Error Control in Numerical Solution of Boundary Integral Equations

G.C. Hsiao and R.E. Kleinman
Center for the Mathematics of Waves
Department of Mathematical Sciences
University of Delaware
Newark, DE 19716
hsiao@math.udel.edu kleinman@math.udel.edu

Abstract

A method of estimating the error committed in numerical solutions of integral equations is presented. It is shown how the error can be computed in spaces other than L^2, why this is reasonable and sometimes necessary. The appropriate function spaces are also shown to lead to bounded condition numbers even for first kind equations.

Introduction

The numerical solution of the integral equations of electromagnetics has been the subject of intense study in the electromagnetic community for over thirty years, following the work of Andreasen [1]. While tremendous strides have been taken, at this time it is fair to say that many problems remain intractable due to the large electrical size of the scattering object and the oscillation of the kernels in the integral equations now in use. Furthermore, except for very simple canonical shapes, there are no computable, mathematically rigorous error estimates available. Here we discuss two aspects of these outstanding problems, deriving a–posteriori error bounds which can be computed without the knowledge of an exact solution and developing a hierarchy of spaces in which to calculate a condition number, which clearly shows how first kind operators can be well conditioned.

Residual Error Estimates - General Remarks

In the numerical solution of integral equations one is invariably faced with the problem of estimating the accuracy of a numerical solution. The significant advances in establishing convergence of Galerkin approximations (moment methods) are based on asymptotic error estimates in terms of mesh width. Here we consider the solution of integral equations using finite element methods, that is, locally supported trial and test functions, e.g. hat functions. Not only can it be shown that the approximate solution converges to the actual solution as the mesh width goes to zero but also the optimal asymptotic orders of convergence can be determined in many cases. However, asymptotic error estimates do not suffice to estimate the actual error for finite mesh widths.

The boundary and domain integral equations that arise in electromagnetics are either of first or second kind while the integral operators are either weakly singular, singular or hypersingular, that is, the kernels have singularities which are integrable, Cauchy, or non–integrable respectively. The equations may be written in the form of an operator equation

$$Lu = f \tag{1}$$

where L denotes a linear integral operator, and where the kind of equation and the singularity of the kernel determine appropriate normed function spaces X and Y for which $L : X \to Y$ is a continuous linear boundedly–invertible mapping of X into Y. These properties imply the existence of constants $c_1, c_2 > 0$ independent of u such that

$$c_1 \|Lu\|_Y \leq \|u\|_X \leq c_2 \|Lu\|_Y \,, \; \forall\, u \in X \tag{2}$$

which means that $\|u\|_X$ and $\|Lu\|_Y$ are equivalent norms, that is, $\|Lu\|_Y$ may be used as a norm in X.

One fruitful approach to measuring the quality of an approximate (numerical) solution of (1) is based on the use of residual errors. For $f \in Y$ let u and $u_a \in X$ be *exact* and *approximate* solutions of equation (1), respectively. Hence, u satisfies (1) and for the error

$$e(u_a) := u - u_a \in X \,, \tag{3}$$

we easily derive the following relation

$$Le(u_a) = Lu - Lu_a = f - Lu_a = r(u_a) \tag{4}$$

where $r(u_a) \in Y$ is referred to as the residual error at u_a. Equations (3) and (4) together with (2) lead to

$$c_1 \|r(u_a)\|_Y \leq \|e(u_a)\|_X \leq c_2 \|r(u_a)\|_Y \,. \tag{5}$$

This relation shows that the error $\|e(u_a)\|_X$ is bounded by a constant times $\|r(u_a)\|_Y$ but more importantly, because of the equivalence of the norms, the quantity $\|r(u_a)\|_Y$ *actually measures* $e(u_a)$. That is, the space X may be equipped either with norm $\|\cdot\|_X$ or $\|L\cdot\|_Y$ and it is with respect to the latter that $e(u_a)$ is measured. Moreover, practical estimates of the constants c_1 and c_2 may be obtained as described below.

