An Iterative Method for Elliptic Problems on Regions Partitioned into Substructures*

By J. H. Bramble, J. E. Pasciak and A. H. Schatz

Abstract. Some new preconditioners for discretizations of elliptic boundary problems are studied. With these preconditioners, the domain under consideration is broken into subdomains and preconditioners are defined which only require the solution of matrix problems on the subdomains. Analytic estimates are given which guarantee that under appropriate hypotheses, the preconditioned iterative procedure converges to the solution of the discrete equations with a rate per iteration that is independent of the number of unknowns. Numerical examples are presented which illustrate the theoretically predicted iterative convergence rates.

1. Introduction. In this paper we will consider as a model problem the Dirichlet problem for a second-order uniformly elliptic equation in two dimensions. Let \( \Omega \) be a bounded domain in \( \mathbb{R}^2 \), with boundary \( \partial \Omega \), which, for the sake of exposition, is the union of two regions \( \Omega_1 \) and \( \Omega_2 \) and a common boundary \( \Gamma \). Examples of such splittings are given in Figure 1.

Thus we shall consider the problem

\[
Lu = f \quad \text{in } \Omega, \\
u = 0 \quad \text{on } \partial \Omega,
\]

where

\[
L = - \sum_{i,j=1}^{2} \frac{\partial}{\partial x_i} \left( a_{ij} \frac{\partial v}{\partial x_j} \right)
\]

with \( a_{ij} \) symmetric, uniformly positive definite, bounded and piecewise smooth on \( \Omega \). The generalized Dirichlet form is given by

\[
A(v, \phi) = \sum_{i,j=1}^{2} \int_{\Omega} a_{ij} \frac{\partial v}{\partial x_i} \frac{\partial \phi}{\partial x_j} \, dx
\]

defined for all \( v \) and \( \phi \) in the Sobolev space \( H^1(\Omega) \) (the space of distributions with square-integrable first derivatives). The \( L^2(\Omega) \) inner product is denoted by

\[
(v, \phi) = \int_{\Omega} v \phi \, dx.
\]

The subspace \( H^1_0(\Omega) \) of \( H^1(\Omega) \) is the completion of the smooth functions with support in \( \Omega \) with respect to the norm in \( H^1(\Omega) \). By integration by parts, the

Received September 25, 1984.
1980 Mathematics Subject Classification. Primary 65N30, 65F10.

*This work was supported in part under the National Science Foundation Grant No. DMS84-05352, under the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, Contract No. DE-AC02-76CH00016, and under the Air Force Office of Scientific Research, Contract No. ISSA86-0026.
problem defined by (1) may be written in weak form: Find \( u \in H_0^1(\Omega) \), such that

\[
A(u, \phi) = (f, \phi)
\]

for all \( \phi \in H_0^1(\Omega) \). This leads immediately to the standard Galerkin approximation. Let \( S_h^0 \) be a finite-dimensional subspace of \( H_0^1(\Omega) \). The Galerkin approximation is defined as the solution of the following problem: Find \( U \in S_h^0 \) such that

\[
A(U, \chi) = (f, \chi)
\]

for all \( \chi \in S_h^0 \). Once a basis \( \{ \chi_i \}_{i=1}^N \) for \( S_h^0 \) is chosen, (3) leads to a system of linear algebraic equations. Write

\[
U = \sum_{i=1}^N \alpha_i \chi_i,
\]

then (3) becomes

\[
\sum_{i=1}^N \alpha_i A(\chi_i, \chi_j) = (f, \chi_j)
\]

\( j = 1, \ldots, N \), which is a linear system for the determination of the coefficients \( \alpha_i, \)
\( i = 1, \ldots, N \).

It is well known that for a wide class of approximation spaces, \( S_h^0, U \) will be a good approximation to \( u \). We shall consider certain spaces \( S_h^0 \) for which we may also develop efficient algorithms for the solution of the underlying linear system (4).

The underlying method which we will consider is a preconditioned iterative method. The choice of particular iterative method within a certain class is not essential, but for the purpose of this paper we may think of the well-known conjugate gradient method which is often used in practice (cf. [1], [6], [9], [10]).

