-smoothing*, respectively, in the literature.

Symmetric multigrid schemes assume that $M_k = N_k$. *Nonsymmetric* multigrid schemes usually assume that $N_k = I$, where I is the identity. However, it is computationally more efficient to assume $M_k = I$ since the residual on level $k-1$ is f_{k-1} and does not need to be recomputed. Only rarely is the complete algorithm analyzed.

The standard V and W cycles correspond to Algorithm MG$(j, \{1, \cdots, 1\}, \cdot, \cdot)$ and Algorithm MG$(j, \{1, 2, \cdots, 2, 1\}, \cdot, \cdot)$, respectively (the definition of the W cycle frequently causes confusion). The F cycle [5] corresponds to something "in between" the V and W cycles.

3 A nonstandard multilevel formulation

In this section, a subtle change is made to Algorithm MG, which produces a simplified analysis for multigrid methods.

To make the notation of this section consistent, a fake (extra) level $j+1$ is introduced. Define

$$\mathcal{M}_{j+1} = \mathcal{M}_j, \qquad P_j = R_{j+1} = Q_{j+1} = I, \qquad A_{j+1} = A_j,$$

and the initial residual on level $j+1$, z_{j+1}, by

$$A_{j+1}x_{j+1}^{(-1)} + f = z_{j+1}.$$

This transforms the problem on all computational levels to one of solving a residual correction problem instead of the real problem on the finest grid and residual correction problems on the coarser grids.

Associated with each level k is a norm $\|\cdot\|_k$, which can be arbitrary. The norms can be different on each level, though the usefulness of this is unclear. For simplicity, the subscript from the norm symbol will be dropped.

Define a k-level (*nonstandard*) correction multigrid scheme using parameters z_{k+1} (the residual on level $k+1$ at some step) and $x_k^{(-1)}$ (the initial guess for level k, which is normally 0, except at the finest level) by

ALGORITHM NSMG($k, z_{k+1}, x_k^{(-1)}$)

(1) Initial residual: $R_{k+1}z_{k+1} \in \mathcal{M}_k$

(2) Pre-Smoothing: $x_k^{(0)} = M_k^{(1)}x_k^{(-1)}$ such that
$A_k x_k^{(0)} + R_{k+1}z_{k+1} = z_k^{(0)}$, where $\|z_k^{(0)}\| \le \rho_k^{(1)}\|z_{k+1}\|$

(3) Let $\hat{x}_k^{(1)} = x_k^{(0)}$, $\hat{z}_k^{(1)} = z_k^{(0)}$, and $\gamma_1^{(1)} = 0$

(4) Repeat $i = 1, \cdots, \mu_k$

(4a) If $i > 1$, then

(4a1) Residual: $A_k x_k^{(i-1)} + R_{k+1}z_{k+1} = \hat{\theta}_k^{(i)}$

(4a2) Pre-Smoothing: $\hat{x}_k^{(i)} = M_k^{(i)}x_k^{(i-1)}$ such that
$A_k\hat{x}_k^{(i)} + R_{k+1}z_{k+1} = \hat{z}_k^{(i)}$,
where
$\|\hat{z}_k^{(i)}\| \le \rho_k^{(i)}\|\hat{\theta}_k^{(i)}\|$

(4b) If $k > 1$, then

(4b1) Correction: $\gamma_k^{(i)} = P_{k-1}\bar{x}_{k-1}^{(i)}$, where
$\bar{x}_{k-1}^{(i)} = \text{NSMG}(k-1, \hat{z}_k^{(i)}, 0)$
and
$A_{k-1}\bar{x}_{k-1}^{(i)} + R_k\hat{z}_k^{(i)} = \bar{z}_{k-1}^{(i)}$

(4c) Residual: $A_k(\hat{x}_k^{(i)} + \gamma_k^{(i)}) + R_{k+1}z_{k+1} = \theta_k^{(i)}$

(4d) Post-Smoothing: $x_k^{(i)} = N_k^{(i)}(\hat{x}_k^{(i)} + \gamma_k^{(i)})$ such that
$A_k x_k^{(i)} + R_{k+1}z_{k+1} = z_k^{(i)}$,
where
$\|z_k^{(i)}\| \le \epsilon_k^{(i)}\|\theta_k^{(i)}\|$

(5) Return $x_k^{(\mu_k)}$

Algorithm MG was defined in §2 in an intentionally imprecise manner. Algorithm NSMG is a precise, but nonstandard definition of Algorithm MG. The first smoothing reduces the norm of the residual on level k by a factor involving the norm of the residual on level $k+1$, which is nonstandard. For subsequent smoothings, this factor involves the norm of the residual on level k instead. The parameters $\{\mu_\ell\}$, which determine how many iterations of the multilevel algorithm to do on each level, can be considered either fixed or adaptively chosen during the course of computation.

