ANALYSIS OF A TWO-LEVEL SCHEME FOR SOLVING FINITE ELEMENT EQUATIONS *

RANDOLPH E. BANK† AND TODD F. DUPONT‡

Abstract. A two-level iterative method for solving linear systems arising from finite element approximations of self-adjoint elliptic boundary value problems is defined and analyzed. Under relatively weak assumptions on the finite element space and differential problem, the number of iterations of this method that are required to reduce the error by a given factor can be bounded independently of the number of unknowns.

1. Introduction. In this work, we analyze a two-level iterative scheme for solving the large sparse linear systems which arise in connection with finite element procedures for solving self-adjoint elliptic boundary value problems. We take as our prototype the Neumann problem

$$\begin{aligned} -\nabla\cdot(a\nabla\psi)+b\psi &= f && \text{in } \Omega, \\ \frac{\partial\psi}{\partial n} &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{1.1}$$

where Ω is a polygonal domain in $I\!R^2$. We assume that a and b are measurable and that there exist positive constants $\underline{a}$, $\bar{a}$, $\underline{b}$, and $\bar{b}$ such that

$$\underline{a} \le a(x) \le \bar{a}, \qquad \underline{b} \le b(x) \le \bar{b}, \qquad \text{for } x \in \bar{\Omega}.$$

Our arguments are applicable, with only minor modifications, to the associated Dirichlet problem

$$\begin{aligned} -\nabla\cdot(a\nabla\psi)+b\psi &= f && \text{in } \Omega, \\ \psi &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{1.2}$$

and we comment on this extension later.

For $\phi, \chi \in \mathcal{H}^1(\Omega)$, let

$$a(\phi,\chi) = \int_\Omega a\nabla\phi\cdot\nabla\chi + b\phi\chi\,dx \tag{1.3}$$

denote the energy inner product associated with the elliptic operator (1.1). Then the Neumann problem can be posed in weak form as follows: Find $\psi \in \mathcal{H}^1(\Omega)$ satisfying

$$a(\psi,\phi) = (f,\phi) \tag{1.4}$$

for all $\phi \in \mathcal{H}^1(\Omega)$, where $(\cdot,\cdot)$ denotes the usual $\mathcal{L}^2(\Omega)$ inner product. It is well known [16] that there exists a unique solution ψ in $\mathcal{H}^1(\Omega)$ for any given f in the dual of $\mathcal{H}^1(\Omega)$.

Let $\mathcal{M}$ be an N-dimensional subspace of $\mathcal{H}^1(\Omega)$. In finite element procedures, $\mathcal{M}$ is typically a space of piecewise polynomials associated with a triangulation $\mathcal{T}$ of Ω. The finite element approximation $u \in \mathcal{M}$ of the solution ψ of (1.4) is given by

$$a(u,\phi) = (f,\phi), \qquad \phi \in \mathcal{M}. \tag{1.5}$$

*Report CNA-159, Center for Numerical Analysis, The University of Texas at Austin, May, 1980.

†Department of Mathematics, University of Texas at Austin, Austin, Texas.

‡Department of Mathematics, University of Chicago, Chicago, Illinois.

Once a suitable basis for $\mathcal{M}$ has been selected, (1.5) represents an $N \times N$ system of linear equations to be solved. Usually N is large, and the matrix associated with (1.5) is sparse.

Our two-level scheme for solving (1.5) involves the decomposition of the space $\mathcal{M}$ as the direct sum $\mathcal{M} = V \oplus W$. This decomposition induces a corresponding block iterative method for the linear system. We show in Section 2 that under rather weak assumptions about $\mathcal{M}$ and the decomposition, the two-level scheme converges at a rate bounded less than one independent of N. In particular, the global convergence of the two-level scheme depends only on the local properties of the triangulation $\mathcal{T}$ and the space $\mathcal{M}$. Our convergence proof does not require a quasi-uniform triangulation, but only a condition on the allowable set of geometries for each individual triangle. The grid may be coarse in some places and refined in others, as long as the transition from coarse to fine triangles is made in a controlled fashion.

In Section 3 we consider some simple extensions and present some examples of classes of spaces to which the method can be successfully applied. Our two-level scheme can be generalized to a k-level scheme for $k > 2$. However, the rate of convergence which our analysis would predict depends on N if k does. We, as well as several others [2, 4, 12, 14], have obtained for various k-level schemes convergence results comparable to our two-level scheme. These multi-level schemes are relatively complicated, and the requirements of the elliptic equation and the space $\mathcal{M}$ are more severe; e.g., the requirement that all the meshes are quasi-uniform. When the domain has sharply re-entrant corners the work estimates for multi-grid methods are worse than in the case of, say, a convex domain. Since the analysis of the two-level scheme does not rely on elliptic regularity, the work estimates are independent of the geometry of the domain.

While the asymptotic analysis of this method is not as promising as that of multi-grid methods, we feel it may be useful in practice. In particular, for examples in which very many finite elements are required to define the geometry of the domain, the power of the multi-grid methods is hard to utilize since little refinement of the defining mesh may be needed to achieve the desired accuracy. In Section 3 we see that in typical applications of this two-level process, the work involved in matrix factorization will be reduced by a factor of eight or more when compared with the direct solution of the full problem.