Let \( A \) be the \( N \times N \) matrix with entries \( A(\chi_i, \chi_j) \), \( \alpha = (\alpha_1, \ldots, \alpha_N) \) and \( F \) the vector with components \( (f, \chi_j) \). Then (4) may be written

\[
A\alpha = F.
\]

Generally, the matrix \( A \) is not well-conditioned so that a direct application of the conjugate gradient method to the symmetric positive definite system (5) will not be a very efficient algorithm. A preconditioned conjugate gradient method can be derived as follows. Let \( B \) be a positive definite symmetric matrix and write

\[
B^{-1}A\alpha = B^{-1}F.
\]
In the context of this paper the matrix $B$ will be associated with another bilinear form $B(\cdot, \cdot)$ defined on $S_h^0 \times S_h^0$. The system (6) is symmetric with respect to the inner product defined by

$$[\alpha, \beta] = \sum_{i,j=1}^{N} B_{ij} \alpha_i \beta_j.$$  

Thus the conjugate gradient method may be applied to (6) with respect to (7). The importance of making a "good" choice for $B$ is well known. The matrix $B$ should have two main properties. First, the solution of the problem

$$B\beta = b$$

should be easy to obtain. This is tantamount to applying the operator $B^{-1}$ to the vector $b$. Secondly, $B$ should be spectrally close to $A$ in the sense that the condition number $K$ of $B^{-1}A$ should not be large. Clearly $K \leq \lambda_1/\lambda_0$, where $\lambda_0$ and $\lambda_1$ are constants such that

$$\lambda_0 [\beta, \beta] \leq [B^{-1}A\beta, \beta] \leq \lambda_1 [\beta, \beta].$$

In terms of the form $B(\cdot, \cdot)$ the first property means that the solution $W$ of

$$B(W, \chi) = (g, \chi), \quad \forall \chi \in S_h^0$$

for a given function $g$ should be easier to obtain than the solution of (3). The spectral condition, in terms of the forms, is

$$\lambda_0 B(V, V) \leq A(V, V) \leq \lambda_1 B(V, V)$$

for all $V \in S_h^0$.

These two properties will guarantee, firstly, that the amount of work per step in applying the conjugate gradient method (as an iterative method) will be small, and, secondly, that the number of steps to reduce the error to a given size will be also small so that an efficient algorithm will result.

In this paper we will describe and analyze a technique for constructing the bilinear form $B(\cdot, \cdot)$ (and hence the preconditioner $B^{-1}$) which only involves solving related Galerkin (or matrix) equations on the subregions $\Omega_1$ and $\Omega_2$. For other works dealing with the solution of boundary value problems via substructuring cf. [4], [5], [7], [13], [14].

2. The Preconditioning Algorithm. In order to present the ideas clearly, we shall specifically consider the case in which the endpoints of $\Gamma$ lie on $\partial \Omega$. This is exemplified by case (a) in Figure 1. The approximation subspace $S_h^0$ of $H_0^1(\Omega)$ is defined by first triangulating $\Omega$, for example as in Figure 2. Then $S_h^0$ is defined to be the collection of functions which are piecewise linear on the triangles, continuous on $\Omega$ and vanish on $\partial \Omega$. Notice that $\Gamma$ is a "mesh line".

In order to construct our form $B(\cdot, \cdot)$, we shall need to define two finite element spaces, related to $S_h^0$. Let $S_h(\Omega_j)$ be the restrictions to $\Omega_j$ of elements in $S_h^0$ and let $S_h^0(\Omega_2)$ consist of those elements of $S_h^0$ which vanish in $\Omega_1$ and in particular on $\Gamma$. We shall also need some related bilinear forms defined on $H_0^1(\Omega) \times H_0^1(\Omega)$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Let

\[ \tilde{A}_k(V, \chi) = \sum_{i,j=1}^{2} \int_{\Omega_k} a_{ij}^{k} \frac{\partial V}{\partial x_i} \cdot \frac{\partial \chi}{\partial x_j} \, dx, \quad k = 1, 2. \]

