AN ENERGY-MINIMIZING INTERPOLATION FOR ROBUST MULTIGRID METHODS

W. L. WAN *, TONY F. CHAN † AND BARRY SMITH ‡

Abstract. We propose a robust interpolation for multigrid based on the concepts of energy minimization and approximation. The formulation is general; it can be applied to any dimensions. The analysis for one dimension proves that the convergence rate of the resulting multigrid method is independent of the coefficient of the underlying PDE, in addition to being independent of the mesh size. We demonstrate numerically the effectiveness of the multigrid method in two dimensions by applying it to a discontinuous coefficient problem and an oscillatory coefficient problem. We also show using a one-dimensional Helmholtz problem that the energy minimization principle can be applied to solving elliptic problems that are not positive definite.

1. Introduction.

Multigrid methods are widely used as efficient solvers for second order elliptic partial differential equations (PDEs) because of their often optimal convergence behavior; that is, their convergence rate is independent of the mesh size. Optimal theory can be found, for example, in [2, 3, 4, 20, 27, 31, 38, 39]. However, the convergence rate may depend on the nature of the coefficients in the PDE. Typically, the convergence deteriorates as the coefficients become rougher. Specifically, if the coefficients are anisotropic [20], have large jumps [1, 5, 10, 11] or are highly oscillatory [17, 26, 34], standard multigrid methods will converge very slowly. Special techniques such as line Gauss-Seidel/block smoothing [5], semi-coarsening [12, 13, 32], algebraic multigrid [6, 28, 30, 33], frequency decomposition [14, 21, 34], and homogenization [17, 26] are used to handle some of these cases. In this paper, we study the design of multigrid methods from the energy minimization point of view, which gives powerful insight into the design of robust multigrid methods.

The success of multigrid hinges on the choice of the coarse grid points, the smoothing procedure, the interpolation operators, and the coarser grid discretization. In standard multigrid, full coarsening, Jacobi or Gauss-Seidel smoothing, and linear interpolation are usually used. Classical convergence theory shows that these simple ingredients are enough to achieve optimal convergence for smooth coefficient problems. In general, however, these choices may lead to slow convergence. In one dimension, to remedy the situation, a more robust interpolation [20, 28, 36] can be used. It is obtained by solving local homogeneous PDEs, which are equivalent to minimizing the energy of the coarse grid basis functions.

The extension of this approach to higher dimensions is not obvious. Nonetheless, many attempts [1, 10, 20, 19, 24, 29, 36] have been made to set up similar local PDEs for defining a robust interpolation. In place of setting up PDEs, we consider an equivalent minimization formulation and derive a so-called energy-minimizing interpolation with special emphasis on its stability and approximation properties, which are essential for optimal convergence. This approach to determining appropriate interpolation operators has also been used for iterative substructuring [15]. It will be made more precise in Section 3.

* Department of Mathematics, University California at Los Angeles, Los Angeles, CA 90095-1555. Email: wlwan@math.ucla.edu. The author has been partially supported by the grants listed under the second author, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W07405-Eng-48, and by the Alfred P. Sloan Foundation as a Doctoral Dissertation Fellow.

† Department of Mathematics, University California at Los Angeles, Los Angeles, CA 90095-1555. Email: chan@math.ucla.edu. This author has been partially supported by the NSF under Contract ASC-9720257.

‡ Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439. Email: bsmith@mcs.anl.gov. This work is supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

Although it is well known that the one-dimensional interpolation mentioned above will produce a robust multigrid method, a convergence analysis has not been given in the literature. In Section 4, we analyze the one-dimensional method derived from the energy-minimizing interpolation. We prove that the convergence rate is independent of the coefficient of the underlying PDE, in addition to the mesh size. In Section 5, we give numerical examples mainly in two dimensions, including a discontinuous coefficient problem, an oscillatory coefficient problem, and a Helmholtz problem. Finally, we summarize our experience by several remarks in Section 6.

We now set up notation to be used in the following sections. Let $V = V^h$, and let $V_1 \subset V_2 \subset \cdots \subset V_J = V$ denote a sequence of nested subspaces of V defined by the span of nodal basis functions, $\{\phi_i^k\}_{i=1}^{n_k}, k = 1, \ldots, J$, at level k. The operator $A : V \to V$ is self-adjoint and induces the A-inner product: $(\cdot,\cdot)_A \equiv (A\cdot,\cdot)$. Also, we define $A_i : V_i \to V_i$ by $(A_i u_i, v_i) = (Au_i, v_i), u_i, v_i \in V_i$. Correspondingly, we have $R_i : V_i \to V_i$, which is an approximate inverse of A_i. Let $Q_i : V \to V_i$ and $P_i : V \to V_i$ be the projection operators with respect to the L^2 and the A inner product, respectively. In the following analysis, the generic constant C is independent of the mesh size h.

2. Stability and Approximation Property.

Before we explain the formulation of the energy-minimizing interpolation, we first discuss our motivation from the classical results of multigrid and domain decomposition methods. Two key properties, stability and approximation, must be satisfied by the coarse subspaces and the smoothers [20] in order to have optimal convergence results. These two terms occur frequently in the literature but often appear in slightly different forms. For example, in the subspace correction framework [39], these two properties are built into the estimate of a constant K_0, which in turn is used to prove optimal convergence together with another constant K_1. The definitions of K_0 and K_1 are as follows:

K_0: For any $v \in V$, there exists a decomposition $v = \sum_{i=1}^J v_i$ for $v_i \in V_i$ such that

$$\sum_{i=1}^{J} (R_i^{-1} v_i, v_i) \le K_0 (Av, v), \tag{1}$$

where R_i is usually known as the smoother in the multigrid context.

K_1: For any $S \subset \{1, \ldots, J\} \times \{1, \ldots, J\}$ and $u_i, v_i \in V$ for $i = 1, \ldots, J$,

$$\sum_{(i,j)\in S} (T_i u_i, T_j u_j)_A \le K_1 \Big(\sum_{i=1}^{J} (T_i u_i, u_i)_A\Big)^{\frac{1}{2}} \Big(\sum_{j=1}^{J} (T_j v_j, v_j)_A\Big)^{\frac{1}{2}}, \tag{2}$$

where $T_i = R_i A_i P_i$.

THEOREM 2.1. *Let E_J be the iteration matrix given by the V-cycle multigrid, namely,*

$$u - u^{k+1} = E_J(u - u^k),$$

where u is the exact solution and u^k and u^{k+1} are two consecutive multigrid iterates. Then

$$E_J = (I - T_J)(I - T_{J-1}) \cdots (I - T_1),$$

and

$$\|E_J\|_A^2 \le 1 - \frac{2 - \omega_1}{K_0(1 + K_1)^2},$$

where $\omega_1 = \max_{1\le i\le J} \rho(R_i A_i)$.

Proof. See [39]. □

By Theorem 2.1, the convergence rate can be improved by producing a smaller K_0 or K_1. In this paper, we propose an interpolation that will potentially decrease the size of the constant K_0 by reducing its dependence on the coefficients of the underlying elliptic PDE.

As shown in [39], the estimate of K_0 relies on two inequalities:

$$\|\tilde{Q}_1 v\|_A^2 + \sum_{k=2}^{J} \|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|_A^2 \leq C_0 \|v\|_A^2, \tag{3}$$

$$\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\| \leq C_1 h_k \|\tilde{Q}_k v\|_A, \qquad \forall k > 1, \tag{4}$$

where $\tilde{Q}_k : V \rightarrow V_k$ is any linear operator onto V_k.

Inequality (3) appears in the Partition Lemma, which is well known in the domain decomposition literature [16, 31]. In the multigrid context, however, this inequality typically is used only implicitly. Intuitively speaking, (3) says that given any $v \in V$, we must be able to decompose v into the subspaces such that the total energy of all the pieces v_i is bounded by a small constant factor of the original energy of v. Besides (3), we also require that functions on the coarser grids approximate those on the finer grids to at least first order accuracy in h_k. This requirement is quantified by the inequality (4). If we have both (3) and (4), we can bound K_0 by a constant independent of the mesh size h.

LEMMA 2.2. *Let $\omega_0 = \min_{2 \leq i \leq J} (\rho(A_i)\lambda_{\min}(R_i))$. Suppose (3) and (4) are satisfied. Then*

$$K_0 \leq \frac{C}{\omega_0},$$

where C is a constant independent of the mesh size.

Proof. See the appendix. □

To summarize, if the stability and the approximation properties (3) and (4) are satisfied, optimal convergence follows. Thus, these two properties characterize a good coarse subspace. It is interesting to note that linear finite element subspaces are not compulsory for the V_k, though they are typically used or assumed in the classical analysis of multigrid methods. Moreover, the $\tilde{Q}_k$ in the approximation inequality (4) need not necessarily be the L^2 projections Q_k. Linear finite element and L^2 projections are simply two convenient and powerful tools for showing the stability and the approximation properties, but are not necessarily the only choice.

Optimal convergence, however, need not mean rapid convergence. The reason is that, in general, K_0 will depend on the PDE coefficients. The implicit dependence of the coefficient of the underlying PDE in the convergence rate may cause the multigrid method to converge very slowly, for example, when the coefficients are not smooth. In the following section, we construct coarse subspaces whose basis functions are, in general, different from piecewise linear finite elements but possess the stability and the approximation properties. In addition, the resulting multigrid algorithm is less sensitive to the coefficients than is the standard multigrid method. Furthermore, we show that these two concepts lead to an optimal convergence for a one-dimensional multigrid method, and we illustrate how they motivate a two-dimensional multigrid algorithm.

3. Energy-minimizing Interpolation.

In this section, we introduce the energy minimization approach to constructing the interpolation. The resulting formulation in the one-dimensional case is well known in the literature [20, 28, 36]. We explain the energy-minimizing interpolation in one dimension first and then in two dimensions.

