MATHEMATICS OF COMPUTATION
VOLUME 60, NUMBER 201
JANUARY 1993, PAGES 1–22

ITERATIVE SCHEMES FOR NONSYMMETRIC AND INDEFINITE ELLIPTIC BOUNDARY VALUE PROBLEMS

JAMES H. BRAMBLE, ZBIGNIEW LEYK, AND JOSEPH E. PASCIAK

ABSTRACT. The purpose of this paper is twofold. The first is to describe some simple and robust iterative schemes for nonsymmetric and indefinite elliptic boundary value problems. The schemes are based in the Sobolev space $H^1(\Omega)$ and require minimal hypotheses. The second is to develop algorithms utilizing a coarse-grid approximation. This leads to iteration matrices whose eigenvalues lie in the right half of the complex plane. In fact, for symmetric indefinite problems, the iteration is reduced to a well-conditioned symmetric positive definite system which can be solved by conjugate gradient iteration. Applications of the general theory as well as numerical examples are given.

1. INTRODUCTION

In the first part of this paper, we shall describe methods based on the normal equations with respect to the Sobolev space $H^1(\Omega)$. Methods of this kind have been suggested in [10, 2]. In [2], a theorem providing bounds for iterative convergence rate was given. Here, we give a somewhat more general version of the above-mentioned result and elaborate on its applicability and implementation. These methods are particularly robust in that preconditioners can often be developed from problems with different boundary conditions, and only limited regularity is required on the solutions of the underlying partial differential equation.

In contrast, iterative schemes for nonsymmetric and indefinite systems have been studied which are based on the normal equations in discrete L^2. The analysis of the resulting iterative schemes seems to require full elliptic regularity. In addition, rapid convergence of the L^2-based algorithms requires more stringent restrictions on the boundary conditions of the problem from which the preconditioner is derived [15].

The $H^1(\Omega)$-based schemes which we shall describe are simple to analyze and robust. Alternative schemes based on generalizations of conjugate gradient and conjugate residual methods have been proposed (cf. [10]). Theoretically, none

Received by the editor March 8, 1991.
1991 *Mathematics Subject Classification.* Primary 65N30; Secondary 65F10.
This manuscript has been authored under contract number DE-AC02-76CH00016 with the U.S. Department of Energy. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. This work was also supported in part under the National Science Foundation Grant No. DMS84-05352 and by the U.S. Army Research Office through the Mathematical Sciences Institute, Cornell University.

0025-5718/93 $1.00 + $.25 per page

of the generalized schemes can be shown to be asymptotically faster than the $H^1(\Omega)$ normal equation method which we shall describe. Extensive comparisons of these methods have been made [10]. These experiments suggest that the methods exhibit similar performance when they converge. However, the generalized approaches may fail to converge in certain applications.

In the second part of this paper, we introduce a technique for reducing certain indefinite or nonsymmetric problems to ones whose spectrum is contained in the right half of the complex plane. We then show how this may be utilized to define an easily computable iterative procedure for solving the resulting reduced problem. In the symmetric indefinite case, the reduced problem is symmetric positive definite and hence, for example, the conjugate gradient method may be applied. In the nonsymmetric case, the reduced problem may be solved using, for example, the GMRES algorithm (cf. [18]).

We shall develop the iterative schemes in an abstract way. To do this, we assume that we are given a Hilbert space $\mathcal{S}$ with norm and inner product respectively denoted by $\|\cdot\|_{H^1}$ and $\widehat{A}(\cdot,\cdot)$. The space $\mathcal{S}$ is compactly and densely contained in a larger space L^2 with norm and inner product denoted by $\|\cdot\|$ and $(\cdot,\cdot)$. We shall be interested in approximating the solution $u \in \mathcal{S}$ of the problem

$$A(u,\theta) = (f,\theta) \quad \text{for all } \theta \in \mathcal{S}, \tag{1.1}$$

for a given function f. The quadratic form A may be nonsymmetric or indefinite, but is assumed to be bounded in the norm on $\mathcal{S}$.

We shall describe the iterative schemes applied to a family of approximations to (1.1). To this end, we assume that we are given finite-dimensional approximation spaces $\mathcal{S}_h \subset \mathcal{S}$ indexed by $h \in (0,1]$. The parameter h corresponds roughly to an approximation grid size, and problems with larger h have fewer unknowns. The approximation $u_h \in \mathcal{S}_h$ is defined to be the Galerkin projection, i.e.,

$$A(u_h,\phi) = (f,\phi) \quad \text{for all } \phi \in \mathcal{S}_h. \tag{1.2}$$

We shall always assume that (1.1) and (1.2) are uniquely solvable. In applications, this is often the case with only minor restriction on the size of h, e.g., $h \le h_0$ for a fixed constant h_0. We shall develop various preconditioned iterative schemes for computing the solution of (1.2) in the remainder of the paper.

In §2, we describe the preconditioned iterative schemes based in H^1 for (1.2). These schemes have appeared in the earlier literature [2, 10]. We give a simple theorem which provides bounds for their convergence.

In §3, we describe iterative schemes in a reduced subspace utilizing a coarse-grid approximation. Alternative ways of using a coarse-grid approximation in these types of problems have appeared, for example, in [7, 8, 20] as well as in the multigrid literature (see [6, 16] and the extended bibliography in [16]). Our approach is unique in that it leads to a problem with lesser rank and, in some applications, a well-conditioned symmetric positive definite system.

We present applications in §4. There, we consider second-order equations with first- and zeroth-order terms as well as oblique derivative boundary conditions. Finally, the results of numerical experiments illustrating the convergence of the proposed iterative schemes are given in §5.

2. A GENERAL ITERATIVE METHOD WITH PRECONDITIONER BASED IN H^1

In this section, we provide iterative schemes for (1.2). These schemes will be defined in terms of a symmetric positive definite quadratic form B_h on $\mathcal{S}_h \times \mathcal{S}_h$. We suppose that there are constants C_0 and C_1 satisfying

$$C_0\widehat{A}(\phi, \phi) \leq B_h(\phi, \phi) \leq C_1\widehat{A}(\phi, \phi) \quad \text{for all } \phi \in \mathcal{S}_h. \tag{2.1}$$

The form B_h will provide the preconditioner for A. For the subsequently described iterative methods to be effective, the ratio C_1/C_0 should remain relatively small, even as h becomes small.

To describe the iterative methods, we first consider the usual computational representation of the finite element problem (1.2). This realization of (1.2) assumes a given basis $\{\phi_h^i\}_{i=1}^{n_h}$ for $\mathcal{S}_h$. The solution u_h of (1.2) is expanded in this basis as

$$u_h = \sum_{i=1}^{n_h} U_h^i \phi_h^i.$$

The unknown coefficient vector U_h satisfies the matrix equation

$$\widetilde{A}_h U_h = F_h, \tag{2.2}$$

where $\widetilde{A}_h$ is the "stiffness matrix" with entries $(\widetilde{A}_h)_{ij} = A(\phi_h^i, \phi_h^j)$ and F_h is the vector with entries $(F_h)_i = (f, \phi_h^i)$. The preconditioning matrix $(\widetilde{B}_h)_{ij} = B_h(\phi_h^i, \phi_h^j)$ is analogously defined. Finally, the preconditioner $\widetilde{M}_h$ is defined to be the inverse of $\widetilde{B}_h$.

The iterative schemes are based on the following simple observation. The solution U_h of (2.2) satisfies

$$\mathscr{A} U_h \equiv \widetilde{M}_h \widetilde{A}_h^t \widetilde{M}_h \widetilde{A}_h U_h = \widetilde{M}_h \widetilde{A}_h^t \widetilde{M}_h F_h, \tag{2.3}$$

where $\widetilde{A}_h^t$ is the transpose of $\widetilde{A}_h$. The matrix $\mathscr{A} = \widetilde{M}_h \widetilde{A}_h^t \widetilde{M}_h \widetilde{A}_h$ is symmetric on R^{n_h} equipped with the inner product

$$[V, W]_h = V^t \widetilde{B}_h W. \tag{2.4}$$

It is immediate that

$$[\mathscr{A} V, W]_h = (\widetilde{A}_h V)^t \widetilde{M}_h (\widetilde{A}_h W) = [V, \mathscr{A} W]_h. \tag{2.5}$$

Inequality (2.1) and the solvability of (1.2) imply that $\mathscr{A}$ is positive definite with respect to this inner product. The preconditioned H^1-based iterative scheme for solving (2.2) is nothing more than the conjugate gradient method applied to (2.3) in the inner product (2.4). This leads to the following algorithm.

