A UNIFIED CONVERGENCE THEORY FOR ABSTRACT MULTIGRID OR MULTILEVEL ALGORITHMS, SERIAL AND PARALLEL*

CRAIG C. DOUGLAS† AND JIM DOUGLAS, JR.‡

Abstract. Multigrid methods are analyzed in the style of standard iterative methods. A basic error bound is derived in terms of residuals on neighboring levels. The terms in this bound derive from the iterative methods used as smoothers 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.

While standard multigrid error analysis typically assumes a specific multigrid cycle (e.g., a V, W, or F cycle), analysis for arbitrary multigrid cycles, including adaptively chosen ones, is provided. This theory applies directly to aggregation-disaggregation methods used to solve systems of linear equations.

Key words. multigrid, aggregation, disaggregation

AMS(MOS) subject classifications. 65N15, 65N10

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.

Multigrid methods combine scaled iterative methods (called *smoothers*) with iterative residual correction on coarser grids to reduce the error on a given (fine) grid. There are similar procedures, known as aggregation-disaggregation methods (see [14], [22], [24], [33], [40] and [46]), when A is not derived from partial differential equations; our theory applies directly to these methods.

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.

Unlike previous papers in the multilevel and multigrid field, the analysis in this paper 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. We allow any multigrid cycle including adaptively chosen ones. Finally, the analysis is not dependent on any specific norm. In fact, different norms can be used on different levels.

The purpose of this paper is to provide a theoretical tool for analyzing nested space multilevel algorithms that are applied to any problem with any set of properties. This

* Draft of March 23, 1992.

† Mathematical Sciences Department, IBM Research Division, Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598-0218 and Department of Computer Science, Yale University, P. O. Box 2158, New Haven, CT 06520-2158. E-mail: *bells@watson.ibm.com* or *douglas-craig@cs.yale.edu*.

‡ Department of Mathematics, Purdue University, Mathematical Sciences Building, West Lafayette, IN 47907. E-mail: *douglas@math.purdue.edu*.

is a *unified* approach to multilevel theory. The approach is simple enough to implement in computer programs without adding an excessive amount of overhead.

We define the basic multigrid algorithms in the traditional recursive style in §2. We then rephrase one of these into a nonstandard form. This leads to the (usually nonsharp) analysis of §3.1 and the examples given in §3.2. The practicality of this analysis is also discussed.

We refine the analysis in §3.1 in §4.

Standard multigrid which is parallelized by domain decomposition is covered by the analysis of §3 and §4. After all, this case is merely a block iterative method used for smoothing.

We define a parallel multigrid algorithm in §5 for the case where multiple coarse spaces are employed. A number of different parallel multigrid methods are covered by this formulation (see [22], [27], and [45]). The simple analysis of §3.1 is extended to this case and this analysis is shown to be sharp for an example method.

The theory in §3 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. Multigrid Algorithms.

2.1. Standard Multilevel Formulation. Suppose that we have 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

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

We assume that there exist mappings between the neighboring spaces:

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

We also that assume there are mappings

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

For partial differential equations, there are natural definitions of Q_k depending on the discretization method and the grids. For finite differences, consider Fig. 1; then let Q_k preserve the nodal values in $\mathcal{M}_{k-1}$.

For finite elements, a number of choices exist. First consider grids with square or rectangular elements. Bilinear basis functions are related to Fig. 1; Q_k again preserves the nodal values related to $\mathcal{M}_{k-1}$ and sets up bilinear functions over the coarse grid. Biquadratics are related to the left picture in Fig. 1. Again, keep the nodal values for the nodes in the coarse grid. Note that some vertex nodes map down to midside values and the interior point value in the coarse grid.

Now consider grids with triangular elements. Linear basis functions are related to Fig. 2, and Q_k preserves the nodal values related to $\mathcal{M}_{k-1}$. Quadratic basis functions are related to Fig. 3 and vertex values in the fine grid go to vertex and midside values in the coarse grid. Cubic basis functions are related to Fig. 4. Note that one triangle in the coarse grid refines to nine triangles in the fine grid, not four. Again, only vertex values in the fine grid are kept to give all node values in the coarse grid.