Residual Error - Smooth Closed Boundaries

To illustrate these ideas more concretely we consider specific integral equations, the usual integral equations which occur in scalar scattering by an acoustically soft (Dirichlet boundary condition) surface, Γ. In two dimensions this corresponds to scattering of TM waves where the electric field has one component and it is parallel to the axis of the infinite cylindrical body. The usual boundary integral equations that arise from Green's theorem are

$$\int_\Gamma g(\mathbf{r},\mathbf{r}')\frac{\partial}{\partial n'}u(\mathbf{r}')\,ds' = u^{\text{inc}}(\mathbf{r}) \quad (6)$$

$$\frac{1}{2}\frac{\partial u}{\partial n}(\mathbf{r}) + \int_\Gamma \frac{\partial}{\partial n}g(\mathbf{r},\mathbf{r}')\frac{\partial u}{\partial n'}(\mathbf{r}')\,ds' = \frac{\partial u^{\text{inc}}}{\partial n}(\mathbf{r}) \quad (7)$$

where

$$g(\mathbf{r},\mathbf{r}') = \frac{e^{ikR}}{4\pi R} \quad \text{in } I\!R^3 \quad (8)$$
$$= \frac{i}{4}H_0^{(1)}(kR) \quad \text{in } I\!R^2,$$

$R = |\mathbf{r}-\mathbf{r}'|$ and the normal to Γ points into the exterior of Γ. To simplify the equations we introduce the shorthand operator notation for equations (6) and (7)

$$S\frac{\partial u}{\partial n} = u^{\text{inc}} \quad (9)$$

$$(\frac{1}{2}I + D')\frac{\partial u}{\partial n} = \frac{\partial u^{\text{inc}}}{\partial n} \quad (10)$$

where the operators S and D' are implicitly defined in (6) and (7). As is well known neither the first nor the second equation is uniquely solvable for all frequencies. However, the combined field equation

$$L\frac{\partial u}{\partial n} := \left(\frac{1}{2}I + D'\right)\frac{\partial u}{\partial n} + i\eta S\frac{\partial u}{\partial n} = \frac{\partial u^{\text{inc}}}{\partial n} + i\eta u^{\text{inc}} \quad (11)$$

is uniquely solvable for all frequencies provided $Re\ \eta \neq 0$. In this example it suffices, provided Γ is smooth, to look for solutions in $L^2(\Gamma)$ and measure the residual in the same space, that is $X = Y = L^2(\Gamma)$. Then equation (5) shows that the residual error in $L^2(\Gamma)$ provides both an upper and a lower bound on the error $e(u_a)$ in $L^2(\Gamma)$. The constants c_1 and c_2 may be estimated in terms of the minimum and maximum singular values of L (see [6]) which themselves may be approximated, when the equation is discretized, by the singular values of the stiffness matrix (impedance matrix). If we look a little more carefully into this simple well known problem, we find that we have not yet told the whole story. The solution we seek is the normal derivative of the total field, a physical quantity. The energy in any bounded subdomain exterior to Γ is measured in terms of the L^2–norm of not only the field quantity u, but also ∇u. Such functions are said to belong to the Sobolev space $H^1_{\text{loc}}(\text{ext}\ \Gamma)$. The boundary values, or "trace", of functions in this space are in $L^2(\Gamma)$ but in fact they have additional integrability properties which are described by the following norm

$$\|u\|^2_{\frac{1}{2}} := \|u\|_0^2 + \int_\Gamma\int_\Gamma \frac{|u(\mathbf{r}) - u(\mathbf{r}')|^2}{R^n}\,ds\,ds' < \infty \quad (12)$$

where $n = 2$ or 3 (depending on the dimension of the Euclidean space) and $\|u\|_0$ denotes the norm in $L^2(\Gamma)$. If the field quantities have higher derivatives which are square integrable in bounded subdomains of the exterior of Γ, then they are said to lie in $H^m_{\text{loc}}(\text{ext}\ \Gamma)$ where $m = 1,2,3,4,\ldots$ denotes the highest order derivatives which are square–integrable. The boundary values of such functions which we denote by $u\big|_\Gamma$, the restriction of u to Γ, will also be smoother and are in the space $H^{m-\frac{1}{2}}(\Gamma)$ which means that the $(m-1)^{\text{st}}$ derivative satisfies the relation (12) and all lower order derivatives (including the order zero) are in $L^2(\Gamma)$.

