Finite Element Analysis of a Scattering Problem

By A. K. Aziz* and R. Bruce Kellogg**

Abstract. A finite element method for the solution of a scattering problem for the reduced wave equation is formulated and analyzed. The method involves a reformulation of the problem on a bounded domain with a nonlocal boundary condition. The space of trial functions includes piecewise polynomial functions and functions arising from spherical harmonics.


Here we present a new numerical approximation technique with the following features:

(i) The subspaces are a combination of finite element subspaces inside the absorbing domain and spherical harmonics in the exterior domain. In addition to introducing a novel way of coupling the spherical harmonics with the field inside the body, the method has the feature that no finite elements are required outside the absorbing body; see, e.g., [9].

(ii) The stiffness matrices are nonsingular and Gaussian elimination without pivoting can be used for the solution of the linear system.

The formulation and the analysis of the present method as applied to the reduced Maxwell equations shall be discussed in a forthcoming paper. The problem considered here is important in many applications. Our work was motivated by a study of the biological effects of microwave radiation. For this problem a computer program, FEMS, has been written which uses the method described in this paper. A discussion of the program and some numerical results will be presented elsewhere.

In Section 2 we formulate our problem and in Section 3 we give the variational formulation and prove a series of lemmas in order to obtain our main result, Theorem 3.1. In Section 4 we describe the finite element procedure for our problem. We also give in this section an error estimate and a brief discussion of the finite-dimensional subspaces involved.

Received August 2, 1979; revised January 26, 1981.
1980 Mathematics Subject Classification. Primary 65N30; Secondary 35J05.
* The work of this author was supported in part by Naval Surface Weapons Center through IR funds.
** The work of this author was supported in part by the Food and Drug Administration under contract FDA 223-77-6005.
2. Formulation of the Problem. Let $\Omega \subset \mathbb{R}^3$ be a bounded domain with smooth boundary $\Gamma$ and $\Omega_0 = \mathbb{R}^3 \setminus \Omega$ be the exterior domain. Let $(u, v)$ and $\|u\|$ denote, respectively, the inner product and norm in $L_2(\Omega)$. Similarly, let $\langle u, v \rangle$ and $|u|$ denote the inner product and norm in $L_2(\Gamma)$. We shall use the Sobolev spaces $H^s(\Omega)$, $H^s(\Gamma)$, and we let $\|u\|_s$ and $|u|_s$, respectively, designate the norms in these spaces. We let $n$ denote the exterior normal to $\Omega$ and let $u_n$ denote the normal derivative of $u$. Suppose $\kappa(x)$ is a bounded, complex valued function on $\mathbb{R}^3$ such that

\[
\Im \kappa(x)^2 > a > 0, \quad x \in \Omega,
\]

\[
k(x) = \kappa_0 > 0, \quad x \in \Omega_0.
\]

Let $f(x)$ be a function on $\mathbb{R}^3$ with $f(x) = 0$ for $x \in \Omega_0$, and let $u_0$ satisfy $\Delta u_0(x) + \kappa_0^2 u_0(x) = 0$, $x \in \mathbb{R}^3$. We shall be concerned with the Problem I: to find $u(x)$ in $H^1_{s0}(\mathbb{R}^3)$ satisfying, with $|x| = r$,

\[
\begin{align*}
\Delta u + \kappa(x)^2 u &= f \quad \text{in } \mathbb{R}^3 \setminus \Gamma, \\
u - u_0 &= O(r^{-1}), \quad r \to \infty, \\
\frac{\partial}{\partial r} (u - u_0) - i\kappa_0 (u - u_0) &= o(r^{-1}), \quad r \to \infty.
\end{align*}
\]

We shall also introduce an auxiliary problem which we call Problem II $\pm$. To this end, let $s \in \mathbb{R}^1$, and $g \in H^s(\Gamma)$ be given. We seek a function $v$ such that

\[
\begin{align*}
\Delta v_\pm + \kappa_0^2 v_\pm &= 0 \quad \text{in } \Omega_0, \\
v_\pm(x) &= g(x), \quad x \in \Gamma, \\
v_\pm(x) &= O(r^{-1}), \quad r \to \infty, \\
\frac{\partial v_\pm}{\partial r} + i\kappa_0 v_\pm &= o(r^{-1}), \quad r \to \infty.
\end{align*}
\]

It is known [11] that the above problem has a unique solution, which can be expressed in terms of a Green's function $G_\pm(x,y)$ by

\[
v_\pm(x) = \int_{\Gamma} \frac{\partial G_\pm}{\partial n_y}(x,y) g(y) \, ds_y.
\]

The function $G_\pm(x,y)$ is smooth for $x \in \Omega_0, y \in \Gamma$, but becomes singular as $x \to y$. It is also known that the normal derivative $\partial v_\pm / \partial n$ is well defined on $\Gamma$, and $\partial v_\pm / \partial n \in H^{s-1}(\Gamma)$. We let $K_\pm : H^s(\Gamma) \to H^{s-1}(\Gamma)$ be the mapping defined by $K_\pm g = \partial v_\pm / \partial n$. Then for each $s \in \mathbb{R}^1$, $K_\pm$ is a bounded map from $H^s(\Gamma)$ into $H^{s-1}(\Gamma)$. Also, $K_\pm$ is a pseudodifferential operator on $H^s(\Gamma)$ of order 1. If, in a neighborhood $N$ of a point $x^* \in \Gamma$, the surface $\Gamma$ coincides with the plane $x_3 = 0$, and, in $N$, $\Omega$ lies in the half space $x_3 < 0$, it may be shown, using the Fourier transformation, that the symbol of $K_\pm$ at $x^*$ is

\[
\sigma_\pm(x^*, \xi_1, \xi_2) = -\left(\xi_1^2 + \xi_2^2\right)^{1/2}.
\]

For the theory of elliptic boundary value problems with pseudodifferential operators, see, for example, [2].