Algorithm 2.1. (1) Let an initial approximation V_0 to U_h be given (e.g., $V_0 = 0$).

(2) Set $R_0 = \widetilde{A}_h^t \widetilde{M}_h (F_h - \widetilde{A}_h V_0)$ and $P_0 = \widetilde{M}_h R_0$.

(3) For $i \geq 0$ define:

$$\begin{aligned}
\alpha_i &= \frac{R_i^t P_i}{(\tilde{A}_h P_i)^t \tilde{M}_h (\tilde{A}_h P_i)}, \\
V_{i+1} &= V_i + \alpha_i P_i, \\
R_{i+1} &= R_i - \alpha_i \tilde{A}_h^t \tilde{M}_h \tilde{A}_h P_i, \\
\beta_i &= \frac{(\tilde{M}_h R_{i+1})^t \tilde{A}_h^t \tilde{M}_h \tilde{A}_h P_i}{(\tilde{A}_h P_i)^t \tilde{M}_h (\tilde{A}_h P_i)}, \\
P_{i+1} &= \tilde{M}_h R_{i+1} - \beta_i P_i.
\end{aligned}$$

Remark 2.1. Note that the inner product (2.4) does not appear in the above algorithm. Thus, it is possible to use a preconditioner $\tilde{M}_h$ without explicitly knowing the corresponding form $B_h(\cdot, \cdot)$ as long as an algorithm for computing the action of $\tilde{M}_h$ is available. In addition, the above algorithm can be programmed so that exactly two evaluations of $\tilde{M}_h$ and one evaluation of $\tilde{A}_h$ and $\tilde{A}_h^t$ are required per iterative step.

Remark 2.2. The above algorithm is not new. Equivalent versions have appeared in the literature (e.g., Algorithm II of [2] and Algorithm 9.3 of [10]).

To bound the rate of convergence of the above conjugate gradient algorithm [17], it suffices to estimate constants satisfying inequalities of the form

$$C_2[V, V]_h \leq [\mathscr{A} V, V]_h \leq C_3[V, V]_h, \tag{2.6}$$

for all vectors $V \in R^{n_h}$. In fact, it is well known that i steps of the above algorithm reduces the initial error by a factor which is less than or equal to

$$\rho_i = 2\left(\frac{\sqrt{C_3} - \sqrt{C_2}}{\sqrt{C_3} + \sqrt{C_2}}\right)^i \tag{2.7}$$

in the norm corresponding to (2.4).

For the purposes of analysis, it is more convenient to deal with operators defined on subspaces of $\mathscr{S}$. We start with the case where $B_h(\cdot, \cdot) = \hat{A}(\cdot, \cdot)$. Let V and W be vectors in R^{n_h}, and set

$$v = \sum_{i=1}^{n_h} V^i \phi_h^i \quad \text{and} \quad w = \sum_{i=1}^{n_h} W^i \phi_h^i.$$

Note that $\tilde{M}_h \tilde{A}_h$ provides a matrix representation of an operator on the subspace $\mathscr{S}_h$. Indeed, if $W = \tilde{M}_h \tilde{A}_h V$, then

$$\hat{A}(w, \phi) = A(v, \phi) \quad \text{for all } \phi \in \mathscr{S}_h. \tag{2.8}$$

We note, in addition, that $[V, W]_h = \hat{A}(v, w)$.

Let $w \in \mathscr{S}_h$ solve (2.8). We can write $w = R_h v$, where $R_h : \mathscr{S} \mapsto \mathscr{S}_h$ is defined by $R_h v = \chi$ with χ the unique function in $\mathscr{S}_h$ satisfying

$$\hat{A}(\chi, \phi) = A(v, \phi) \quad \text{for all } \phi \in \mathscr{S}_h. \tag{2.9}$$

Note that $R_h = \hat{A}_h^{-1} A_h$, where $A_h : \mathscr{S}_h \mapsto \mathscr{S}_h$ is defined by

$$(A_h v, \theta) = A(v, \theta) \quad \text{for all } \theta \in \mathscr{S}_h$$

and $\widehat{A}_h$ is defined analogously. It immediately follows that

$$[\mathscr{A}V, V]_h = \widehat{A}(R_h v, R_h v).$$

Thus, (2.6) can be rewritten as

$$C_2\|v\|_{H^1}^2 \leq \|R_h v\|_{H^1}^2 \leq C_3\|v\|_{H^1}^2 \quad \text{for all } v \in \mathscr{S}_h. \tag{2.10}$$

Of course, (2.6) is not the same as (2.10) when $B_h(\cdot, \cdot) \neq \widehat{A}(\cdot, \cdot)$. In that case, (2.1) and (2.10) can be combined to show that

$$C_2 C_1^{-2}[V, V]_h \leq [\mathscr{A}V, V]_h \leq C_3 C_0^{-2}[V, V]_h. \tag{2.11}$$

Thus, in either case, estimates for the rate of convergence of the above conjugate gradient algorithm will follow from estimates of the form of (2.10) provided that B_h is a good preconditioner for $\widehat{A}$ on $\mathscr{S}_h$.

Remark 2.3. As we have seen above, in the case when $B_h(\cdot, \cdot) = \widehat{A}(\cdot, \cdot)$, $\widetilde{M}_h\widetilde{A}_h$ is a matrix representation of the operator R_h. Similarly, $\widetilde{M}_h\widetilde{A}_h^t$ is a matrix representation of the adjoint of the operator R_h with respect to the inner product $\widehat{A}(\cdot, \cdot)$. Thus, the reformulation (2.3) is developed by first preconditioning the system (i.e., applying $\widetilde{M}_h$) and subsequently multiplying by the $\widehat{A}(\cdot, \cdot)$ adjoint of the preconditioned system. In many applications, $\widehat{A}(\cdot, \cdot)$ corresponds to the inner product in $H^1(\Omega)$.

Let the operator $R: \mathscr{S} \mapsto \mathscr{S}$ be defined by $Rv = \chi$, where χ is the unique function in $\mathscr{S}$ satisfying

$$\widehat{A}(\chi, \theta) = A(v, \theta) \quad \text{for all } \theta \in \mathscr{S}. \tag{2.12}$$

The following theorem provides some hypotheses for proving the lower estimate in (2.10). Note that the upper estimate follows immediately from the boundedness assumption on the form $A(\cdot, \cdot)$. For the purposes of this theorem and elsewhere in the remainder of this paper, the character C will be used to denote a generic positive constant which may assume different values in different places. The constant C will always be independent of the mesh parameter h.

Theorem 2.1. *Let R_h and R be defined respectively by* (2.9) *and* (2.12). *Assume that the following estimates hold*:

$$\|\theta\|_{H^1} \leq C\{\|R_h\theta\|_{H^1} + \|\theta\|\} \quad \textit{for all } \theta \in \mathscr{S}_h; \tag{2.13}$$

there exists a fixed positive number γ with

$$\|(R - R_h)\theta\| \leq Ch^\gamma\|\theta\|_{H^1} \quad \textit{for all } \theta \in \mathscr{S}; \tag{2.14}$$

for any $\varepsilon > 0$ there exists a constant C_ε such that

$$\|\theta\| \leq C_\varepsilon\|R\theta\| + \varepsilon\|\theta\|_{H^1} \quad \textit{for all } \theta \in \mathscr{S}. \tag{2.15}$$

Then there exists a constant $h_0 > 0$ such that for $h \leq h_0$,

$$\|\theta\|_{H^1} \leq C\|R_h\theta\|_{H^1} \quad \textit{for all } \theta \in \mathscr{S}_h.$$

This theorem was essentially given in [2]. Its proof is obvious. Applications of the theorem are given in detail in §4.

3. A COARSE-GRID REDUCTION TECHNIQUE

We develop a method which utilizes a "coarse-grid solve" in this section. This results in a system with a reduced number of unknowns which is either positive definite in the case when A is symmetric indefinite, or has a positive symmetric part in the general case.

To describe this technique, we assume that along with our approximation space $\mathcal{S}_h$, we are given a "coarser" subspace $\mathcal{S}_H \subset \mathcal{S}_h$ with mesh parameter $H > h$. We are interested in solving (1.2).