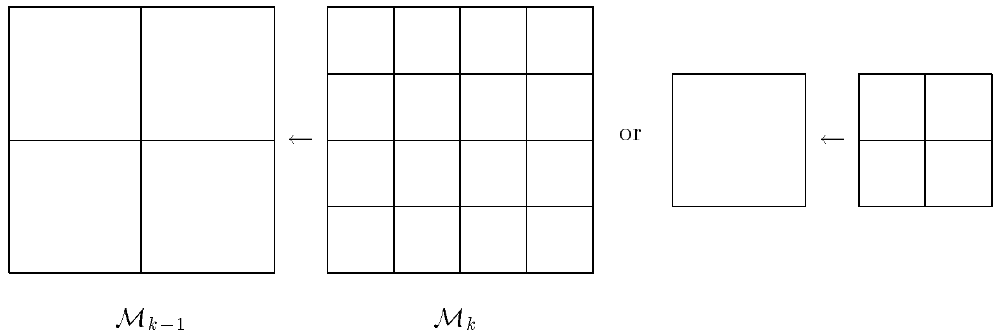

Fig. 1. *Square grid elements*

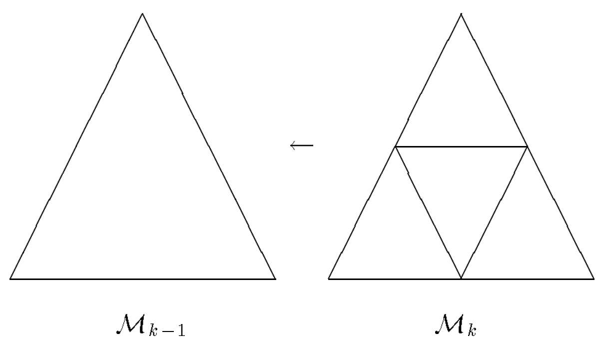

Fig. 2. *Triangular linear elements*

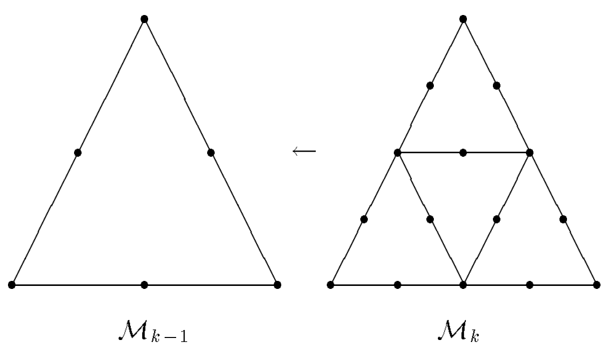

Fig. 3. *Triangular quadratic elements*

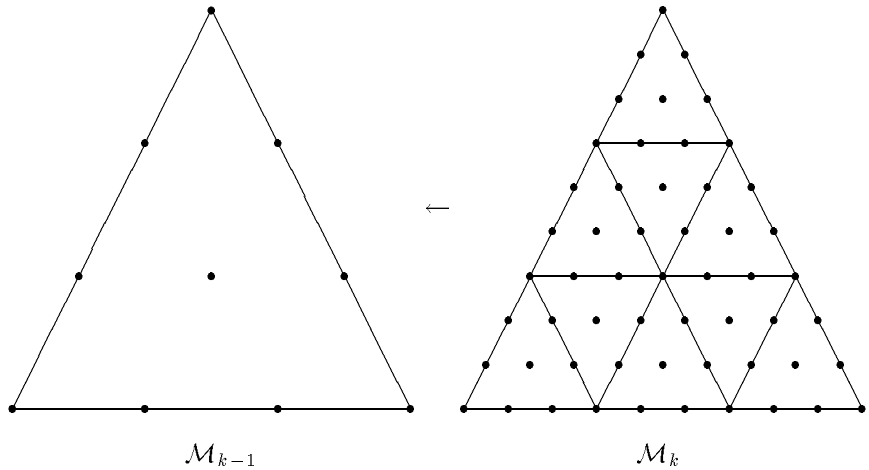

FIG. 4. *Triangular cubic elements*

Quite similar concepts can be used for defining Q_k for the p-version of the finite element method and for three dimensional problems.

For $k \geq 1$, we 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, we 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 iterative methods M_k and N_k. However, it is a term which has been abused frequently by many authors, including one of us. The term was used in [9] to describe the effect of one iteration of an iterative 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 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. We will use the term smoother in the traditional multigrid sense, 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 leads to an analysis which assumes a fixed ℓ_k throughout the multigrid iterations. We do not require either assumption in §3.

There are two principal variants of multigrid algorithms. The first 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. Define a k-level (*standard*) correction multigrid scheme by

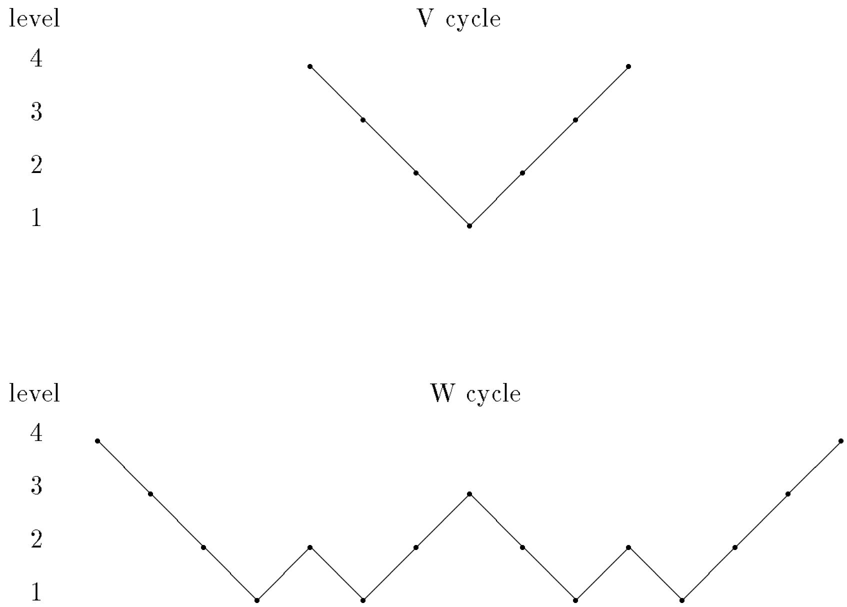

FIG. 5. *V and W cycles*

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:
$$x_k \leftarrow x_k + P_{k-1}\ \mathrm{MG}(k-1, \{\mu_\ell\}_{\ell=1}^{k-1}, 0, R_k(A_k x_k + f_k))$$
(2c) Smoothing: $x_k \leftarrow N_k^{(i)}(x_k, f_k)$
(3) Return x_k

This definition requires that $\mu_1 = 1$.

Symmetric multigrid schemes assume that $M_k = N_k$ (see [3], [5]–[8], [11], [12], [34], [39], and [49]) and *nonsymmetric* multigrid schemes usually assume that $N_k = I$, where I is the identity (see [1], [2], [4], [6], [9], [25]–[32], [38], [41], [42], [47], [48], [50]). However, it is computationally more efficient to assume $M_k = I$ (see [15]) 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 (see [15]–[17], [35]–[37], and [43]).

The second multigrid 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. Define a k-level (*standard*) nested iteration multigrid scheme by

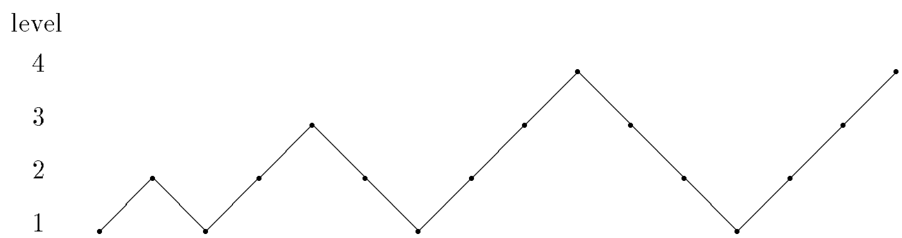

FIG. 6. *Nested Iteration (NI) V cycle*

ALGORITHM NI($j, \{\mu_k\}_{k=1}^{j}, x_1, \{f_k\}_{k=1}^{j}$)

(1) For $k = 1, \cdots, j$, do

(1a) If $k > 1$, then $x_k \leftarrow P_{k-1} x_{k-1}$

(1b) $x_k \leftarrow$MG($k, \{\mu_\ell\}_{\ell=1}^{k}, x_k, f_k$)

(2) Return x_j

Analyses of Algorithm NI can be found in [4], [15]–[17] and [30].

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 [10] corresponds to something "in between" the V and W cycles. Full Multigrid corresponds to Algorithm NI($j, \{\{1, \cdots, 1\}, \cdot, \cdot\}$). Figures 5 and 6 illustrate some of these cycles.

2.2. Nonstandard Multilevel Formulation. In this section, we make a subtle change to Algorithm MG, which allows us to provide a greatly simplified theory for multigrid methods.

To make the notation of this section consistent, we introduce a fake (extra) level $j + 1$. (This is the mother of all problems.) 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

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

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, we will drop the subscript from the norm symbol.

Define a k-level (*nonstandard*) correction multigrid scheme using parameters

z_{k+1}	The residual on level $k + 1$ at some step.
$x_k^{(-1)}$	The initial guess for level k; this 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) Smoothing: $x_k^{(0)} = M_k^{(1)} x_k^{(-1)}$ such that

$$A_k x_k^{(0)} + R_{k+1}z_{k+1} = z_k^{(0)}, \text{ where } \|z_k^{(0)}\| \le \rho_k^{(1)} \|z_{k+1}\|$$

(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) 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) 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 §2.1 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., [3]), 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., [15]). 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.