3.1. One Dimension.

We consider the following model problem:

$$\begin{aligned} -\frac{d}{dx}a(x)\frac{d}{dx}u(x) &= f \qquad \text{in } (0,1) \\ u &= 0 \qquad \text{at } x = 0 \text{ and } x = 1, \end{aligned} \tag{5}$$

where $a(x)$ and $f(x)$ are integrable and $a(x)$ is uniformly positive.

Let $H^1(0,1)$ be the standard Sobolev space on the interval $[0,1]$ and $H_0^1(0,1)$ its subspace whose functions vanish at $x = 0$ and $x = 1$. Then the variational formulation of (5) is to find $u \in H_0^1(0,1)$ such that

$$a(u, v) = (f, v) \qquad \forall v \in H_0^1(0,1),$$

where

$$a(u,v) = \int_0^1 a(x)u'(x)v'(x)dx, \quad (f,v) = \int_0^1 f(x)v(x)dx.$$

Given a uniform grid with grid size $h = 1/n$, let $x_j^h = jh, j = 0, \ldots, n$. Define the fine grid linear finite element space to be

$$V^h = \{v^h \in H_0^1(0,1) : \ v^h \text{ is linear on } [x_j^h, x_{j+1}^h], j = 0, \ldots, n-1\},$$

and denote the set of nodal basis by $\{\phi_j^h\}_{j=1}^n$. The finite element approximation to the solution of (5) is the function $u^h \in V^h$, so that

$$a(u^h, v^h) = (f, v^h) \qquad \forall v^h \in V^h. \tag{6}$$

Let $u^h = \sum_{j=1}^n \mu_j \phi_j^h$ and $f = \sum_{j=1}^n \beta_j \phi_j^h$. Then (6) is equivalent to a linear system:

$$\mathcal{A}^h \mu = \mathcal{M}^h b,$$

where $\mu = (\mu_1, \ldots, \mu_n)^T$, $b = (\beta_1, \ldots, \beta_n)^T$, $\mathcal{A}^h$ is the stiffness matrix, and $\mathcal{M}^h$ is the mass matrix. Define $\tilde{\mathcal{A}}^h$ to be the augmented stiffness matrix that includes also the boundary points. Thus, $\tilde{\mathcal{A}}^h$ is singular with the null space consisting of constant functions, and $\mathcal{A}^h$ is a submatrix of it.

Let $x_i^H = x_{2i}^h, i = 0, \ldots, n/2$ be the set of coarse grid points. Now we define a coarse subspace V^H for multigrid by defining the coarse grid nodal basis functions $\{\phi_i^H\}$. That is,

$$V^H = \operatorname{span}\{\phi_i^H : i = 1, \ldots, m\},$$

and $m = n/2-1$. Since $\{\phi_i^H\}$ are nodal basis functions on the coarse grid, $\phi_i^H(x_{2i}^h) = 1$ and $\phi_i^H(x_{2i-2}^h) = \phi_i^H(x_{2i+2}^h) = 0$. We need only to define $\phi_i^H(x_{2i-1}^h)$ and $\phi_i^H(x_{2i+1}^h)$ (see Figure 1). For example, if we let them equal $1/2$, the basis functions $\{\phi_i^H\}$ are just linear finite elements, implying that the interpolation from the coarse grid to the fine grid is piecewise linear.

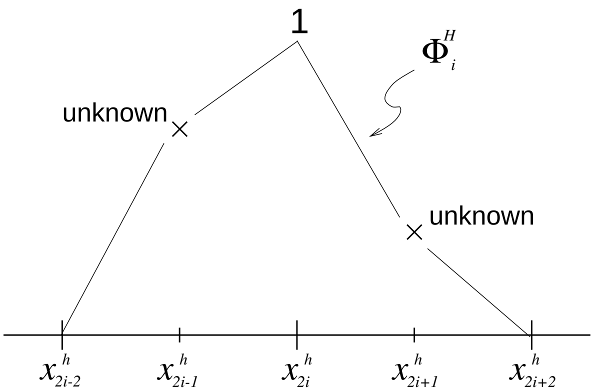

FIG. 1. *1D coarse grid basis function* ϕ_i^H *on its support* $[x_{2i-2}^h, x_{2i+2}^h]$.

Since $\{\phi_i^H\}$ is a basis of V^H, which is a subspace of V^h, there exists a unique matrix $\mathcal{I}_h^H$ of size $n \times m$ such that

$$[\phi_1^H \cdots \phi_m^H] = [\phi_1^h \cdots \phi_n^h]\mathcal{I}_h^H.$$

The matrix $\mathcal{I}_h^H$ is usually known as the prolongation (or interpolation) matrix and its transpose $(\mathcal{I}_h^H)^T = \mathcal{I}_H^h$ as the restriction matrix in the multigrid context. Hence, the set of coarse grid basis functions defines an interpolation and vice versa. In the following, instead of deriving an interpolation method directly, we construct an energy-minimizing basis.

Formulation. As noted above, the interpolation is uniquely defined if the coarse grid basis functions $\{\phi_i^H\}$ are known. We can define $\phi_i^H(x)$ by solving the following local PDE problem in $[x_{i-1}^H, x_i^H] = [x_{2i-2}^h, x_{2i}^h]$:

$$
\begin{aligned}
-\frac{d}{dx}a(x)\frac{d}{dx}\phi_i^H &= 0 \qquad\qquad \text{in } [x_{2i-2}^h, x_{2i}^h], \\
\phi_i^H(x_{2i-2}^h) = 0, \;\; \phi_i^H(x_{2i}^h) &= 1.
\end{aligned}
\tag{7}
$$

We observe that the PDE formulation of the basis functions has a "physical" meaning attached to it. Specifically, it looks for basis functions that have small energy. It is best illustrated by the following result.

LEMMA 3.1. *An equivalent formulation of (7) is*

$$
\begin{aligned}
&\min \;\; a(\phi_i^H, \phi_i^H) \qquad\qquad \textit{in } [x_{2i-2}^h, x_{2i}^h], \\
\textit{subject to} \quad &\phi_i^H(x_{2i-2}^h) = 0, \;\; \phi_i^H(x_{2i}^h) = 1.
\end{aligned}
\tag{8}
$$

Thus, the solution of the local PDE minimizes the energy of the coarse grid basis functions. This observation turns out to be very convenient for extending the idea to higher dimensions.

The solution of $\phi_i^H(x)$ on $[x_{2i-2}^h, x_{2i}^h]$ defines $\phi_i^H(x_{2i-1}^h)$ implicitly. We can do the same for $\phi_i^H(x_{2i+1}^h)$ in $[x_{2i}^h, x_{2i+2}^h]$. The local PDE formulation calculates the "harmonic" function ϕ_i^H which minimizes the energy on its support. If $a(x) \equiv 1$, ϕ_i^H is a linear function and we get back linear interpolation, that is, $\phi_i^H(x_{2i-1}^h) = \phi_i^H(x_{2i+1}^h) = 1/2$. In fact, in this case, ϕ_i^H is harmonic in the usual sense, and it has minimum energy. In general, instead of 1/2, we have

$$
\phi_i^H(x_{2i-1}^h) = -\frac{a(\phi_{2i-1}^h, \phi_{2i}^h)}{a(\phi_{2i-1}^h, \phi_{2i-1}^h)} = -\frac{\mathcal{A}^{(}_{\in\rangle-\infty,\in\rangle}}{\mathcal{A}^{(}_{\in\rangle-\infty,\in\rangle-\infty}},
\tag{9}
$$

where $(\mathcal{A}^{(}_{\rangle|})$ is the stiffness matrix. Since our interpolation depends on the matrix $\mathcal{A}^{(}$, sometimes it is called a matrix-dependent interpolation in the algebraic multigrid context. The resulting interpolation was also described in [20, 28, 36] but from a different point of view. Ours is novel in the sense that we interpret it from the energy-minimization principle, which provides a clue to developing similar interpolation operators in higher dimensions.

The approximation property (4) is closely related to preserving constant functions. In fact, the coarse space V^H constructed in this way automatically contains constant functions on the fine grid.

LEMMA 3.2.

$$
\sum_{i=1}^{m} \phi_i^H(x) = 1.
$$

Proof. Let $\psi^H(x) = \sum_{i=1}^m \phi_i^H(x)$. By (7), for $i = 1, \ldots, m$, ψ^H satisfies the following:

$$
\begin{aligned}
-\frac{d}{dx}a(x)\frac{d}{dx}\psi^H &= 0 \qquad\qquad \text{in } [x_{2i-2}^h, x_{2i}^h], \\
\psi^H(x_{2i-2}^h) = 1, \;\; \psi^H(x_{2i}^h) &= 1.
\end{aligned}
$$

By uniqueness, $\psi^H \equiv 1$ on $[x_{2i-1}^h, x_{2i}^h]$, and hence the result follows. $\square$

Thus, the interpolation derived from the energy-minimizing coarse grid basis functions preserves constants.

Remarks: (1) If $a(x)$ is piecewise constant, this interpolation preserves the continuity of the flux, $a(x)\nabla u$, at the discontinuities [20]. (2) If red-black Gauss-Seidel is used as a smoother, the resulting multigrid method coincides with the cyclic reduction method in the numerical linear algebra context.

3.2. Higher Dimensions. The construction of the energy-minimizing interpolation described in this section is valid for two and three dimensions. However, to facilitate understanding, we focus on the standard structured grid on the square domain Ω: $[0,1]\times[0,1]$ in two dimensions. The model problem is

$$
\begin{aligned}
-\nabla\cdot a(x,y)\nabla u(x,y) &= f(x,y), \qquad \text{in } \Omega \\
u &= 0 \qquad\qquad\quad \text{on } \partial\Omega,
\end{aligned}
\tag{10}
$$

with the same assumptions on $a(x,y)$ and $f(x,y)$ as before. Again, we use a finite element method to discretize (10).