The algorithm is developed in terms of the projector $P_H: \mathcal{S} \mapsto \mathcal{S}_H$ which is defined by $P_H v = w$, where $w \in S_H$ is the solution of

$$A(w, \phi) = A(v, \phi) \quad \text{for all } \phi \in \mathcal{S}_H. \tag{3.1}$$

Subsequently, we shall impose sufficient conditions such that the solvability of (3.1) is guaranteed provided that H is sufficiently small. Throughout this development, we shall assume that the solution of the coarse-grid problems (e.g., (3.1)) is relatively inexpensive. We then write the solution of (1.2) as

$$u_h = P_H u_h + (I - P_H)u_h. \tag{3.2}$$

We next provide a technique for computing $(I - P_H)u_h$. Let Q_H denote the L^2 projection onto $\mathcal{S}_H$, i.e., $Q_H v$ is the unique function in $\mathcal{S}_H$ satisfying

$$(Q_H v, w) = (v, w) \quad \text{for all } w \in \mathcal{S}_H.$$

Note that $(I - P_H)u_h$ satisfies the equation

$$A((I - P_H)u_h, \phi) = (f, \phi) - A(P_H u_h, \phi) \quad \text{for all } \phi \in \mathcal{S}_h. \tag{3.3}$$

Let $\mathcal{S}_h^{\perp}$ be defined as the image of $\mathcal{S}_h$ under the operator $I - Q_H$. Then it is easy to see that $(I - P_H)u_h = (I - P_H)v$ for any function $v \in \mathcal{S}_h$ satisfying

$$A((I - P_H)v, \phi) = (f, \phi) - A(P_H u_h, \phi) \quad \text{for all } \phi \in \mathcal{S}_h^{\perp}. \tag{3.4}$$

We will later impose conditions which guarantee the existence of a unique function $v \in \mathcal{S}_h^{\perp}$ satisfying (3.4). The algorithm for the reduced equations is, essentially, a scheme for solving (3.4).

Before stating the reduced algorithm, we provide a theorem which guarantees existence and uniqueness of solutions to (3.1) and (3.4).

Theorem 3.1. *Assume that the form A satisfies a Gårding inequality of the form*

$$C_4\widehat{A}(v, v) - C_5\|v\|^2 \le A(v, v) \quad \textit{for all } v \in \mathcal{S}. \tag{3.5}$$

Assume that there is a fixed $\gamma > 0$, such that functions $w \in \mathcal{S}_H$ satisfying (3.1) *also satisfy*

$$\|v - w\| \le CH^{\gamma}\|v - w\|_{H^1} \quad \textit{for all } v \in \mathcal{S}. \tag{3.6}$$

Then there exists a positive constant H_0 such that for $H \le H_0$, (3.1) *is uniquely solvable and* (3.4) *is uniquely solvable in $\mathcal{S}_h^{\perp}$. Moreover, there exists a positive constant C_6 such that*

$$C_6\widehat{A}((I - P_H)v, (I - P_H)v) \le A((I - P_H)v, (I - P_H)v) \tag{3.7}$$

for all $v \in \mathcal{S}_h$.

Proof. The unique solvability of (3.1) under the above assumptions follows after applying an argument given in [19]. Inequality (3.7) follows combining (3.5) and (3.6).

We need only show the unique solvability of (3.4) on $\mathcal{S}_h^\perp$. Inequality (3.7) implies that the quadratic form on the right-hand side of (3.4) has a nonnegative symmetric part which vanishes only on functions v with $(I - P_H)v = 0$. For $v \in \mathcal{S}_h^\perp$,

$$\|v\| = \|(I - Q_H)(I - P_H)v\| \leq \|(I - P_H)v\|.$$

This completes the proof of the theorem.

Theorem 3.1 justifies the following three-step algorithm for computing the solution u_h of (1.2).

Algorithm 3.1. (1) Compute $P_H u_h$ and the data on the right-hand side of (3.3).
(2) Find the unique function $v \in \mathcal{S}_h^\perp$ satisfying (3.4).
(3) Compute $(I - P_H)v$ and set $u_h = P_H u_h + (I - P_H)v$.

We propose to solve problem (3.4) by preconditioned iteration where the preconditioner is defined in all of $\mathcal{S}_h$. We note that the function $v \in \mathcal{S}_h^\perp$ satisfying (3.4) is the solution of the operator equation

$$A_h^\perp v \equiv A_h(I - P_H)v = Q_h f - A_h P_H u_h, \tag{3.8}$$

where Q_h is the L^2 projection onto $\mathcal{S}_h$. By (3.7), $A_h^\perp$ is an operator from $\mathcal{S}_h^\perp$ onto $\mathcal{S}_h^\perp$ with positive definite symmetric part. As in §2, let B_h be a symmetric positive definite preconditioning form, and define the corresponding preconditioning operator $M_h: \mathcal{S}_h \mapsto \mathcal{S}_h$ by

$$B_h(M_h\chi, \theta) = (\chi, \theta) \quad \text{for all } \theta \in \mathcal{S}_h.$$

Then, $M_h^\perp = (I - Q_H)M_h$ is a symmetric positive definite operator on $\mathcal{S}_h^\perp$.

Clearly, the solution v of (3.8) satisfies

$$\begin{aligned} M_h^\perp A_h^\perp v &= M_h^\perp(Q_h f - A_h P_H u_h) \\ &= (I - Q_H)M_h(Q_h f - A_h P_H u_h) \equiv (I - Q_H)g_h. \end{aligned} \tag{3.9}$$

To analyze the rate of convergence of iterative algorithms applied to (3.9), we must provide some estimates for the eigenvalues of the operator $M_h^\perp A_h^\perp$. It is natural to introduce the inner product $\|\cdot\|_{B_h^\perp} = ((M_h^\perp)^{-1}\cdot, \cdot)^{1/2}$ and prove the following lemma.

Lemma 3.1. *Assume that* (2.1) *and* (3.7) *hold. Then for any* $u, v \in \mathcal{S}_h^\perp$,

$$|(A_h^\perp u, v)| \leq C_U \|u\|_{B_h^\perp} \|v\|_{B_h^\perp} \tag{3.10}$$

and

$$C_L \|u\|_{B_h^\perp}^2 \leq (A_h^\perp u, u). \tag{3.11}$$

Proof. Let $P_B: \mathcal{S}_h \mapsto \mathcal{S}_H$ be defined by

$$B_h(P_B v, \theta) = B_h(v, \theta) \quad \text{for all } \theta \in \mathcal{S}_H. \tag{3.12}$$

For $w, v \in \mathscr{S}_h^{\perp}$,

$$((M_h^{\perp})^{-1}w, M_h^{\perp}v) = (w, v) = ((I - P_B)w, v)$$
$$= B_h((I - P_B)w, M_h v) = B_h((I - P_B)w, M_h^{\perp}v).$$

Thus,

$$((M_h^{\perp})^{-1}w, w) = B_h((I - P_B)w, w) \quad \text{for all } w \in \mathscr{S}_h^{\perp}. \tag{3.13}$$

First we prove (3.10). By the boundedness of A,

$$|A((I - P_H)u, v)| = |A((I - P_H)u, (I - P_B)v)|$$
$$\leq C\|(I - P_H)u\|_{H^1}\|(I - P_B)v\|_{H^1}. \tag{3.14}$$

Using (3.7), we get

$$\|(I - P_H)u\|_{H^1}^2 \leq CA((I - P_H)u, (I - P_H)u) = CA((I - P_H)u, (I - P_B)u)$$
$$\leq C\|(I - P_H)u\|_{H^1}\|(I - P_B)u\|_{H^1}.$$

Dividing by $\|(I - P_H)u\|_{H^1}$ and using (2.1), (3.13), and (3.14) proves inequality (3.10).

We next prove (3.11). By (2.1), (3.7), and the minimization property of P_B,

$$\|u\|_{B_h^{\perp}}^2 \leq B_h((I - P_H)u, (I - P_H)u) \leq CA((I - P_H)u, u).$$

This completes the proof of the lemma.

The simplest application of the above lemma is the analysis of the linear iterative scheme applied to (3.9). It is assumed that an initial iterate $v_0 \in \mathscr{S}_h^{\perp}$ is provided. For example, one could take $v_0 = 0$. For $i = 1, 2, \ldots,$ we define

$$v_i = v_{i-1} + \tau((I - Q_H)g_h - M_h^{\perp}A_h^{\perp}v_{i-1}). \tag{3.15}$$

Here, τ is a positive scaling factor. A trivial calculation gives that the error $e_i = v - v_i$ satisfies

$$\|e_i\|_{B_h^{\perp}}^2 \leq \rho\|e_{i-1}\|_{B_h^{\perp}}^1 \leq \rho^i\|e_0\|_{B_h^{\perp}}^2,$$

where $\rho = (1 - 2\tau C_L + \tau^2 C_U^2)$. Thus, if we take $\tau = C_L/C_U^2$, then $\rho = (1 - C_L^2/C_U^2)$. As usual, the linear iterative method requires estimation of a parameter. Note, however, that a smaller value of τ above will give rise to convergent iteration with a somewhat slower rate.