2. The Two-Level Iteration. Let $\mathcal{T}$ be a triangulation of Ω. For each triangle $T \in \mathcal{T}$, denote by h_T the diameter of the circumscribing circle for T, and by d_T the diameter of the inscribing circle divided by h_T. Let $h = \max_{T \in \mathcal{T}} h_T$. We let d_0 be a positive constant such that $d_0 \leq d_T$ for all $T \in \mathcal{T}$. It is only through d_0 that the shape regularity of $\mathcal{T}$ will enter the constants in our estimates.

Let $\mathcal{S}$ denote the set of triangles T having $h_T = 1$, $d_0 \leq d_T$, and one vertex at the origin. Designate a particular triangle $T_r \in \mathcal{S}$ as the reference triangle; T_r can be mapped onto any triangle $T \in \mathcal{S}$ using a linear transformation. Let $\mathcal{A}$ be the set of linear transformations in correspondence with triangles in $\mathcal{S}$:

$$\mathcal{A} = \{A_T | A_T \text{ is linear, } A_T(T_r) = T \in \mathcal{S}\}.$$

Any triangle in $\mathcal{T}$ can be generated by scaling and translating a triangle in $\mathcal{S}$.

Let $\mathcal{M}$ be an N-dimensional finite element subspace of $\mathcal{H}^1(\Omega)$ defined over the triangulation $\mathcal{T}$. We decompose $\mathcal{M}$ as the direct sum $\mathcal{M} = V \oplus W$, where V and W are non-trivial subspaces. For $u \in \mathcal{M}$, we systematically use $u = v + w$ where $v \in V$ and $w \in W$. Denote by $u_T = v_T + w_T$, V_T, W_T the restrictions of u, V, and W,

respectively, to $T \in \mathcal{T}$. Let V_r and W_r denote reference spaces of functions defined on T_r. We require that V and W satisfy the following conditions for all $T \in \mathcal{T}$:

(A1) If u_T is constant, then $v_T = 0$.
(A2) The space V_T contains constant functions.
(A3) There exists a mapping $\mathcal{B}_T$, consisting og a linear map $A_T \in \mathcal{A}$, followed by a scaling and translation, such that $\mathcal{B}_T$ carries T_r onto T and that $\mathcal{B}_T^{-*}$ defined by $\mathcal{B}_T^{-*}(z) = z \circ \mathcal{B}_T^{-1}$ is an onto map of V_r to V_T and W_r to W_T.

In Section 3 we give examples of spaces satisfying these hypotheses.

Consider the following iteration for approximating for approximating the solution u of (1.5): Let $u_0 \in \mathcal{M}$ be given, and define a sequence $u_k = v_k + w_k$, where $v_k \in V$ and $w_k \in W$, by

$$a(v_{k+1} - v_k, \chi) = (f, \chi) - a(u_k, \chi), \qquad \chi \in V, \tag{2.1}$$

$$a(w_{k+1} - w_k, \chi) = (f, \chi) - a(u_k, \chi), \qquad \chi \in W. \tag{2.2}$$

The sequence $\{u_k\}$ will be seen to converge to u of (1.5); further, if we have a family of spaces $\{\mathcal{M}\}$ corresponding to various triangulations of Ω, the rate of convergence (in the energy norm associated with $a(\cdot, \cdot)$) will be independent of the particular space $\mathcal{M}$, provided V_r, W_r and d_0 are the same for all $\mathcal{M}$. The iteration (2.1)-(2.2) is easy to define and analyze, but the second step can be awkward to carry out. We indicate in Section 3 how to replace (2.2) by a more easily computable process.

Let $\{\phi_i\}_{i=1}^N$ be a basis for $\mathcal{M}$ such that

$$V = \operatorname{span}\{\phi_i\}_{i=1}^{N_V}, \qquad W = \operatorname{span}\{\phi_i\}_{i=N_V+1}^{N}. \tag{2.3}$$

Define the symmetric, positive definite $N \times N$ matrix M by $M_{ij} = a(\phi_j, \phi_i)$. The solution of (1.5) then reduces to solving the linear system of equations for $U = (U^1, \ldots, U^N)^T$

$$MU = F, \tag{2.4}$$

where $F_i = (f, \phi_i)$ and $u = \sum_{i=1}^N U^i \phi_i$. Corresponding to the decomposition $\mathcal{M} = V \oplus W$, the matrix M can be partitioned as

$$M = \begin{pmatrix} A & C \\ C^T & B \end{pmatrix} \tag{2.5}$$

where A is $N_V \times N_V$ with $A_{ij} = a(\phi_j, \phi_i)$ and B is $(N - N_V) \times (N - N_V)$ with $B_{ij} = a(\phi_{N_V+j}, \phi_{N_V+i})$. The iteration (2.1)-(2.2) can then be generalized to the following:

$$\hat{M}(U_{k+1} - U_k) = \omega(F - MU_k), \qquad k = 0, 1, \ldots, \tag{2.6}$$

where U_0 is given,

$$\hat{M} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \tag{2.7}$$

and ω is a scalar relaxation parameter (in (2.1)-(2.2), $\omega = 1$).