3.2.1. Formulation. The extension to higher dimensions of the local PDE approach is difficult because there is no natural analog between one dimension and higher dimensions. For instance, in one dimension, the coarse grid points form the boundaries of the local subdomains so that well-posed PDEs can be easily defined. In higher dimensions, however, the boundaries consist of both coarse grid and noncoarse grid points, and hence local boundary value problems apparently do not exist. Nevertheless, several possibilities for setting up local PDEs are discussed in the literature, for instance, the stencil or the so-called black-box multigrid approach [1, 10, 11, 20, 19, 23, 24, 36, 40], the Schur complement approach [18, 25, 29], and the algebraic multigrid approach [8, 9, 35], each of which mimics the one-dimensional case in some way.

Our approach is based on the observation (8). The coarse grid basis functions $\{\phi_i^H\}$ should possess the least amount of energy while preserving constant functions. The precise mathematical formulation is explained in the following.

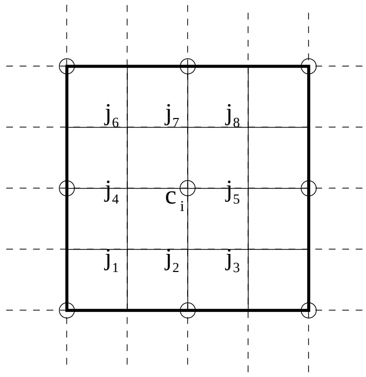

FIG. 2. *2D coarse grid basis function ϕ_i^H on its support. ϕ_i^H is a linear combination of fine grid basis functions ϕ_j^h, $j=j_1,\ldots,j_8$ and c_i.*

Suppose a maximal independent set of the vertices of the finer grid is selected as coarse grid points, and denote the index set by $M=\{c_1,\ldots,c_m\}$, $m=(n/2+1)^2$. Write the coarse grid nodal basis function ϕ_i^H at node x_{c_i} as a linear combination of the fine grid ones:

$$
\phi_i^H = \sum_{j\in\{l:\tilde{A}^h_{c_i,l}\neq 0\}\setminus M} \varphi_j^i\phi_j^h + \phi_{c_i}^h. \tag{11}
$$

Thus, ϕ_i^H is a local combination of the fine grid basis functions whose corresponding node is adjacent to node x_{c_i} but not itself a coarse grid point. Figure 2 shows the support of ϕ_i^H in two dimensions. The indices j in the sum on the right-hand side of (11) correspond to $j_1,\ldots,j_8$. Since ϕ_i^H is a nodal basis function, the coefficient of $\phi_{c_i}^h$ is equal to 1. We define the interpolation by solving a constrained minimization problem for $\{\varphi_j^i\}$:

$$
\min\ \frac{1}{2}\sum_{i=1}^m \|\phi_i^H\|_A^2 \qquad \text{subject to} \qquad \sum_{i=1}^m \phi_i^H(x) = 1 \text{ in } \bar{\Omega}. \tag{12}
$$

Notice that the minimization problem is solved up to and including the boundary of Ω. Usually, the grid points on the boundary with Dirichlet boundary condition are treated separately, and no coarse grid point is placed there. However, in our formulation, we compute all ϕ_i^H including the ones at the boundary, but only those not on the boundary with Dirichlet condition are used in the interpolation.

LEMMA 3.3. *An equivalent formulation of (7) and (8) is the global minimization*

$$\min \ \frac{1}{2}\sum_{i=1}^{m} \|\phi_i^H\|_A^2 \qquad \textit{subject to} \qquad \sum_{i=1}^{m} \phi_i^H(x) = 1 \ \textit{on} \ [0,1].$$

Thus, we see a way to naturally generalize the approach for generating a robust interpolation from one dimension to multiple dimensions.

Remarks: (1) The values of the basis functions are defined implicitly by the solution of (12) and are not known explicitly in general. However, for the Laplacian, we recover exactly the bilinear interpolation on tensor-product grids, which is known to lead to optimal multigrid convergence for Poisson equations.

LEMMA 3.4. *The solution of (12) gives the bilinear interpolation if $a(x) \equiv 1$.*

Proof. See the appendix. □

We also remark that if triangular grids are used, the linear interpolation is almost recovered; numerical experiments show that the interpolation values are close to 1/2.

(2) Like algebraic multigrid, the construction of the interpolation operator is purely algebraic. In other words, geometry and in particular the grid information are not needed. Besides, the formulation of the interpolation is still valid if the coarse grid points do not form an independent set. Independent sets are certainly beneficial to efficiency but are not necessary. In some situations, we may want to remove this requirement, for example, when semi-coarsening is used.

(3) Finally, we remark that we may generalize the formulation further by putting in positive weights θ_i in front of $\|\phi_i^H\|_A^2$. Similarly, we have the following equivalence.

LEMMA 3.5. *An equivalent formulation of (7) and (8) is the global weighted minimization*

$$\min \ \frac{1}{2}\sum_{i=1}^{m} \theta_i \|\phi_i^H\|_A^2 \qquad \textit{subject to} \qquad \sum_{i=1}^{m} \phi_i^H(x) = 1 \ \textit{on} \ [0,1],$$

for any sets of positive θ_i.

In our experience, special scalings, for instance, $\theta_i = 1/\tilde{\mathcal{A}}^h_{c_i,c_i}$, may improve the performance for problems such as discontinuous coefficient PDEs where the discontinuities do not align with any coarser grids. However, an optimal choice of θ_i has not yet been fully analyzed, and hence we shall not discuss this generalization further in the present paper.

3.2.2. Solution of the Minimization Problem. We describe a solution procedure for the minimization problem (12) below. For each i, write $\phi_i^H = \sum_{j=1}^n \varphi_j^i \phi_j^h$ and $\varphi^i = (\varphi_1^i, \ldots, \varphi_n^i)^T$. By (11), ϕ^i is a sparse vector. For example, in two dimensions, φ^i has at most 9 nonzeros. For structured triangular grids, φ^i has at most 7 nonzeros. Let $\Phi = [\varphi^1; \cdots; \varphi^m]$ be an $mn \times 1$ vector obtained by appending all the φ's. Note that $\|\phi_i^H\|_A^2 = \|\sum_{j=1}^n \varphi_j^i \phi_j^h\|_A^2 = (\varphi^i)^T \tilde{\mathcal{A}}^h \varphi^i$. (Recall that $\tilde{A}^h$ is the augmented stiffness matrix on the fine grid without incorporating any Dirichlet boundary condition.) Thus, (12) can be written as the following equivalent discrete linear constrained quadratic minimization problem.

$$\min \ \frac{1}{2}\Phi^T \mathcal{Q} \Phi \quad \text{s.t.} \quad \mathcal{B}^T \Phi = \mathbf{1}. \tag{13}$$

The symbol $\mathbf{1}$ denotes a vector of all 1's. The $mn \times mn$ SPD matrix $\mathcal{Q}$ is block diagonal with each block equal to $\tilde{\mathcal{A}}_i^h$ which is defined as

$$(\tilde{\mathcal{A}}_i^h)_{kl} = \begin{cases} \tilde{\mathcal{A}}_{kl}^h & \text{if } \varphi_k^i \neq 0 \text{ and } \varphi_l^i \neq 0 \\ \delta_{kl} & \text{otherwise.} \end{cases}$$

The $n \times mn$ rectangular matrix $\mathcal{B}^{\mathcal{T}} = [\mathcal{J}_\infty^{\mathcal{T}} \cdots \mathcal{J}_{\Updownarrow}^{\mathcal{T}}]$, where $\mathcal{J}_\rangle = \mathcal{J}_\rangle^{\mathcal{T}}$ is a matrix corresponding to the restriction operator that maps v to v_i such that $(v)_k = (v_i)_k$ on $\mathrm{supp}(\phi_i^H)$ and $(v_i)_k = 0$ otherwise. More precisely,

$$(\mathcal{J}_\rangle)_{\|\updownarrow} = \begin{cases} 1 & \text{if } k = l \text{ and } \varphi_k^i \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that $\mathcal{J}_\rangle^{\mathcal{T}} \varphi^\rangle = \varphi^\rangle$ and hence $\mathcal{B}^{\mathcal{T}} \oplus = \sum_{\rangle=\infty}^{\Updownarrow} \mathcal{J}_\rangle^{\mathcal{T}} \varphi^\rangle = \sum_{\rangle=\infty}^{\Updownarrow} \varphi^\rangle = \mathbf{1}$. We solve the discrete linearly constrained minimization problem (13) by the Lagrange multiplier formulation, which is equivalent to

$$\begin{bmatrix} \mathcal{Q} & \mathcal{B} \\ \mathcal{B}^{\mathcal{T}} & 0 \end{bmatrix} \begin{bmatrix} \Phi \\ \Lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}, \tag{14}$$

where Λ is an $n \times 1$ vector of Lagrange multipliers. If Λ is known, Φ can be computed by solving

$$\mathcal{Q}\Phi = -\mathcal{B}\Lambda. \tag{15}$$

Since $\mathcal{Q}$ is block diagonal and inverting each block corresponds to solving a matrix of at most 9×9 in size, it is trivial to compute Φ once Λ is known. Thus, the entire minimization procedure is reduced to solving for the Lagrange multipliers Λ via

$$(\mathcal{B}^T \mathcal{Q}^{-1} \mathcal{B})\Lambda = -\mathbf{1}. \tag{16}$$

Note that $\mathcal{B}$ and $\mathcal{Q}^{-\infty}$ are sparse matrices. We can solve the linear system by conjugate gradient (CG).