Standard multigrid theory analyzes the case when a certain number of smoothing steps are used. This may be explicitly stated (e.g., [1]), or it may be phrased as to require the choice of a constant number of smoothing iterations such that some error reduction condition is satisfied (e.g., [7]). This is worst case analysis and rarely models the behavior seen in practice. However, it allows the proof of certain complexity results of optimal order.

The nonstandard formulation allows two interpretations of smoothing: first as the standard form, and second as fixing the factors $\epsilon_k^{(i)}$ and $\rho_k^{(i)}$ and letting the number of smoothing steps vary per iteration.

4 Analysis

In this section, assume that $\{\mathcal{M}_k\}$ is nested and analyze $z_j^{(i)}$ under minimal assumptions. Two flavors of analysis are considered. The first is a trivial analysis that should not be used when anything is really known about the problem. The second is an affine space decomposition analysis that is somewhat sharper than the first treatment.

The first result assumes only a simple property about each of the restrictions R_k: there exists a constant, $\delta_k \in \mathbb{R}$, such that

$$\|(I - Q_k^{-1}R_k)u\| \leq \delta_k \|u\|, \quad u \in \mathcal{M}_k. \tag{3}$$

Since normally $\dim(Range(Q_k^{-1})) < \dim(\mathcal{M}_k)$, $\delta_k \geq 1$. In many cases it is possible to choose norms for which $\delta_k = 1$ and which are meaningful for the underlying elliptic problem.

The problem is to determine conditions for $\{\rho_k^{(i)}, \epsilon_k^{(i)}\}$ in order to guarantee convergence of Algorithm NSMG. The results do not depend directly on properties of the A_k and f_k.

The basic theorem is as follows.

Theorem 1 *Assume that z_{j+1} is the residual on level $j+1 \geq 2$ and that the prolongation operators P_k, $1 \leq k \leq j$, are imbeddings and the inverse of the operator restrictions Q_k^{-1}, $2 \leq k \leq j+1$, are embeddings:*

$$P_k \equiv {}^{i}_{\mathcal{M}_k \to \mathcal{M}_{k+1}} \quad \textit{and} \quad Q_k^{-1} \equiv {}^{i}_{\mathcal{M}_{k-1} \to \mathcal{M}_k}. \tag{4}$$

Let

$$E_1^{(1)} = \epsilon_1^{(1)}\rho_1^{(1)} \quad \text{and} \quad E_k^{(\mu_k)} = \prod_{i=1}^{\mu_k}\left(\epsilon_k^{(i)}\rho_k^{(i)}\left[\delta_k + E_{k-1}^{(\mu_{k-1})}\right]\right), \quad k > 1.$$

Then,

$$\|Q_j^{-1} z_j^{(\mu_j)}\| \leq E_j^{(\mu_j)}\|z_{j+1}\|.$$

The proof of Theorem 1 is a double induction argument and can be found in [12].

Remark 1 *In some instances, different restriction operators $R_k^{(i)}$ are used during a multigrid cycle. Substituting $\delta_k^{(i)}$ for δ_k covers this case.*

Remark 2 *For the V cycle with $\epsilon_j^{(i)} = \epsilon_j$ and $\rho_j^{(i)} = \rho_j$, $j = 1, \cdots, k$, the definition of $E_k^{(1)}$, $k > 1$, simplifies to*

$$E_k^{(1)} = \sum_{\ell=1}^{k}\left(\prod_{m=0}^{\ell-1}\epsilon_{k-m}\rho_{k-m}\right)\delta_{k-\ell} + \rho_1\prod_{m=2}^{k}\epsilon_m\rho_m.$$

Remark 3 *When adaptively choosing when to change levels, the error term for the coarser level will be different each time a correction step is performed. Substituting $E_k^{(\mu_k^{(i)})}$ for $E_k^{(\mu_k)}$ covers this case.*

Remark 4 *For numerous problems, $\delta_k \geq 1$ guarantees that Theorem 1 is not sharp nor even realistic. See §5 for another interpretation of δ_k that is computationally useful since for specific residual vectors u in (3), δ_k can be much less than* 1.