The energy norm associated with $a(\cdot,\cdot)$ is denoted by $\|\!|\cdot|\!\|$, and the M-norm of the vector x is defined by $\|x\|_M^2 = x^T M x$. For the particular M in (2.4), $\|x\|_M^2 = a(z,z) = \|\!|z|\!\|^2$, where $z \in \mathcal{M}$ is the function associated with the coefficients x. To analyze the convergence of (2.6), we use the following theorem, whose proof can be found in [8, 9, 11].

THEOREM 2.1. *Let M and $\hat{M}$ be symmetric and positive definite. Let μ_1 and μ_2 be positive constants such that, for all $x \neq 0$, $x^T M x / x^T \hat{M} x \in [\mu_1, \mu_2]$. Then for $0 \leq \omega \leq 2/\mu_2$, the sequence $\{U_k\}$ defined in (2.6) converges to $M^{-1}F$. Further, for $\omega = 2/(\mu_1+\mu_2)$, the M-norm of the error is reduced by a factor of at least $(\mu_2-\mu_1)/(\mu_2+\mu_1)$ in each iteration.*

To estimate the convergence of (2.6), we are led to study the Rayleigh quotient

$$\frac{x^T M x}{x^T \hat{M} x} = \frac{\|\!|v+w|\!\|^2}{\|\!|v|\!\|^2 + \|\!|w|\!\|^2} \tag{2.8}$$

where $v = \sum_{i=1}^{N_V} x_i \phi_i \in V$ and $w = \sum_{i=N_V+1}^{N} x_i \phi_i \in W$.

LEMMA 2.2. *Let $\mathcal{M} = V \oplus W$ satisfy assumptions A1-A3. Then there exists a positive number γ, $0 \leq \gamma < 1$, $\gamma = \gamma(\bar{a}/\underline{a}, \bar{b}/\underline{b}, d_0, V_r, W_r)$, such that the strengthened Cauchy inequality*

$$|a(v,w)| \leq \gamma \|\!|v|\!\| \, \|\!|w|\!\| \tag{2.9}$$

holds for all $v \in V$ and $w \in W$.

Proof. It is sufficient to prove (2.9) triangle by triangle for $T \in \mathcal{T}$; to see this, note that if

$$|a(v,w)_T| \leq \gamma_T \|\!|v|\!\|_T \|\!|w|\!\|_T, \tag{2.10}$$

where $a(\cdot,\cdot)_T$ denotes the restriction of $a(\cdot,\cdot)$ to T and $\|\!|\cdot|\!\|_T$ the associated norm, then

$$\begin{aligned} |a(v,w)| &= \left| \sum_T a(v,w)_T \right| \\ &\leq \sum_T \gamma_T \|\!|v|\!\|_T \|\!|w|\!\|_T \\ &\leq \gamma \left(\sum_T \|\!|v|\!\|_T^2 \right)^{1/2} \left(\sum_T \|\!|w|\!\|_T^2 \right)^{1/2}, \end{aligned}$$

where $\gamma = \max_T \gamma_T$.

We prove (2.10) by showing the existence σ_T and ν_T satisfying

$$|a_1(v,w)| = \left| \int_T a \nabla v \cdot \nabla w \, dx \right| \leq \nu_T \|\!|v|\!\|_{1,T} \|\!|w|\!\|_{1,T}, \tag{2.11}$$

$$|a_0(v,w)| = \left| \int_T bvw \, dx \right| \leq \sigma_T \|\!|v|\!\|_{0,T} \|\!|w|\!\|_{0,T}, \tag{2.12}$$

where $\|\!|\cdot|\!\|_{i,T}$ is the (semi) norm associated with $a_i(\cdot,\cdot)$, $i = 0, 1$. If (2.11)-(2.12) hold, then for $\gamma_T = \max(\sigma_t, \nu_T)$,

$$a(v,w)_T^2 = \left(a_0(v,w) + a_1(v,w)\right)^2$$

$$
\begin{aligned}
&\le \gamma_T^2 \left(|\!|\!| v |\!|\!|_{0,T} |\!|\!| w |\!|\!|_{0,T} + |\!|\!| v |\!|\!|_{1,T} |\!|\!| w |\!|\!|_{1,T} \right)^2 \\
&\le \gamma_T^2 \left(|\!|\!| v |\!|\!|_{0,T}^2 + |\!|\!| v |\!|\!|_{1,T}^2 \right) \left(|\!|\!| w |\!|\!|_{0,T}^2 + |\!|\!| w |\!|\!|_{1,T}^2 \right) \qquad (2.13) \\
&= \gamma_T^2 |\!|\!| v |\!|\!|_T^2 |\!|\!| w |\!|\!|_T^2 .
\end{aligned}
$$

We now consider the proof of (2.11). Note that $|\!|\!| \cdot |\!|\!|_{1,T}$ defines a norm on W_T, but only a semi-norm on V_T, since $|\!|\!| c |\!|\!|_{1,T} = 0$ for any constant c. However, it suffices to consider only those functions in V_T with average value zero. Thus let $V_T' = \{ v \in V_T | \int_T v \, dx = 0 \}$. Then

$$
a_1(v, v,) = a_1(v - c, v - c); \qquad a_1(v, w) = a_1(v - c, w) \qquad (2.14)
$$

for all $v \in V_t$, $w \in W_T$, and $c = \frac{1}{|T|} \int_T v \, dx \in V_T$. In view of (2.14), we need prove (2.11) only for $v \in V_T'$ and note $|\!|\!| \cdot |\!|\!|_{1,T}$ defines a norm on $V_T' \oplus W_T$.