The solution process of (16) could be costly. Depending on the conditional number of $\mathcal{Q}^{-\infty}$, the CG iteration may converge slowly. We shall discuss how to speed up the process. First, we need not compute $(\mathcal{B}^T \mathcal{Q}^{-1} \mathcal{B})^{-1}\mathbf{1}$ exactly because we are merely computing the interpolation to be used in the multigrid method. In fact, the numerical results in Section 5 indicate that Λ is usually accurate enough when the relative residual of (16) is less then 10^{-2}.

Besides, we have a readily obtainable initial guess for Λ. Consider equation (15). Multiplying both sides by $\mathcal{B}^T$, we have

$$\Lambda = -(\mathcal{B}^T \mathcal{B})^{-1} \mathcal{Q}\Phi.$$

As shown in the proof of Lemma A.2 in the appendix, $\mathcal{B}^T\mathcal{B}$ is a diagonal matrix. Hence, this gives an easy way to compute an initial guess for Λ from Φ. Since the interpolation weights are between 0 and 1, the solution Φ usually is not very far from the linear interpolation. It may be advantageous to use the linear interpolation as an initial guess for Φ, which in turn provides an initial guess for Λ.

It is interesting to note that $\tilde{\mathcal{A}}^h$ is a free and natural preconditioner for $\mathcal{B}\mathcal{Q}^{-\infty}\mathcal{B}$. By the definition by $\mathcal{B}$ and $\mathcal{Q}$, rewrite the product $\mathcal{B}^T \mathcal{Q}^{-\infty} \mathcal{B}$ as a sum of matrices:

$$\mathcal{B}^T \mathcal{Q}^{-\infty} \mathcal{B} = \sum_{i=1}^{m} \mathcal{J}_i^T (\tilde{\mathcal{A}}_i^h)^{-1} \mathcal{J}_i = \sum_{i=1}^{m} \mathcal{R}_i^T (\mathcal{R}_i \tilde{\mathcal{A}}^h \mathcal{R}_i^T)^{-1} \mathcal{R}_i,$$

where $\mathcal{R}_i$ is the submatrix of the nonzero rows of $\mathcal{J}_i$ and it is sometimes known as the restriction matrix in the domain decomposition context. Clearly, $\mathcal{B}^T \mathcal{Q}^{-\infty} \mathcal{B}$ is an overlapping additive Schwarz preconditioner of $\tilde{\mathcal{A}}^h$. Unfortunately, $\tilde{\mathcal{A}}^h$ is singular in our case. A simple remedy is to use $\tilde{\mathcal{A}}^h + \eta\mathcal{I}$ instead as the preconditioner.

Because of the potentially high cost of computing Φ, the energy-minimizing interpolation is aimed at problems for which linear interpolation does not work well. Quite often, we may need to solve the same system many times, for instance, in time-dependent problems. The expensive setup cost can be compensated by the rapid convergence of each multigrid solve.

3.2.3. Connections to Other Approaches. As noted above, the entire procedure of constructing the interpolation is algebraic, and so it can be considered as a type of algebraic multigrid. In fact, it is related to the one derived by Vanek, Mandel, and Brezina [35]. In their approach, groups of fine grid elements are agglomerated to form larger elements, or macroelements. In each agglomerated region (which can be thought of a subdomain in the domain decomposition context), a value of 1 is assigned to each node as an initial guess of the coarse grid basis. Because of the high energy of the piecewise constant basis functions, they are *smoothed* by a few steps of Jacobi iteration. Our energy-minimizing coarse grid basis can also be thought of being formed by agglomerating nearby fine grid elements, but the agglomeration only occurs at elements whose node is a coarse grid node. Also there are overlaps among agglomerated regions, while there is none in the approach of Vanek et al. Moreover, the support of their basis functions will increase when the Jacobi "smoothing" steps are applied to the basis functions. In our approach, the supports are fixed and the energy is minimized by solving the minimization problem (12).

Because of the agglomeration view of the construction, our approach is also related to the one derived by Chan et al. [8, 9]. They explicitly form the macroelements by agglomeration using standard graph theoretical techniques. Then they have several way of defining the coarse grid basis functions. One way is the following. The noncoarse grid points on the edge of a macroelement are assigned a value using the graph distance, and those noncoarse grid points in the interior are obtained by solving a local homogeneous PDE. Our approach does not prescribe a value on the edges of the macroelements first and then solve for the interior points. Rather, we take all the unknowns together and solve for all the values simultaneously by solving the minimization problem.

4. Convergence Analysis. Much of the classical multigrid convergence analysis cannot be applied directly to the proposed multigrid algorithm because the coarse spaces defined by the basis functions are not standard finite element spaces. The one-dimensional analysis is complete and is presented here.

First, we show the stability property (3). The proof is based on the observation that the coarse grid basis functions contain a hierarchy of A-orthogonal basis functions; in other words, they are orthogonal in the A-inner product. Note that the coarse grid points are chosen to be the even fine grid points (see Section 3.1), that is, $x_i^{k-1} = x_{2i}^k$.

LEMMA 4.1. *For any $l < k, i = 1, \ldots, n_l, j = 1, \ldots, n_k/2$, we have*

$$a(\phi_i^l, \phi_{2j-1}^k) = 0. \tag{17}$$

Proof. Let k be fixed. We first prove the case $l = k-1$ using a technique suggested by Xu [37]. In this case, (17) is just the direct consequence of the fact that the equivalent variational formulation of (7) implies that $a(\phi_i^{k-1}, \phi_{2i-1}^k) = 0$, and the support of ϕ_i^{k-1} is only on $[x_{2i-2}^k, x_{2i+2}^k]$.

Now suppose it is true for $l = \bar{k}$. By definition,

$$\phi_i^{\bar{k}-1} = \sum_{j=2i-1}^{2i+1} \alpha_j \phi_j^{\bar{k}},$$

where $\alpha_{2i} = 1$, $\alpha_{2i-1} = \phi_i^{\bar{k}-1}(x_{2i-1}^{\bar{k}})$, and $\alpha_{2i+1} = \phi_i^{\bar{k}-1}(x_{2i+1}^{\bar{k}})$ are given by (9). Thus

$$a(\phi_i^{\bar{k}-1}, \phi_{2j-1}^k) = \alpha_{2i-1} a(\phi_{2i-1}^{\bar{k}}, \phi_{2j-1}^k) + a(\phi_{2i}^{\bar{k}}, \phi_{2j-1}^k) + \alpha_{2i+1} a(\phi_{2i+1}^{\bar{k}}, \phi_{2j-1}^k) = 0,$$

since all the terms vanish by assumption. The result follows from induction. □

Lemma 4.1 implies that the interpolation algorithm generates implicitly a set of A-orthogonal hierarchical basis functions. The orthogonality property immediately implies the stability of the nested subspaces.

LEMMA 4.2. *For any $v \in V$, there is a nontrivial decomposition $v = \sum_{k=1}^J v_k$ with $v_k \in V_k$ such that*

$$\sum_{k=1}^J (v_k, v_k)_A = (v, v)_A. \tag{18}$$

Proof. For any $v \in V$, Lemma 4.1 implies that there exists an orthogonal hierarchical decomposition of v constructed as follows. We first define v_1 to be the nodal value interpolant of v at the coarsest level V_1. Then we subtract v_1 from v to obtain w_2. Because of the nodal interpolation, the values of w_2 at $x_i^1, i = 1, \ldots, n_1$, are zero. We proceed similarly by defining v_2 to be the nodal value interpolant of w_2 and so on. Formally, we have the following:

$$v_1 = \sum_{i=1}^{n_1} v(x_i^1)\phi_i^1 \qquad \text{and} \qquad v_k = \sum_{i=1}^{n_k} w_k(x_i^k)\phi_i^k, \qquad k = 2, \ldots, J, \tag{19}$$

where $w_k = v - \sum_{i=1}^{k-1} v_i$. Our decomposition implies that $v_k(x_j^k) = w_k(x_j^k) = 0, j$ even. Therefore, by Lemma 4.1, the v_k's are A-orthogonal, since if $l < k$, we have

$$\begin{aligned} a(v_l, v_k) &= a(\sum_{i=1}^{n_l} w_l(x_i^l)\phi_i^l, \sum_{j=1}^{n_k} w_k(x_j^k)\phi_j^k) \\ &= \sum_{i=1}^{n_l}\sum_{j=1}^{n_k/2} w_l(x_i^l) w_k(x_{2j-1}^k) a(\phi_i^l, \phi_{2j-1}^k) \\ &= 0. \end{aligned}$$

The equality (18) follows immediately from the orthogonality of v_k's. □

COROLLARY 4.3. *Let $W_1 = V_1$ and $W_k = V_k \ominus V_{k-1}, k = 2, \ldots, J$, in the A-inner product. Then V can be expressed as a direct sum of W_k's:*

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_J.$$

Corollary 4.3 induces a projection operator $\tilde{Q}_k : V \to V_k$ defined by

$$\tilde{Q}_k v = v_1 + v_2 + \cdots + v_k, \tag{20}$$

where $v = v_1 + \cdots + v_J, v_k \in W_k$, is the unique representation of v defined in (19). This operator $\tilde{Q}_i$ will be used to prove the approximation property (4). Here we do not use the L^2 projection Q_k because $\tilde{Q}_k$ is a more natural and convenient choice in the one-dimensional case. In view of Lemma 4.2 and Corollary 4.3, the stability property (3) is satisfied.

In the literature, the approximation property (4) is typically proved by making use of the fact that the interpolation preserves constant functions. In the two level case, we have shown in Lemma 3.2 that constant functions are indeed preserved by the coarse grid basis functions in our case. Using the same proof technique, we can easily show that it is also true for the multilevel case.