Remark 5 *Many papers have been written analyzing multigrid using a variational point of view instead of an algebraic one. Rewrite (2) as*

$$\textit{find } u_k \in \mathcal{M}_k \textit{ such that} \quad a_k(u_k, v_k) + f_k(v_k) = 0, \quad \forall v_k \in \mathcal{M}_k.$$

Then Theorem 1 can be rewritten in a variational form.

Now consider an affine space analysis. Each space $\mathcal{M}_j$ is decomposed approximately into the parts which are corrected by the residual correction steps, and the parts which are relatively unaffected. This theory is considerably more complicated, but sharper than that in Theorem 1.

Each space $\mathcal{M}_j$ is assumed to be decomposable into a smooth part $\mathcal{S}_j$ and a rough part $\mathcal{T}_j$, e.g.,

$$\mathcal{M}_j = \mathcal{S}_j \oplus \mathcal{T}_j, \quad \text{where} \quad \mathcal{T}_j = \mathcal{M}_{j-1} \quad \text{and} \quad \mathcal{S}_j = \mathcal{M}_{j-1}^{\perp} \cap \mathcal{M}_j. \tag{5}$$

So, $\mathcal{S}_j$ contains the high frequency components and $\mathcal{T}_j$ contains the low frequency ones. Note that other definitions for $\mathcal{S}_j$ and $\mathcal{T}_j$ can be used.

Let $1 \leq k \leq j$. Assume that $v_k \in \mathcal{M}_k$. Let

$$|||v_k||| \equiv |||v_k|||_k \equiv \|v_k|_{\mathcal{S}_k}\|$$

and

$$< v_k > \equiv < v_k >_k \equiv \|v_k|_{\mathcal{T}_k}\|.$$

If v_k are w_k are the residuals before and after a post-smoothing iteration using N_k, and $\|w_k\|^2 = \epsilon_k^2 \|v_k\|^2$, then there exist $\epsilon_{k,S}$ and $\epsilon_{k,T}$ such that

$$\|w_k\|^2 = \epsilon_{k,S}^2 |||v_k|||^2 + \epsilon_{k,T}^2 < v_k >^2. \tag{6}$$

Similarly, if v_k are $\bar{w}_k$ are the residuals before and after a pre-smoothing iteration using M_k, and $\|\bar{w}_k\|^2 = \rho_k^2 \|v_k\|^2$, then there exist $\rho_{k,SS}$, $\rho_{k,ST}$, $\rho_{k,TT}$, and $\rho_{k,TS}$ such that

$$\begin{aligned} |||\bar{w}_k|||^2 &= \rho_{k,SS}^2 |||v_k|||^2 + \rho_{k,ST}^2 < v_k >^2 \quad \text{and} \\ < \bar{w}_k >^2 &= \rho_{k,TS}^2 |||v_k|||^2 + \rho_{k,TT}^2 < v_k >^2. \end{aligned} \tag{7}$$

As was noted at the end of §3, these parameters will probably only be bounded with estimates of some form.

The result here requires more precise knowledge than (3), namely that for any $u \in \mathcal{M}_k$, there exist constants $\delta_{k,S}$ and $\delta_{k,T} \in \mathbb{R}$ such that

$$|||(I - Q_k^{-1} R_k)u|||^2 \leq \delta_{k,S}^2 |||u|||^2 \quad \text{and} \quad < (I - Q_k^{-1} R_k)u >^2 \leq \delta_{k,T}^2 < u >^2.$$

The problem is to determine conditions for $\{\rho_{k,XY}^{(i)}, \epsilon_{k,X}^{(i)}\}$, $X, Y \in \{S, T\}$, in order to guarantee convergence of Algorithm NSMG. As before, the results do not depend directly on properties of the A_k and f_k.