3. Variational Formulations. To give a finite element procedure for the approximate solution of (2.2) and (2.3), we reformulate the problem by introducing a bilinear form. To this end let $\gamma$ denote the trace operator, restricting a function to
Thus, \( \gamma: H^1(\Omega) \rightarrow H^{1/2}(\Gamma) \) is a bounded operator. Let \( \langle g, h \rangle \) denote the inner product in \( L^2(\Gamma) \), and also the pairing between \( g \in H^2(\Gamma) \) and \( h \in H^{-1}(\Gamma) \). We define a bilinear form on \( H^1(\Omega) \times H^1(\Omega) \) by

\[
B(u, w) = - (\nabla u, \nabla w) + (\kappa^2 u, w) + \langle K \gamma u, \gamma w \rangle.
\]

Clearly \( B \) is a bounded form. Using \( B \), we formulate Problem III: find \( u \in H^1(\Omega) \) such that

\[
B(u, w) = B(u_0, w) + (f - \Delta u_0 - \kappa^2 u_0, w), \quad w \in H^1(\Omega).
\]

Using the following lemma, we see that Problem III gives a reformulation of Problem I.

**Lemma 3.1.** Let \( u \) solve Problem I, and suppose \( u \in H^2(\Omega) \). Then \( u \) solves III.

**Proof.** Set \( z = u - u_0 \). From Green’s formula,

\[
(f - \Delta u_0 - \kappa^2 u_0, w) = (\Delta z + \kappa^2 z, w) = - (\nabla z, \nabla w) + (\kappa^2 z, w) + \langle z, w \rangle.
\]

In \( \Omega_0 \), \( z \) solves Problem II+ with \( g = \gamma z \). Hence \( z = K \gamma z \), and we have (3.1).

Let

\[
H^2_{\pm}(\Omega) = \{ u \in H^2(\Omega): u_n = K \gamma u \}.
\]

Then \( H^2_{\pm}(\Omega) \) is a closed subspace of \( H^2(\Omega) \). We define mappings

\[
A_{\pm}: H^2_{\pm}(\Omega) \rightarrow H^0(\Omega), \quad A_u = \Delta u + \kappa^2 u, \quad A_{-u} = \Delta u + \kappa^2 u.
\]

Using Green’s formula we have

\[
(A_{+} u, w) = B(u, w), \quad u \in H^2_{+}(\Omega), w \in H^1(\Omega).
\]

We show

**Lemma 3.2.**

\[
(A_{+} u, w) = (u, A_{-} w), \quad u \in H^2_{+}(\Omega), w \in H^2_{-}(\Omega),
\]

\[
\|u\| \leq \|A_{\pm} u\|, \quad u \in H^2_{\pm}(\Omega).
\]

**Proof.** Let \( u \in H^2_{+}(\Omega), w \in H^2_{-}(\Omega) \), so

\[
(A_{+} u, w) = - (\nabla u, \nabla w) + (\kappa^2 u, w) + \langle u_n, w \rangle.
\]

Let \( U \) denote the solution of Problem II+ with \( g = \gamma u \), and let \( W \) denote the solution of Problem II± with \( g = \gamma w \).

Since \( u_n = K \gamma u \), we have \( U_n = u_n \) on \( \Gamma \). Let \( B_R \) denote the ball with center 0 and radius \( R \), and with surface \( S_R \). Choose \( R \) so that \( B_R \supset \Omega \), and let \( \Omega_{0,R} = B_R \cap \Omega_0 \).

Then, since \( \Delta u + \kappa^2 u = 0 \) in \( \Omega_0 \),

\[
\langle u_n, w \rangle = \int_{\Omega_{0,R}} \left[ - \nabla U \cdot \nabla W + \kappa^2 U \bar{W} \right] dx + \int_{S_R} \bar{W} \frac{\partial u}{\partial r} ds.
\]

Inserting this formula in (3.5) and using (2.6), we obtain

\[
(A_{+} u, w) = - (\nabla u, \nabla w) + (\kappa^2 u, w)
\]

\[
+ \lim_{R \to \infty} \left\{ \int_{S_R} \left[ - \nabla U \cdot \nabla W + \kappa^2 U \bar{W} \right] dx + i \kappa \int_{S_R} U \bar{W} \, dS \right\}.
\]
Similarly, if \( w \in H^2_\pm(\Omega) \) and \( u \in H^2_\pm(\Omega) \), we obtain
\[
(A_w, u) = - (\nabla w, \nabla u) + (\kappa^2 w, u) + \lim_{R \to \infty} \left\{ \int \left[ - \nabla W \cdot \nabla \bar{U} + \kappa^2 W \bar{U} \right] dx - i \kappa_0 \int_{S_R} W \bar{U} ds \right\}.
\]

Comparing (3.6) and (3.7), we get (3.3). Setting \( u = w \) in (3.6) or (3.7), taking the imaginary part, and using (2.1), we obtain (3.4).