LEMMA 4.4. *For any $k = 1, \ldots, J$,*

$$\sum_{i=1}^{n_k} \phi_i^k(x) = 1.$$

With this result, we can now prove the approximation property (4).

LEMMA 4.5. *For any $v \in V$ and any $k = 2, \ldots, J$,*

$$\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\| \le C h_k \|\tilde{Q}_k v\|_A. \tag{21}$$

Proof. We compute the quantities on both sides explicitly to see how preserving constant functions comes into play. Since $\tilde{Q}_k$ is a projection, we can always change v to $\tilde{Q}_k v$ in the left-hand side of (21). Without loss of generality, we assume $v \in V_k$, that is, $v = \sum_{i=1}^{n_k} \nu_i \phi_i^k$. This means we need only prove

$$\|v - \tilde{Q}_{k-1} v\| \le C h_k \|v\|_A.$$

By the definition of $\tilde{Q}_{k-1}$ in (20),

$$\tilde{Q}_{k-1}v = \sum_{i=1}^{n_{k-1}} \nu_{2i}\phi_i^{k-1}.$$

Let $w = v - \tilde{Q}_{k-1}v = \sum_{i=1}^{n_k} \omega_i \phi_i^k$. Then we can verify that

$$\begin{aligned}
\omega_{2i} &= 0, \qquad i = 1, \ldots, n_k/2, \\
\omega_{2i-1} &= \nu_{2i-1} - (\alpha\nu_{2i-2} + \beta\nu_{2i}),
\end{aligned}$$

where $\alpha = \phi_{i-1}^{k-1}(x_{2i-1}^k)$ and $\beta = \phi_i^{k-1}(x_{2i-1}^k)$. Because the coarse grid basis functions preserve constant, we have $\alpha + \beta = 1$ and hence

$$\begin{aligned}
\omega_{2i-1} &= (\alpha+\beta)\nu_{2i-1} - (\alpha\nu_{2i-2} + \beta\nu_{2i}), \\
&= \alpha(\nu_{2i-1} - \nu_{2i-2}) - \beta(\nu_{2i} - \nu_{2i-1}).
\end{aligned}$$

Now we estimate the L^2 norm of w on $[x_{2i-2}^k, x_{2i}^k]$:

$$\begin{aligned}
\int_{x_{2i-2}^k}^{x_{2i}^k} w^2 dx &= \int_{x_{2i-2}^k}^{x_{2i}^k} (\omega_{2i-1}\phi_{2i-1}^k)^2 dx \\
&= \omega_{2i-1}^2 \int_{x_{2i-2}^k}^{x_{2i}^k} (\phi_{2i-1}^k)^2 dx \\
&= \omega_{2i-1}^2 \mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty},
\end{aligned}$$

where $\mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}$ is the $(2i-1, 2i-1)$ entry of $\mathcal{M}_{\|}$, which is the mass matrix with respect to $\{\phi_i^k\}_{i=1}^{n_k}$. Using the formula for ω_{2i-1} and the elementary inequality: $(\alpha A - \beta B)^2 \le \alpha A^2 + \beta B^2$ for $\alpha + \beta = 1$, we have that

$$\begin{aligned}
\int_{x_{2i-2}^k}^{x_{2i}^k} w^2 dx &= \mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}[\alpha(\nu_{\in\rangle-\infty} - \nu_{\in\rangle-\in}) - \beta(\nu_{\in\rangle} - \nu_{\in\rangle-\infty})]^{\in} \\
&\le \mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}[\alpha(\nu_{\in\rangle-\infty} - \nu_{\in\rangle-\in})^{\in} + \beta(\nu_{\in\rangle} - \nu_{\in\rangle-\infty})^{\in}].
\end{aligned} \tag{22}$$

But the A-norm of v is given by

$$\begin{aligned}
&\int_{x_{2i-2}^k}^{x_{2i}^k} a(x)(v')^2 dx \\
= &\int_{x_{2i-2}^k}^{x_{2i-1}^k} a(x)(\nu_{2i-2}{\phi_{2i-2}^k}' + \nu_{2i-1}{\phi_{2i-1}^k}')^2 dx + \int_{x_{2i-1}^k}^{x_{2i}^k} a(x)(\nu_{2i-1}{\phi_{2i-1}^k}' + \nu_{2i}{\phi_{2i}^k}')^2 dx \\
= &-(\nu_{2i-1} - \nu_{2i-2})^2 \int_{x_{2i-2}^k}^{x_{2i-1}^k} a(x){\phi_{2i-2}^k}'{\phi_{2i-1}^k}' dx - (\nu_{2i} - \nu_{2i-1})^2 \int_{x_{2i-1}^k}^{x_{2i}^k} a(x){\phi_{2i-1}^k}'{\phi_{2i}^k}' dx,
\end{aligned}$$

since $\phi_{2i-2}^k(x) + \phi_{2i-1}^k(x) = 1$ on $[x_{2i-2}^k, x_{2i-1}^k]$, which implies ${\phi_{2i-2}^k}'(x) + {\phi_{2i-1}^k}'(x) = 0$. A similar argument also holds for the second integral. Together with the formulas for α and β in (9), we have

$$\int_{x_{2i-2}^k}^{x_{2i}^k} a(x)(v')^2 dx = [\alpha(\nu_{2i-1} - \nu_{2i-2})^2 + \beta(\nu_{2i} - \nu_{2i-1})^2]\mathcal{A}^{\|}_{\in\rangle-\infty,\in\rangle-\infty},$$

where $\mathcal{A}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}$ is the $(2i-1, 2i-1)$ entry of $\mathcal{A}_{\|}$ which is the stiffness matrix at level k with respect to the basis $\{\phi_i^k\}_{i=1}^{n_k}$. Combining with (22), we have

$$\int_{x_{2i-2}^k}^{x_{2i}^k} w^2 dx \le \frac{\mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}}{\mathcal{A}^{\|}_{\in\rangle-\infty,\in\rangle-\infty}} \int_{x_{2i-2}^k}^{x_{2i}^k} a(x)(v')^2 dx. \tag{23}$$

It is easy to show that

$$\mathcal{M}^{\|}_{\in\rangle-\infty,\in\rangle-\infty} = \int_{\S^{\|}_{\in\rangle-\in}}^{\S^{\|}_{\in\rangle}} (\phi^{\|}_{\in\rangle-\infty})^{\in} \lceil \S \leq \mathcal{O}(\langle_{\|}),$$

and

$$\mathcal{A}^{\|}_{\in\rangle-\infty,\in\rangle-\infty} = \int_{\S^{\|}_{\in\rangle-\in}}^{\S^{\|}_{\in\rangle}} \dashv(\S)(\phi^{\|}_{\in\rangle-\infty}{}')^{\in} \lceil \S \geq \mathcal{O}(\langle^{-\infty}_{\|}).$$

Thus, (23) becomes

$$\int_{x^k_{2i-2}}^{x^k_{2i}} w^2 dx \leq Ch_k^2 \int_{x^k_{2i-2}}^{x^k_{2i}} a(x)(v')^2 dx,$$

where C is independent of h_k. Summing over i, we obtain the approximation property. $\square$

Hence, by Lemma 2.2, K_0 is bounded by a constant independent of the mesh size h, although the constant may depend on the coefficient $a(x)$. The coefficient dependence comes from the bound given by the approximation property (see Lemma 4.5). It turns out that we can eliminate the coefficient dependence by estimating K_0 directly from its definition (1) if the damped Jacobi or Gauss-Seidel smoothings are used.

LEMMA 4.6. *Let R_k^{DJ} and R_k^{GS} be the approximate inverses of A_k given by the damped Jacobi method and the Gauss-Seidel method, respectively:*

$$\begin{aligned} (R_k^{DJ} v_k, v_k) &= \frac{1}{\rho((\mathcal{D}^k)^{-1}\mathcal{A}^k)} (\nu^k)^T (\mathcal{D}_k)^{-1} \nu^k, \\ (R_k^{GS} v_k, v_k) &= (\nu^k)^T (\mathcal{D}^k - \mathcal{L}^k)^{-1} \nu^k, \end{aligned}$$

where $\mathcal{A}^k = \mathcal{D}^k - \mathcal{L}^k - (\mathcal{L}^k)^T, \mathcal{D}^k =$ diagonal of $\mathcal{A}^k$ and $v_k = \sum_{j=1}^{n_k} \nu_j^k \phi_j^k$. Then

$$K_0 \leq 3,$$

for the damped Jacobi smoothing, and

$$K_0 = 1,$$

for the Gauss-Seidel smoothing, and they are both independent of the coefficient of the PDE.