A sharper convergence result than Theorem 1 is as follows.

Theorem 2 *Assume that z_{j+1} is the residual on level $j+1 \geq 2$ and that P_k, $1 \leq k \leq j$, and Q_k^{-1}, $2 \leq k \leq j+1$, satisfy* (4). *Let*

$$E_1^{(1)} = \epsilon_{1,S}^{(1)} \rho_{1,S}^{(1)} \equiv E_{1,SS}^{(1)} \quad \text{and} \quad E_{1,ST}^{(1)} = E_{1,TS}^{(1)} = E_{1,TT}^{(1)} = 0.$$

For $1 < k \leq j$, let

$$E_{k,SS}^{(i)} = \epsilon_{k,S}^{(i)} \left[\left(\delta_{k,S} + E_{k-1,SS}^{(\mu_{k-1})} \right) \rho_{k,SS}^{(i)} + E_{k-1,ST}^{(\mu_{k-1})} \rho_{k,ST}^{(i)} \right],$$

$$E_{k,TS}^{(i)} = \epsilon_{k,T}^{(i)} \left[\left(\delta_{k,T} + E_{k-1,TT}^{(\mu_{k-1})} \right) \rho_{k,TS}^{(i)} + E_{k-1,TS}^{(\mu_{k-1})} \rho_{k,SS}^{(i)} \right],$$

$$E_{k,TT}^{(i)} = \epsilon_{k,T}^{(i)} \left[\left(\delta_{k,T} + E_{k-1,TT}^{(\mu_{k-1})} \right) \rho_{k,TT}^{(i)} + E_{k-1,TS}^{(\mu_{k-1})} \rho_{k,ST}^{(i)} \right],$$

and

$$E_{k,ST}^{(i)} = \epsilon_{k,S}^{(i)} \left[\left(\delta_{k,S} + E_{k-1,SS}^{(\mu_{k-1})} \right) \rho_{k,ST}^{(i)} + E_{k-1,ST}^{(\mu_{k-1})} \rho_{k,TT}^{(i)} \right].$$

Then,

$$\|Q_j^{-1} z_j^{(\mu_j)}\| \leq \prod_{i=1}^{\mu_j} \max\left\{ E_{j,SS}^{(i)} + E_{j,TS}^{(i)},\ E_{j,ST}^{(i)} + E_{j,TT}^{(i)} \right\} \|z_{j+1}\|. \tag{8}$$

The proof of (8) is a double induction argument and can be found in [12].

Remark 6 *For a symmetric multilevel algorithm (see §2), all of the terms in Theorem 2 exist. It is possible to to see that whenever an individual term is large, there is another term multiplying it that is small.*

Remark 7 *For nonsymmetric multilevel algorithms, the expressions simplify since some of the individual terms are either* 0 *or* 1.

Remark 8 *For simple enough the* $\delta_{j,T} \approx 0$ *and* $\delta_{j,S} \approx 1$.

Special care is required when using this theory since it is, in some sense, too general. It is quite easy to calculate various terms in the two theorems using incompatible norms, resulting in nonsensical results.

5 Examples

In this section, δ_k is computed for several examples. The first is for Dirichlet problems on $\mathbb{R}^2$ with simple, but not entirely trivial meshes. While the estimates are rather pessimistic, some advice is offered on practical uses of the simple theory in §4. Next, an example is presented where Theorem 1 is sharp. Finally, two problems arising in attempting to numerically simulate flames are examined.

Assume that for each k, $k = 1, \cdots, j$, the spaces $\mathcal{M}_k$ has a bilinear hat function basis over uniform squares of side length h_k. This does not imply that the domain Ω is either rectangular or convex, just polygonal (possibly with holes) with boundary segments either parallel to the axes or inclined 45° to the axes (which requires appropriate modifications to some of the basis functions).