3. Simple Analysis.

3.1. Theory. In this section, we assume that $\{\mathcal{M}_k\}$ is nested and analyze $z_j^{(i)}$ under minimal assumptions. Our results assume 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 for most applications, $\dim(\mathcal{M}_k) < \dim(\mathcal{M}_{k+1})$, Q_k cannot be inverted; thus, we must explain the interpretation of Q_k^{-1}. In the examples for finite element methods discussed in §2.1, $\mathcal{M}_k$ represented 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 was true for both the h-version and the p-version of the method.

The same relation holds for refinements in the finite difference case. Hence, we can take Q_k^{-1} 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. We note that a Moore-Penrose type pseudoinverse may not be the best choice; a Drazin type pseudoinverse may be better.

Since $\dim(Range(Q_k^{-1})) < \dim(\mathcal{M}_k)$, $\delta_k \geq 1$. 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:*

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

Let

$$E_1^{(1)} = \epsilon_1^{(1)} \rho_1^{(1)} \quad \text{and} \quad E_k^{(\mu_k)} = \prod_{i=1}^{\mu_k} \left(\epsilon_k^{(i)} \rho_k^{(i)} \left[\delta_k + E_{k-1}^{(\mu_{k-1})} \right] \right), \quad k > 1. \tag{5}$$

Then,

$$\|Q_j^{-1} z_j^{(\mu_j)}\| \leq E_j^{(\mu_j)} \|z_{j+1}\|. \tag{6}$$

Remark 1. *Normally, there is no smoother N_1. In this case, $\epsilon_1^{(1)} = 1$.*

Remark 2. *In some instances, different restriction operators $R_k^{(i)}$ are used during a multigrid cycle. Substituting $\delta_k^{(i)}$ for δ_k in the proof covers this case.*

Remark 3. *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 4. *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* $E_k^{(\mu_k)}$ *in the proof covers this case.*

Proof: The proof of (6) is a double induction argument. The result is trivial when $j = 1$. Assume that the result is true for all levels $k < j$.

We first assume that $\mu_j = 1$. Since

$$\|z_{j-1}^{(1)}\| \le \rho_j^{(1)} \|z_j^{(0)}\|,$$

it follows that

$$\begin{aligned} \|Q_j^{-1} z_j^{(1)}\| &\le \epsilon_j^{(1)} \|\theta_j^{(1)}\| \\ &= \epsilon_j^{(1)} \|A_j(x_j^{(0)} + \gamma_j^{(1)}) + R_{j+1} z_{j+1}\| \\ &= \epsilon_j^{(1)} \|z_j^{(0)} + A_j \gamma_j^{(1)}\| \\ &= \epsilon_j^{(1)} \|z_j^{(0)} + A_j P_{j-1} \bar{x}_{j-1}^{(1)}\|. \end{aligned}$$

Note that $A_j P_{j-1} \bar{x}_{j-1}^{(1)} = Q_j^{-1} A_{j-1} \bar{x}_{j-1}^{(1)} = Q_j^{-1}(\bar{z}_{j-1}^{(1)} - R_j z_j^{(0)})$. Hence,

$$\begin{aligned} \|Q_j^{-1} z_j^{(1)}\| &\le \epsilon_j^{(1)} \|(I - Q_j^{-1} R_j) z_j^{(0)} + Q_j^{-1} \bar{z}_{j-1}^{(1)}\| \\ &\le \epsilon_j^{(1)} \delta_j \|z_j^{(0)}\| + \epsilon_j^{(1)} E_{j-1}^{(\mu_{j-1})} \|z_j^{(0)}\| \\ &\le \epsilon_j^{(1)} \rho_j^{(1)} [\delta_j + E_{j-1}^{(\mu_{j-1})}] \|z_{j+1}\|. \end{aligned}$$

Now assume that the result is true for $\mu_j = 1, \cdots, i-1$. Then,

$$\begin{aligned} \|Q_j^{-1} z_j^{(i)}\| &\le \epsilon_j^{(i)} \|A_j(\hat{x}_j^{(i)} + \gamma_j^{(i)}) + R_{j+1} z_{j+1}\| \\ &= \epsilon_j^{(i)} \|\hat{z}_j^{(i)} + A_j \gamma_j^{(i)}\| \\ &= \epsilon_j^{(i)} \|\hat{z}_j^{(i)} + A_j P_{j-1} \bar{x}_{j-1}^{(i)}\| \\ &\le \epsilon_j^{(i)} \|(I - Q_j^{-1} R_j) \hat{z}_j^{(i)} + Q_j^{-1} \bar{z}_{j-1}^{(i)}\| \\ &\le \epsilon_j^{(i)} \delta_j \|\hat{z}_j^{(i)}\| + \epsilon_j^{(i)} E_{j-1}^{(\mu_{j-1})} \|\hat{z}_j^{(i)}\| \\ &\le \epsilon_j^{(i)} \rho_j^{(i)} [\delta_j + E_{j-1}^{(\mu_{j-1})}] \|z_j^{(i-1)}\| \\ &\le \prod_{\ell=1}^{i} \left(\epsilon_j^{(\ell)} \rho_j^{(\ell)} [\delta_j + E_{j-1}^{(\mu_{j-1})}] \right) \|z_{j+1}\|. \end{aligned}$$

□

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

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

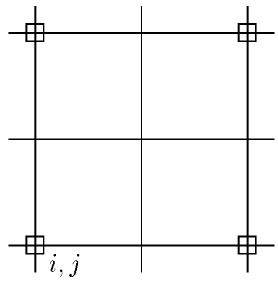

FIG. 7. *Part of example grid*

Then Theorem 1 can be rewritten in a variational form, and the proof is a trivial modification of the one stated.

To analyze Algorithm NI, assume that (2) is approximated by some ξ_k such that

$$A_k \xi_k + f_k = \theta_k,$$

starting from some initial guess $x_k = P_{k-1}\xi_{k-1}$. Given some $\{\sigma_k\}_{k=1}^{j}$, we want

$$(7) \qquad \begin{cases} \|\theta_1\| & \leq & \sigma_1 \|A_1 x_1 + f_1\|, \\ \|\theta_k\| & \leq & \sigma_k \|\theta_{k-1}\|, \quad 1 < k \leq j. \end{cases}$$

In most applications of multigrid to partial differential equations, $\sigma_1 = C h_1^{-q}$, where $C \in \mathbb{R}$, h_1 is the mesh spacing (or diameter of the triangles) on the coarsest grid, and q is related to the order of the truncation error. Then, $\sigma_k = h_k^q / h_{k-1}^q$, $k = 2, \cdots, j$.