Remark. Using (3.2), (3.4), and a limiting argument, we obtain
\[
a \|u\|^2 < |B(u, u)|, \quad u \in H^1(\Omega).
\]
The density of \( H^2_\pm(\Omega) \) in \( H^1(\Omega) \) follows from Lemma 3.4 below and density properties of interpolation spaces.

We regard \( A_\pm \) as an unbounded operator on \( L^2(\Omega) \) with domain \( H^2_\pm(\Omega) \). With this stipulation we have

Lemma 3.3. \( A_\pm \) is a closed, densely defined, invertible operator on \( L^2(\Omega) \) and
\[
(A_\pm)^* = A_\mp.
\]

Proof. Since \( H^2_\pm(\Omega) \) contains smooth functions which vanish near \( \Gamma, H^2_\pm(\Omega) \) is dense in \( L^2(\Omega) \). From (3.4) it follows that \( A_\pm \) is (1-1). We show that the boundary condition \( u_\pm = \mathcal{K}_\pm u \) covers the operator \( A_\pm \) in \( \Gamma \). For suppose that, in a neighborhood \( N \) of a point \( x^* \in \Gamma \), the surface \( \Gamma \) coincides with the plane \( x_3 = 0 \), and, in \( N \), \( \Omega \) lies in the half space \( x_3 < 0 \). The covering condition at \( x^* \) requires that if \( z(t) \) satisfies the equations
\[
z'' - (\xi_1^2 + \xi_2^2)z = 0, \quad t < 0,
z(t) \to 0, \quad t \to -\infty,
z'(0) = \sigma_\pm(x^*, \xi_1, \xi_2)z(0),
\]
then \( z(t) \equiv 0 \). Using (2.8), we easily verify this, so the covering condition is satisfied at \( x^* \). Since the covering condition is preserved under a change of independent variables, it follows that the covering condition holds at each point of \( \Gamma \). Hence the a priori inequality holds [2, p. 101], and, using (3.4), we obtain
\[
\|u\|_2 < c \|A_\pm u\|, \quad u \in H^2_\pm(\Omega).
\]

From (3.9) it easily follows that \( A_\pm \) is a closed operator and that the range of \( A_\pm \) is a closed subspace of \( L^2(\Omega) \). Since Problem I has a solution for all smooth \( f \), the range of \( A_\pm \) is dense in \( L^2(\Omega) \). Hence the range of \( A_\pm \) is \( L^2(\Omega) \), and \( (A_\pm)^{-1} \) is a bounded operator on \( L^2(\Omega) \). Setting \( A_\pm u = f, A_\pm w = g \) in (3.3), we obtain
\[
(f, A^{-1}_\pm g) = (A^{-1}_\pm f, g), \quad f, g \in L^2(\Omega).
\]
Hence \( (A^{-1}_\pm)^* = A^{-1}_\mp \), so, from [17, Chapter 8, Theorem 6.2], \( (A_\pm)^* = A_\mp \).

Lemma 3.4. \([H^2_\pm(\Omega), L^2(\Omega)]_{1/2} = H^1(\Omega)\).

Proof. Since \( H^2_\pm(\Omega) \subset H^2(\Omega) \), we see that the inclusion \( \subset \) holds. We must show the converse inclusion. We make use of the spaces \( H^{s_\pm} \) defined in [8]. Given \( f \in H^1(\Omega) \) we must show that there is a function \( u(x, y) \in H^{2,1}(\Omega, R^1) \) such that
\[
u(x, 0) = f(x), \quad x \in \Omega,
\]
\[
u_\pm(x, y) = \mathcal{K}_\pm u(x, y), \quad x \in \Gamma, y \in R^1.
\]
To this end we must first discuss the operator $K_\pm$ on the space $H^{r}\Gamma, R^1)$. Let $-\Lambda$ be the Laplace-Beltrami operator on $\Gamma$; then $I + \Lambda$ is positive definite and

$$(I + \Lambda)^{\pm 1/2}: H^r(\Gamma) \to H^{r+\pm 1}(\Gamma).$$

Therefore $(I + \Lambda)^{-1/2}K_\pm$ is a bounded operator on $H^r(\Gamma)$ and may be defined as a bounded operator on $H^r(\Gamma \times R^1)$. Since $\Lambda$ is a second order differential operator, by [8, Lemma 4.2],

$$\Lambda: H^{s-2, -2s/r}(\Gamma, R^1) \to H^{r-s, -2s/r}(\Gamma, R^1)$$

is a bounded operator. Hence, by interpolation,

$$(I + \Lambda)^{1/2}: H^{r+s}(\Gamma, R^1) \to H^{r-1, -s/r}(\Gamma, R^1)$$

is a bounded operator. Writing $K_\pm = (I + \Lambda)^{1/2} (I + \Lambda)^{-1/2}K_\pm$,

$$K_\pm: H^{s-2, -2s/r}(\Gamma, R^1) \to H^{r-1, -s/r}(\Gamma, R^1)$$

is a bounded operator.