Proof. This proof is a modification of the proof of Lemma A.1. For any $v \in V$, let $v = \sum_{k=1}^J v_k$ be a decomposition of v given by (3), that is, $v_k = (\tilde{Q}_k - \tilde{Q}_{k-1})v$. We first estimate K_0 for the damped Jacobi smoothing and then for Gauss-Seidel smoothing. In view of the definition of K_0, for each $k > 1$, we consider

$$((R_k^{DJ})^{-1} v_k, v_k) = \rho((\mathcal{D}^k)^{-1}\mathcal{A}^k)\ (\nu^k)^T \mathcal{D}^k \nu^k.$$

From the calculations in the proof of Lemma 4.5, we showed that $\nu_j^k = 0$, j even. Thus

$$(\nu^k)^T \mathcal{D}^k \nu^k = (\nu^k)^T \mathcal{A}^k \nu^k = \|v_k\|_A^2.$$

The estimate of $\rho((\mathcal{D}^k)^{-1}\mathcal{A}^k) = \rho((\mathcal{D}^k)^{-1/2}\mathcal{A}^k(\mathcal{D}^k)^{-1/2})$ is purely algebraic. Note that the product $(\mathcal{D}^k)^{1/2}\mathcal{A}^k(\mathcal{D}^k)^{1/2}$ is simply the matrix obtained by the diagonal scaling of $\mathcal{A}^k$. Hence, it is still tridiagonal, and its diagonal elements are all 1's. The element of the $(i, i+1)$th entry is given by $\mathcal{A}^k_{i,i+1}/\sqrt{\mathcal{A}^k_{i,i}\mathcal{A}^k_{i+1,i+1}}$. Since $\mathcal{A}^k$ is SPD, it is easy to show that the $(i, i+1)$th entry is bounded by 1 in size. By the Gershgorin Circle Theorem, $\rho((\mathcal{D}^k)^{-1}\mathcal{A}^k) \le 3$. Hence

$$\begin{aligned}\sum_{k=1}^{J}((R_k^{DJ})^{-1}v_k, v_k) &= \|v_1\|_A^2 + \sum_{k=2}^{J}((R_k^{DJ})^{-1}v_k, v_k) \\ &\le \|v_1\|_A^2 + 3\sum_{k=2}^{J}\|v_k\|_A^2 \\ &\le 3\|v\|_A^2.\end{aligned}$$

By the definition of K_0, the estimate follows.

Similarly, for the Gauss-Seidel smoothing, we consider

$$\begin{aligned}((R_k^{GS})^{-1}v_k, v_k) &= (\nu^k)^T(\mathcal{D}^k - \mathcal{L}^k)\nu^k \\ &= \frac{1}{2}(\nu^k)^T\mathcal{D}^k\nu^k + \frac{1}{2}(\nu^k)^T\mathcal{A}^k\nu^k \\ &= (\nu^k)^T\mathcal{A}^k\nu^k \\ &= \|v_k\|_A^2.\end{aligned}$$

Thus, $K_0 = 1$, since

$$\sum_{k=1}^{J}((R_k^{GS})^{-1}v_k, v_k) = \sum_{k=1}^{J}\|v_k\|_A^2 = \|v\|_A^2.$$

□

For the estimate for K_1, instead of the V_k, we consider the W_k defined in Corollary 4.3. It is not hard to see that all the previous results still hold. In addition, we have $P_iP_j = 0$, for any $i \neq j$.

LEMMA 4.7. *Let ω_1 be the smallest constant such that*

$$(A_kv_k, v_k) \le \omega_1(R_k^{-1}v_k, v_k) \qquad \forall v_k \in V_k.$$

Then

$$K_1 \le \omega_1.$$

If $R_k = R_k^{DJ}$, then $\omega_1 = 1$. If $R_k = R_k^{GS}$, then $\omega_1 < 2$.

Proof. The bound for K_1 is a direct consequence of Lemma 4.6 in [39] and the fact that $P_iP_j = 0$ for $i \neq j$. If $R_k = R_k^{DJ}$, then

$$\omega_1 = \rho(R_k^{DJ}A_k) = \frac{1}{\rho((\mathcal{D}^k)^{-1}\mathcal{A}^k)}\rho((\mathcal{D}^k)^{-1}\mathcal{A}^k) = 1.$$

If $R_k = R_k^{GS}$, then

$$\frac{(A_kv_k, v_k)}{((R_k^{GS})^{-1}v_k, v_k)} = \frac{(\nu^k)^T\mathcal{A}^k\nu^k}{\frac{1}{2}(\nu^k)^T\mathcal{D}^k\nu^k + \frac{1}{2}(\nu^k)^T\mathcal{A}^k\nu^k} < 2.$$

Hence $\omega_1 < 2$. □

5. Numerical Results. In this section, we present results of numerical experiments mainly in two dimensions to verify that the multigrid algorithm resulting from the energy-minimizing interpolation has optimal convergence behavior and is robust with respect to the coefficients of the PDEs. In all the numerical examples, the computational domain is $\Omega = [0,1] \times [0,1]$ with homogeneous Dirichlet boundary condition. In the multigrid procedure, a V-cycle is used with two pre- and two post-pointwise Gauss-Seidel smoothings. The iteration was terminated when the relative residual norm was less than 10^{-6}. The number of multigrid levels is such that the coarsest grid is a single point, or as otherwise stated.

In Section 3.2.1, we mentioned that it is not necessary to compute the Lagrange multipliers to machine precision. In all cases discussed below, we used piecewise linear or bilinear interpolation as our initial guess for the minimization problem. In the numerical results, we show how the accuracy of the Lagrange multipliers affect the efficiency and convergence of the resulting multigrid method. Moreover, as discussed in Section 3.2.2, the augmented stiffness matrix $\tilde{\mathcal{A}}^h$, or more precisely, $\tilde{\mathcal{A}}^h + \eta\mathcal{I}$, is a free preconditioner for solving the Lagrange multiplier equation (16). In the numerical examples, this preconditioner is used with η chosen as 10^{-3}.

Example 1: In the appendix, we prove that the energy-minimizing interpolation recovers the bilinear interpolation if $a(x) \equiv 1$ in the case when the structured square grid is used. But linear interpolation is not exactly obtained in the triangular grid case. In this example, we solve the Poisson equation

$$-\Delta u = 1,$$

on the triangular grid. The result is shown in Table 1. We vary the grid size from $h = 1/16$ to $h = 1/64$ and the number of multigrid levels from 3 to 6. We see that both the linear and the energy-minimizing interpolations give a convergence rate independent of the mesh size and the number of multigrid level.

	Linear				Energy-min			
h	3	4	5	6	3	4	5	6
1/16	7	7	-	-	7	7	-	-
1/32	6	7	7	-	6	7	7	-
1/64	6	7	7	7	6	7	7	7

Table 1

Number of V-cycles using linear and energy-minimizing interpolations when $a(x) \equiv 1$.

Example 2: In this example, we verify numerically that the convergence rate does not depend on the number of levels. Here we consider the following PDE with a smooth coefficient:

$$-\nabla \cdot (1 + x\exp(y))\nabla u = 1.$$

Table 2 shows the number of multigrid iterations to convergence. We denote the multigrid method with bilinear interpolation by MGBL and our energy-minimizing multigrid method by MGE(ϵ), where ϵ specifies the stopping criterion for the conjugate gradient (CG) method applied to the Lagrange multiplier equation (16). More precisely, the CG iteration is stopped when the relative residual norm is less than ϵ. We see that when the optimization problem is effectively solved ($\epsilon = 10^{-12}$), the convergence rate is independent of the mesh size h and the number of levels. In fact, we observe that same convergence rate can be achieved even if the optimization problem is solved approximately ($\epsilon = 10^{-1}$). Thus, we may reduce the cost by applying significantly fewer number of CG iterations as shown in Table 3, which gives the number of conjugate gradient iterations at each multigrid level to solve (16).

We remark that this example is used to illustrate the optimal convergence of MGE(ϵ) and the effect of varying ϵ only. It is not cost effective to use energy-minimizing interpolation when bilinear interpolation works well.

Example 3: We compare the multigrid method using bilinear interpolation with that using energy-minimizing interpolation by solving the following discontinuous coefficient problem [1, modified Example I]:

$$-\nabla \cdot a(x,y)\nabla u = 1,$$

where

$$a(x,y) = \begin{cases} a^+ & 0.25 \le x \le 0.75 \quad \& \quad 0.25 \le y \le 0.75 \\ a^- & \text{otherwise.} \end{cases}$$

h	MGBL				MGE(10^{-1})				MGE(10^{-12})			
	4	5	6	7	4	5	6	7	4	5	6	7
1/16	5	-	-	-	5	-	-	-	5	-	-	-
1/32	5	5	-	-	5	5	-	-	5	5	-	-
1/64	5	5	5	-	5	5	5	-	5	5	5	-
1/128	5	5	5	5	5	5	5	5	5	5	5	5

TABLE 2

Number of V-cycles using bilinear and energy-minimizing interpolations when $a(x) = 1 + x\exp(y)$.

h	level	MGE(10^{-1})	MGE(10^{-12})
1/16	4	1	53
	2	1	22
1/32	5	1	98
	3	1	33
	2	1	22
1/64	6	1	180
	4	1	53
	2	1	22
1/128	7	1	309
	5	1	98
	3	1	33
	2	1	22

TABLE 3

Number of CG iterations at each multigrid level with varying ϵ when $a(x) = 1 + x\exp(y)$.

We fix $a^- = 1$ and vary a^+ from 10 to 10^4. The convergence results are given in Table 4. Same notations are used as in Example 1. Here $*$ denotes convergence beyond 100 multigrid iterations. Consistent with the classical theory, the convergence rate of the standard multigrid does not depend on the mesh size h. However, the convergence rate deteriorates substantially as the jump of the discontinuity increases. On the other hand, the convergence of the energy-minimizing multigrid method does not depend both on the mesh size and the size of the jump. Again, MGE(10^{-1}) shows similar convergence as MGE(10^{-12}).

Table 5 shows the average number of CG iterations on the fine grid, in place of the number of CG iterations on each grid level shown in Table 3. It is computed as follows. One CG iteration on the first coarse grid is counted as 1/2 CG iteration on the fine grid and so on. By applying only three extra CG iterations to construct the energy-minimizing interpolation, the convergence of the multigrid is improved significantly. This result demonstrates that extra cost of solving the minimization problem is justified by the much faster convergence of the multigrid method.

h	MGBL				MGE(10^{-1})				MGE(10^{-12})			
	10	10^2	10^3	10^4	10	10^2	10^3	10^4	10	10^2	10^3	10^4
1/16	14	$*$	$*$	$*$	6	5	6	6	6	5	5	5
1/32	14	$*$	$*$	$*$	6	6	6	6	6	6	6	6
1/64	14	$*$	$*$	$*$	6	6	7	7	6	6	6	6
1/128	14	$*$	$*$	$*$	7	7	7	7	7	6	6	6

TABLE 4

Number of V-cycles using bilinear and energy-minimizing interpolations for the discontinuous coefficient problem. The jump $a^+ = 10, 10^2, 10^3, 10^4$. $*$ More than 100 V-cycles required for convergence.