Given $\{\delta_k\}_{k=1}^{j}$ from (3), Theorem 1 tells us how to choose $\{\rho_k, \epsilon_k\}_{k=1}^{j}$ so that a μ cycle reduces the residual sufficiently.

THEOREM 2. *Under the same assumptions as Theorem 1, with* $\{\sigma_k\}_{k=1}^{j}$ *given,* (7) *is satisfied for* $\{\{\rho_k^{(i)}, \epsilon_k^{(i)}\}_{i=1}^{\mu_k}\}_{k=1}^{j}$ *chosen to satisfy*

$$E_k^{(\mu_k)} \leq \sigma_k \quad \text{for } 1 \leq k \leq j$$

in (5).

3.2. Examples. In this section, we compute δ_k for Dirichlet problems on $\mathbb{R}^2$. While the estimates are rather pessimistic, we offer some advice on practical uses of the simple theory in §3.1. Finally, we present an example where Theorem 1 is sharp.

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). Then, locally, each interior section of the grid is as in Fig. 7.

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 $(i+1, j+1)$ (see Fig. 7) 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.

If the real convergence rate of Algorithm MG (or NSMG) for trivial problems like Laplace's equation in one or two dimensions was really what is predicted by Theorem 1 using the values $\delta^{(5)}$ or $\delta^{(9)}$, multigrid methods would never have been developed.

Besides motivating §4, 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, we calculated

$$ \delta(v) = \frac{\|(I - XR_2)v\|}{\|v\|}, $$

where this was calculated first using $X = Q_2^{-1}$ then by approximating Q_2^{-1} by $X = R_2^T$. The following was observed.

$\delta(v)$	X	
	Q_2^{-1}	R_2^T
Minimum	0.3444	0.5610
Maximum	0.9312	0.9724
Average	0.7126	0.8310

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

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.)

4. Sharper Analysis. In this section, we analyze $z_j^{(i)}$ under minimal assumptions. Here, we decompose each space $\mathcal{M}_j$ 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 of §3.1.

We assume that each space $\mathcal{M}_j$ can be decomposed into a smooth part $\mathcal{S}_j$ and a rough part $\mathcal{T}_j$:

$$\mathcal{M}_j = \mathcal{S}_j \oplus \mathcal{T}_j, \quad \text{where} \quad \mathcal{T}_j = \mathcal{M}_{j-1} \quad \text{and} \quad \mathcal{S}_j = \mathcal{M}_{j-1}^{\perp} \cap \mathcal{M}_j. \tag{8}$$

So, $\mathcal{S}_j$ contains the high frequency components and $\mathcal{T}_j$ contains the low frequency ones.

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 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

$$\|w_k\|^2 = \epsilon_{k,S}^2 |||v_k|||^2 + \epsilon_{k,T}^2 < v_k >^2. \tag{9}$$

Similarly, if v_k are $\bar{w}_k$ are the residuals before and after a 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

$$\begin{aligned} |||\bar{w}_k|||^2 &= \rho_{k,SS}^2 |||v_k|||^2 + \rho_{k,ST}^2 < v_k >^2 \quad \text{and} \\ < \bar{w}_k >^2 &= \rho_{k,TS}^2 |||v_k|||^2 + \rho_{k,TT}^2 < v_k >^2. \end{aligned} \tag{10}$$

As was noted at the end of §2.2, we will probably only be able to bound these parameters with estimates of some form.

The results in this section require 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 3. *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*

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

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,

$$\|Q_j^{-1} z_j^{(\mu_j)}\| \leq \prod_{i=1}^{\mu_j} \max \left\{ E_{j,SS}^{(i)} + E_{j,TS}^{(i)},\ E_{j,ST}^{(i)} + E_{j,TT}^{(i)} \right\} \|z_{j+1}\|. \tag{11}$$

Proof: The proof of (11) is a double induction argument.

The result is trivial for level 1. Now assume that the result is true for all levels $k < j$.

First, we assume that $\mu_j = 1$.

$$\|Q_j^{-1} z_j^{(1)}\| \leq \epsilon_{j,S}^{(1)} |||\theta_j^{(1)}||| + \epsilon_{j,T}^{(1)} < \theta_j^{(1)} >. \tag{12}$$

Using simple algebra, similar to that contained in the proof of Theorem 1, we get

$$\begin{aligned}
|||\theta_j^{(1)}||| &= |||(I - Q_j^{-1} R_j) z_j^{(0)} + Q_j^{-1} \bar{z}_{j-1}^{(1)}||| \\
&\leq \delta_{j,S} |||z_j^{(0)}||| + |||Q_j^{-1} \bar{z}_{j-1}^{(1)}||| \\
&\leq \delta_{j,S} |||z_j^{(0)}||| + (E_{j-1,SS}^{(\mu_{j-1})}) |||z_j^{(0)}||| + (E_{j-1,ST}^{(\mu_{j-1})}) < z_j^{(0)} > \\
&= \left(\delta_{j,S} + E_{j-1,SS}^{(\mu_{j-1})} \right) |||z_j^{(0)}||| + E_{j-1,ST}^{(\mu_{j-1})} < z_j^{(0)} >.
\end{aligned}$$

Now,

$$\begin{aligned}
|||z_j^{(0)}||| &\leq \rho_{j,SS} |||z_{j+1}||| + \rho_{j,ST} < z_{j+1} >, \\
< z_j^{(0)} > &\leq \rho_{j,TS} |||z_{j+1}||| + \rho_{j,TT} < z_{j+1} >.
\end{aligned}$$

So,

$$
\begin{aligned}
|||\theta_j^{(1)}||| &\le \left(\delta_{j,S} + E_{j-1,SS}^{(\mu_{j-1})}\right)\left\{\rho_{j,SS}^{(1)}|||z_{j+1}||| + \rho_{j,ST}^{(1)} < z_{j+1} >\right\} + \\
&\quad E_{j-1,ST}^{(\mu_{j-1})}\left\{\rho_{j,TS}^{(1)}|||z_{j+1}||| + \rho_{j,TT}^{(1)} < z_{j+1} >\right\} \\
&= \left\{\left(\delta_{j,S} + E_{j-1,SS}^{(\mu_{j-1})}\right)\rho_{j,SS}^{(1)} + E_{j-1,ST}^{(\mu_{j-1})}\rho_{j,TS}^{(1)}\right\}|||z_{j+1}||| + \\
&\quad \left\{\left(\delta_{j,S} + E_{j-1,SS}^{(\mu_{j-1})}\right)\rho_{j,ST}^{(1)} + E_{j-1,ST}^{(\mu_{j-1})}\rho_{j,TT}^{(1)}\right\} < z_{j+1} > \\
&= \hat{E}_{j,SS}^{(1)}|||z_{j+1}||| + \hat{E}_{j,ST}^{(1)} < z_{j+1} >.
\end{aligned}
$$