Set

$$D_{ij} = \{(i+1, j), (i-1, j), (i, j+1), (i, j-1)\}$$

and

$$\hat{D}_{ij} = \{(i+1, j+1), (i+1, j-1), (i-1, j+1), (i-1, j-1)\}.$$

Let ${R_k}^{(9)} v_{ij}$ be the following weighted sum of v_{ij} and its eight neighbors from level k:

$${R_k}^{(9)} v_{ij} = \frac{1}{4}\left[v_{ij} + \frac{1}{2} \sum_{(k,\ell)\in D_{ij}} v_{k\ell} + \frac{1}{4} \sum_{(k,\ell)\in \hat{D}_{ij}} v_{k\ell} \right]$$

We approximate $\delta_k^{(9)} = \delta_k({R_k}^{(9)})$ using a piecewise bilinear hat function v on level $k-1$ which is centered at some point $(i+1, j+1)$ on level k. Note that,

if $v_{ij} = (-1)^{i+j}$, then $R_k v_{ij} = 0$ at any interior point of the $(k-1)$-grid. Thus, $\delta_k \geq 1$; since R_k satisfies a maximum principal, it then follows that

$$\|(I - Q_k^{-1} R_k{}^{(9)})v\|_{\ell^\infty} \leq \|v\|_{\ell^\infty}$$

and that

$$\delta_k^{(9)} = 1.$$

Let $R_k{}^{(5)} v_{ij}$ be the following weighted sum of v_{ij} and its four neighbors from level k:

$$R_k{}^{(5)} v_{ij} = \frac{1}{4}\left[2v_{ij} + \frac{1}{2}\sum_{(k,\ell)\in D_{ij}} v_{k\ell}\right]$$

Again, the same argument shows that, with respect to the ℓ^∞,

$$\delta_k^{(5)} = 1.$$

If there are boundary elements associated with the edges at 45° to the axes, $R_k{}^{(9)}$ and $R_k{}^{(5)}$ can be mixed to form R_k.

Besides motivating the affine space analysis, the theory of this section can actually be used in computer programs to adaptively change the parameter choices on coarser levels k (μ_k and the number of iterations in the smoothers). Consider Laplace's equation on the unit interval, two levels, a uniform mesh, a central difference discretization, linear interpolation and projection, and one Jacobi iteration as the smoother. Sharp theory says that the convergence rate is bounded by 0.5. In a strictly nonrigorous exercise, 5000 randomly chosen problems were generated. In theory, $\delta_2^{(3)} = 1$, where $\delta_2^{(3)}$ is derived using a three point restriction operator R_2. However, for individual residual vectors v, the following was calculated:

$$\delta(v) = \frac{\|(I - Q_2^{-1} R_2)v\|}{\|v\|}.$$

The following was observed.

Statistic	$\delta(v)$
Minimum	0.3444
Maximum	0.9312
Average	0.7126

Further, there was a direct correlation between the size of the estimated $\delta(v)$ and the actual error reduction produced by one multigrid iteration.

Now consider the affine space analysis. Assume that only post smoothing is performed; this causes many of the terms in Theorem 2 to be either 1 or 0. In this case, Theorem 2 predicts that the convergence rate is bounded by 0.5, which is sharp. Unfortunately, Theorem 2 predicts an overly pessimistic convergence rate

when two post smoothing steps are used (c.f., [1] which gets the right bound in both cases).

For some problems, multigrid with particular smoothers is known to be a terrible method. For example, let $q \geq 5$ in

$$\begin{cases} -10^q u_{xx} - 10^{-q} u_{yy} &= f \text{ in } (0,1)^2, \\ u &= 0 \text{ on } \partial(0,1)^2, \end{cases}$$

and choose a central difference discretization on a uniform mesh and Jacobi as the smoother. Then the coarse grid corrections do not necessarily improve the approximation to the solution. In this case, Theorem 1 actually is sharp. (The fix to making multigrid work well for this problem is to use either a line relaxation or a conjugate gradient method as the smoother or rougher.)

The examples given so far were not the of interest to the authors of [12] when this theory was developed, however. Two problems which are currently being studied arise in numerical simulation of flames. These are complicated nonlinear coupled partial differential equations which are amenable to solution by multigrid methods provided that the right solvers are used on each level. The first is a flame sheet model (see [13]) while the second is a laminar, axisymmetric diffusion flame model (see [16]).