Here \( a_{ij}^{k} \) is a positive definite matrix for each \( k \), which may differ from \( a_{ij} \). For example, if the coefficients \( a_{ij} \) are variable we may want to choose \( a_{ij}^{k} = a_{ij}(x_k) \), where \( x_k \) is some point of \( \Omega_k \). In this case the resulting subproblems may be efficiently solved. Set \( \tilde{A}(V, \chi) = \tilde{A}_1(V, \chi) + \tilde{A}_2(V, \chi) \) on \( S_0^0 \times S_0^0 \). Let us now consider an arbitrary function \( V \in S_0^0 \). We decompose \( V \) on \( \Omega_2 \) as follows. Let \( V = V_H + V_P \), where \( V_P \in S_0^0(\Omega_2) \) and satisfies

\[ \tilde{A}_2(V_P, \chi) = \tilde{A}_2(V, \chi), \quad \forall \chi \in S_0^0(\Omega_2). \]

Notice that \( V_P \) is determined on \( \Omega_2 \) by the values of \( V \) on \( \Omega_2 \) and that

\[ \tilde{A}_2(V_H, \chi) = 0, \quad \forall \chi \in S_0^0(\Omega_2). \]

Thus, on \( \Omega_2 \), \( V \) is decomposed into a function \( V_P \) which vanishes on \( \partial \Omega_2 \) and a function \( V_H \) which satisfies the above homogeneous equations. With a slight abuse of terminology we shall refer to such a function \( V_H \) as "discrete harmonic". We now define the bilinear form \( B(\cdot, \cdot) \) on \( S_0^0(\Omega_2) \) by

\[ B(V, \phi) = \tilde{A}_1(V, \phi) + \tilde{A}_2(V_P, \phi_P). \]

We shall show that the corresponding equation

\[ B(W, \chi) = (g, \chi), \quad \forall \chi \in S_0^0 \]

can be solved by solving related Galerkin equations on \( \Omega_1 \) and \( \Omega_2 \). This is done as follows: Consider \( \chi \in S_0^0(\Omega_2) \). Then (10) reduces to

\[ \tilde{A}_2(W_P, \chi) = (g, \chi), \quad \forall \chi \in S_0^0(\Omega_2). \]

Since \( W_P \in S_0^0(\Omega_2) \) this is just the solution of a discrete Dirichlet problem on \( \Omega_2 \). With \( W_P \) now known, we write (10) as

\[ \tilde{A}_1(W, \chi) = (g, \chi) - \tilde{A}_2(W_P, \chi_P) = (g, \chi) - \tilde{A}_2(W_P, \chi). \]

The last equality follows since \( \tilde{A}_2(W_P, \chi_P) = 0 \). The equations (11) uniquely determine \( W \in S_0^0(\Omega_1) \). In fact, \( W \) is the discrete solution of a mixed Neumann-Dirichlet problem on \( \Omega_1 \). Having now determined \( W \) on \( \Omega_1 \) and, in particular, on \( \Gamma \), we determine \( W_H \) as the discrete harmonic function on \( \Omega_2 \) with values \( W \) on \( \Gamma \) and zero on the rest of \( \partial \Omega_2 \). This involves solving another discrete Dirichlet problem on \( \Omega_2 \).
The spectral equivalence of the form $A(\cdot, \cdot)$ and $B(\cdot, \cdot)$ (and hence of the matrices $A$ and $B$) will now be demonstrated. We shall show that the condition number $K$ is bounded independent of the dimension of $S_h^0$. In particular, we shall prove the following

**Theorem.** Let $A(\cdot, \cdot)$, $S_h^0$ and $B(\cdot, \cdot)$ be defined as above. Then there are positive constants $\lambda_0$ and $\lambda_1$ independent of $h$ such that

$$\lambda_0 B(V, V) \leq A(V, V) \leq \lambda_1 B(V, V), \quad \forall V \in S_h^0.$$ 

**Proof.** Because of the uniform positive definiteness of the $2 \times 2$ matrix $\{a_{ij}\}$ and the positive definiteness of the constant matrices $\{a_{ij}^k\}, k = 1, 2$, there are positive constants $\alpha_0$ and $\alpha_1$ such that

$$\alpha_0 A(v, v) \leq A(v, v) \leq \alpha_1 A(v, v)$$

for all $v \in H_0^1(\Omega)$. Thus, it remains to be shown that there exist positive constants $\beta_0$ and $\beta_1$ such that