Using similar algebraic manipulations,

$$
\begin{aligned}
< \theta_j^{(1)} > &\le \left\{\left(\delta_{j,T} + E_{j-1,TT}^{(\mu_{j-1})}\right)\rho_{j,TS}^{(1)} + E_{j-1,TS}^{(\mu_{j-1})}\rho_{j,SS}^{(1)}\right\}|||z_{j+1}||| + \\
&\quad \left\{\left(\delta_{j,T} + E_{j-1,TT}^{(\mu_{j-1})}\right)\rho_{j,TT}^{(1)} + E_{j-1,TS}^{(\mu_{j-1})}\rho_{j,ST}^{(1)}\right\} < z_{j+1} > \\
&= \hat{E}_{j,TS}^{(1)}|||z_{j+1}||| + \hat{E}_{j,TT}^{(1)} < z_{j+1} >.
\end{aligned}
$$

Substituting the bounds for $|||\theta_j^{(1)}|||$ and $< \theta_j^{(1)} >$ into (12) gives us

$$
\begin{aligned}
||Q_j^{-1}z_j^{(1)}|| &\le \left\{\epsilon_{j,S}^{(1)}\hat{E}_{j,SS}^{(1)} + \epsilon_{j,T}^{(1)}\hat{E}_{j,TS}^{(1)}\right\}|||z_{j+1}||| + \\
&\quad \left\{\epsilon_{j,T}^{(1)}\hat{E}_{j,TT}^{(1)} + \epsilon_{j,S}^{(1)}\hat{E}_{j,ST}^{(1)}\right\} < z_{j+1} > \\
&= \left\{E_{j,SS}^{(1)} + E_{j,TS}^{(1)}\right\}|||z_{j+1}||| + \left\{E_{j,TT}^{(1)} + E_{j,ST}^{(1)}\right\} < z_{j+1} > \\
&\le \max\left\{E_{j,SS}^{(1)} + E_{j,TS}^{(1)}, E_{j,TT}^{(1)} + E_{j,ST}^{(1)}\right\}||z_{j+1}||.
\end{aligned}
$$

Now assume that the result is true for $\mu_j = 1, \cdots, i-1$. A similar argument shows that

$$
\begin{aligned}
||Q_j^{-1}z_j^{(i)}|| &\le \left\{E_{j,SS}^{(i)} + E_{j,TS}^{(i)}\right\}|||z_j^{(i-1)}||| + \left\{E_{j,TT}^{(i)} + E_{j,ST}^{(i)}\right\} < z_j^{(i-1)} > \\
&\le \max\left\{E_{j,SS}^{(i)} + E_{j,TS}^{(i)}, E_{j,TT}^{(i)} + E_{j,ST}^{(i)}\right\}||z_j^{(i-1)}|| \\
&\le \prod_{\ell=1}^{i}\max\left\{E_{j,SS}^{(\ell)} + E_{j,TS}^{(\ell)}, E_{j,TT}^{(\ell)} + E_{j,ST}^{(\ell)}\right\}||z_{j+1}||. \qquad \square
\end{aligned}
$$

At this point, a number of examples ought to be demonstrated. However, the possibilities are almost endless. For example, how does the order of approximation (in the projection and interpolation processes) affect the convergence rate (see [19])? Do these results contradict known results using local mode analysis (they do)? How does use of different iterative methods as smoothers (or roughers) affect the convergence rate? Answering each of these questions is beyond the scope of this paper and will be answered in sequels.

5. Multiple Coarse Space Methods. Standard multigrid which is parallelized by domain decomposition is already covered by the analysis of the previous two sections. After all, this is merely a block iterative method used for smoothing.

The concept of using multiple subspaces to solve a problem whose solution lies in a particular space is hardly new. In fact, no one is alive from the era in which it was invented. We will never know who really invented it, but it was certainly introduced no later than in 1869 (see [44]).

Assume a rooted tree of problems (2) which are arbitrarily numbered. For a given problem k, it either has a set C_k of coarse space correction problems or it has none at all (i.e., $C_k = \emptyset$). When $C_k \neq \emptyset$, there are restriction and prolongation operators for each coarse space problem $\ell \in C_k$ such that

$$R_\ell : \mathcal{M}_k \mapsto \mathcal{M}_\ell \quad \text{and} \quad P_\ell : \mathcal{M}_\ell \mapsto \mathcal{M}_k.$$

We also assume that there are mappings

$$Q_\ell : \mathcal{M}_k \to \mathcal{M}_\ell \quad \text{such that} \quad A_\ell = Q_\ell A_k P_\ell.$$

How these are defined is very similar to the earlier cases. The discussion in §2.1 applies to this case as well.

Define a k-level (*standard*) multiple coarse space correction multigrid scheme by

ALGORITHM PMG$(k, \mu_k, C_k, x_k, f_k)$

(1) If $C_k = \emptyset$, then solve $A_k x_k = f_k$ exactly or by smoothing

(2) If $C_k \neq \emptyset$, then repeat $i = 1, \cdots, \mu_k$:

(2a) Smoothing: $x_k \leftarrow M_k^{(i)}(x_k, f_k)$

(2b) Residual Correction:

$$x_k \leftarrow x_k + \sum_{\ell \in C_k} P_\ell \ \mathrm{PMG}(\ell, \mu_\ell, C_\ell, 0, R_\ell(A_k x_k + f_k))$$

(2c) Smoothing: $x_k \leftarrow N_k^{(i)}(x_k, f_k)$

(3) Return x_k

Ta'asan [45] introduced this method to the multigrid community when standard multigrid failed to converge for a class of problems with highly oscillatory solutions. He uses standard interpolation and projection methods and a Kaczmarz relaxation method in his examples.

Hackbusch [31] developed a variant of Ta'asan's method using a different set of interpolation and projection methods. By doing this, he can use standard smoothers. Ta'asan's and Hackbusch's methods are both referred to as *robust multigrid*, which adds confusion to the field.

Frederickson and McBryan [27] are interested in keeping all of the processors busy on a massively parallel single instruction, multiple data machine (SIMD). They used standard interpolation and projection methods and an elaborate smoother on each level. Unless great care is taken, this method computes the correction in one of the correction spaces while the corrections in the remaining spaces add up (pointwise) to zero. Their method is referred to as *parallel superconvergent multigrid*.

Each of the methods introduced in this section so far uses interleaved grids. For a problem on a square, this translates into the following, where the numbers refer to which subproblem the unknowns belong:

3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2

(This is similar to the motivation for multicolored orderings for standard iterative methods.) Each of these methods requires that all of the matrices associated with the spaces be generated, except in trivial cases, thus doubling the memory requirements expected for solving boundary value problems. In addition, the coarse space operators are more difficult to compute using Hackbusch's variant than for either of the other two methods. In general, these are all space wasteful methods.