If $m \geq 2$, then the normal derivative of u, which of course will not in general be as smooth as $u\big|_\Gamma$, will lie in the Sobolev space $H^{m-\frac{3}{2}}(\Gamma)$. In such cases both u and $\frac{\partial u}{\partial n}$ will also lie in $L^2(\Gamma)$ so it makes sense to look for solutions of (11) in $L^2(\Gamma)$ and also to measure error in this space as well. However when $m = 1$, the normal derivative lies in a larger space than $L^2(\Gamma)$ and in fact lies in the dual of $H^{\frac{1}{2}}(\Gamma)$, the space of all bounded linear functionals on $H^{\frac{1}{2}}(\Gamma)$, which is denoted by $H^{-\frac{1}{2}}(\Gamma)$. The norm in this space is defined as

$$\|\frac{\partial u}{\partial n}\|_{-\frac{1}{2}} := \sup_{\substack{v\in H^{1/2}(\Gamma)\\ v\neq 0}} \frac{|\int_\Gamma \frac{\partial u}{\partial n}v\,ds|}{\|v\|_{\frac{1}{2}}} \quad (13)$$

Of course, while the norm in $H^{\frac{1}{2}}(\Gamma)$, (12), is more difficult to compute than that in $L^2(\Gamma)$, it is still possible. However, the norm in $H^{-\frac{1}{2}}(\Gamma)$ is immensely more difficult to actually compute without further modifications.

Residual Error - Open Surfaces

Historically the preferred formulation of a problem involves a second kind equation, with a weakly singular integral operator if possible, as in the above example.

For such equations the space of square integrable functions, L^2, is a suitable setting since both the data and the solution lie in this space. Moreover this is the space in which computations are most simply performed. However first kind equations occur as naturally as second kind equations, and work over the past twenty years [3,5,8,9] has shown that many of the *operators* involved are boundedly–invertible (i.e., isomorphisms), just as for second kind equations when considered in the function spaces appropriate to the physical problem involved.

In fact, for open surfaces the only naturally arising intĕgral equations are of first kind. In this case the appropriate solution space may not be L^2 no matter how smooth the data f is! For example, when the scatterer is a conducting strip of width $2a$, then the solution of the EFIE is singular of order $1/\sqrt{x^2-a^2}$ at the end points and is *not* in L^2.

If we consider the Dirichlet problem as above in the case when Γ has no interior, that is the scatterer collapses to a strip or plate, then Green's theorem yields only the first kind equation

$$Lw := \int_\Gamma g(\mathbf{r},\mathbf{r}')\, w(\mathbf{r}')\, ds' = u^{\text{inc}}(\mathbf{r}), \quad \mathbf{r}\in\Gamma, \tag{14}$$