Example 4: We solve another PDE to demonstrate the robustness of the energy-minimizing multigrid

h	MGE(10^{-1})		MGE(10^{-12})	
	10	10^4	10	10^4
1/16	3.00	3.00	3.00	5.50
1/32	2.50	2.50	2.50	3.63
1/64	2.25	2.25	2.25	2.88
1/128	2.13	2.13	2.13	2.44

TABLE 5

Average number of CG iterations on the fine grid for the discontinuous coefficient problem. The jump $a^+ = 10, 10^4$.

method. The coefficient is oscillatory, and the equation is [22, Example 7.4]:

$$-\nabla \cdot \frac{1}{(2+P\sin(x/\epsilon))(2+P\sin(y/\epsilon))}\nabla u = 1.$$

We chose $P = 1.99$ and ϵ=0.1 and 0.01. The results are shown in Tables 6 and 7. This time, the coefficient is very rough, and the minimization problem is more difficult to solve. In the case, MGE(10^{-1}) is not accurate enough to have good convergence. However, with a slight increase in the accuracy, MGE(10^{-2}) recovers the same rapid convergence of MGE(10^{-12}).

We remark that the nonuniform number of V-cycles to convergence for the case $\epsilon = 0.01$ may be because the mesh size h is not small enough to resolve the coefficient $a(x, y)$ for the first couple of values of h.

h	MGBL		MGE(10^{-1})		MGE(10^{-2})		MGE(10^{-12})	
	0.1	0.01	0.1	0.01	0.1	0.01	0.1	0.01
1/16	$*$	4	7	5	7	5	7	5
1/32	51	$*$	23	14	7	14	7	14
1/64	65	58	$*$	11	7	7	7	7
1/128	66	$*$	$*$	11	7	10	7	10

TABLE 6

Number of V-cycles using bilinear and energy-minimizing interpolations for the oscillatory coefficient problem. $\epsilon = 0.1, 0.01$. $*$ *More than 100 V-cycles required for convergence.*

h	MGE(10^{-1})		MGE(10^{-2})		MGE(10^{-12})	
	0.1	0.01	0.1	0.01	0.1	0.01
1/16	38.75	1.75	42.25	1.75	90.75	56.75
1/32	13.63	63.88	62.38	71.00	124.13	184.13
1/64	3.81	82.94	115.94	127.94	228.06	308.69
1/128	2.22	211.22	177.94	316.22	388.78	664.22

TABLE 7

Average number of CG iterations on the fine grid for the oscillatory coefficient problem. $\epsilon = 0.1, 0.01$.

Example 5: We show by a one-dimensional Helmholtz equation that the energy minimization principle is not restricted to positive definite second order elliptic PDEs. The model equation is

$$\Delta u + \alpha u = 1, \tag{24}$$

where α is a positive constant. This operator is indefinite.

We use multigrid to solve the linear system $\mathcal{A}^h$. For this problem, we obtained ϕ_i^H from solving the local PDEs (7), not from the minimization problem (12), since constant functions are not in the

h	Linear	Energy
1/32	$*$	5
1/64	$*$	5
1/128	$*$	5

TABLE 8

Number of V-cycles using linear and energy-minimizing interpolations for the Helmholtz problem.

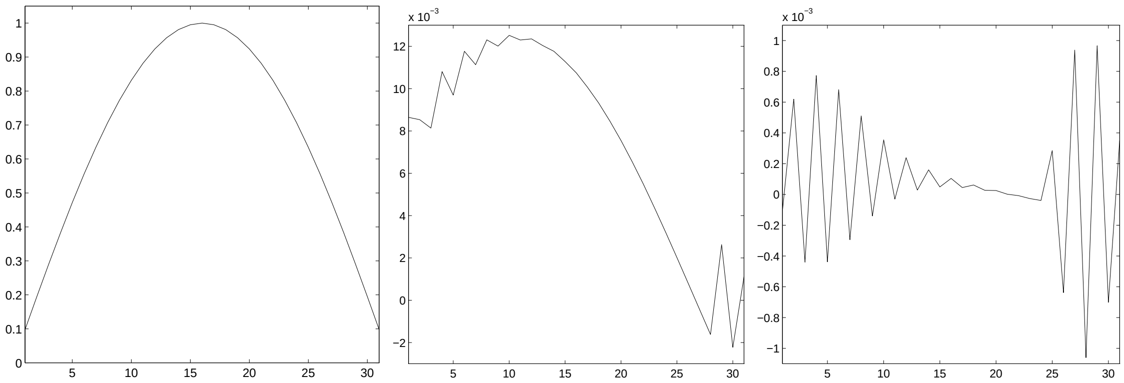

FIG. 3. *Left to right: errors after 0, 4 and 8 Gauss-Seidel iterations when $\mathcal{A}^h$ is the Helmholtz operator.*

kernel of $\mathcal{A}^h$. The convergence results of the multigrid methods using linear and energy-minimizing interpolations are shown in Table 8. The $*$ in the first column indicates that standard multigrid takes more than 100 V-cycles to convergence. The poor convergence comes from the effect of smoothing and the way the interpolation is done. The eigenfunctions of the operator $\mathcal{A}^h$ corresponding to small energy are oscillatory, whereas those corresponding to large energy are relatively smooth. As a result of standard relaxation smoothings, the errors become more oscillatory. Figure 3 shows the effect of 4 and 8 iterations of Gauss-Seidel smoothing applied to a smooth initial error. Such a phenomenon was also discussed in [7]. Hence, if we use linear interpolation, it will not be able to approximate the oscillatory error on the coarser subspaces. This fact causes the failure of the standard multigrid method.

On the other hand, the multigrid method using energy-minimizing interpolation works fine and shows no deterioration, because the energy minimization captures the property of this type of operators and produces oscillatory coarse grid basis functions (see Figure 4). This consistency enables a good approximation on the coarser subspaces, and hence the multigrid convergence is much better.

Remark: The coarse grid basis functions obtained by solving the local PDEs do not preserve constants, an approach that is natural because the operator $\mathcal{A}$ does not annihilate constant functions. If we were to extend our minimization formulation to this case in higher dimensions, we would have to modify the constraint in (12).

6. Concluding Remarks. Through the analytical and numerical results, we have demonstrated that energy-minimizing and constant preserving are two key properties of the coarse grid interpolation required to have a robust multigrid method. An obvious drawback to the construction of the robust interpolation is the expensive solve of the minimization problem. An inexact preconditioned conjugate gradient method with the linear interpolation as initial guess is proposed to overcome this problem. The numerical results show that the setup cost is not too expensive, especially when the system is to be solved many times. Nevertheless, more efficient methods to solve the minimization problem need to be derived and studied.

Finally, because of the algebraic nature of the construction of the interpolation, our method is also applicable to complicated geometries, for instance, unstructured grids, but these cases are not discussed in the present paper.

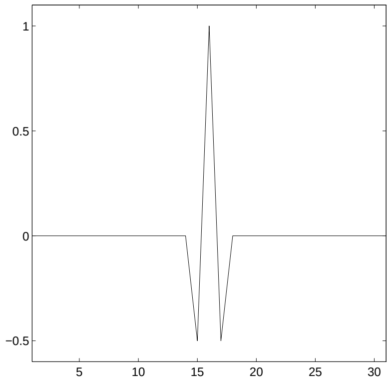

FIG. 4. *A coarse grid basis function obtained by the energy minimization when $\mathcal{A}^h$ is the Helmholtz operator.*

A. Appendix. In the following, we give the proofs of some of the previous results. In Section 2, we claimed that if the stability and the approximation properties (3), (4) are satisfied, the parameter K_0 can be bounded by a constant independent of the mesh size. Note that the following result is stated without proof in [39].

LEMMA A.1. *Let $\omega_0 = \min_{2\le i\le J}\,(\rho(A_i)\lambda_{\min}(R_i))$. Suppose (3) and (4) are satisfied. Then*

$$K_0 \le \frac{C}{\omega_0}.$$

Proof. For any $v \in V$, let $v = \sum_{k=1}^{J} v_k$ be a decomposition of v given by (3), namely $v_k = (\tilde{Q}_k - \tilde{Q}_{k-1})v$. In view of the definition of K_0 and, for each $k > 1$, we consider

$$(R_k^{-1}v_k, v_k) \le \frac{1}{\lambda_{\min}(R_k)}(v_k, v_k) \le \frac{\rho(A_k)}{\omega_0}\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|^2 \le \frac{C}{\omega_0}\|\tilde{Q}_k v\|_A^2,$$

by (4) and $\rho(A_k) = O(h_k^{-2})$. Substituting v by $\tilde{v} = (\tilde{Q}_k - \tilde{Q}_{k-1})v$ in (4) and rewriting $\tilde{v} = v$, we have

$$\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\| \le Ch_k\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|_A.$$

Hence,

$$(R_k^{-1}v_k, v_k) \le \frac{C}{\omega_0}\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|_A^2.$$

For $k = 1$, since $R_1 = A_1^{-1}$, we get instead

$$(R_1^{-1}v_1, v_1) = ((R_1A_1)^{-1}v_1, A_1v_1) = \|v_1\|_A^2 = \|\tilde{Q}_1 v\|_A^2.$$

Combining with (3), we have

$$\begin{aligned}
\sum_{k=1}^{J}(R_k^{-1}v_k, v_k) &= \|\tilde{Q}_1 v\|_A^2 + \sum_{k=2}^{J}(R_k^{-1}v_k, v_k) \\
&\le \|\tilde{Q}_1 v\|_A^2 + \frac{C}{\omega_0}\sum_{k=2}^{J}\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|_A^2 \\
&\le \max\left\{1, \frac{C}{\omega_0}\right\}\left(\sum_{k=1}^{J}\|(\tilde{Q}_k - \tilde{Q}_{k-1})v\|_A^2\right) \\
&= \max\left\{1, \frac{C}{\omega_0}\right\}\|v\|_A^2.
\end{aligned}$$

By the definition of K_0, the estimate follows. $\square$

We next prove a result in Section 3.2.1 that states that if $a(x) \equiv 1$, the minimization problem will recover the bilinear interpolation in the square grid case.