A simple homogeneity argument showns ν_T is independent of h_T, hence h. Let $\hat{x} = (x - x_0)/h_T$ where x_0 is a vertex of T. Under this change of variables (2.11) becomes

$$
\left| \int_{\hat{T}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{w} \, d\hat{x} \right| \le \nu_T \left(\int_{\hat{T}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{v} \, d\hat{x} \right)^{1/2} \left(\int_{\hat{T}} \hat{a} \nabla \hat{w} \cdot \nabla \hat{w} \, d\hat{x} \right)^{1/2}, \qquad (2.15)
$$

$\hat{T} \in \mathcal{S}$, $v(x) = \hat{v}(\hat{x})$, $w(x) = \hat{w}(\hat{x})$, and $a(x) = \hat{a}(\hat{x})$. In view of (2.15) and assumptions A1-A3, we can restrict attention to the reference triangle T_r, the reference spaces V_r and W_r, and the compact set of linear transformations $\mathcal{A}$. Let $\mathcal{B} \in \mathcal{A}$ be the linear map taking T_r to $\hat{T}$. Then, with $a^*(x) = \hat{a}(\mathcal{B}\hat{x})$, $w^*(x) = \hat{w}(\mathcal{B}\hat{x})$, $v^*(x) = \hat{v}(\mathcal{B}\hat{x})$,

$$
\int_{\hat{T}} \hat{a} \nabla \hat{v} \cdot \nabla \hat{w} \, d\hat{x} = |\det \mathcal{B}| \int_{T_r} a^* (\mathcal{B}^{-T} \nabla v^*) \cdot (\mathcal{B}^{-T} \nabla w^*) \, dx. \qquad (2.16)
$$

Because both T_r and $\hat{T}$ are in $\mathcal{S}$, there is a positive constant $\kappa = \kappa(d_0)$ such that, for all non-zero $z \in \mathcal{M}_r' \equiv V_r' \oplus W_r$,

$$
\underline{a} \kappa^{-1} \le \frac{\langle z, z \rangle}{[z, z]} \le \bar{a} \kappa, \qquad (2.17)
$$

where $\langle \cdot, \cdot \rangle$ is the inner product on the right-hand side of (2.16) and $[\cdot, \cdot]$ is the corresponding inner product with $\mathcal{B} =$identity and $a^* \equiv 1$. It now follows from Lemma 4.1 in the Appendix that the angle θ between V_r' and W_r, as determined using the $\langle \cdot, \cdot \rangle$ inner product, is bounded away from zero. In fact if Θ is the corresponding angle with respect to the $[\cdot, \cdot]$ inner product, then,

$$
\sin \theta \ge \left(\frac{\kappa^2 \bar{a}}{\underline{a}} \right)^{-2} \sin \Theta. \qquad (2.18)
$$

Since Θ is positive and depends only on the spaces V_r and W_r, we see that $\nu_T < 1$ where $\cos \theta < \nu_T$, can be taken to depend only on d_0, $\bar{a}/\underline{a}$, V_r and W_r.

The argument used to prove (2.12) is similar but one works directly with V_T rather than V_T'. □

Note that if $a(x)$ is Lipschitz in each triangle T, then the ratio $\bar{a}/\underline{a}$ in (2.18) can be replaced by $1 + ch_T$; a similar modification can be made in the estimation of σ_T.

THEOREM 2.3. *Let M and $\hat{M}$ be defined by (2.5) and (2.7), respectively. Then there exists a real number γ, $0 \le \gamma < 1$, $\gamma = \gamma(\bar{a}/\underline{a}, \bar{b}/\underline{b}, d_0, V_r, W_r)$, such that, for all $x \neq 0$,*

$$1 - \gamma \le \frac{x^T M x}{x^T \hat{M} x} \le 1 + \gamma \tag{2.19}$$

Thus the iterations (2.1)-(2.2) and (2.6) converge; the optimal value of ω in (2.6) is one, and the energy norm of the error is reduced by at least a factor of γ per iteration.

Proof. From (2.8), we see that (2.19) is equivalent to

$$1 - \gamma \le \frac{|\!|\!| v + w |\!|\!|^2}{|\!|\!| v |\!|\!|^2 + |\!|\!| w |\!|\!|^2} = 1 + \frac{2a(v,w)}{|\!|\!| v |\!|\!|^2 + |\!|\!| w |\!|\!|^2} \le 1 + \gamma \tag{2.20}$$

for $v \in V$ and $w \in W$. But (2.20) is an immediate consequence of Lemma 2.2. The remaining conclusions follow directly from Theorem 2.1. □

Note that if we have a family of triangulations with $h \to 0$, for which d_0, V_r and W_r are fixed, then γ does not depend on h. Dirichlet boundary conditions can be treated using the two-level scheme just as easily as Neumann. If we were solving the problem associated with (1.2), then we would require assumption A2 only for triangles T whose closures do not meet $\partial\Omega$. The proof of (2.11) is simplified for boundary triangles, since the reduction to V_T' is not necessary.