The above algorithm requires exactly one evaluation each of $M_h^{\perp}$ and $A_h^{\perp}$ per iterative step. This means that the action of the operator $(I - Q_H)$ must also be computed. To avoid the computation of Q_H, we develop a variation of the above procedure.

Note that Algorithm 3.1 only requires the computation of $(I - P_H)v$. With this in mind, we consider the following variation of (3.15). Let $\overline{v}_0 \in \mathscr{S}_h$ be given, and set $\hat{v}_0 = (I - P_H)\overline{v}_0$. For $i = 1, 2, \ldots$ define

$$\hat{v}_i = \hat{v}_{i-1} + \tau(I - P_H)(g_h - M_h A_h \hat{v}_{i-1}). \tag{3.16}$$

Set $v_0 = (I - Q_H)\overline{v}_0$. By applying $(I - P_H)$ to (3.15), it is easy to see that $\hat{v}_i = (I - P_H)v_i$, where v_i is the iterate in (3.15). Thus, $\hat{v}_i$ converges to $(I - P_H)v$ with a rate which is bounded by the convergence rate of (3.15). In addition, (3.16) avoids the computation of the action of the operator $(I - Q_H)$.

For completeness, we describe (3.16) in terms of the computational basis for $\mathcal{S}_h$ already used in §2 as well as a computational basis $\{\phi_H^i\}$ for $\mathcal{S}_H$. Let $\widetilde{A}_H$ denote the stiffness matrix for the form A with respect to the basis, and $\widetilde{I}_H$ denote the matrix operator which transforms coefficients corresponding to a function in $\mathcal{S}_H$ into the coefficients which represent the function in the basis for $\mathcal{S}_h$. We note that $\widetilde{I}_H\widetilde{A}_H^{-1}\widetilde{I}_H^t\widetilde{A}_h$ is the matrix representation for the operator P_H. Similarly, $\widetilde{M}_h\widetilde{A}_h$ is the matrix representation of the product M_hA_h. For convenience, we shall denote by $\widetilde{P}_H^{\perp}$ the matrix operator $I - \widetilde{I}_H\widetilde{A}_H^{-1}\widetilde{I}_H^t\widetilde{A}_h$ and by G_h the vector with entries $\{(g_h, \phi_h^i)\}$. Let $\overline{W}_0$ denote the coefficients of the function $\overline{v}_0$ in the basis for $\mathcal{S}_h$ and set $W_0 = \widetilde{P}_H^{\perp}\overline{W}_0$. Then the vector of coefficients W_i corresponding to the function $\hat{v}_i$ is given by the iteration

$$W_i = W_{i-1} + \tau\widetilde{P}_H^{\perp}(G_h - \widetilde{M}_h\widetilde{A}_hW_{i-1}). \tag{3.17}$$

The linear iteration (3.15) is also related to a natural two-grid multiplicative iterative method for (1.2). Given an initial iterate $u_0 \in \mathcal{S}_h$, we define for $i = 1, 2, \ldots,$

(1) $w_i = u_{i-1} + A_H^{-1}Q_H(f - A_hu_{i-1})$.
(2) $u_i = w_i + \tau M_h(Q_hf - A_hw_i)$.

It is easy to see that the error $\tilde{e}_i = u_h - u_i$ satisfies the product formula (cf. [4])

$$\tilde{e}_i = (I - \tau M_hA_h)(I - P_H)\tilde{e}_{i-1}.$$

Let $v_0 = (I - Q_H)u_0$. A simple mathematical induction shows that u_i and v_i (generated by (3.15)) differ by a function in $\mathcal{S}_H$. Consequently,

$$\begin{aligned}\|\tilde{e}_n\|_{H^1} &= \|(I - \tau M_hA_h)(I - P_H)\tilde{e}_{n-1}\|_{H^1}\\ &= \|(I - \tau M_hA_h)(I - P_H)e_{n-1}\|_{H^1}\\ &\le C\|e_{n-1}\|_{B_h^{\perp}} \le C\rho^{n-1}\|e_0\|_{B_h^{\perp}} \le C\rho^{n-1}\|e_0\|_{H^1}.\end{aligned} \tag{3.18}$$

Thus, the multiplicative algorithm converges at the same rate as the simple linear algorithm.

An analysis of an unscaled variation of the multiplicative algorithm was presented in [20]. There are two disadvantages of the unscaled approach. The first is that it essentially requires the use of a sufficiently accurate preconditioner, i.e., the ratio C_1/C_0 needs to be less than $1+\varepsilon$ for a sufficiently small value of ε. The second disadvantage of the unscaled approach is that the analysis only applies if the nonsymmetric part of the equation is of lower order (see Remark 4.1).

Remark 3.1. It is obvious that the operator M_h in (3.16) can be replaced by any other operator $\overline{M}_h$ provided that $\overline{M}_h\phi - M_h\phi$ is in $\mathcal{S}_H$ for all $\phi \in \mathcal{S}_h$. This means that it is possible to skip the coarse-grid solves when M_h is defined by domain decomposition or multigrid. The same replacement is valid in the case of the multiplicative algorithm above, since the functions u_i and v_i will still differ by a function in $\mathcal{S}_H$ and hence (3.18) will still hold.

The linear algorithms given above require estimation of the iteration parameter. As an alternative, the H^1-normal equation method of §2 can be applied to the reduced equation (3.8). This is equivalent to applying conjugate gradient iteration in the inner product $((M_h^{\perp})^{-1}\cdot, \cdot)$ to the preconditioned equation

$$M_h^{\perp}(A_h^{\perp})^tM_h^{\perp}A_h^{\perp}v = M_h^{\perp}(A_h^{\perp})g_h. \tag{3.19}$$

Applying Theorem 3.1 and Lemma 3.1 shows that the condition number of the operator appearing in (3.19) is bounded by $(C_U/C_L)^2$. Thus, m steps of conjugate gradient results in an error reduction bounded by

$$2\left(\frac{1-C_L/C_U}{1+C_L/C_U}\right)^m$$

in an appropriate norm. The conjugate gradient iteration shows a theoretical acceleration of convergence over the linear method. As in the case of linear iteration, it is possible to implement an equivalent cg-like iteration for $(I-P_H)v$ which avoids evaluation of the operator Q_H.

As in the symmetric case, there are several descent methods which can be directly applied to (3.9). One such method suggested by many authors is the GMRES method [18]. Mathematically, GMRES provides an algorithm for computing a best approximation in a certain Krylov space. In our application, one assumes an initial approximation $v_0 \in \mathcal{S}_h^{\perp}$ and defines the Krylov space $\mathcal{K}_m$ to be the span of the vectors $r_0, M_h^{\perp}A_h^{\perp}r_0, \ldots, (M_h^{\perp}A_h^{\perp})^{m-1}r_0$, where $r_0 = M_h^{\perp}(Q_h f - A_h P_H u_h - A_h v_0)$. The improved approximation v_m is equal to $v_0 + \chi$, where χ is the unique function in $\mathcal{S}_h^{\perp}$ which minimizes the residual error $\|r_0 - M_h^{\perp}A_h^{\perp}\theta\|_{B_h^{\perp}}$ over all functions $\theta \in \mathcal{K}_m$.

We develop the computational version of GMRES following the development of (3.17), the computational version of (3.15). We start with the algorithm on $\mathcal{S}_h^{\perp}$, then introduce a version of the algorithm for computing $(I - P_H)v$ directly, and finally present its implementation in terms of matrix operators. Details of the two intermediate algorithms are left to the reader. We present the implementation form of GMRES which provides the coefficient vector of a function converging to $(I - P_H)v$:

Algorithm 3.2. (1) Let $\overline{W}_0$ be given and set $W_0 = \widetilde{P}_H^{\perp}\overline{W}_0$.

(2) Set $R_0 = F_h - \widetilde{A}_h \widetilde{I}_H \widetilde{A}_H^{-1} \widetilde{I}_H^t F_h - \widetilde{A}_h W_0$ and $V_1 = \widetilde{P}_H^{\perp}\widetilde{M}_h R_0/(R_0^t \widetilde{P}_H^{\perp}\widetilde{M}_h R_0)^{1/2}$.