A very different fourth approach has been developed by C. Douglas et al and is referred to as either *constructive interference* or, more recently, *domain reduction*. The reason cited in [22] for using a multiple coarse space parallel multigrid algorithm was to eliminate the smoothing step from standard multigrid algorithms. Smoothing takes most of the computational time, but contributes almost nothing to the convergence rate, whereas coarse grid corrections take little time and reduce the error substantially. A general theory and simple examples for multiple coarse space methods was provided in [22] using no smoothing on the fine grid and mutually orthogonal subspaces which covered all of the error components of the original space (thus, $\delta_j = 0$ in (13). This leads to very efficient direct methods instead of the expected iterative ones.

A side benefit of this theory is that the fine grid problem does not have to be generated (see [21], [20], and [23]) and, by using a trick, most of the coarse space matrices do not have to be generated either (see [18]). This method can use substantially less memory than a standard iterative or multigrid algorithm.

An additional note about domain reduction is that it leads naturally to more than 2^d subspaces for a d dimensional problem. In [13], an eight way decomposition of a problem on a square is constructed, leading to problems defined on squares, rectangles, and triangles. Both 60 and 64 way decompositions of a problem on a cube are constructed in [20].

Note that, in theory, a 192 way decomposition of a problem on a cube is possible. If each of the 192 subproblems is solved by sparse Gaussian elimination, the entire problem would be solved directly 2660 times faster than if the original one is solved by the same method (the latter is not advised, however).

To see how each of the variants operates, consider various projection operators in one dimension (where $x_0 = c_{N+1} = 0$):

- Linear projection: for $1 \leq i \leq N$,

$$y_i = x_{i-1} + 2x_i + x_{i+1}.$$

  Let one subspace consist of the odd numbered y_i's and another subspace for the even numbered y_i's. This is used by both Ta'asan and Frederickson–McBryan.

- Linear–linear orthogonal complement projection: for $1 \leq i \leq N$,

$$y_i = \begin{cases} x_{i-1} + 2x_i + x_{i+1} & i\ even, \\ \\ -x_{i-1} + 2x_i - x_{i+1} & i\ odd. \end{cases}$$

The subspaces are defined as in the linear projection case. The values at odd numbered y_i's correspond to the orthogonal complement of the traditional space defined by linear projection. This is used by Hackbusch.

- Symmetric–antisymmetric projection: for $1 \le i \le N/2$,

$$y_i = \frac{x_i + x_{N-i+1}}{2} \quad \textit{and} \quad \hat{y}_i = \frac{x_i - x_{N-i+1}}{2}.$$

  One subspace consists of y_i's and the other of $\hat{y}_i$'s. This defines a domain folding (or reduction) where even and odd functions are annihilated in exactly one subspace and exactly reproduced in the other. This is used by C. Douglas et al.

Two and three dimensional definitions are defined in the obvious manner using tensor products.

Now, define a k-level (*nonstandard*) multiple coarse space correction multigrid scheme by

ALGORITHM NSPMG($k, p, z_p, x_k^{(-1)}$)

(1) Initial residual: $R_k z_p \in \mathcal{M}_k$

(2) Smoothing: $x_k^{(0)} = M_k^{(1)} x_k^{(-1)}$ such that
$$A_k x_k^{(0)} + R_k z_p = z_k^{(0)}, \text{ where } \|z_k^{(0)}\| \le \rho_k^{(1)} \|z_p\|$$

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

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

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

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

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

(4b) If $C_k \ne \emptyset$, then

(4b1) Correction: $\gamma_k^{(i)} = \sum_{\ell \in C_k} P_\ell \bar{x}_\ell^{(i)}$, where
$$\bar{x}_\ell^{(i)} = \text{NSPMG}(\ell, k, \hat{z}_k^{(i)}, 0)$$
and
$$A_\ell \bar{x}_\ell^{(i)} + R_\ell \hat{z}_k^{(i)} = \bar{z}_\ell^{(i)}$$

(4c) If $C_k = \emptyset$, then $\gamma_k = 0$

(4d) Residual: $A_k(\hat{x}_k^{(i)} + \gamma_k^{(i)}) + R_k z_p = \theta_k^{(i)}$

(4e) Smoothing: $x_k^{(i)} = N_k^{(i)}(\hat{x}_k^{(i)} + \gamma_k^{(i)})$ such that
$$A_k x_k^{(i)} + R_k z_p = z_k^{(i)},$$
where
$$\|z_k^{(i)}\| \le \epsilon_k^{(i)} \|\theta_k^{(i)}\|$$

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

Again, $z_j^{(i)}$ is analyzed under minimal assumptions. The results assume only a simple property: there exists a constant, $\delta_k \in \mathbb{R}$, such that

$$\|(I - \sum_{\ell \in C_k} Q_\ell^{-1} R_\ell)u\| \le \delta_k \|u\|, \quad u \in \mathcal{M}_k. \tag{13}$$

As before, the same types of hypotheses are needed to explain Q_ℓ^{-1}.

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

In the case that the coarse spaces $\{\mathcal{M}_\ell\}$ are nested inside of $\mathcal{M}_k$, the basic theorem is as follows.

THEOREM 4. *Assume that z_p is the residual for problem p, R_k maps $\mathcal{M}_p$ into $\mathcal{M}_k$, and that all P_ℓ and Q_ℓ^{-1}, $\ell \in C_k$, satisfy* (4). *Let*

$$E_k^{(\mu_k)} = \begin{cases} \epsilon_k^{(1)} \rho_k^{(1)}, & \text{if } C_k = \emptyset \\ \prod_{i=1}^{\mu_k} \left(\epsilon_k^{(i)} \rho_k^{(i)} \left[\delta_k + \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)} \right] \right), & \text{otherwise.} \end{cases} \tag{14}$$

Then,

$$\|Q_k^{-1} z_k^{(\mu_j)}\| \le E_k^{(\mu_k)} \|z_p\|. \tag{15}$$

Proof: The proof of (15) is a double induction argument. The result is trivial when $C_j = \emptyset$. Assume that the result is true for all problems corresponding to indices in $C_j \ne \emptyset$.