3. Extensions and Examples. We can refine the iterative process of Section 2 in many ways. For example, we can consider replacing the block Jacobi iteration (2.6) with a corresponding two-level Gauss-Seidel iteration [17]

$$\bar{M}(U_{k+1} - U_k) = F - MU_k, \tag{3.1}$$

where

$$\bar{M} = \begin{pmatrix} A & 0 \\ C^T & B \end{pmatrix}. \tag{3.2}$$

Letting $u_{k+1/2} = v_{k+1} + w_k$, the analogues of (2.1)-(2.2) are

$$a(v_{k+1} - v_k, \chi) = (f, \chi) - a(u_k, \chi), \qquad \chi \in V, \tag{3.3}$$

$$a(w_{k+1} - w_k, \chi) = (f, \chi) - a(u_{k+1/2}, \chi), \qquad \chi \in W. \tag{3.4}$$

Let ϵ_k and δ_k denote the errors in v_k and w_k, respectively. Then from (3.3)-(3.4) we have

$$a(\epsilon_{k+1} + \delta_k, \chi) = 0, \qquad \chi \in V, \tag{3.5}$$

$$a(\epsilon_{k+1} + \delta_{k+1}, \chi) = 0, \qquad \chi \in W, \tag{3.6}$$

Taking $\chi = \epsilon_{k+1} \in V$ in (3.5), and using Lemma 2.2, we have

$$|\!|\!| \epsilon_{k+1} |\!|\!| \le \gamma |\!|\!| \delta_k |\!|\!|. \tag{3.7}$$

Similarly, taking $\chi = \delta_{k+1} \in W$ in (3.6) yields

$$|\!|\!| \delta_{k+1} |\!|\!| \le \gamma |\!|\!| \epsilon_{k+1} |\!|\!|. \tag{3.8}$$

Combining (3.7)-(3.8) we get

$$\begin{aligned} |||\epsilon_{k+1}||| &\le \gamma^2 |||\epsilon_k|||, \\ |||\delta_{k+1}||| &\le \gamma^2 |||\delta_k|||. \end{aligned} \tag{3.9}$$

Since the overall error at the k-th step is $\epsilon_k + \delta_k$, we have from Lemma 2.2 and (3.9)

$$\begin{aligned} |||\epsilon_k + \delta_k|||^2 &\le (1+\gamma)\left(|||\epsilon_k|||^2 + |||\delta_k|||^2\right) \\ &\le \gamma^{4k}(1+\gamma)\left(|||\epsilon_0|||^2 + |||\delta_0|||^2\right) \\ &\le \gamma^{4k}\left(\frac{1+\gamma}{1-\gamma}\right)|||\epsilon_0 + \delta_0|||^2. \end{aligned} \tag{3.10}$$

Asymptotically, this implies an error reduction of γ^2 per iteration.

We could also consider the use of (2.1)-(2.2) in connection with a conjugate gradient procedure [1, 6, 7]; this yields an error reduction of at least $\gamma/(1+\sqrt{1-\gamma^2})$ per iteration.

A more subtle refinement involves the concept of inner iterations. We note that (2.6) requires the solution of linear systems involving the matrices A and B at each step, which may be relatively costly in terms of numerical computations. This is especially true with respect to linear systems involving the matrix B. For example, suppose $\mathcal{M}$ is the space of $\mathcal{C}^0$ piecewise polynomials of degree $s > 1$; we can take V to be the space of $\mathcal{C}^0$ piecewise linear polynomials. The dimension of V is then $N_V \approx N/s^2$. Under these conditions, it may be reasonable to solve linear systems involving A directly, using sparse matrix methods based on Gaussian elimination [10, 13, 15]. It is important, however, to devise efficient methods for solving, approximately, $Bx = y$, which can be incorporated in a convenient fashion into the two-level scheme.

Let $B = E - (E - B)$ be a splitting of B and define $G = I - E^{-1}B$. We assume $\lim_{k\to\infty} G^k = 0$. We solve, approximately, $Bx = y$ using the m-step iteration

$$E(x_{k+1} - x_k) = y - Bx_k, \qquad k = 0, 1 \ldots, m-1, \tag{3.11}$$

for a fixed value of m and some initial guess x_0. A simple induction argument establishes that (3.11) is equivalent to solving

$$B(I - G^m)^{-1}x_m = B(I - G^m)^{-1}G^m x_0 + y. \tag{3.12}$$

We shall analyze the two level scheme (2.6) using the inner iteration (3.11). If the initial guess for inner iteration is taken as the latest estimate of the solution (corresponding to $x_0 = 0$ in (3.11)-(3.12)), the two level scheme with inner iterations may be summarized as

$$M'(U_{k+1} - U_k) = \omega(F - MU_k), \tag{3.13}$$

where

$$M' = \begin{pmatrix} A & 0 \\ 0 & B(I - G^m)^{-1} \end{pmatrix}.$$

Note that if E is symmetric, then so is M'.