Returning to the solution of (3.9), (3.10) we set $f_0(x)w = f(x)$. Since $\gamma f \in H^{1/2}(\Gamma)$, by [8, Lemma 4.1], there is a $g_0(x, y) \in H^{3/2, 1/4}(\Gamma, R^1)$ such that $g_0(x, 0) = f(x), x \in \Gamma$. Then we find a $u(x, y) \in H^{2, 1}(\Omega, R^1)$ such that

$$u(x, 0) = f_0(x), \quad x \in \Omega,$$

$$u(x, y) = g_0(x, y), \quad x \in \Gamma, y \in R^1,$$

$$u(x, y) = g_1(x, y), \quad x \in \Gamma, y \in R^1.$$

Hence $u$ satisfies (3.9), (3.10) and the proof is complete.

We now establish the “inf sup condition” for the form $B$.

**Theorem 3.1.** There is a $c > 0$ such that for each $u \in H^1(\Omega)$ there is a $v \in H^1(\Omega)$ such that $|B(u, v)| > c\|u\|_1\|v\|_1$.

**Proof.** Regarding $A_\pm$ as a bounded invertible map of $H^2(\Omega) \to L^2(\Omega)$, we extend $A_+$ to a map of $L^2(\Omega) \to H^{-2}(\Omega)'$ as follows. If $f \in L^2(\Omega)$, we consider $A_+ f$ as a linear functional on $H^2(\Omega)$ according to the formula

$$(A_+ f)(w) = (f, A_+ w), \quad w \in H^2(\Omega).$$

In particular, if $f \in H^2(\Omega)$, we see that

$$(A_+ f)(w) = (A_+ f, w),$$

so this definition agrees with the previous definition of $A_+$. The extended map is also easily seen to be (1-1) and invertible. By interpolation we find that

$$A_+ : \left[ H^2(\Omega), L^2(\Omega) \right]_{1/2} \to \left[ L^2(\Omega), H^{-2}(\Omega)' \right]_{1/2}$$

is a bounded, (1-1) invertible map. From Lemma 3.4 and the dual properties of interpolation [5] it follows that $A_+ : H^1(\Omega) \to H^1(\Omega)'$ is a bounded, (1-1) invertible map. By taking the limit in (3.2), we find that for $u \in H^1(\Omega), w \in H^1(\Omega)$,

$$(A_+ u)(w) = B(u, w), \quad u, w \in H^1(\Omega).$$

Now let $w \in H^1(\Omega)$ be given. Select $\xi \in H^1(\Omega)'$ such that

$$\xi(w) = \|w\|_1, \quad \|\xi\|_{H^1(\Omega)} = 1.$$
Let \( u = A^{-1}_+ \xi \in H^1(\Omega) \). Then from (3.11) we have
\[
B(u, w) = \|w\|_1 > c\|w\|_1 - \|u\|_1,
\]
where \( c = \|A^{-1}_+\|^{-1} \), and where the norm refers to the map \( A^{-1}_+: H^1(\Omega) \to H^1(\Omega) \).

For our estimate we also require

**Lemma 3.5.** There are constants \( c_i > 0 \) (\( i = 1, 2 \)) such that for \( u \in H^1(\Omega) \)
\[
|B(u, u)| > c_1\|u\|_1^2 - c_2\|u\|_0^2.
\]

**Proof.** We have
\[
|B(u, u)| > -\text{Re} B(u, u) = (\nabla u, \nabla u) - \text{Re}(\kappa^2 u, u) - \text{Re}<K_+ \gamma u, \gamma u>
\]
\[
> c_3\|u\|_1^2 - c_4\|u\|_0^2 - \text{Re}<K_+ \gamma u, \gamma u>.
\]

Let \( \Omega_1 \) be the region between \( \Omega \) and \( |x| < R \), where \( R > \text{diameter of } \Omega \), and denote by \( U \) the extension of \( u \) to \( \Omega_{\text{ext}} \). Then
\[
\int_{\Omega_1} [ -|\nabla U|^2 + \kappa_0^2 |U|^2 ] \, dx = -\int_{|x| = R} \frac{\partial U}{\partial r} \, ds + \langle K_+ \gamma u, \gamma u \rangle.
\]

Hence
\[
(3.13) \quad \text{Re}<K_+ \gamma u, \gamma u> < \kappa_0^2 \int_{\Omega_1} |U|^2 \, dx + \int_{|x| = R} |\nabla U| \, ds.
\]

Now we select \( R \) so that \( \kappa_0^2 \) is not an eigenvalue of the problem
\[
\Delta z + \lambda z = 0 \quad \text{in } \Omega_1, \quad z(x) = 0, \quad x \in \partial \Omega_1.
\]

Then we may consider \( U \) as the solution of the well-posed Dirichlet problem
\[
(3.14) \quad \Delta U + \kappa_0^2 U = 0 \quad \text{in } \Omega_1, \quad U \text{ given on } \partial \Omega_1.
\]

We now show that, for any real \( s \), there is a constant \( c(s) > 0 \) such that
\[
(3.15) \quad \|U\|_{H^s(\Omega_1)} \leq c(s)\|U\|_{H^{s-1/2}(\partial \Omega_1)}.
\]