$$\beta_0 B(V, V) \leq \tilde{A}(V, V) \leq \beta_1 B(V, V)$$

for all $V \in S_h^0$. Clearly,

$$B(V, V) \leq \tilde{A}(V, V),$$

so that $\beta_0 = 1$. Thus, we need to prove that

$$\tilde{A}(V, V) \leq \beta_1 B(V, V),$$

which will obviously follow from the inequality

$$\tilde{A}_2(V_H, V_H) \leq \gamma \tilde{A}_1(V, V)$$

with $\gamma$ independent of $h$. The inequality (12) is proved as follows: Let $v_H$ be the restriction to $\Omega_2$ of a function in $H_0^1(\Omega)$ which satisfies

$$v_H = V \quad \text{on } \Gamma$$

and

$$\tilde{A}_2(v_H, \phi) = 0 \quad \forall \phi \in H_0^1(\Omega_2).$$

This is the “harmonic” function in $\Omega_2$ taking the values $V$ on $\Gamma$. To estimate $V_H$, we compare it with $v_H$ and use known estimates. Clearly,

$$\tilde{A}_2(V_H, V_H) \leq 2 \tilde{A}_2(V_H - v_H, V_H - v_H) + 2 \tilde{A}_2(v_H, v_H).$$

Since $v_H$ vanishes on $\partial \Omega_2 \setminus \Gamma$, we have the a priori estimate (cf. [12], [8])

$$\tilde{A}_2(v_H, v_H) \leq C|v_H|_{H^{1/2}(\Omega_2)}^2$$

$$\leq C|V|_{H^{1/2} (\Gamma)}^2 \leq C|V|_{H^{1/2} (\partial \Omega)}^2.$$ 

The last two inequalities follow from the definition of $H^{1/2}(\Gamma)$ (cf. [11]) and the fact that $V$ vanishes on $\partial \Omega$.

Now from the definition of $V_H$ and $v_H$ it follows easily that

$$\tilde{A}_2(V_H - v_H, V_H - v_H) \leq \inf \tilde{A}_2(\chi - v_H, \chi - v_H)$$

with the infimum taken over functions $\chi \in S_h^0$ with $\chi = V_H$ on $\Gamma$. By well-known properties of $S_h^0$ we see that for $0 < \epsilon < 1/2$,

$$\inf \tilde{A}_2(\chi - v_H, \chi - v_H) \leq C h^{2\epsilon} \|v_H\|_{H^{1/2} (\Omega_2)}^2.$$
Now, using a well-known a priori inequality (cf. [12], [8]) and an "inverse property" of $S_h^0$ (cf. [2]), we see that

$$h^{2k} \| v_h \|_{H^{k+1}(\Omega_2)} \leq Ch^{2k} \| v_h \|_{H^{k+1}(\partial \Omega_2)} \leq C|V|_{H^{k+1}(\partial \Omega_2)}.$$

Combining the above estimates yields

$$\tilde{A}_2(V_H, V_H) \leq C|V|_{H^{k+1}(\partial \Omega_2)} \leq C\tilde{A}_1(V, V),$$

the last inequality following by a trace theorem (cf. [12]). This proves (12) which in turn completes the proof of the theorem.

3. Matrix Representation of the Operators. In this section we will describe the preconditioner $B$ in terms of block matrices. It will be shown that $B$ has a special structure and that the process for solving $B\alpha = b$ previously described may also be seen to be a block Gauss elimination process.

We shall suppose that we have the usual nodal basis for $S_h^0$ and that the nodes are partitioned into three subsets corresponding to those in $\Gamma$, $\Omega_1$ and $\Omega_2$. We shall order the corresponding vectors as follows:

$$\alpha = \begin{pmatrix} v_0 \\ v_1 \\ v_2 \end{pmatrix}$$

with $v_0$, $v_1$ and $v_2$ corresponding to the nodes on $\Gamma$, $\Omega_1$ and $\Omega_2$, respectively. In terms of block matrices the system corresponding to $B$ is

$$\begin{pmatrix} B_{00} & B_{01} & B_{02} \\ B_{01}^T & B_{11} & 0 \\ B_{02}^T & 0 & B_{22} \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}.$$

Now the first step of the solution process described in Section 2 consists of writing the equivalent system

$$\begin{pmatrix} B_{00} - B_{02}B_{22}^{-1}B_{02}^T & B_{01} & 0 \\ B_{01}^T & B_{11} & 0 \\ B_{02}^T & 0 & B_{22} \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} b_0 - B_{02}B_{22}^{-1}b_2 \\ b_1 \\ b_2 \end{pmatrix}. \tag{13}$$