The functions in the space W are all quite oscillatory, since V contains local constants. Thus the solution of the equations involving B should be easy, because

on such an oscillatory space, the differential operator behaves very much like a large multiple of the identity; this is made precise below.

Suppose that there is a basis $\{\Phi_j\}_{j=1}^J$ for W_r such that, for each triangle $T \in \mathcal{T}$, and for each basis function $\phi_i \in W$ that is nontrivial on T, ϕ_i restricted to T is given by $\mathcal{B}^{-*}\Phi_j$ for some $j = 1, \ldots, J$, where $\mathcal{B}^{-*}$ is defined in assumption A3. This is a very natural assumption in the case of nodal finite elements for which the nodal parameters are function values. Under this assumption we have the following lemma.

LEMMA 3.1. *Let $D = diag(B_{ii})$ be the diagonal matrix with the same diagonal as B. Then there exist positive constants μ_1 and μ_2, depending on $\bar{a}/\underline{a}$, $\bar{b}/\underline{b}$, d_0 and $\{\Phi_j\}_{j=1}^J$, such that for $x \neq 0$*

$$\mu_1 \leq \frac{x^T B x}{x^T D x} \leq \mu_2. \tag{3.14}$$

Proof. We first note that

$$\frac{x^T B x}{x^T D x} = \frac{\|\!|w|\!\|^2}{\left(\sum_{i=N_V+1}^N \|\!|w_i|\!\|^2\right)}, \tag{3.15}$$

where $w_i = x_{i-N_V}\phi_i$ and $w = \sum_{i=N_V+1}^N w_i \in W$. Next observe that it is sufficient to prove (3.14) triangle by triangle; if

$$\mu_1 \sum_i \|\!|w_i|\!\|_T^2 \leq \|\!|w|\!\|_T^2 \leq \mu_2 \sum_i \|\!|w_i|\!\|_T^2, \tag{3.16}$$

then (3.14) follows by summing over all $T \in \mathcal{T}$. The homogeneity argument used in proving (2.11)-(2.12) shows μ_1 and μ_2 do not depend on h_T. Changing variables as in (2.15)-(2.16), and using the equivalence of four particular norms on W_r gives the result. When treating the gradient terms, the equivalence of norms on W_r is used only for the norms

$$\left[\sum c_j\phi_j, \sum c_j\phi_j\right]^{1/2}, \qquad \left\{\sum c_j^2[\phi_j, \phi_j]\right\}^{1/2},$$

where $[\cdot,\cdot]$ is defined after (2.17). □

From Lemma 3.1 and Theorem 2.1, it follows that the iteration (3.13) converges if $E = D/\tilde{\omega}$, $0 < \tilde{\omega} < 2/\mu_2$; in particular note this is true even in the case of only one inner iteration ($m = 1$). It may be advantageous, however, to take $m > 1$, if the increased cost per outer iteration is offset by a reduction in spectral radius sufficient to render a net saving in the cost of solving the problem.

If $w = \sum_{i=N_v+1}^N c_i\phi_i$, frequently c_i correspond to various derivatives of w evaluated at grid points. In such a case, we can set $W = W_1 \oplus W_2 \oplus \ldots \oplus W_\ell$, where each space W_i can be associated with a particular derivative or set of derivatives. This decomposition of W induces block iterative methods based on the corresponding partition of B. Methods of this type can be analyzed as above.

We now return to the example of $\mathcal{C}^0$ piecewise polynomials of degree $\leq s$ and consider the computational aspects of the two-level scheme. As above take V to be the space of $\mathcal{C}^0$ piecewise linear polynomials. Then the work for factoring A using a good ordering algorithm, such as minimal degree, will be $O(N^{3/2}/s^3)$. The work involved in factoring A is then $1/s^3$ times the work needed to factor the matrix $\tilde{A}$ that

would result if piecewise linear functions were used on triangles formed by dividing each of the given triangles into s^2 congruent ones. Since the matrix M has even more nonzeroes that the matrix $\tilde{A}$, the factorization of A is less than $1/s^3$ times the work to factor M.

Because the factorization is the leading order term in the work estimate, it is worthwhile putting it in perspective by considering an example. Foe $s = 3$, multiplication by M takes about $8.5N$ multiplies, and using a nested dissection ordering (on a regular grid) the factorization of A takes about $9.5(N/9)^{3/2}$ multiplications [10]. Thus the factorization is as much work as two multiplications by M when N is 2300. For many practical problems, 2300 unknowns using piecewise cubics provides far greater accuracy than is needed.