(3) For $j = 1, 2, \ldots, m$ define:

for $i = 1, 2, \ldots, j$ define

$$h_{ij} = V_i^t \widetilde{A}_h V_j;$$

$$\widehat{V}_{j+1} = \widetilde{P}_H^{\perp}\widetilde{M}_h\widetilde{A}_h V_j - \sum_{i=1}^{j} h_{ij}V_i;$$

$$h_{j+1,j} = (\widehat{V}_{j+1}^t \widetilde{A}_h V_j)^{1/2};$$

$$V_{j+1} = \widehat{V}_{j+1}/h_{j+1,j}.$$

(4) Define $W_m = W_0 + \chi$, where $\chi = \sum_{i=1}^{m} y_{mi}V_i$, and the coefficients y_{mi} are chosen to minimize the quantity $(R_0 - \widetilde{A}_h\chi)^t \widetilde{M}_h (R_0 - \widetilde{A}_h\chi)$.

Remark 3.2. Steps (2) and (3) above are essentially the Arnoldi algorithm for implementing a Gram-Schmidt orthogonalization of the Krylov space $\mathcal{K}_m$. In general, the use of such an orthogonal basis results in a more stable numerical algorithm.

Remark 3.3. A particular efficient technique for implementing the minimization process in step (4) is given in [18].

Remark 3.4. The above algorithm becomes somewhat inefficient if m becomes too large. Let n denote the number of unknowns. Then implementation of the above algorithm requires storage on the order of mn and operations on the order of nm^2. Accordingly, it is often convenient to fix m and repetitively restart the algorithm.

The use of the coarse-grid solve is fundamental for the application of GMRES in that it results in a system with positive definite symmetric part. In fact, GMRES may fail as an iterative procedure when applied to general nonsymmetric problems. It is known (see [18]) that the rate of convergence of GMRES when applied to a problem with a positive definite symmetric part can be bounded in terms of the smallest eigenvalue of the symmetric part and the norm of the original operator. Let $e_m = (I - P_H)v - \hat{v}_m$, where $\hat{v}_m$ is the function with coefficients W_m. Applying the above-mentioned analysis shows that

$$(M_h a_h^{\perp} e_m, A_h^{\perp} e_m) \leq (1 - \alpha^2/\beta^2)^m (M_h A_h^{\perp} e_0, A_h^{\perp} e_0).$$

The constant α above is the smallest eigenvalue of the symmetric part of the operator $M_h^{\perp} A_h^{\perp}$. By Theorem 3.1 and Lemma 3.1, $\alpha \geq C_L$. The constant β above is the norm of the operator $M_h^{\perp} A_h^{\perp}$. By Theorem 3.1 and Lemma 3.1, $\beta \leq C_U$. Thus, the theoretical convergence rate for GMRES is no better than the simple linear algorithm and worse than the H^1-normal method.

In the case when A_h is symmetric, Theorem 3.1 and Lemma 3.1 imply that $A_h^{\perp}$ is symmetric positive definite. Consequently, we can apply the conjugate gradient to the preconditioned equation (3.9) (in the inner product $((M_h^{\perp})^{-1}\cdot, \cdot)$). Applying Theorem 3.1 and Lemma 3.1 gives that

$$C_L((M_h^{\perp})^{-1}u, u) \leq (A_h^{\perp}u, u) \leq C_U((M_h^{\perp})^{-1}u, u) \quad \text{for all } u \in \mathcal{S}_h^{\perp}.$$

Hence, the condition number of the operator $M_h^{\perp} A_h^{\perp}$ is bounded by C_U/C_L, and the error reduction for m steps of this conjugate gradient algorithm is bounded by

$$2\left(\frac{1 - \sqrt{C_L/C_U}}{1 + \sqrt{C_L/C_U}}\right)^m$$

in an appropriate norm.

The following iteration for computing the coefficients of the vector $(I - P_H)v$ can be developed from the conjugate gradient algorithm discussed above.

Algorithm 3.3. (1) Let $\overline{W}_0$ be given and set $W_0 = \tilde{P}_H^{\perp} \overline{W}_0$.

(2) Set $R_0 = F_h - \tilde{A}_h \tilde{I}_H \tilde{A}_H^{-1} \tilde{I}_H^t F_h - \tilde{A}_h W_0$ and $P_0 = \tilde{P}_H^{\perp} \tilde{M}_h R_0$.

(3) For $i \geq 0$ define:

$$\alpha_i = \frac{R_i^t P_i}{P_i^t \tilde{A}_h P_i},$$

$$W_{i+1} = W_i + \alpha_i P_i,$$

$$R_{i+1} = R_i - \alpha_i \tilde{A}_h P_i,$$

$$\beta_i = \frac{(\tilde{P}_H^{\perp} \tilde{M}_h R_{i+1})^t \tilde{A}_h P_i}{P_i^t \tilde{A}_h P_i},$$

$$P_{i+1} = \tilde{P}_H^{\perp} \tilde{M}_h R_{i+1} - \beta_i P_i.$$

Remark 3.5. Algorithm 3.3 can be programmed in such a way as to only require one application of each of the matrices $\widetilde{A}_h$, $\widetilde{P}_H^{\perp}$, and $\widetilde{M}_h$ per iterative step.

4. Applications

In this section, we consider applications of the results of the previous sections to second-order elliptic boundary value problems. Let Ω be a domain in d-dimensional Euclidean space, and consider the problem

$$\mathscr{L}u = f \quad \text{in } \Omega, \tag{4.1}$$

$$u = 0 \quad \text{on } \Gamma_D, \tag{4.2}$$

$$\frac{\partial u}{\partial \nu} + \beta(x)u = 0 \quad \text{on } \Gamma_N. \tag{4.3}$$

Here, $\partial\Omega = \Gamma_D \cup \Gamma_N$ and $\frac{\partial}{\partial \nu}$ denotes the outward conormal derivative on $\partial\Omega$. The operator $\mathscr{L}$ is given by

$$\mathscr{L}u = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i} a_{ij}(x) \frac{\partial u}{\partial x_j} + \sum_{i=1}^{d} b_i(x) \frac{\partial u}{\partial x_i} + c(x)u.$$

We assume that the matrix $\{a_{ij}(x)\}$ is symmetric, uniformly positive definite, and bounded. We also assume that β is in $L^{\infty}(\partial\Omega)$.

We shall also consider oblique derivative problems when $d = 2$. In this case, we require that $\partial\Omega$ be piecewise smooth and let t denote the (positively oriented) tangential direction along $\partial\Omega$. The condition (4.3) is replaced by

$$\frac{\partial u}{\partial \nu} + \beta(x)u + \sigma(x)\frac{\partial u}{\partial t} = 0 \quad \text{on } \Gamma_N. \tag{4.4}$$

We assume that $\sigma(x)$ is smooth on the smooth arcs of Γ_N. In addition, we assume that each smooth part of Γ_N has endpoints on $\partial\Gamma_D$. This makes the problem variational (cf. §4.4.3 of [12]).

To develop and analyze this example, we shall need to use Sobolev spaces on Ω and $\partial\Omega$. For a nonnegative l, the Sobolev space of order l on Ω will be denoted $H^l(\Omega) = W_2^l(\Omega)$ with norm $\|\cdot\|_l$ (see Definition 1.3.2.1 of [12]). On the boundary, the space of order l will be denoted $H^l(\partial\Omega)$ with norm $|\cdot|_l$.

We next provide the weak formulation of (4.1)–(4.4). Define $\mathscr{S}$ to be the functions in $H^1(\Omega)$ whose trace vanish on Γ_D. Let $(\cdot, \cdot)$ denote the $L^2(\Omega)$ inner product and $\langle \cdot, \cdot \rangle$ denote the $L^2(\partial\Omega)$ inner product. The weak formulation of (4.1)–(4.4) is: Find $u \in \mathscr{S}$ such that

$$A(u, \psi) = (f, \psi) \quad \text{for all } \psi \in \mathscr{S}. \tag{4.5}$$

Here, the form A is defined by

$$\begin{aligned} A(\psi, \theta) = &\sum_{i,j=1}^{d} \left(a_{ij} \frac{\partial \psi}{\partial x_j}, \frac{\partial \theta}{\partial x_i} \right) + \sum_{i=1}^{d} \left(b_i \frac{\partial \psi}{\partial x_i}, \theta \right) \\ &+ (c\psi, \theta) + \left\langle \beta\psi + \sigma \frac{\partial \psi}{\partial t}, \theta \right\rangle. \end{aligned} \tag{4.6}$$

By convention, we take $\sigma = 0$ when $d \neq 2$. The bilinear form A is bounded with respect to the norm in $H^1(\Omega)$ (cf. §4.4.3 of [12]).