LEMMA A.2. *The solution of (12) gives the bilinear interpolation if $a(x) \equiv 1$.*

Proof. Let $\Phi_0 = [\varphi_0^1; \cdots; \varphi_0^m]$ be the vector corresponding to the bilinear interpolation. Thus, the $n \times 1$ sparse vector φ_0^i, corresponding to the coefficients of ϕ_j^h in the expansion of ϕ_i^H, has nonzeros 1/4, 1/2, and 1 only. We verify by direct substitution that Φ_0 satisfies the Euler-Lagrange equation (14) with an appropriately defined Λ_0.

Since φ_0^i is sparse, we may consider the nonzeros of φ_0^i only when computing the product $\tilde{A}_i^h \varphi_0^i$. Define

$$\begin{aligned}
\Omega_1 &= \{x_k^h : x_k^h \text{ is an interior noncoarse grid point which does not connect to any} \\
&\qquad \text{coarse grid points on the mesh.}\} \\
\Omega_2 &= \{x_k^h : x_k^h \text{ is an interior noncoarse grid point which connects to exactly 2} \\
&\qquad \text{coarse grid points on the mesh.}\} \\
\Omega_3 &= \{x_k^h : x_k^h \text{ is a noncoarse grid boundary point.}\}
\end{aligned}$$

By the definition of $\tilde{A}_i^h$, after some calculation, we can verify that

$$(\tilde{\mathcal{A}}_i^h \varphi_0^i)_k = \begin{cases} 0 & \text{if } x_k^h \in \Omega_1 \\ 3/2 & \text{if } x_k^h \in \Omega_2 \\ 3/4 & \text{if } x_k^h \in \Omega_3. \end{cases} \tag{25}$$

Here $\tilde{\mathcal{A}}_i^h$ is the nine point stencil

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

The values at the coarse grid points are not considered because the solution Φ must have a value of 1 there.

Let Λ_0 be the vector of Lagrange multipliers defined by

$$(\Lambda_0)_k = \begin{cases} 0 & \text{if } x_k^h \in \Omega_1 \\ -3/2 & \text{if } x_k^h \in \Omega_2 \\ -3/4 & \text{if } x_k^h \in \Omega_3. \end{cases}$$

First, from (25), we see that the values of $\tilde{A}_i^h \varphi_0^i$ depend not on the location of φ_0^i but on the location corresponding to the component k only. Second, $\mathcal{B}$ is a column of restriction matrices that have either 1 or 0 on the diagonal. It is not hard to see that $\mathcal{Q}\Phi_0 + \mathcal{B}\Lambda_0 = 1$. Hence, the result follows. $\square$

REFERENCES

[1] R. E. Alcouffe, A. Brandt, J. E. Dendy, Jr., and J. W. Painter. The multi-grid method for the diffusion equation with strongly discontinuous coefficients. *SIAM J. Sci. Stat. Comput.*, 2(4):430–454, 1981.

[2] J. Bramble. *Multigrid Methods*. Longman Scientific & Technical, Essex, UK, 1993.

[3] J. Bramble, J. Pasciak, J. Wang, and J. Xu. Convergence estimates for multigrid algorithms without regularity assumptions. *Math. Comp.*, 57:23–45, 1991.

[4] J. Bramble, J. Pasciak, J. Wang, and J. Xu. Convergence estimates for product iterative methods with applications to domain decomposition and multigrid. *Math. Comp.*, 57:1–21, 1991.

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

[6] A. Brandt, S. McCormick, and J. W. Ruge. Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations. Manuscript, 1983.

[7] R. H. Chan, T. F. Chan, and W. L. Wan. Multigrid for differential-convolution problems arising from image processing. Technical Report CAM 97-20, Department of Mathematics, UCLA, 1997.

[8] T. F. Chan, S. Go, and L. Zikatanov. Lecture notes on multilevel methods for elliptic problems on unstructured grids. Lecture Course 28th Computational Fluid Dynamics, March 3–7, 1997, von Karman Institute for Fluid Dynamics, Belgium, 1997.

[9] T. F. Chan, J. Xu, and L. Zikatanov. Agglomeration strategies in multigrid method. In preparation, 1997.

[10] J. E. Dendy, Jr. Black box multigrid. *J. Comp. Phys.*, 48:366–386, 1982.

[11] J. E. Dendy, Jr. Black box multigrid for nonsymmetric problems. *Appl. Math. Comput.*, 13:261–283, 1983.

[12] J. E. Dendy, Jr., M. P. Ida, and J. M. Rutledge. A semi-coarsening multigrid algorithm for SIMD machines. *SIAM J. Sci Stat. Comput.*, 13:1460–1469, 1992.

[13] J. E. Dendy, Jr., S. F. McCormick, J. W. Ruge, T. F. Russell, and S. Schaffer. Multigrid methods for three-dimensional petroleum reservoir simulation. In *10th Symposium on Reservoir Simulation, Houston*, pages 19–25. Society of Petroleum Engineers (SPE), Inc., February 6–8 1989.

[14] J. E. Dendy, Jr. and C. C. Tazartes. Grandchild of the frequency decomposition multigrid method. *SIAM J. Sci. Comput.*, 16(2):307–319, 1994.

[15] M. Dryja, B. Smith, and O. Widlund. Schwarz analysis of iterative substructuring algorithms for problems in three dimensions. *SIAM J. Numer. Anal.*, 31:1662–1694, 1994.

[16] M. Dryja and O. B. Widlund. Towards a unified theory of domain decomposition algorithms for elliptic problems. In Tony F. Chan, Roland Glowinski, Jacques Periaux, and Olof B. Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 3–21. SIAM, Philadelphia, 1990.

[17] B. Engquist and E. Luo. Convergence of a multigrid method for elliptic equations with highly oscillatory coefficients. *SIAM J. Numer. Anal.*, 34:2254–2273, 1997.

[18] J. Fuhrmann. A modular algebraic multilevel method. In *Virtual Proceedings of the Ninth International GAMM-Workshop on Parallel Multigrid Methods*, 1996.

[19] W. Hackbusch. Ein iteratives Verfahren zur schnellen Auflosung elliptischer Randwert probleme. Technical Report 76-12, Universitat zu Koln, 1976.

[20] W. Hackbusch. *Multi-grid Methods and Applications.* Springer-Verlag, Berlin, 1985.

[21] W. Hackbusch. The frequency decomposition multigrid method, part I: Application to anisotropic equations. *Numer. Math.*, 56:229–245, 1989.

[22] T. Y. Hou, X. H. Wu, and Z. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. Preprint, 1996.

[23] R. Kettler. Analysis and comparison of relaxation schemes in robust multigrid and preconditioned conjugate gradient methods. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, pages 502–534. Springer-Verlag, November 1981.

[24] R. Kettler and J.A. Meijerink. A multigrid method and a combined multigrid-conjugate gradient method for elliptic problems with strongly discontinuous coefficients in general domains. Shell Publ. 604, KSEPL, Rijswijk, 1981.

[25] S. Knapek. Matrix-dependent multigrid-homogenization for diffusion problems. In T. Manteuffal and S. McCormick, editors, *The Preliminary Proceedings of the Copper Mountain Conference on Iterative Methods, volume I*, 1996.

[26] E. Luo. *Multigrid Method for Elliptic Equation with Oscillatory Coefficients.* PhD thesis, Department of Mathematics, UCLA, Los Angeles, 1993.

[27] S. McCormick. *Multigrid Methods.* SIAM, Philadelphia, 1988.

[28] A. Reusken. Multigrid with matrix-dependent transfer operators for a singular perturbation problem. *Computing*, 50:199–211, 1993.

[29] A. Reusken. Multigrid with matrix-dependent transfer operators for convection-diffusion problems. In *Multigrid Methods, Vol. IV, Springer Lectures in Mathematics*. Springer-Verlag, 1993.

[30] J. W. Ruge and K. Stuben. Algebraic multigrid. In S. McCormick, editor, *Multigrid Methods*, pages 73–130. SIAM, 1987.

[31] B. Smith, P. Bjorstad, and W. Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, Cambridge, 1996.

[32] R. A. Smith and A. Weiser. Semicoarsening multigrid on a hypercube. *SIAM J. Sci Stat. Comput.*, 13:1314–1329, 1992.

[33] K. Stuben. Algebraic multigrid: Experience and comparisons. *Appl. Math. Comput.*, 13:419–451, 1983.

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

[35] P. Vanek, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56:179–196, 1996.

[36] P. Wesseling. *An Introduction to Multigrid Methods*. Wiley, Chichester, 1992.

[37] J. Xu. Private communication. 1996.

[38] J. Xu. *Theory of Multilevel Methods*. PhD thesis, Cornell University, Ithaca, N. Y., 1989.

[39] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34:581–613, 1992.

[40] P. M. De Zeeuw. Matrix-dependent prolongations and restrictions in a blackbox multigrid solver. *J. Comp. Appl. Math.*, 33:1–27, 1990.