(See, e.g., [3, Theorem 2.4.2] for a related assertion.) To prove this, we first note that if \( \phi \) satisfies
\[
\Delta \phi + \kappa_0^2 \phi = f \quad \text{in } \Omega_1, \quad \phi = g \quad \text{on } \partial \Omega_1,
\]
then
\[
(3.16) \quad \|\phi\|_{H^{s+2}(\Omega_1)} \leq c_1(s)\|f\|_{H^s(\Omega_1)} + \|g\|_{H^{s+1/2}(\partial \Omega_1)} , \quad s > 0.
\]

Setting \( \phi = U \), we obtain (3.15) for \( s > 2 \). Next, we choose \( \phi \) so that \( g = 0 \). Then, by Green's formula,
\[
\int_{\Omega_1} U f \, dx = -\int_{\partial \Omega_1} U \frac{\partial \phi}{\partial n} \, ds.
\]

Hence, using (3.16) and the trace inequality
\[
\|\frac{\partial \phi}{\partial n}\|_{H^{s+1/2}(\partial \Omega_1)} \leq c_2(s)\|\phi\|_{H^{s+2}(\Omega_1)} , \quad s > 0,
\]
we obtain
\[
\int_{\Omega_1} U f \, dx \leq c_3(s)\|U\|_{H^{s-1/2}(\partial \Omega_1)}\|f\|_{H^s(\Omega_1)} , \quad s > 0.
\]
Dividing both sides by the norm of \( f \) and taking the supremum over smooth \( f \), we obtain (3.15) for \( s < 0 \). Now let \( A \) be the solution operator for the problem (3.14). Thus, we have shown that

\[
A : H^{s-1/2}(\partial \Omega) \to H^s(\Omega)
\]

is a bounded operator for \( s < 0 \) and \( s > 2 \). By interpolation we conclude that (3.17) is a bounded operator for all \( s \), so (3.15) holds for all \( s \).

Using (3.15) with \( s = \frac{1}{2} \), and (3.13), we obtain

\[
\text{Re} \langle K + \gamma u, \gamma u \rangle < c_5 \left\{ |u|^2_0 + \int_{|x| = R} \left( |U|^2 + |U_r|^2 \right) \, ds \right\}.
\]

We now use (2.7) and the fact that \( S_R \) is at a positive distance from \( \Gamma \), so \( \partial G \pm (x, y)/\partial \gamma \) is a smooth function for \( x \in S_R, y \in \Gamma \). We obtain

\[
\int_{|x| = R} \left( |U|^2 + |U_r|^2 \right) \, ds \leq c_6 |u|^2_0.
\]

Using these in (3.18),

\[
\text{Re} \langle K + \gamma u, \gamma u \rangle < c_7 |u|^2_0 < c_7 |u|^2_{\theta^{-1/2}} < c_8 |u|^2_{\theta}, \quad \frac{1}{2} < \theta < 1.
\]

From (3.12) and the inequality \( |u|^2_{\theta} \leq \varepsilon |u|^2_1 + c(\varepsilon) |u|^2_0 \), we obtain the result.

4. The Discrete Problem. To formulate our discrete approximation to Problem I, we specify a finite-dimensional subspace \( S \subset H^1(\Omega) \), and, in analogy with (3.1), we seek a \( \tilde{u} \in S \) such that

\[
B(\tilde{u}, w) = B(u_0, w) + \left( f - \Delta u_0 - \kappa^2 u_0, w \right), \quad w \in S.
\]

We shall refer to \( \tilde{u} \) as the approximate solution of Problem I, using the subspace \( S \). We first show that the approximate solution \( \tilde{u} \) is well defined.

**Lemma 4.1.** There is exactly one \( \tilde{u} \in S \) satisfying (4.1).

**Proof.** The equation (4.1) when written in terms of a basis for \( S \), comprise a finite system of linear equations. To show that the system is nonsingular, it suffices to show that if \( z \in S \) and if \( B(z, w) = 0 \) for all \( w \in S \), then \( z = 0 \). Choosing \( w = z \) and applying (3.8), we obtain

\[
a||z||^2 < |B(z, z)| = 0.
\]

To analyze the discretization error \( u - \tilde{u} \), we shall show that \( B \) satisfies a discrete form of the inf sup condition. For this we first prove a weak form of the inf sup condition that may also be of use in other problems. (See Schatz [14] for a similar result.)

**Lemma 4.2.** Let \( H_i, i = 0, 1, 2 \), be three Hilbert spaces. Suppose \( H_0 \subset H_1 \) with compact injection. Let \( B \) be a bounded bilinear form on \( H_1 \times H_2 \) which satisfies: if \( u \in H_1 \) and \( B(u, v) = 0, v \in H_2 \), then \( u = 0 \). For \( n = 1, 2, \ldots \), let \( M_{in} \subset H_i, i = 1, 2 \), be two finite-dimensional subspaces of equal dimension. Suppose \( M_{2n} \subset M_{2n+1} \) and \( \bigcup_n M_{2n} \) is dense in \( H_2 \). Suppose \( B \) satisfies the “weak inf sup condition”: there are \( c_i > 0, i = 1, 2 \), such that for \( u \in M_{1n}, \) there is a \( v \in M_{2n} \) such that

\[
|B(u, v)| \geq |c_1||u||_{H_1} - c_2||u||_{H_0}, \quad ||v||_{H_2}.
\]
Then there is an integer \( N > 0 \) and a constant \( c_3 > 0 \) such that for \( n > N \), if \( u \in M_{1n} \), there is a \( v \in M_{2n} \) such that
\[
|B(u, v)| \geq c_3 \|u\|_{H^1} \|v\|_{H^1}.
\]