The second step in the process corresponds to the solution of

$$\begin{pmatrix} B_{00}^{(1)} & B_{01} \\ B_{01}^T & B_{11} \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} = \begin{pmatrix} b_0 - B_{02}B_{22}^{-1}b_2 \\ b_1 \end{pmatrix},$$

where the entries of $B_{00}^{(1)}$ are given by $\tilde{A}_1(\phi_i, \phi_j)$ with $i$ and $j$ corresponding to nodes on $\Gamma$. Consequently, $B_{00} = B_{00}^{(1)} + B_{02}B_{22}^{-1}B_{02}^T$. Thus, we see that the $B$ has the form

$$B = \begin{pmatrix} B_{00}^{(1)} + B_{02}B_{22}^{-1}B_{02}^T & B_{01} & B_{02} \\ B_{01}^T & B_{11} & 0 \\ B_{02}^T & 0 & B_{22} \end{pmatrix},$$

and the process is just that of block Gauss elimination. The final step in the process corresponds to backsolving (13) for $v_2$, once $v_0$ and $v_1$ are known.
Remark. In the case that the forms $A(\cdot, \cdot)$ and $\tilde{A}(\cdot, \cdot)$ coincide, so that the subproblems for the original form are solved exactly, the present method coincides with that recently studied by Bjørstad and Widlund [3]. In contrast, the general method presented here only assumes the solvability of the subproblems corresponding to the form $\tilde{A}(\cdot, \cdot)$ which we are free to choose. We emphasize that the purpose here is to construct a preconditioner for the original problem which in our case is more general.

4. Applications and Numerical Experiments. In this section we shall present some results of numerical experiments which illustrate the convergence of the iterative algorithm discussed in Section 2. To this end we shall measure $K$ (the condition number of the preconditioned system), the number of iterations required to reduce the iteration error in the $L^2$-norm of the residual by some factor $\varepsilon$ and the average reduction per iteration.

In the two examples considered in this section we shall use subspaces $\{S_h^0\}$ of piecewise linear functions on a rectangular grid of size $h$. In both examples we shall use the algorithm to solve the finite element equations approximating an elliptic problem of the form

$$-\nabla \cdot (a(x, y)\nabla u) = f \quad \text{in} \ \Omega,$$

$$u = g \quad \text{on} \ \partial \Omega,$$

where $\Omega$ is the "U" shaped domain given in the following figure.

Example 1. For our first example we chose

$$a(x, y) = 1 + x/2 + y/3.$$ 

The functions $f$ and $g$ were taken to correspond to the solution $u = \sin x \sin y$. The domain $\Omega$ was split into two domains $\Omega_1$ and $\Omega_2$ and the algorithm of Section 2 was applied where the coefficients of the preconditioning form (9) were taken to be piecewise constant in $\Omega_1$ and constant in $\Omega_2$.

Table 1 gives the average error reduction per iteration and the number of iterations necessary to reduce the initial error by a factor of $10^{-4}$. The table clearly indicates that the error reduction is independent of the mesh parameters as theoretically predicted.
Example 2. In this example we study the condition number of the preconditioned system of equations for a problem with discontinuous coefficients. More precisely, we consider solving the above problem with

\[ a(x, y) = \begin{cases} 
1 & \text{in } \Omega_1, \\
y & \text{in } \Omega_2,
\end{cases} \]

where \( \gamma \) is a constant. The functions \( f \) and \( g \) are chosen so that the solution \( u \) is given by

\[
\begin{align*}
(x + y)(1 - y)^2 + 3\gamma xy + 3(1 - y)x & \quad \text{in } \Omega_1, \\
(x^2 + y^2)(1 - y)^2 + 3xy & \quad \text{in } \Omega_2.
\end{align*}
\]

Table 2 lists the condition number of the preconditioned system for various values of \( \gamma \). The results are given for \( h = 1/12 \); almost identical results were obtained for \( h = 1/3 \) and \( h = 1/6 \). Note the improved condition number as \( \gamma \) becomes small.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>[ K = \frac{\lambda_1}{\lambda_0} ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>.5</td>
<td>1.5</td>
</tr>
<tr>
<td>.1</td>
<td>1.1</td>
</tr>
<tr>
<td>.05</td>
<td>1.05</td>
</tr>
</tbody>
</table>