We first assume that $\mu_j = 1$. Then

$$\begin{aligned} \|Q_j^{-1} z_j^{(1)}\| &\le \epsilon_j^{(1)} \|\theta_j^{(1)}\| \\ &= \epsilon_j^{(1)} \|A_j(x_j^{(0)} + \gamma_j^{(1)}) + R_j z_p\| \\ &= \epsilon_j^{(1)} \|z_j^{(0)} + A_j \gamma_j^{(1)}\| \\ &= \epsilon_j^{(1)} \|z_j^{(0)} + \sum_{\ell \in C_k} A_j P_\ell \bar{x}_\ell^{(1)}\| \end{aligned}$$

Note that $\sum A_j P_\ell \bar{x}_\ell^{(1)} = \sum Q_\ell^{-1} A_\ell \bar{x}_\ell^{(1)} = \sum Q_\ell^{-1}(\bar{z}_\ell^{(1)} - R_\ell z_j^{(0)})$. Hence,

$$\begin{aligned} \|Q_j^{-1} z_j^{(1)}\| &\le \epsilon_j^{(1)} \|(I - \sum_{\ell \in C_k} Q_\ell^{-1} R_\ell) z_j^{(0)} + \bar{z}_\ell^{(1)}\| \\ &\le \epsilon_j^{(1)} \delta_j \|z_j^{(0)}\| + \epsilon_j^{(1)} \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)} \|z_j^{(0)}\| \\ &\le \epsilon_j^{(1)} \rho_j^{(1)} [\delta_j + \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)}] \|z_p\|. \end{aligned}$$

Now assume that the result is true for $\mu_j = 1, \cdots, i-1$. Then,

$$\begin{aligned}
\|Q_j^{-1} z_j^{(i)}\| &\leq \epsilon_j^{(i)} \|A_j(\hat{x}_j^{(i)} + \gamma_j^{(i)}) + R_j z_p\| \\
&= \epsilon_j^{(i)} \|\hat{z}_j^{(i)} + A_j \gamma_j^{(i)}\| \\
&= \epsilon_j^{(i)} \|\hat{z}_j^{(i)} + \sum_{\ell \in C_k} A_j P_\ell \bar{x}_\ell^{(i)}\| \\
&\leq \epsilon_j^{(i)} \|(I - \sum_{\ell \in C_k} Q_\ell^{-1} R_\ell) \hat{z}_j^{(i)} + Q_\ell^{-1} \bar{z}_\ell^{(i)}\| \\
&\leq \epsilon_j^{(i)} \delta_j \|\hat{z}_j^{(i)}\| + \epsilon_j^{(i)} \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)} \|\hat{z}_j^{(i)}\| \\
&\leq \epsilon_j^{(i)} \rho_j^{(i)} [\delta_j + \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)}] \|z_j^{(i-1)}\| \\
&\leq \prod_{\ell=1}^{i} \left(\epsilon_j^{(\ell)} \rho_j^{(\ell)} [\delta_j + \sum_{\ell \in C_k} E_\ell^{(\mu_\ell)}] \right) \|z_p\|. \qquad \square
\end{aligned}$$

Theorem 4 can be made somewhat sharper by using an approach in [20].

For an important special case Theorem 4 can be made sharper.

Corollary 1. *In addition to the assumptions of Theorem 4, assume that* $\|\cdot\|$ *is derived from an inner product and that the* $\mathcal{M}_\ell$, *all* $\ell \in C_k$, *are mutually orthogonal in this inner product. Then*

$$E_k^{(\mu_k)} = \begin{cases} \epsilon_k^{(1)} \rho_k^{(1)}, & \text{if } C_k = \emptyset \\ \prod_{i=1}^{\mu_k} \left(\epsilon_k^{(i)} \rho_k^{(i)} \left[\delta_k + \max_{\ell \in C_k} E_\ell^{(\mu_\ell)} \right] \right), & \text{otherwise.} \end{cases}$$

The proof is a trivial modification of the one for Theorem 4.

Consider the domain reduction method (see [13], [18], [20], and [23]). One of the basic assumptions of this method implies that $\delta_k = 0$ for all k. Only two levels (with many coarse space correction problems), no smoothing on the fine grid, and $\mu_k = 1$ makes computational sense for this method. Further, mutually orthogonal coarse spaces can be constructed for elliptic boundary value problems, so that Corollary 1 applies for the fine grid problem j. Hence,

$$E_j^{(1)} = \max_{\ell \in C_j} \{\epsilon_\ell^{(1)} \rho_\ell^{(1)}\},$$

which is sharp.

Unfortunately, these methods have not been directly compared on a nontrivial example problem. The closest is a simple problem from the literature (we apologize in advance to each originator of a method compared here):

$$\begin{cases} -10^5 u_{xx} - 10^{-5} u_{yy} & = f \text{ in } (0,1)^2, \\ u & = 0 \text{ on } \partial(0,1)^2. \end{cases}$$

To make things comparable with known results, a uniform mesh is used, a standard (simple) discretization, the energy norm, one Jacobi iteration in the analysis for multigrid (MG) and robust multigrid, and a direct solve on the coarsest level(s). The contraction factors are:

Method	Contraction factor	
MG	0.97	(17×17 grid)
Robust MG	0.33	(h independent)
Domain reduction (iterative)	ϵ	(h independent)
Domain reduction (direct)	0.00	(h independent)
Parallel superconvergent	small	(h independent)

The solver in the domain reduction is either iterative (solving each problem to an accuracy of ϵ) or direct. If the smoother called for in parallel superconvergent can be constructed, then the contraction factor missing above will be very small, on the order of 0.05. Note that a line relaxation method, instead of point Jacobi, would make multigrid work quite well.

We conclude this section by noting that the approach of §4 can also be applied to the algorithms in this section. A sharper convergence result than Theorem 4 is as follows.

THEOREM 5. *Assume that z_p is the residual for problem p, R_k maps $\mathcal{M}_p$ into $\mathcal{M}_k$, and that all P_ℓ and Q_ℓ^{-1}, $\ell \in C_k$, satisfy* (4). *If $C_k = \emptyset$, then*

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

If $C_k \neq \emptyset$, then

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

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

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

and

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

Then,

$$\|Q_j^{-1} z_j^{(\mu_j)}\| \leq \prod_{i=1}^{\mu_j} \max \left\{ E_{j,SS}^{(i)} + E_{j,TS}^{(i)},\ E_{j,ST}^{(i)} + E_{j,TT}^{(i)} \right\} \|z_{j+1}\|. \tag{16}$$

Further, if the norm $\|\cdot\|$ is derived from an inner product and all $\mathcal{M}_\ell$, $\ell \in C_k$, are mutually orthogonal in this inner product, then the $\sum$ symbols in the definitions of $E_{k,XY}$, $X, Y \in \{S, T\}$, can be changed to max *symbols.*

6. 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.

To get error bounds which are sharp enough to be really useful, we need to provide more delicate analysis based on splitting each solution space $\mathcal{M}_k$ into two parts.

However, in the multiple coarse space case, the simple analysis for Algorithm PMG can be sharp.

REFERENCES

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

[2] 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.

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

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

[5] D. Braess and W. Hackbusch, *A new convergence proof for the multigrid method including the V cycle*, SIAM J. Numer. Anal., 20 (1983), pp. 967–975.

[6] J. H. Bramble and J. E. Pasciak, *New convergence estimates for multigrid algorithms*, Math. Comp., 49 (1987), pp. 311–329.

[7] J. H. Bramble, J. E. Pasciak, and J. Xu, *The analysis of multigrid algorithms for nonsymmetric and indefinite problems*, Math. Comp., 51 (1988), pp. 389–414.

[8] ———, *The analysis of multigrid algorithms with nonnested and noninherited forms*, Math. Comp., 56 (1991), pp. 1–34.