**Proof.** Let \( M_{2n}' \) denote the space of linear functionals on \( M_{2n} \) and define a map \( L : M_{1n} \to M_{2n}' \) by
\[
(Lu)(v) = \overline{B(u, v)}, \quad v \in M_{2n}.
\]
Then there is an integer \( N > 0 \) and a constant \( c_4 > 0 \) such that if \( n > N \)
\[
(4.3) \quad \|u\|_{H^1} \leq c_4 \|Lu\|_{M_{2n}'}, \quad u \in M_{1n}.
\]
For, if (4.3) does not hold, there are sequences \( n_j \to \infty \) and \( u_j \in M_{1n} \) such that
\[
\|u_j\|_{H^1} = 1, \quad \|Lu_j\|_{M_{2n}'} \to 0.
\]
Let \( v_j \in M_{2n} \) be chosen to satisfy (4.2) and normalized so \( \|v_j\|_{H^2} = 1 \). Then from (4.2) it follows that
\[
\|Lu_j\|_{M_{2n}'} \geq |(Lu_j)(v_j)| = |B(u_j, v_j)| \geq c_1 \|u_j\|_{H^1} - c_2 \|u_j\|_{H^0}.
\]
Hence
\[
c_1 \|u_j\|_{H^1} \leq c_2 + O(1),
\]
so \( u_j \) is a bounded sequence in \( H^1 \). Hence, selecting a subsequence, we may assume that \( u_j \to u \) in \( H^1 \), \( u_j \to u_0 \) in \( H^0 \). Let \( v \in H_2 \) be arbitrary and, from density, let \( \tilde{v}_j \in M_{2j} \) be chosen so that \( \|v - \tilde{v}_j\|_{H^2} \to 0 \). Then
\[
|B(u_j, v)| \leq |B(u_j, \tilde{v}_j)| + |B(u_j, v - \tilde{v}_j)|
\leq \|Lu_j\|_{M_{2n}'} \|\tilde{v}_j\|_{H^2} + c \|u_j\|_{H^1} \|v - \tilde{v}_j\|_{H^2} \to 0.
\]
Hence \( B(u, v) = 0 \) for all \( v \in H_2 \). By our hypothesis, \( u = 0 \) which is a contradiction and proves (4.3). Now, for \( u \in M_{1n} \), choose \( v \in M_{2n} \) to satisfy (4.2) and with \( \|v\|_{2n} = 1 \). Then
\[
\|Lu\|_{M_{2n}'} \geq |B(u, v)| \geq c_1 \|u\|_{H^1} - c_2 \|u\|_{H^0},
\]
so we obtain from (4.3)
\[
(4.4) \quad \|u\|_{H^1} \leq c_5 \|Lu\|_{M_{2n}'}, \quad n \geq N.
\]
Now let \( n \geq N, u \in M_{1n} \), and let \( v \in M_{2n} \) satisfy
\[
(Lu)(v) = \|Lu\|_{M_{2n}'} , \quad \|v\|_{H^2} = 1.
\]
Then, using (4.4), we have
\[
|B(u, v)| = \|Lu\|_{M_{2n}'} \geq c_5^{-1} \|u\|_{H^1} \|v\|_{H^1},
\]
and the proof is complete.

We now show that our approximate method gives, in a quasi-optimal sense, as good an approximation to the solution as can be expected from the subspace that is being used.

**Theorem 4.1.** Let \( S_j \subset H^1(\Omega) \) be an increasing family of finite-dimensional subspaces of \( H^1(\Omega) \) such that \( \bigcup S_j \) is dense in \( H^1(\Omega) \). Let \( u_j \in S_j \) be the approximate solution of Problem 1 using the subspace \( S_j \). Then there is a constant \( c > 0 \),
independent of \( j \) but depending on the family \( \{ S_j \} \), such that, if \( u \) is the solution of Problem I, 
\[
\| u - u_j \|_1 \leq c \inf \{ \| u - z \|_1 : z \in S_j \}.
\]

**Proof.** In Lemma 4.2, we set \( H_1 = H_2 = H^1(\Omega) \), \( H_0 = H^0(\Omega) \). Then, using Lemma 3.5, we see that the hypotheses of Lemma 4.2 hold. Using Lemmas 4.1 and 4.2, we find that there is a \( J > 0 \) such that, for \( j > J \), the hypotheses of [3, Theorem 6.2.1] hold. Hence there is a \( c > 0 \) such that, for \( j > J \), (4.5) holds. Since the cases \( j = 1, 2, \ldots, J - 1 \) are finite in number, we see that (4.5) holds for all \( j \), which proves the theorem.