In the flame sheet model, the chemical reactions are described with a single one step irreversible reaction corresponding to infinitely fast conversion of reactants into stable products. This reaction is assumed to be limited to a very thin exothermic reaction zone located at the locus of stoichiometric mixing of fuel and oxidizer, where temperature and products of combustion are maximized. To further simplify the governing equations, one neglects thermal diffusion effects, assumes constant heat capacities and Fick's law for the ordinary mass diffusion velocities, and takes all the Lewis numbers equal to unity. With these approximations, the energy equation and the major species equations take on the same mathematical form and by introducing Schvab-Zeldovich variables, one can derive a source free convective-diffusive equation for a single conserved scalar. Although no information can be recovered about minor or intermediate species in the flame sheet limit, the temperature and the stable major species profiles in the system can be obtained from the solution of the conserved scalar equation coupled to the flow field equations. Further, the location of the physical spatially distributed reaction zone and its temperature distribution can be adequately predicted by the flame sheet model for many important fuel-oxidizer combinations and configurations. Since being studied as a means of obtaining an approximate solution to use as an initial iterate for a one dimensional detailed kinetics computation in [19], flame sheets have been routinely employed to initialize multidimensional diffusion flames.

A schematic of the physical configuration is given in Figure 1 (though not drawn to scale). It consists of an inner cylindrical fuel jet (radius R_I =0.2cm), an outer co-flowing annular oxidizer jet (radius R_O =2.5cm) and a dead zone extending to R_{max} =7.5cm. The inlet velocity profile of the fuel and oxidizer are

a plug flow of 35cm/s. This yields a typical value for the Reynolds number of 550. Further, the flame length is approximately L_f =3cm and the length of the computational domain is set to L =30cm. Although the fuel and oxidizer reservoirs are at room temperature (300°Kelvin), we need to assume, in the flame sheet model, that the temperature already reaches the peak temperature value along the inlet boundary at $r = R_I$. This peak temperature is estimated for a methane-air configuration to be 2050°K. Hence, the inlet profile of the conserved scalar, $S^0(r)$, is specified in such a way that the resulting temperature distribution blends the room temperature reservoirs and the peak temperature by means of a narrow Gaussian centered at R_I. The narrowness of the Gaussian profile has a relevant influence on the calculated flame length, so that its parameters have to be determined appropriately.

A damped Newton multilevel solver is used (see [3] and [18]). Due to the model used, nonstaggered grids can be used, though they are tensor product grids with quite variable mesh spacings. The linear problems solved on each level are 36 point operators. We found that GMRES with a Gauss-Seidel preconditioner was a very good solver for each level. The code uses a left preconditioned residual norm to determine when the solutions are adequate. In calculating $\delta_k^{(i)}$ in this norm, we found it to be in the interval $[10^6, 10^8]$ frequently. This required that the ϵ's and ρ's be quite small in order to achieve convergence. However, $\delta_k^{(i)} \ll \|z_{k+1}\|$ so that this is not really an imposition. Even so, we saw speed ups of a factor of 10.5 on an IBM RS6000-560 workstation over the unigrid solution approach (see [13]).

While $\delta_k^{(i)}$ was reduced dramatically by using a semi-coarsening approach, the overall run time increased by 50% over the traditional multigrid approach.

We used a damped Newton multilevel approach instead of a full approximation scheme (see [20]) because experiments us to believe that in the full chemistry case, FAS will be too expensive.

The second flame numerical simulation is of a laminar, axisymmetric, methane-air diffusion flame using nonlinear damped Newton multigrid (see [16]). The physical configuration is based on an inner cylindrical fuel stream surrounded by a coflowing oxidizer jet and the inlet velocities are high enough to produce a lifted flame with a triple flame ring structure at its base. Computationally, we solve the total mass, momentum, energy, and species conservation equations with complex transport and finite rate chemistry submodels. The velocity field is predicted using a vorticity-velocity formulation and the governing partial differential equations are discretized on a nonstaggered grid. The numerical solution involves a pseudo transient process and a steady-state Newton iteration combined with nonlinear damped Newton multigrid. Coarse grid information is used to provide initial starting estimates for the Newton iteration on the finest level and also to form correction problems, thus yielding significant savings in the execution times.