We now turn to the consideration of the two-level scheme for the piecewise linear case ($s = 1$). We can still apply the two-level scheme through an appropriate choice of basis functions. Suppose $\mathcal{T}_{2h}$ is a triangulation of Ω; construct $\mathcal{T} \equiv \mathcal{T}_h$ by dividing each triangle in $\mathcal{T}_{2h}$ into four congruent triangles by pairwise connecting the midpoints of the edges. Then $\mathcal{M} \equiv \mathcal{M}_h = \mathcal{M}_{2h} \oplus W$, where $\mathcal{M}_{2h}$ is the space of $\mathcal{C}^0$ piecewise linear polynomials over $\mathcal{T}_{2h}$, and W is the span of the usual nodal basis functions associated with nodes in $\mathcal{T}_h$ which are not in $\mathcal{T}_{2h}$. The two-level iteration can now be applied with $\mathcal{M}_{2h}$ playing the role of V, since the restrictions A1-A3 will be satisfied with respect to $\mathcal{T}_{2h}$.

It is now a small step, at least conceptually, to generalize to more than two levels. Suppose that $\mathcal{T}_{2h}$ has arisen from a refinement of $\mathcal{T}_{4h}$; then we can define $V \equiv \mathcal{M}_{4h}$, Z as the span of the nodal basis functions corresponding to nodes in $\mathcal{T}_{2h}$ which are not in $\mathcal{T}_{4h}$, and W as above. Then $\mathcal{M}_h = \mathcal{M}_{2h} \oplus W = V \oplus Z \oplus W$ and we obtain a simple three-level iteration, for $u_k = v_k + z_k + w_k$, $v_k \in V$, $z_k \in Z$, $w_k \in W$:

$$
\begin{aligned}
a(v_{k+1} - v_k, \chi) &= (f, \chi) - a(u_k, \chi), \qquad && \chi \in V, \\
a(z_{k+1} - z_k, \chi) &= (f, \chi) - a(u_k, \chi), \qquad && \chi \in Z, \\
a(w_{k+1} - w_k, \chi) &= (f, \chi) - a(u_k, \chi), \qquad && \chi \in W.
\end{aligned} \tag{3.17}
$$

This is a block Jacobi iteration with three blocks; to obtain convergence results using Theorem 2.1, we must bound the Rayleigh quotient $\|v+z+w\|^2/(\|v\|^2+\|z\|^2+\|w\|^2)$. An easy bound can be obtained using Lemma 2.2; noting that

$$
\frac{\|v + z + w\|^2}{\|v\|^2 + \|z\|^2 + \|w\|^2} = \left(\frac{\|v + z + w\|^2}{\|v + z\|^2 + \|w\|^2} \right) \left(\frac{\|v + z\|^2 + \|w\|^2)}{\|v\|^2 + \|z\|^2 + \|w\|^2} \right)
$$

we have

$$
(1 - \gamma)^2 \leq \frac{\|v + z + w\|^2}{\|v\|^2 + \|z\|^2 + \|w\|^2} \leq (1 + \gamma)^2. \tag{3.18}
$$

While the obvious extension of this argument will work for any fixed number of levels, one would like the number of levels to depend on N. In this case the above analysis will fail to show that the rate of convergence is bounded less than one independent of h.

However, such results have been obtained for multilevel schemes [2, 4, 5, 12, 14]. To do so, the concept of simple block iteration has been abandoned in favor of recursively defined algorithms. Furthermore, all presently known proofs explicitly or implicitly require some elliptic regularity, that the meshes $\mathcal{T}_{h_j}$ all be quasi-uniform,

and that the spaces $\mathcal{M}_{h_j}$ satisfy certain approximation properties more severe than A1-A3.

The results of Section 2 can be extended, under appropriate hypotheses (corresponding to A1-A3), to other types of finite element spaces, e.g., those defined on rectangles, quadrilaterals, macro-triangles, or classes of elements with curved edges. For example, we briefly consider the case of tensor product spaces defined on rectangles. For the tensor product $\mathcal{C}^0$ quadratic space, we can take V to be the tensor product $\mathcal{C}^0$ linear space. The natural basis induced by this choice of V and be associated with the derivatives w_{xx}, w_{yy} and w_{xxyy} of a function $w \in W$. This induces natural block 3×3 inner iterations of the form (3.11).

The case of tensor product $\mathcal{C}^1$ (Hermite) cubics is similar. Here we work with the natural interpolation basis; the space V can then be taken as the span of the value-tensor-value basis functions. Basis functions in W can be associated with the derivatives w_x, w_y and w_{xy} of $w \in W$. Again this induces natural block 3×3 inner iterations. The tensor product $\mathcal{C}^0$ linear case can be treated in an analogous fashion to the linear/triangle case described above. For a more complete discussion of the tensor product case see [3].

Finally, we remark that the two-level scheme is applicable to three-dimensional problems; here its advantages can be more fully exploited. Consider the case of $\mathcal{C}^0$ piecewise polynomials of degree $s > 1$ over a triangulation based on tetrahedrons. Here, as before, V can be taken as the space of $\mathcal{C}^0$ piecewise linear polynomials, but now $N_V \approx N/s^3$, so that the cost of solving the linear system involving A is relatively less significant than in the case of two-dimensional problems.

4. Appendix. In Section 2, we used the fact that if two inner products give rise to comparable norms, then the angles measured by those norms are also comparable. This follows from the following lemma.