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

[10] ———, *Guide to multigrid development*, in Multigrid Methods, W. Hackbusch and U. Trottenberg, eds., Springer–Verlag, New York, 1982, pp. 220–312.

[11] S. C. Brenner, *An optimal order multigrid for P1 nonconforming finite elements*, Math. Comp., 52 (1989), pp. 1–5.

[12] ———, *An optimal order nonconforming multigrid method for the biharmonic equation*, SIAM J. Numer. Anal., 26 (1989), pp. 1124–1138.

[13] F. Brezzi, C. C. Douglas, and L. D. Marini, *A parallel domain reduction method*, Numer. Meth. for PDE, 5 (1989), pp. 195–202.

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

[15] 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.

[16] ———, *Abstract multi–grid with applications to elliptic boundary–value problems*, in Elliptic Problem Solvers II, G. Birkhoff and A. Schoenstadt, eds., Academic Press, New York, 1984, pp. 453–466.

[17] ———, *Multi–grid algorithms with applications to elliptic boundary–value problems*, SIAM J. Numer. Anal., 21 (1984), pp. 236–254.

[18] ———, *A tupleware approach to domain decomposition methods*, Appl. Numer. Math., 8 (1991), pp. 353–373.

[19] ———, *The effect of interpolation on the rate of convergence in two level algorithms for elliptic partial differential equations*, in Preliminary version of the Proceedings of the Copper Mountain Conference on Iterative Methods, T. Manteuffel, ed., Computational Mathematics Group, University of Colorado at Denver, Denver, Colorado, 1992. Also available as IBM Research Report RC 17739, Yorktown Heights, NY, 1992.

[20] C. C. Douglas and J. Mandel, *The domain reduction method: high way reduction in three dimensions and convergence with inexact solvers*, in Fourth Copper Mountain Conference

on Multigrid Methods, J. Mandel, S. F. McCormick, J. E. Dendy, C. Farhat, G. Lonsdale, S. V. Parter, J. W. Ruge, and K. Stüben, eds., Society for Industrial and Applied Mathematics, Philadelphia, 1989, pp. 149–160.

[21] ———, *A group theoretic approach to the domain reduction method*, Tech. Rep. 17173, IBM Research Division, Yorktown Heights, New York, 1991. To appear in Computing (1992).

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

[23] C. C. Douglas and B. F. Smith, *Using symmetries and antisymmetries to analyze a parallel multigrid algorithm*, SIAM J. Numer. Anal., 26 (1989), pp. 1439–1461.

[24] L. M. Dudkin and E. B. Yershov, *Interindustries input–output models and the material balances of separate products*, Planned Economy, 5 (1965), pp. 54–63.

[25] R. P. Federenko, *A relaxation method for solving elliptic difference equations*, Z. Vycisl. Mat. i. Mat. Fiz., 1 (1961), pp. 922–927.

[26] ———, *The speed of convergence of one iteration process*, Z. Vycisl. Mat. i. Mat. Fiz., 4 (1964), pp. 559–563.

[27] P. Frederickson and O. McBryan, *Parallel superconvergent multigrid*, in Multigrid Methods: Theory, Applications, and Supercomputing, S. F. McCormick, ed., Marcel Dekker, New York, 1988, pp. 195–210.

[28] A. Greenbaum, *Analysis of a multigrid method as an iterative technique for solving linear systems*, SIAM J. Numer. Anal., 21 (1984), pp. 473–485.

[29] W. Hackbusch, *A multi–grid method applied to a boundary value problem with variable coefficients in a rectangle*, Tech. Rep. 77–17, Mathematisches Institut, Universitat zu Koln, Cologne, 1977.

[30] ———, *Multigrid Methods and Applications*, Springer–Verlag, Berlin, 1985.

[31] ———, *A new approach to robust multi–grid solvers*, in ICIAM'87: Proceedings of the First International Conference on Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, 1988, pp. 114–126.

[32] J. F. Maitre and F. Musy, *Multigrid methods: convergence theory in a variational framework*, SIAM J. Numer. Anal., 21 (1984), pp. 657–671.

[33] J. Mandel, *Convergence of an iterative method for the system $ax + y = x$ using aggregation*, Ekonom.–Mat. Obzor., 17 (1981), pp. 159–169.

[34] ———, *Multigrid convergence for nonsymmetric, indefinite variational problems and one smoothing step*, Appl. Math. Comput., 19 (1986), pp. 201–216.

[35] J. Mandel, S. F. McCormick, and R. E. Bank, *Variational multigrid theory*, in Multigrid Methods, S. F. McCormick, ed., Society for Industrial and Applied Mathematics, Philadelphia, 1987, pp. 131–177.

[36] J. Mandel, S. F. McCormick, and J. Ruge, *An algebraic theory for multigrid methods for variational problems*, SIAM J. Numer. Anal., 25 (1988), pp. 91–110.

[37] J. Mandel and S. Parter, *On the multigrid F–cycle*, Appl. Math. Comput., 37 (1990), pp. 19–36.

[38] S. F. McCormick, *An algebraic interpretation of multigrid methods*, SIAM J. Numer. Anal., 19 (1982), pp. 548–560.

[39] ———, *Multigrid methods for variational problems: general theory for the V–cycle*, SIAM J. Numer. Anal., 22 (1985), pp. 634–643.

[40] W. L. Miranker and V. Y. Pan, *Methods of aggregation*, Lin. Alg. Appl., 29 (1980), pp. 231–257.

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

[42] ———, *On multigrid convergence in the indefinite case*, Math. Comp., 32 (1978), pp. 1082–1086.

[43] S. V. Parter, *Estimates for multigrid methods based on red–black Gauss–Seidel smoothings*, Numer. Math., 52 (1988), pp. 701–723.

[44] H. A. Schwarz, *Über einige Abbildungsaufgaben*, Ges. Math. Abh., 11 (1869), pp. 65–83.

[45] S. Ta'asan, *Multigrid Methods for Highly Oscillatory Problems*, PhD thesis, Weizmann Institute of Science, Rehovot, Israel, 1984.

[46] I. Y. Vakhutinsky, L. M. Dudkin, and A. A. Ryvkin, *Iterative aggregation – a new approach to the solution of large–scale problems*, Econometrics, 47 (1979), pp. 821–841.

[47] J. R. VanRosendale, *Rapid Solution of Finite Element Equations on Locally Refined Grids by Multi–Level Methods*, PhD thesis, University of Illinois, Urbana–Champaign, 1980.

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

[49] J. Xu, *Theory of multilevel methods*, PhD thesis, Cornell University, 1989. Also, Department

of Mathematics, Penn State, Technical Report AM48.

[50] H. Yserentant, *The convergence of multi–level methods for solving finite element equations in the presence of singularities*, Tech. Rep. 14, Rheinische–Westfalische Technische Hochschule, Aachen, West Germany, 1982.