The physical configuration consists of an inner methane-nitrogen jet (with radius 0.2cm), an air coflow (with radius 2.5cm), and the computational domain is $[0, 7.5] \times [0, 30]$ (all units are centimeters). The temperature and species mass

fractions values for the surrounding air are the same as the ones for the dead zone. This physical configuration was chosen because experimental data and a numerical solution using primitive variables were already available for this problem.

Once again, a variable width tensor product set of grids was used. Due to the high number of chemical species in the calculation, the discrete Jacobians were 270 point operators. In the left preconditioned norm, $\delta_k^{(i)}$ was frequently in the interval $[10^6, 10^{10}]$. However, $\delta_k^{(i)} \ll \|z_{k+1}\|$ so that this is not really an imposition. Still, a factor of 9.7 speed up was achieved on a 57×73 fine grid over a unigrid approach. In this example, $\delta_k^{(i)}$ was not reduced dramatically by using a semi-coarsening approach.

6 Multiple coarse grid methods

In [12], the Theorems 1 and 2 are extended to a multiple coarse space model. In this case, there are multiple δ's for each level, the quantity depending on the number of coarse level correction problems that are associated with each level.

While the theorems of §4 may not be satisfactory for simple problems, the multiple coarse space theory is for these problems. This style of analysis is much more accurate due to the fact that we can show that the δ's can be quite small, including being 0 for the case of the domain reduction method (see [9], [14], and [15]).

7 Conclusions

It is possible to prove a convergence result for multigrid and aggregation-disaggregation methods with minimal knowledge about the problem. By treating multigrid as a simple iterative method, almost nothing needs to be known about the grids, solution spaces, linear systems of equations, iterative methods used as smoothers (or roughers), restriction and prolongation operators, or the norms used on each level.

Being able to prove such a result is much easier than showing that it is useful all of the time. In fact, this theory is normally not sharp enough to satisfy theoreticians. It should be used in computational settings in which almost nothing is known about the convergence rate a priori.

One of the advantages of this theory is that all of the parameters are available during execution of a computer program. Hence, adaptively changing levels can be achieved with certainty of success.

Code availability

A series of codes, Madpack (see [11] and its references), are available from MGNet [10] which are compatible with the philosophy applied here and with the earlier theory in [8].

Acknowledgments

I am indebted to Professor Jim Douglas, Jr., Alexandre Ern, and Professor Mitchell Smooke for helpful discussions.

References

[1] R. E. Bank and C. C. Douglas. Sharp estimates for multigrid rates of convergence with general smoothing and acceleration. *SIAM J. Numer. Anal.*, 22:617–633, 1985.

[2] R. E. Bank and T. Dupont. An optimal order process for solving elliptic finite element equations. *Math. Comp.*, 36:35–51, 1981.

[3] R. E. Bank and D. J. Rose. Analysis of a multilevel iterative method for nonlinear finite element equations. *Math. Comp.*, 39:453–465, 1982.

[4] A. Brandt. Multi–level adaptive solution to boundary–value problems. *Math. Comp.*, 31:333–390, 1977.

[5] A. Brandt. Guide to multigrid development. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, pages 220–312. Springer–Verlag, New York, 1982.

[6] F. Chatelin and W. L. Miranker. Acceleration by aggregation of successive approximation methods. *Lin. Alg. Appl.*, 43:17–47, 1982.

[7] C. C. Douglas. *Multi–grid algorithms for elliptic boundary–value problems.* PhD thesis, Yale University, May 1982. Also, Computer Science Department, Yale University, Technical Report 223.

[8] C. C. Douglas. Multi–grid algorithms with applications to elliptic boundary–value problems. *SIAM J. Numer. Anal.*, 21:236–254, 1984.

[9] C. C. Douglas. A tupleware approach to domain decomposition methods. *Appl. Numer. Math.*, 8:353–373, 1991.

[10] C. C. Douglas. MGNet: a multigrid and domain decomposition network. *ACM SIGNUM Newsletter*, 27:2–8, 1992.

[11] C. C. Douglas. Implementing abstract multigrid or multilevel methods. Technical Report YALEU/DCS/TR-952, Department of Computer Science, Yale University, New Haven, 1993. To appear in the Proceedings of the Sixth Copper Mountain Multigrid Conference, NASA, 1993.