To find the approximate solution \( \bar{u} \) using a subspace \( S \), we select a basis \( \{ z_i \} \), \( 1 \leq i \leq m \), of \( S \). Setting
\[
A = \begin{bmatrix} a_{ij} \end{bmatrix}, \quad a_{ij} = B(z_i, z_j), \quad F = \begin{bmatrix} f_i \end{bmatrix},
\]
and writing \( \bar{u}(x) = \sum u_i z_i(x) \), \( U = [u_i] \), we see that (4.1) may be written as the matrix system \( AU = F \). From Lemma 4.1, this matrix equation always has a solution. It is important to be able to handle large matrices. In finite element programs this is frequently done with sparse matrix routines. The aim of the next lemma is to show that the matrix \( A \) can be factored without pivoting, so the unknowns can be arranged to minimize the storage requirements of the matrix.

**Lemma 4.3.** We may write \( A = LU \), where \( L \) and \( U \) are, respectively, left and right triangular.

**Proof.** Let \( I \) be a subset of \( \{1, \ldots, m\} \), and let \( A_I \) be the principal minor of \( A \) obtained by removing column \( j \) and row \( j \) for each \( j \notin I \). Let \( S_I \subset S \) denote the subspace spanned by \( \{ z_i, i \in I \} \). Then \( A_I \) is the matrix used in finding the approximate solution \( u_I \) using the subspace \( S_I \). From Lemma 4.1, \( A_I \) is nonsingular. Hence [13], [15] the factorization \( A = LU \) may be accomplished.

Our approximate method has a potential difficulty, in that the operator \( K_+ \), and hence the bilinear form \( B \), is difficult to evaluate. We overcome this difficulty by a judicious choice of subspaces, which we now describe. Let \( V_N \subset H^1_{\text{loc}}(\Omega_0) \) be a finite-dimensional collection of functions which satisfy
\[
\Delta v + \kappa_0^2 v = 0 \quad \text{in } \Omega_0,
\]
(4.6)
\[
\begin{align*}
\gamma(x) &= O(r^{-1}), & r \to \infty, \\
\frac{\partial v}{\partial r}(x) - i\kappa_0 v(x) &= o(r^{-1}), & r \to \infty.
\end{align*}
\]
(4.7)

A specific choice of \( V_N \) arises, for example, from the separation of variables in spherical coordinates applied to (4.6). Suppose \( 0 \in \Omega \). Let \( Y_{m,n}(\theta, \phi) \) be a surface harmonic, and let \( h_{n}^{1}(\rho) \) be a spherical Bessel function [1, Chapter 10]. Then \( v(r, \theta, \phi) = h_{n}^{1}(\kappa r) Y_{m,n}(\theta, \phi) \) is a particular solution of (4.6), (4.7). We may take \( V_N \) to be the collection of all such solutions with \( 1 \leq n \leq N \).

Regarding the subspace \( V_N \) we shall make the following

**Assumption 1.** The set \( \gamma(\bigcup_N V_N) \) is dense in \( H^{1/2}(\Gamma) \). For the spherical harmonic subspaces described above, the density of \( \gamma(\bigcup_N V_N) \) in \( L^2(\Gamma) \) has been recently proved in [12]. See also [19].
We also require a collection of functions of finite element type. Let there be given a decomposition of $\mathbb{R}^3$ into simplices of maximum size $h$. Let $W_h$ be the set of restrictions to $\Omega$ of continuous piecewise linear functions on this triangulation. Let $W_{h0} = W_h \cap H^1_0(\Omega)$.

The subspace of functions used in our variational principle is formed by combining the spaces $W_{h0}$ and $V_N$. We describe two ways in which this can be done. For the first way, we pick a smooth function $\xi$ such that $\xi \equiv 1$ near $\Gamma$ and $\xi \equiv 0$ near $0$. If $v \in V_N$, then $\xi v$ is a smooth function in $\Omega$, so the resulting restriction of $\xi v$ to $\Omega$ is in $H^1(\Omega)$. We also let $\xi v$ denote this restriction. We then set $S_{hN}^1 = W_{h0} + \xi V_N$. We have

**Lemma 4.4.** If the subspaces $V_N$ satisfy Assumption 1, then the collection of all the subspaces $S_{hN}^1$, $h > 0$, $N = 1, 2, \ldots$, is dense in $H^1(\Omega)$.

**Proof.** Supposing the contrary, we have, for some $z \in H^1(\Omega)$, $z \neq 0$,

\[
\int \int_\Omega \left[ \nabla z \cdot \nabla u + zu \right] \, dx = 0, \quad u \in W_{h0} + \xi V_N.
\]

Setting $u \in W_{h0}$ and using the fact that the union of these spaces is dense in $H^1(\Omega)$, we find that (4.8) holds for all $u \in H^1_0(\Omega)$. Hence $-\Delta z + z = 0$ in $\Omega$, and $z$ has a normal derivative $z_\nu$ on $\Gamma$ with $z_\nu \in H^{-1/2}(\Gamma)$. Set $u = \xi v$, $v \in V_N$. Then from (4.8) we obtain

\[
\int_\Gamma z_\nu v \, d\sigma = 0.
\]

By Assumption 1, (4.9) holds for all $v \in H^{1/2}(\Gamma)$. Hence $z_\nu = 0$, so $z = 0$, which is a contradiction.

Using Lemma 4.4, we may apply Theorem 4.1 to obtain an error estimate when the subspace $S_{hN}^1$ is used. This subspace, however, has a certain disadvantage. The support of the functions in $\xi V_N$ depends on the support of $\xi$, and hence is independent of $h$. As a result, the number of nonzero matrix elements in the stiffness matrix is $O(h^{-3})$. To avoid this problem, we now give a second choice of subspace and a modified variational principle which commits a “variational crime”.