Remark 4.1. Many analyses of iterative schemes for the solution of the discrete systems which result from the numerical approximation of nonsymmetric and indefinite boundary value problems require that the nonsymmetric part of the form A be of lower order (e.g., [6–8, 14, 15, 20]). This means that $A_N(v, w) = (A(v, w) - A(w, v))/2$ can be bounded by

$$|A_N(v, w)| \leq C\|v\|_{1-\alpha}\|w\|_\alpha \tag{4.7}$$

for some $\alpha \in [0, 1]$. Inequality (4.7) does not hold in the case of boundary condition (4.4). In contrast, the theory presented in this paper is applicable and shows that the corresponding iterative schemes remain effective.

To verify the hypotheses of the theorems in the previous sections, we shall use the equivalent inner product

$$\hat{A}(\psi, \theta) = \sum_{i,j=1}^{d} \left(a_{ij} \frac{\partial \psi}{\partial x_j}, \frac{\partial \theta}{\partial x_i} \right) + (\psi, \theta) \tag{4.8}$$

on $H^1(\Omega)$.

We note that the Gårding inequality (3.5) holds for the form A. In fact, the term associated with the tangential derivative results in no difficulty, since

$$\left\langle \sigma \frac{\partial \psi}{\partial t}, \psi \right\rangle = -\left\langle \frac{\partial \sigma}{\partial t} \psi, \psi \right\rangle - \left\langle \sigma \psi, \frac{\partial \psi}{\partial t} \right\rangle$$

and hence

$$\left\langle \sigma \frac{\partial \psi}{\partial t}, \psi \right\rangle = -\frac{1}{2} \left\langle \frac{\partial \sigma}{\partial t} \psi, \psi \right\rangle. \tag{4.9}$$

The remaining terms are handled by standard perturbation arguments.

Remark 4.2. Many approaches for providing estimates for nonsymmetric and indefinite problems require that the form differs from a "nice" form (e.g., $\hat{A}(\cdot, \cdot)$) by a lower-order perturbation. Our theory only requires that the symmetric part of the form differs from a nice form by a lower-order perturbation. Note that the oblique derivative term in (4.4) has the same strength as the second-order derivative terms in (4.1). However, (4.9) shows that the term $\langle \sigma \frac{\partial \psi}{\partial t}, \psi \rangle$ is weaker than $\hat{A}(\psi, \psi)$.

We shall assume that (4.5) and its adjoint have unique solutions. In addition, we assume that solutions of adjoint problems have a modest amount of elliptic regularity. Specifically, given $g \in L^2(\Omega)$, let v solve

$$A(\phi, v) = (\phi, g) \quad \text{for all } \phi \in \mathcal{S}.$$

We assume that for some $\gamma \in (0, 1]$, there is a constant C not depending on $g \in L^2(\Omega)$ such that

$$\|v\|_{1+\gamma} \leq C\|g\|. \tag{4.10}$$

Estimates of the form (4.10) for many applications fitting into the general framework of problem (4.1)–(4.4) can be found in various references. We shall quote some of these results in the Appendix. In addition, we shall provide a proof of (4.10) in the case of boundary condition (4.4).

We approximate the solution of (4.5) by using the finite element method. This involves the use of a sequence of approximation spaces $\{\mathcal{S}_h\}$ which are

subspaces of $\mathcal{S}$ and indexed by $h \in (0, 1]$. Many examples of such constructions can be found in [1, 9]. We put very little restriction on these spaces other than the requirement that they satisfy standard approximation properties. In particular, we allow mesh refinement in all of our applications. In this case, the parameter h corresponds to the size of the largest triangle or finite element.

For the purposes of developing iterative algorithms, we only require that the subspaces satisfy

$$\inf_{\phi \in \mathcal{S}_h} \|v - \phi\|_1 \leq Ch\|v\|_2 . \tag{4.11}$$

The inequality (4.11) holds with fixed C and for all $v \in H^2(\Omega) \cap \mathcal{S}$. This does not exclude the use of higher-order spaces and refinement in order to obtain better solution accuracy.

The Galerkin approximation $u_h \in \mathcal{S}_h$ to the solution u of (4.5) is defined by (1.2). Inequality (3.6) follows applying the standard finite element duality technique [1, 9], (4.10), and (4.11). Thus, we have shown that the hypotheses of Theorem 3.1 hold for this application. The following theorem shows that the hypotheses of Theorem 2.1 also hold for this application.

Theorem 4.1. *Under the above assumptions and definitions of* $\mathcal{S}$, $\mathcal{S}_h$, $\widehat{A}_h$, *and* A, (2.13)–(2.15) *hold, i.e., the hypotheses of Theorem* 2.1 *hold.*

We shall use the following lemma in the proof of Theorem 4.1. Its proof will be given in the Appendix.

Lemma 4.1. *Let* $0 < \gamma < 1/2$ *be fixed. Given* $\varepsilon > 0$, *there exists a constant* $C_{\varepsilon,\gamma}$ *such that for all* $\phi \in H^1(\Omega)$ *and* $\psi \in H^{1+\gamma}(\Omega)$,

$$\widehat{A}(\phi, \psi) \leq (C_{\varepsilon,\gamma}\|\phi\| + \varepsilon\|\phi\|_1)\|\psi\|_{1+\gamma} . \tag{4.12}$$

Proof of Theorem 4.1. We first prove (2.13). By (3.5), for $\theta \in \mathcal{S}_h$,

$$\begin{aligned} \widehat{A}(\theta, \theta) &\leq C_4^{-1}(C_4\widehat{A}(\theta, \theta) - C_5\|\theta\|^2 + C_5\|\theta\|^2) \\ &\leq C_4^{-1}(A(\theta, \theta) + C_5\|\theta\|^2) = C_4^{-1}(\widehat{A}(R_h\theta, \theta) + C_5\|\theta\|^2) . \end{aligned} \tag{4.13}$$

Inequality (2.13) follows from (4.13) and obvious manipulations.

We next prove (2.14). Let P_h denote the elliptic projection onto the space $\mathcal{S}_h$, i.e., $P_h : \mathcal{S} \mapsto \mathcal{S}_h$ be defined analogously to P_H in (3.1). From the definition, it is immediate that $R_h = P_hR$. Inequality (3.6) with h replacing H gives

$$\|(R - R_h)\phi\| \leq Ch^\gamma\|R\phi\|_1 .$$

Note that R is defined by the relation

$$\widehat{A}(Rv, \phi) = A(v, \phi) \quad \text{for all } \phi \in \mathcal{S} .$$

Taking $\phi = Rv$ and using the boundedness of $A(\cdot, \cdot)$ gives

$$\|Rv\|_1 \leq C\|v\|_1 . \tag{4.14}$$

Inequality (2.14) follows.

Before proving (2.15), we first define some additional notation. Let $\mathcal{S}^{-1}$ denote the dual of $\mathcal{S}$, i.e., $\mathcal{S}^{-1}$ is the set of distributions on Ω for which the norm

$$\|u\|_{-1} = \sup_{\phi \in \mathcal{S}} \frac{(u, \phi)}{\|\phi\|_1}$$

is finite. In addition, we define the following operators:

(1) $A: \mathcal{S} \mapsto \mathcal{S}^{-1}$ by $Aw = v$, where v is the unique function in $\mathcal{S}^{-1}$ satisfying
$$(v, \phi) = A(w, \phi) \quad \text{for all } \phi \in \mathcal{S}.$$

(2) $\hat{A}: \mathcal{S} \mapsto \mathcal{S}^{-1}$ by $\hat{A}w = v$, where v is the unique function in $\mathcal{S}^{-1}$ satisfying
$$(v, \phi) = \hat{A}(w, \phi) \quad \text{for all } \phi \in \mathcal{S}.$$

(3) $T^*: \mathcal{S}^{-1} \mapsto \mathcal{S}$ by $T^*w = v$, where v is the unique function in $\mathcal{S}$ satisfying
$$A(\phi, v) = (w, \phi) \quad \text{for all } \phi \in \mathcal{S}.$$

(4) $\hat{T}: \mathcal{S}^{-1} \mapsto \mathcal{S}$ by $\hat{T}w = v$, where v is the unique function in $\mathcal{S}$ satisfying
$$\hat{A}(v, \phi) = (w, \phi) \quad \text{for all } \phi \in \mathcal{S}.$$

Note that $R = \hat{T}A$.