where Γ denotes the open surface (strip or plate) and

$$w(\mathbf{r}') = \left[\frac{\partial u}{\partial n}\right]. \tag{15}$$

We denote by $\left[\frac{\partial u}{\partial n}\right]$ the difference between the values of $\frac{\partial u}{\partial n}$ on each side of Γ. There is no second kind equation in this case. Nevertheless this first kind equation is uniquely solvable for $\operatorname{Im} k \geq 0$, but w possesses singularities at the edges of Γ which means that w is not in $L^2(\Gamma)$. As before $u \in H^1_{\text{loc}}(\text{ext}\,\Gamma)$ and $u\big|_\Gamma \in H^{1/2}(\Gamma)$ where $H^{1/2}(\Gamma)$ is defined as before except that Γ is now the open surface. But for open surfaces the definition of the trace must be clarified to allow different limits depending from which side the surface is approached. In general the trace may have different values depending on the direction of approach. The jump in the trace, the difference between the two (possibly different) limiting values of u, is better behaved than the individual traces since the singular behavior at edges will be lessened. For the Dirichlet problem for the strip, it is well known that the field itself is not singular at the edges but behaves as the square root of distance to the edge. However the jump in the traces will be even smoother. The appropriate space for the jump of the traces is thus not $H^{1/2}(\Gamma)$ but a smoother (smaller) space, $\tilde{H}^{1/2}(\Gamma)$. This space is defined as follows. Let $\tilde{\Gamma}$ be an extension of Γ to a closed surface so that $\Gamma \subset \tilde{\Gamma}$. The norm in $H^{1/2}(\tilde{\Gamma})$ is defined exactly as before, (12), with $\tilde{\Gamma}$ replacing Γ. $\tilde{H}^{1/2}(\Gamma)$ is the subspace of $H^{1/2}(\tilde{\Gamma})$ which contains all functions in $H^{1/2}(\tilde{\Gamma})$ which vanish outside of Γ. The norm of $\tilde{H}^{1/2}(\Gamma)$ is therefore

$$\|u\|_{\tilde{H}^{1/2}(\Gamma)} = \|u_0\|_{H^{1/2}(\tilde{\Gamma})} \quad \text{where} \quad u_0 = \begin{cases} u & \text{on } \Gamma \\ 0 & \text{on } \tilde{\Gamma}\setminus\Gamma \end{cases} \tag{16}$$

Having defined the two spaces, $H^{1/2}(\Gamma)$ and the proper subspace $\tilde{H}^{1/2}(\Gamma)$, the corresponding dual spaces may be defined as

$$(H^{1/2}(\Gamma))' := \tilde{H}^{-1/2}(\Gamma) \tag{17}$$

and

$$(\tilde{H}^{1/2}(\Gamma))' := H^{-1/2}(\Gamma). \tag{18}$$

The norms in these negative order Sobolev spaces may be defined in a similar way as (13) but they are both very difficult to compute. The important fact is that w, the solution of the integral equation (14), lies in $\tilde{H}^{-1/2}(\Gamma)$, which is the dual of $H^{1/2}(\Gamma)$, not in $H^{-1/2}(\Gamma)$. The error, $e(w_a)$, will also lie in $\tilde{H}^{-1/2}(\Gamma)$, so that even if the exact solution were known, it would be very difficult to calculate $e(w_a)$. However it has been shown [11] (see also [4]) that the operator L defined in (14) maps $\tilde{H}^{-1/2}(\Gamma)$ to $H^{1/2}(\Gamma)$ and is one–to–one, onto and continuous. Thus the residual error $r(w_a)$ will lie in $H^{1/2}(\Gamma)$ and hence is computable. Moreover the residual error may also be computed in $L^2(\Gamma)$ but this will always be an underestimate.

Condition Number

Another aspect of error analysis which we wish to address is the role of condition number as an indicator of accuracy in numerical solutions of integral equations. The conventional approach to numerical solutions involves discretization of the integral equation and solution of the resulting linear system using any of a variety of methods. The condition number of the system is usually used as an indicator of its stability.

The condition number of an operator L, or its discretized matrix form, is given by

$$\text{Cond}(L) = \|L\|\,\|L^{-1}\| \tag{19}$$

but this requires that the proper definition of the operator norms be used. If $L : X \to Y$ then

$$\|L\| = \sup_{\substack{v\in X\\ v\neq 0}} \frac{\|Lv\|_Y}{\|v\|_X} \quad \text{and} \quad \|L^{-1}\| = \sup_{\substack{w\in Y\\ w\neq 0}} \frac{\|L^{-1}w\|_X}{\|w\|_Y}. \tag{20}$$