For given $h$ and $N$, let $P_{hN} : V_N \to \gamma(W_h)$ be a linear map such that $P_{hN} v - v$ is small on $\Gamma$. To be precise, we assume that $\|\gamma(P_{hN} v - v)\|_{H^{1/2}(\Gamma)}$ is small. For example, if the surface $\Gamma$ were a polyhedron, and if the triangulation conformed with $\Gamma$, we could define $P_{hN}$ by piecewise linear interpolation. We let $S_{hN}^2$ be the collection of all functions $v \in W_h$ such that $\gamma v \in P_{hN}(V_N)$. In particular, $W_{h0} \subset S_{hN}^2$. We let $Q_{hN} : \gamma(W_h) \to V_N$ be a map such that $P_{hN} Q_{hN} = I$. Thus, $Q_{hN}$ is a right inverse of $P_{hN}$. We define a bilinear form $\tilde{B}$ on $S_{hN}^2$ by

\[
\tilde{B}(u, w) = - (\nabla u, \nabla w) + (\kappa^2 u, w) + \langle K, \gamma Q_{hN} u, \gamma w \rangle, \quad u, w \in S_{hN}^2.
\]

In analogy with (3.1) we define an approximate method as follows. We seek a $\tilde{u} \in S_{hN}^2$ such that

\[
\tilde{B}(\tilde{u}, w) = B(u_0, w) + \langle f - \Delta u_0 - \kappa^2 u_0, w \rangle, \quad w \in S_{hN}^2.
\]

We remark that the subspace $S_{hN}^2$ depends not only on the spaces $W_h$ and $V_N$, but also on the choice of the approximation operator $P_{hN}$. The bilinear form $\tilde{B}$
depends not only on $S_{hN}^2$, but on the choice of the right inverse $Q_{hN}$. It is not
evident that the system (4.11) has a solution $\tilde{u}$. If there is a solution, the following
theorem gives an error estimate for it. To state the theorem, we need another
assumption on the family of subspaces $V_N$ and approximation operators $P_{hN}$.

Assumption 2. The set $\gamma(\bigcup N P_{hN}V_N)$ is dense in $H^{1/2}(\Gamma)$.

Theorem 4.2. Suppose that the subspaces $V_N$ and maps $P_{hN}$ satisfy Assumption 2.
Then there is a constant $c > 0$, which does not depend on $h$ or $N$, such that if $u$
satisfies (3.1) and $\tilde{u}$ satisfies (4.11), then

$$
(4.12) \quad \|u - \tilde{u}\|_1 \leq c \inf \{\|u - u^*\|_1 : u^* \in S_{hN}^2\} + c \|I - P_{hN}\|Q_{hN}\tilde{u}\|_{H^{1/2}(\Gamma)}.
$$

Proof. Using Assumption 2, the proof of Lemma 4.4 shows that $\bigcup S_{hN}^2$ is dense
in $H^1(\Omega)$. From Lemma 4.2, $B$ satisfies the inf sup condition on $S_{hN}^2$. Let $u^* \in S_{hN}^2$
be arbitrary. Then there is a $w \in S_{hN}^2$ such that

$$
c\|\tilde{u} - u^*\|_1 \leq B(\tilde{u} - u^*, w), \quad \|w\|_1 = 1.
$$

Hence

$$
c\|\tilde{u} - u^*\|_1 \leq B(\tilde{u} - u^*, w) = B(\tilde{u} - u, w) + B(u - u^*, w)
\leq B(\tilde{u}, w) - \tilde{B}(\tilde{u}, w) + \tilde{B}(\tilde{u}, w) - B(u, w) + c_1\|u - u^*\|_1.
$$

Using (3.1) and (4.11), we see that the middle two terms of this expression combine
to vanish. Also, using the properties of $K_+$,

$$
B(\tilde{u}, w) - \tilde{B}(\tilde{u}, w) = \langle K_+ \gamma \tilde{u} - K_+ \gamma Q_{hN}\tilde{u}, w \rangle \leq c_2\|\tilde{u} - Q_{hN}\tilde{u}\|_{H^{1/2}(\Gamma)}.
$$

Hence we obtain

$$
c\|\tilde{u} - u^*\| \leq c_2\|\tilde{u} - Q_{hN}\tilde{u}\|_{H^{1/2}(\Gamma)} + c_1\|u - u^*\|_1.
$$

Since $u - \tilde{u} = u - u^* + u^* - \tilde{u}$, we may now use the triangle inequality to obtain
the asserted result.

Remark. The last term on the right side of (4.12) is due to the "variational crime"
that has been incorporated into the bilinear form $\tilde{B}$. It would be of interest to
estimate the size of this term.

We acknowledge some useful suggestions of J. Bramble.

Institute for Physical Science and Technology
University of Maryland
College Park, Maryland 20742

and

Department of Mathematics
University of Maryland Baltimore County Campus
Baltimore, Maryland 21228

Institute for Physical Science and Technology
University of Maryland
College Park, Maryland 20742

3. I. Babuška & A. K. Aziz, "Survey lectures on the mathematical foundations of the finite element
method," The Mathematical Foundations of the Finite Element Methods with Applications to Partial