[12] C. C. Douglas and J. Douglas. A unified convergence theory for abstract multigrid or multilevel algorithms, serial and parallel. *SIAM J. Numer. Anal.*, 30:136–158, 1993.

[13] C. C. Douglas and A. Ern. Numerical solution of flame sheet problems with and without multigrid methods. *submitted to Advances in Comp. Math.*, 1993. Also available as Yale University Department of Computer Science Report YALEU/DCS/TR-955, New Haven, CT, 1993.

[14] C. C. Douglas and J. Mandel. A group theoretic approach to the domain reduction method. *Computing*, 48:73–96, 1992.

[15] C. C. Douglas and W. L. Miranker. Constructive interference in parallel algorithms. *SIAM J. Numer. Anal.*, 25:376–398, 1988.

[16] A. Ern, C. C. Douglas, and M. D. Smooke. Numerical simulation of laminar diffusion flames with multigrid methods. Technical Report YALEU/DME/TR-xxx, Department of Mechanical Engineering, Yale University, New Haven, 1993. In preparation.

[17] W. Hackbusch. *Multigrid Methods and Applications.* Springer–Verlag, Berlin, 1985.

[18] W. Hackbusch and A. Reusken. Analysis of a damped nonlinear multilevel method. *Numer. Math.*, 55:225–246, 1989.

[19] D. E. Keyes and M. D. Smooke. Flame sheet starting estimates for counterflow diffusion flame problems. *J. Comput. Phys.*, 73:267–288, 1987.

[20] C. Liu, Z. Liu, and S. F. McCormick. Multigrid methods for numerical simulation of laminar diffusion flames. *AIAA*, 93-0236:1–11, 1993.

[21] P. Wesseling. *An Introduction to Multigrid Methods.* John Wiley & Sons, Chichester, 1992.

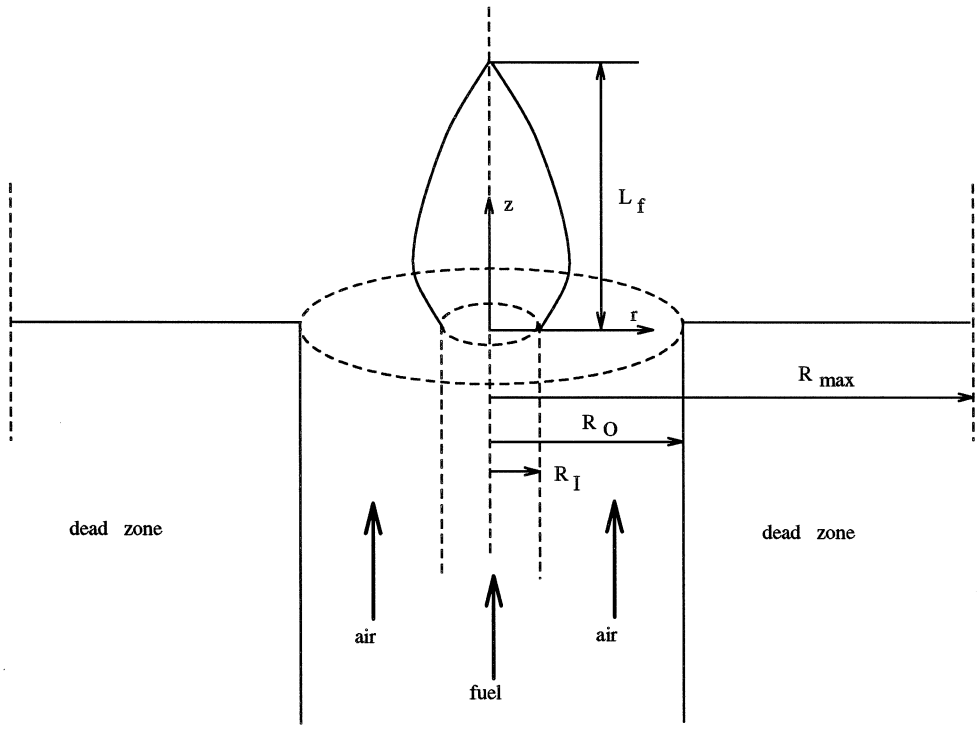

Figure 1: Flame sheet physical configuration