We now prove (2.15). For $u \in \mathcal{S}$,
$$\|u\|^2 = A(u, T^*u) = (Au, T^*u) = \hat{A}(\hat{T}Au, T^*u).$$
Applying Lemma 4.1 and (4.10) and noting that $R = \hat{T}A$ gives
$$\|u\| \leq C(C_{\varepsilon,\gamma}\|Ru\| + \varepsilon\|Ru\|_1). \tag{4.15}$$
Inequality (2.15) follows immediately combining (4.14) and (4.15). This completes the proof of the theorem.

5. Numerical experiments

In this section, we provide the results of numerical examples illustrating the theory developed in earlier sections. We shall consider a model problem in two-dimensional space. Specifically, we consider problem (4.1)–(4.4) where
$$\Omega = [0, 1] \times [0, 1] \quad \text{and} \quad \mathcal{L}u = -\Delta u + au_x + bu_y - cu.$$
We shall consider three numerical examples. The first is the symmetric indefinite case $(a = b = 0)$. The second example adds nonsymmetry using nonzero a, b. The final example illustrates the oblique derivative problem (i.e., boundary condition (4.4) with $\sigma \neq 0$).

The sequence of subspaces are the usual finite element approximation spaces. Specifically, the domain Ω is first partitioned into $m \times m$ square subdomains of side length $1/m$. Each smaller square is then divided into two triangles by one of the diagonals. The approximation space $\mathcal{S}_h$ is defined to be the set of functions which are continuous on Ω, piecewise linear with respect to the triangulation, and vanish on Γ_D. In the first two examples, Γ_D is all of $\partial\Omega$. In the last example, $\Gamma_D = \partial\Omega/\Gamma_N$, where $\Gamma_N = \{(0, y) | y \in (0, 1)\}$.

We seek the Galerkin solution $u_h \in \mathcal{S}_h$ satisfying (1.2), where $A(\cdot, \cdot)$ is defined by (4.6). For all of our examples, the preconditioner for (1.2) is developed from the form corresponding to the Laplace operator,
$$D(u, \phi) = \int_\Omega (u_x\phi_x + u_y\phi_y + c_1 u\phi)\, dx\, dy, \tag{5.1}$$

TABLE 5.1. Algorithm 2.1 applied to Example 1 with $c_1 = 0$

$1/h$	c	$K(\mathscr{A})$	NI
32	115	1143	51
64	115	1024	51
128	115	1011	51
32	150	2339	59
64	150	2294	61
128	150	2379	61

TABLE 5.2. Algorithm 2.1 applied to Example 1 with $c_1 = c$

$1/h$	c	$K(\mathscr{A})$	NI
32	115	209	59
64	115	195	59
128	115	191	59
32	150	214	70
64	150	159	71
128	150	171	72

where c_1 is a nonnegative constant. Specifically, the preconditioning operator M_h is defined by applying one V-cycle of a multigrid procedure corresponding to the form $D(\cdot, \cdot)$ on the subspace $\mathscr{S}_h$ (cf. [3, 5]). The problem corresponding to (5.1) is solved exactly on the coarsest grid of mesh size $1/4$. In addition, one sweep of Gauss-Seidel smoothing was used on the way down with a sweep in the opposite direction on the way up. This results in a symmetric preconditioning form $B_h(\cdot, \cdot)$.

One factor which can be used to interpret the efficiency of the proposed iterative schemes is the number of iterations required to achieve a certain accuracy. Specifically, we define NI to be the number of steps required to reduce the initial error by a factor of 10^{-6}, i.e.,

$$\|e_{NI}\| \leq 10^{-6}\|e_0\|,$$

where $e_i = U_h - V_i$, U_h is the solution to (2.2), and V_i is the ith iterate in the given algorithm.

Example 1. For our first example, we set $a = b = 0$, let $\Gamma_D = \partial\Omega$, and consider various values of c. We compare three methods for solving this problem. The first uses Algorithm 2.1 with preconditioner defined taking $c_1 = 0$ in (5.1).

TABLE 5.3. Algorithm 3.3 applied to Example 1 with $c_1 = 0$

$1/h$	$1/H$	c	$K(M_h^{\perp} A_h^{\perp})$	NI
32	8	115	*	*
32	16	115	1.84	7
64	16	115	2.21	8
128	16	115	2.33	8
32	16	150	2.26	7
64	16	150	2.81	8
128	16	150	2.94	8

TABLE 5.4. Algorithm 2.1 applied to Example 2

$1/h$	c	$K(\mathscr{A})$	NI
32	115	2723	52
64	115	2738	52
128	115	2892	52
32	150	4466	66
64	150	5643	67
128	150	5948	69

The second again uses Algorithm 2.1 but with $c_1 = c$ as suggested in [11]. The final uses the subspace reduction technique of §3 and Algorithm 3.3.

Tables 5.1 and 5.2 illustrate the effect that the parameter c_1 in (5.1) has on the convergence behavior of Algorithm 2.1. The reported condition numbers $K(\mathscr{A})$ are significantly larger in the case when $c_1 = 0$ than in the case of $c_1 = c$. However, the number of iterations required for a given accuracy is less in the case of $c_1 = 0$. The large condition numbers in the case of $c_1 = 0$ are due to a few large eigenvalues. In such instances, the conjugate gradient algorithm converges faster than predicted by the condition number alone.

Table 5.3 illustrates the effect which the coarse-grid reduction technique has on the convergence rate of the resulting scheme. The $*$'s in the first row ($H = 1/8$) indicate that the preconditioned system failed to be positive definite for that value of H. This shows the necessity of taking H sufficiently small. The remaining rows show the substantial benefit resulting from the coarse-grid solve once the threshold $H \leq H_0$ has been reached. Since the reduced system is positive definite, conjugate gradient iteration can be applied and half the number of operator evaluations are required per step. Note that we get drastically

TABLE 5.5. Conjugate gradient applied to (3.19); Example 2

$1/h$	$1/H$	c	$K(M_h^{\perp}(A_h^{\perp})^t M_h^{\perp} A_h^{\perp})$	NI
32	8	115	1750	28
32	16	115	3.42	10
64	16	115	4.93	12
128	16	115	5.46	12
32	16	150	5.36	11
64	16	150	8.49	13
128	16	150	9.43	13

TABLE 5.6. Algorithm 3.2 applied to Example 2

$1/h$	$1/H$	c	NI
32	8	115	16
32	16	115	7
64	16	115	8
128	16	115	9
32	16	150	7
64	16	150	8
128	16	150	9

smaller condition numbers, and far fewer iterations are required for convergence (compared to the results in Tables 5.1 and 5.2).

Example 2. This example illustrates the convergence behavior of the iterative algorithms developed earlier when applied to a nonsymmetric and indefinite problem. This time we take $a = 1$, $b = 2$, vary c, and, once again, take $\Gamma_D = \partial\Omega$. All of the remaining results are for $c_1 = 0$.

Tables 5.4 and 5.5 illustrate the effect that the coarse-grid reduction has on this problem.

Table 5.4 gives results for Algorithm 2.1, whereas Table 5.5 gives similar results but for the conjugate gradient method applied to the reduced equations (3.19). Once again we see that the coarse-grid reduction technique leads to algorithms with significantly lower condition numbers which converge to a given accuracy in far fewer iterations. The large condition number in the first row of Table 5.5 is in agreement with the failure of convergence observed in the first row of Table 5.3. However, the method of Table 5.5 (based on the normal

TABLE 5.7. Algorithm 2.1 applied to Example 3

$1/h$	σ	$K(\mathscr{A})$	NI
32	1	3.50	12
64	1	3.63	12
128	1	3.66	13
32	10	159	43
64	10	169	68
128	10	170	88
32	50	2380	72
64	50	2514	128
128	50	2515	202

TABLE 5.8. GMRES applied to (5.2); Example 3

$1/h$	σ	NI
32	1	16
64	1	17
128	1	18
32	10	148
64	10	154
128	10	160
32	50	820
64	50	855
128	50	860

equations) is robust and will converge with any value of H but at a possibly slower rate.