LEMMA 4.1. *Suppose $\langle \cdot, \cdot \rangle$ and $[\cdot, \cdot]$ are inner products that define norms $|\cdot|$ and $\|\cdot\|$, respectively, on a space X. Suppose that there exist $\underline{\mu}$ and $\bar{\mu}$ such that, for all nonzero $z \in X$,*

$$0 < \underline{\mu} \leq \frac{\langle z, z \rangle}{[z, z]} \leq \bar{\mu}. \tag{4.1}$$

For any non-trivial $x, y \in X$, let

$$\beta = \frac{\langle x, y \rangle}{|x|\,|y|}, \qquad \gamma = \frac{[x, y]}{\|x\|\,\|y\|}. \tag{4.2}$$

Then

$$1 - \beta^2 \geq \left(\frac{\underline{\mu}}{\bar{\mu}} \right)^4 (1 - \gamma^2). \tag{4.3}$$

Proof. We can assume $\|x\| = \|y\| = 1$. Note that

$$\begin{aligned} 1 - \beta^2 &= (1 + \beta)(1 - \beta) \\ &= \frac{1}{4} \left| \frac{x}{|x|} + \frac{y}{|y|} \right|^2 \left| \frac{x}{|x|} - \frac{y}{|y|} \right|^2 \\ &= \frac{1}{4|x|^4} |x + \sigma y|^2 |x - \sigma y|^2, \end{aligned}$$

where $\sigma = |x|/|y|$. From (4.1) we see that

$$1 - \beta^2 \geq \frac{1}{4}\left(\frac{\underline{\mu}}{\bar{\mu}}\right)^2 \|x + \sigma y\|^2 \|x - \sigma y\|^2. \tag{4.4}$$

This inequality and the relations

$$\|x \pm \sigma y\|^2 = 1 + \sigma^2 \pm 2\sigma\gamma \tag{4.5}$$

imply that

$$1 - \beta^2 \geq \left(\frac{\underline{\mu}}{\bar{\mu}}\right)^2 \left\{\sigma^2(1 - \gamma^2) + \frac{(1 - \sigma^2)^2}{4}\right\}. \tag{4.6}$$

Discard the $(1 - \sigma^2)^2/4$ term in (4.6) to see that

$$1 - \beta^2 \geq \left(\frac{\underline{\mu}}{\bar{\mu}}\right)^2 \sigma^2(1 - \gamma^2). \tag{4.7}$$

The conclusion (4.3) now follows since $\sigma \geq \underline{\mu}/\bar{\mu}$.
□

REFERENCES

[1] O. Axelsson, *On preconditioning and convergence acceleration in sparse matrix problems*, tech. rep., CERN European Organization for Nuclear Research, Geneva, 1974.

[2] N. S. Bahkvalov, *On the convergence of a relaxation method with natural constraints on the elliptic operator*, Zh. Vychisl. Mat. mat. Fiz., 6 (1966), pp. 861–885.

[3] R. E. Bank, *Efficient algorithms for solving tensor product finite element equations*, Numer. Math., 31 (1978), pp. 49–61.

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

[5] A. Brandt, *Multi-level adaptive techniques I: the multigrid method*, tech. rep., IBM Thomas J. Watson Research Center, Yorktown Heights, New York.

[6] P. Concus, G. H. Golub, and D. P. O'Leary, *A generalized conjugate gradient method for the numerical solution of elliptic differential equations*, in Sparse Matrix Computations, (J. R. Bunch and D. J. Rose, eds.), Academic Press, New York, 1976, pp. 309–332.

[7] J. Douglas, Jr. and T. Dupont, *Preconditioned conjugate gradient iteration applied to Galerkin methods for mildly nonlinear Dirichlet problems*, in Sparse Matrix Computations, (J. R. Bunch and D. J. Rose, eds.), Academic Press, New York, 1976, pp. 309–332.

[8] T. Dupont, R. P. Kendall, and H. H. Rachford, *An approximate factorization procedure for self-adjoint elliptic difference equations*, SIAM J. Numer. Anal., 5 (1968), pp. 559–573.

[9] E. G. D'Yakanov, *On an iterative method for the solution of finite difference equations*, Dok. Akad. Nauk. SSSR., 138 (1961), pp. 522–525.

[10] J. A. George, *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal., 10 (1973), pp. 345–363.

[11] J. E. Gunn, *The solution of difference equations by semi-explicit iterative techniques*, SIAM J. Numer. Anal., 2 (1965), pp. 24–45.

[12] W. Hackbusch, *On the convergence of a multi-grid iteration applied to finite element equations*, tech. rep., Report 77-8, Universität Köln, 1977.

[13] A. J. Hoffmann, M. S. Martin, and D. J. Rose, *Complexity bounds for regular finite difference and finite element grids*, SIAM J. Numer. Anal., 10 (1973), pp. 364–369.

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

[15] A. H. Sherman, *On the Efficient Solution of Saprse Linear Systems of of Linear and Nonlinear Equations*, PhD thesis, Yale University, 1975.

[16] G. Stampucchia, *Équations Elliptiques du Second Ordre à Coefficients Discontinus*, Les Presses de l'University de Montreal, 1965.

[17] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.