This definition presumes the boundedness of L and L^{-1} in the appropriate norms. Indeed the Sobolev spaces introduced above provide the appropriate spaces for the integral operators of scattering theory. However the spaces differ depending on the particular operators. That is, the operators S and $\frac{1}{2}I + D'$ have different mapping properties, and this has been a topic of intensive investigations over the last two decades [3,8,9]. If X and Y are

both $L^2(\Gamma)$ then we refer to the condition number as the L^2 condition number. If the integral operator maps the Sobolev $H^s(\Gamma)$ into $H^t(\Gamma)$ continuously, then the order of the operator is defined to be the number $\alpha := s - t$. It has been shown [5,8] that the L^2 condition number of the discretized form of an integral operator of order α varies as $O(h^{-|\alpha|})$ where h is the mesh width of the discretization. Since second kind operators of the form identity plus compact perturbation have order $\alpha = 0$, $(s = t)$, the condition number of second kind operators is bounded $(= O(h^0))$ and this is one reason for their popularity. However many numerical investigators have found that despite their high condition numbers, some first kind operators give rise to linear systems whose solutions are better behaved than equivalent second kind equations, e.g. [2,10]. The Sobolev space setting described previously provides a first step toward understanding the cause of this unexpected behavior.

First we observe that when the operator is discretized (projected onto a finite dimensional subspace) the condition number is expressed in terms of the singular values of the discretized operator (the stiffness or impedance matrix) as

$$\mathrm{cond}(L_N) = \frac{\max\limits_{0\le n\le N} \Lambda_n}{\min\limits_{0\le n\le N} \Lambda_n} \tag{21}$$

where L_N is the matrix form of L, N is the dimension of the subspace and Λ_n are the singular values of L_N, and Λ_n^2 are the eigenvalues of $L_N^* L_N$. If $\{u_n\}_1^N \subset H^s(\Gamma)$ and $\{v_n\}_1^N \subset H^t(\Gamma)$, then the stiffness matrix L_N has the following elements:

$$(Lu_n, v_m)_{H^t(\Gamma)}\,. \tag{22}$$

Since these inner products will vary with t, we write $\mathrm{cond}_t(L_N)$ to indicate the inner product used to form the matrix elements. Once the matrix is defined the condition number is given by (21). For $t = 0$, $H^0(\Gamma) = L^2(\Gamma)$, and $\mathrm{cond}_0(L_N)$ is the L^2 condition number.

Now we consider a very simple example where the condition number may be found explicitly. Let $k = 0$, and Γ be a sphere of radius one in $\mathbb{R}^3$. Further choose u_n and v_n to be spherical harmonics. It has been shown [7] that the eigenvalues λ_n and singular values of Λ_n of the operators (9) (10) and (11) (with $\eta = 1$) are given in the table:

L	λ_n	Λ_n
S	$\frac{1}{2n+1}$	$\frac{1}{2n+1}$
$\frac{1}{2}I + D'$	$\frac{n}{2n+1}$	$\frac{n}{2n+1}$
$\frac{1}{2}I + D' + iS$	$\frac{n+i}{2n+1}$	$\frac{\sqrt{n^2+1}}{2n+1}$

It is then easy to find the L^2 condition numbers according to the definition (19):

L_N	$\max \Lambda_n$	$\min \Lambda_n$	$\mathrm{cond}_0(L_N)$
S	1	$\frac{1}{2N+1}$	$2N+1$
$\frac{1}{2}I + D'$	$\frac{N}{2N+1}$	0	∞
$\frac{1}{2}I + D' + iS$	1	$\frac{1}{2\sqrt{2}}$	$2\sqrt{2}$

This table illustrates a number of points. First, the L^2 condition number of S is unbounded as $N \to \infty$, hence first kind equations were traditionally avoided. Second, zero is an eigenvalue of $\frac{1}{2}I + D'$, so L_N is singular and $\mathrm{cond}_0(L_N) = \infty$. This phenomenon is well known even for nonzero values of k where the second kind equation must be augmented at interior resonances to restore uniqueness. One way to accomplish this is to employ the combined operator $\frac{1}{2}I + D' + iS$ which is seen to have a bounded condition number. All of these results are based on the L^2 condition number. If however we consider the operators as maps from $H^s(\Gamma)$ to $H^t(\Gamma)$ then the discretized operator is given by (22) . For the above example it was shown [7] that singular values of L now depend explicitly on t as follows:

L_N	λ_n	Λ_n	$\mathrm{cond}_t(L_N)$
S	$\frac{(1+n^2)^t}{2n+1}$	$\frac{(1+n^2)^t}{2n+1}$	$O(N^{\lvert 2t-1\rvert})$
$\frac{1}{2}I + D'$	$\frac{(1+n^2)^t\, n}{2n+1}$	$\frac{(1+n^2)^t\, n}{2n+1}$	∞
$\frac{1}{2}I + D' + iS$	$\frac{(1+n^2)^t(n+i)}{2n+1}$	$\frac{(1+n^2)^t\sqrt{n^2+1}}{2n+1}$	$O(N^{\lvert 2t\rvert})$

From this table it is clear that if L_N is the discretized form of S, then the condition number will be bounded as $N \to \infty$ only if $t = 1/2$. This means that by forming the elements of the stiffness matrix, using the inner product in $H^{1/2}(\Gamma)$ will result in a well conditioned matrix. Similarly the discretized form of $\frac{1}{2}I + D' + iS$ will have a bounded condition number only if $t = 0$, that is, the matrix elements formed using the L^2 inner product. Thus we see that the discretized form of both first and second kind operators may have bounded condition numbers if the elements are computed in the appropriate space. However, it is not at all clear that condition number is a good indicator of numerical error in solving the discrete system of equations. There is a growing body of evidence to the contrary. The question of how to use information about the condition number of the stiffness matrix in error and stability analysis of a numerical solutions is still open. In our opinion, a large condition number is insufficient evidence on which to base a decision as to whether to accept or reject a particular numerical procedure.

Acknowledgment: This work was partially supported under AFOSR Grant No. F49620-94-1-0219

References

1. M. G. Andreason, Scattering from parallel metallic cylinders with arbitrary cross-section, *IEEE Trans. Antennas and Propagation*, vol. AP–14, pp. 746–754, 1964.

2. P. M. van den Berg, A. P. M. Zwamborn, G. C. Hsiao and R. E. Kleinman, Iterative solutions of first kind integral equations, in *Direct and Inverse Boundary Value Problems*, R. Kleinman, R. Kress and E. Martensen, eds., Peter Lang, Frankfurt, 213–232, 1991.

3. R. Dautray and J. L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol 4, Springer–Verlag, 1990.

4. M. Feistauer, G. C. Hsiao and R. E. Kleinman, Asymptotic and a posteriori error estimates for boundary element solutions of hypersingular integral equations, *SIAM J. Numer. Anal.*, to appear, **33**, 1996.

5. G. C. Hsiao, On the stability of boundary element methods for integral equations of the first kind, in *Boundary Elements IX* 1, C. A. Brebbia, W. L. Wendland and G. Kuhn, eds., Springer–Verlag, 177–192, 1987.

6. G. C. Hsiao and R. E. Kleinman, Feasible error estimates in boundary element methods in *Boundary Element Technology VII*, C. A. Brebbia and M. S. Ingber, eds., Computational Mechanics Publications, Southampton, 875–886, 1992.

7. G. C. Hsiao and R. E. Kleinman, Error analysis in numerical solution of acoustic integral equations, *Int. J. for Num. Meth. in Eng.* **37**, 2921–2933, 1994.

8. G. C. Hsiao and W. L. Wendland, A finite element method for some integral equations of the first kind, *J. Math. Anal. Appl.* **58**, 449–481, 1977.

9. G. C. Hsiao and W. L. Wendland, The Aubin-Nitsche lemma for integral equations, *J. Integral Equations* **3**, 299–315, 1981.

10. M. Ingber and T. J. Rudolphi, Solution of potential problems using combinations of the regular and derivative boundary integral equations, *Appl. Math. Modelling* **14**, 536–543, 1990.

11. E. P. Stephan and W. L. Wendland, An augmented Galerkin procedure for the boundary integral method applied to two–dimensional screen and crack problems, *Applicable Analysis* **18**, 185–219, 1984.