Finally, Table 5.6 gives iteration counts for GMRES (Algorithm 3.2) applied to this example. Note that the iteration counts of Table 5.6 are somewhat better than those of Table 5.5. This is due to the fact that the nonsymmetry of this example is very weak.

Example 3. For the final example, we consider the oblique derivative problem. Specifically, we set $a = b = c = 0$, $\Gamma_N = \{(0, y) | y \in (0, 1)\}$, and $\sigma(x) = \sigma$ for various values of σ. In this case, the equations always have a positive

definite symmetric part, and hence the coarse-grid reduction technique will not be applied.

Tables 5.7 and 5.8 (see p. 19) illustrate the convergence behavior of Algorithm 2.1 and GMRES directly applied to

$$\widetilde{M}_h \widetilde{A}_h U_h = \widetilde{M}_h F_h . \tag{5.2}$$

To keep storage requirements manageable, GMRES was restarted every ten iterative steps. Algorithm 2.1 always converged faster than GMRES for this application. In fact, for larger σ, the H^1-normal approach provides a much more effective algorithm.

6. Appendix

In this Appendix, we shall discuss the estimates (4.10) and Lemma 4.1. We first provide a proof that (4.10) is satisfied for some examples with boundary condition (4.4). We finish by proving Lemma 4.1.

Inequality (4.10) in the case of Dirichlet boundary conditions and domains with polygonal boundaries has been proved in [13]. We will prove this estimate with boundary condition (4.4) under some further assumptions. Specifically, we consider the problem

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega , \\ u &= 0 \quad \text{on } \Gamma_D , \\ \frac{\partial u}{\partial \nu} + \sigma(x)\frac{\partial u}{\partial t} &= 0 \quad \text{on } \Gamma_N , \end{aligned} \tag{6.1}$$

where Ω has a polygonal boundary and σ is constant on each connected component of Γ_N. If Γ_D is empty, we require that

$$\int_\Omega f\,dx = 0 \quad \text{and} \quad \int_\Omega u\,dx = 0 .$$

The first step is to provide an L^p estimate for the solution u of (6.1). For nonnegative s and $1 \le p < \infty$, we denote by $W_p^s(\Omega)$ the usual Sobolev space with norm $\|\cdot\|_{s,p}$ given, for example, by Definition 1.3.2.1 of [12]. Combining Theorems 4.3.2.4 and 4.4.4.13 of [12], using the uniqueness of solutions to (6.1), and examining the behavior of the singular functions associated with the boundary edge angles and the boundary conditions shows that there is a p with $1 < p \le 2$ such that

$$\|u\|_{2,p} \le C\|f\|_{0,p} . \tag{6.2}$$

We next note the following imbedding inequality (see Theorem 1.4.4.1 and the remark thereafter in [12]): For each $1 < p \le 2$, there exists a constant C_p such that

$$\|\phi\|_{3-2/p} \le C_p\|\phi\|_{2,p} \quad \text{for all } \phi \in W_p^2(\Omega) . \tag{6.3}$$

Combining (6.2) and (6.3) and using the fact that $\|f\|_{0,p} \le C\|f\|$ proves (4.10) with $\gamma = 2 - 2/p$.

Remark 6.1. There is no difficulty in replacing the Laplacian in (6.1) by the Laplacian plus lower-order terms. Generalizations to variable-coefficient operators and boundary conditions seem possible but are tedious and hence will not be treated here.

We now provide a proof of Lemma 4.1. We first consider the derivative terms in (4.8). Let E denote the extension operator given by Theorem 1.4.3.1 of [12]. For a function v defined on Ω, let $\tilde{v}$ denote the extension of v by zero to R^d. Since $\gamma < 1/2$, Corollary 1.4.4.5 of [12] gives that the norm $\|\tilde{v}\|_{W_2^\gamma(R^2)}$ is equivalent to the norm $\|v\|_\gamma$ for all $v \in H^\gamma(\Omega)$. Thus, for $\phi \in H^1(\Omega)$ and $\psi \in H^{1+\gamma}(\Omega)$,

$$\left(\frac{\partial \phi}{\partial x_i}, \frac{\partial \psi}{\partial x_j}\right) = \left(\mathcal{F}\left(\frac{\partial (E\phi)}{\partial x_i}\right), \mathcal{F}\left(\widetilde{\frac{\partial \psi}{\partial x_j}}\right)\right),$$

where $\mathcal{F}$ denotes the Fourier transform. By the Schwarz inequality,

$$\left(\frac{\partial \phi}{\partial x_i}, \frac{\partial \psi}{\partial x_j}\right) \le C\left(\int_{R^d} \frac{|\zeta|^2}{(1+|\zeta|^2)^\gamma} |\mathcal{F}(E\phi)(\zeta)|^2 \, d\zeta\right)^{1/2} \|\psi\|_{1+\gamma} \tag{6.4}$$

$$\le (C_{\delta,\gamma}\|\phi\| + \delta\|\phi\|_1)\|\psi\|_{1+\gamma},$$

where $\delta > 0$ is arbitrary. The lemma follows by summing (6.4) and obvious manipulations.

Bibliography

1. A. K. Aziz and I. Babuška, *Part* I, *survey lectures on the mathematical foundations of the finite element method*, The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations (A. K. Aziz, ed.), Academic Press, New York, 1972, pp. 1–362.
2. J. H. Bramble and J. E. Pasciak, *Preconditioned iterative methods for nonselfadjoint or indefinite elliptic boundary value problems*, Unification of Finite Element Methods (H. Kardestuncer, ed.), Elsevier Science/North-Holland, New York, 1984, pp. 167–184.
3. ____, *New convergence estimates for multigrid algorithms*, Math. Comp. **49** (1987), 311–329.
4. J. H. Bramble, J. E. Pasciak, J. Wang, and J. Xu, *Convergence estimates for product iterative methods with applications to domain decomposition*, Math. Comp. **57** (1991), 1–21.
5. J. H. Bramble, J. E. Pasciak, and J. Xu, *The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms*, Math. Comp. **56** (1991), 1–34.
6. ____, *The analysis of multigrid algorithms for nonsymmetric and indefinite elliptic problems*, Math. Comp. **51** (1988), 389–414.
7. X.-C. Cai and O. Widlund, *Domain decomposition algorithms for indefinite elliptic problems*, SIAM J. Sci. Statist. Comput. **13** (1992), 243–258.
8. X.-C. Cai and J. Xu, *A preconditioned GMRES method for nonsymmetric and indefinite problems*, preprint.
9. P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, New York, 1978.
10. H. C. Elman, *Iterative methods for large, sparse, nonsymmetric systems of linear equations*, Yale Univ. Dept. of Comput. Sci. Rep. 229, 1982.
11. C. I. Goldstein, *Analysis and application of multigrid preconditioners for singularly perturbed boundary value problems*, SIAM J. Numer. Anal. **26** (1989), 1090–1123.
12. P. Grisvard, *Elliptic problems in nonsmooth domains*, Pitman, Boston, 1985.
13. R. B. Kellogg, *Interpolation between subspaces of a Hilbert space*, Univ. of Maryland, Inst. Fluid Dynamics and Appl. Math., Tech. Note BN-719, 1971.
14. J. Mandel, *Multigrid convergence for nonsymmetric, indefinite variational problems and one smoothing step* (Proc. 2nd Copper Mtn. Conf. Multigrid Methods), Appl. Math. Comput. **19** (1986), 201–216.

15. T. A. Manteuffel and S. V. Parter, *Preconditioning and boundary conditions*, SIAM J. Numer. Anal. **27** (1990), 656–694.
16. S. McCormick, ed., *Multigrid methods*, SIAM, Philadelphia, PA, 1987.
17. W. M. Patterson, 3rd, *Iterative methods for the solution of a linear operator equation in Hilbert space—a survey*, Lecture Notes in Math., vol. 394, Springer-Verlag, New York, 1974.
18. Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput. **7** (1986), 856–869.
19. A. H. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Math. Comp. **28** (1974), 959–962.
20. J. Xu, *A new class of iterative methods for nonsymmetric boundary value problems*, preprint.

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NEW YORK 14853
E-mail address: bramble@mssun7.msi.cornell.edu

STATISTICS RESEARCH STATION, AUSTRALIAN NATIONAL UNIVERSITY, CANBERRA, ACT 2601, AUSTRALIA
E-mail address: leyk@vila.anw.edu.au

DEPARTMENT OF APPLIED SCIENCE, BROOKHAVEN NATIONAL LABORATORY, UPTON, NEW YORK 11973
E-mail address: pasciak@bnl.gov