Title

ON THE CONVERGENCE OF THE CONJUGATE GRADIENT METHOD FOR SINGULAR CAPACITANCE MATRIX EQUATIONS

Author

Shieh, A. S.L.

Publication Date

1977-03-01

Submitted for Publication

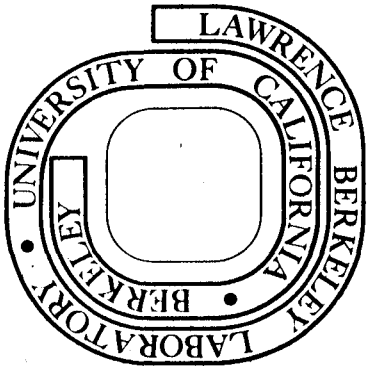
LBL-4668
Preprint c.

ON THE CONVERGENCE OF THE CONJUGATE GRADIENT METHOD FOR SINGULAR CAPACITANCE MATRIX EQUATIONS

A. S. L. Shieh

March 1977

Prepared for the U. S. Energy Research and Development Administration under Contract W-7405-ENG-48



DISCLAIMER

ON THE CONVERGENCE OF THE CONJUGATE GRADIENT METHOD FOR SINGULAR CAPACITANCE MATRIX EQUATIONS[1]

A. S. L. Shieh*

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

ABSTRACT

It is shown analytically in this work that the conjugate gradient method is an efficient means of solving the singular capacitance matrix equations arising from the Neumann problem of the Poisson equation. The total operation counts of the algorithm does not exceed constant times $n^2(\log n)^2$ $(n = 1/h)$ for any bounded domain with sufficiently smooth boundary.

---

*Now at Mathematics Research Center, Univ. of Wisconsin-Madison.

[1]Issued also as Mathematics Research Center Report #1730, University of Wisconsin-Madison. Supported in part by the Energy Research and Development Administration under Contract No. W-7405-ENG-48 and by the United States Army under Contract No. DAAG29-75-C-0024 and the National Science Foundation.

§ 1. Introduction

Over the past decade, very fast direct methods have been developed to solve Poisson's equation on certain simple regions with Dirichlet, Neumann or periodic boundary conditions. See e.g. [2], [7], [10], [17], [19], and [29]. The capacitance matrix methods are developed recently to solve Poisson's equation on arbitrary bounded regions with smooth boundaries by imbedding the discrete problem into a region where these fast direct methods are applicable. See section 7 of this work for a brief survey of previous work on capacitance matrix methods. In this work it is shown mathematically that by making the correct Ansatz guided by classical potential theory, the convergence of the conjugate gradient method for solving the capacitance matrix equations is essentially independent of the mesh size. The total operation counts of the algorithm do not exceed constant $n^2(\log n)^2$ where $h = 1/n$ is the mesh size. Only numerical schemes of first order accuracy for the interior Neumann problem of the Poisson equation on bounded two dimensional regions with smooth boundaries are considered here. See [28] for a similar treatment of the Dirichlet problem.

§2. Certain results from classical potential theory

We give only a very brief review of a few results of classical potential theory. For a detailed exposition see e.g. [8], [12], [22] and [25]. We define the potential γ resulting from a charge distribution ρ on a smooth boundary curve $\partial\Omega$ by

$$\gamma(x) = (1/\pi) \int_{\partial\Omega} \rho(\xi) \log r \, ds(\xi) .$$

Here $x = (x_1, x_2)$, $\xi = (\xi_1, \xi_2)$ and $r^2 = (x_1 - \xi_1)^2 + (x_2 - \xi_2)^2$. The Green's function $(1/2\pi) \log r$ which we shall denote by G^* satisfies

$$\Delta(1/2\pi) \log r = \delta(r) ,$$

where $\delta(r)$ is the delta function. For the interior Neumann problem, we make the Ansatz

$$u(x) = (1/2\pi) \iint_{\Omega} f(\xi) G^* d\xi + \gamma(x) \tag{2.1}$$

for the solution of

$$\begin{aligned} \Delta u &= f, \qquad x \in \Omega \\ \partial u/\partial \nu &= g, \qquad x \in \partial\Omega . \end{aligned} \tag{2.2}$$

Here ν denotes the outer normal to the boundary curve $\partial\Omega$. The first term on the right hand side of (2.1) is a space potential term and will be denoted by u_S. The boundary condition is satisfied by choosing ρ such that

$$\rho - (1/\pi) \int_{\partial\Omega} \rho \partial G^*/\partial \nu_x ds = g + (\partial/\partial\nu) u_S \Big|_{\partial\Omega} . \tag{2.3}$$

This equation can be written as

$$(I - K)\rho = \tilde{g} , \tag{2.4}$$

where K is a compact operator defined by the integral above. The equation is a Fredholm integral equation of the second kind and thus a well posed problem. It has a simple zero eigenvalue and is solvable if $\tilde{g}$ has a zero mean value. We remark that G^* in equations (2.1) and (2.3) can be replaced by the Green's function on a rectangle with zero Dirichlet boundary conditions or any other Green's function of the Laplacian.

§ 3. The capacitance matrix method for the Neumann problem

In this section we develop a similar formal potential theory for the discrete problems arising from the original Neumann problem (2.2). See also sections 3 and 5 of [26] for a similar discussion. We shall assume that uniform mesh sizes in both coordinate directions are used.

We replace the Laplace operator by the five-point formula. The Green's function G^* used in section 2 will then be replaced by the discrete Green's function on a rectangular region S with Dirichlet boundary conditions. We denote this Green's function by B^{-1} where B is the matrix representing the discrete Laplacian $h^2\Delta_h$, employing undivided differences, on S and zero boundary values on the grid points of ∂S.

We imbed Ω in S as follows. The set of mesh points is decomposed into three disjoint sets Ω_h, $\partial\Omega_h$ and $(C\Omega)_h$. The set $\partial\Omega_h$ contains all the irregular mesh points in Ω, i.e. mesh points that do not have all four neighbors within the open set Ω. Ω_h is the set of regular mesh points inside Ω and $(C\Omega_h)$ contains the remaining, the exterior mesh points. We further require that Ω is bounded away from ∂S uniformly in h. We then set up the matrix equation

$$Au = v \tag{3.1}$$

that we are solving as follows. We require that B and A differ only on the rows that corresponds to the irregular mesh points. On these rows we combine the discrete Laplacian with difference approximations to the normal derivative. We must, however, be sure that the solution on $\Omega_h \cup \partial\Omega_h$ is independent of the solution or data on $(C\Omega)_h$. This is achieved by eliminating from the discrete Laplacian, centered at an irregular mesh point, the values of the solution at its exterior neighbors. We write

$$A = B - UV^T .$$

The matrices U and V have m columns where m is the number of points in $\partial\Omega_h$. U represents an extension operator that retains the values of mesh function on $\partial\Omega_h$ and makes the remaining values equal to zero. The rows of V^T are simply the differences between the corresponding rows of B and A. After a suitable permutation, the matrix A is reducible,

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} . \tag{3.2}$$

The submatrix A_{11} is the matrix for the linear system of equations of the original discrete problem arising from discretization of the original problem (2.2). It is easily seen that the restriction to $\Omega_h \cup \partial\Omega_h$ of any solution of $Au = v$ must be a solution to the original discrete problem. V^T will be chosen so that the row sums of A_{11} and V^T vanish and A_{11} has a simple zero eigenvalue. The matrix A_{22} is nonsingular since it represents a finite difference approximation to a Dirichlet problem on $C\Omega$. It is then easily verified that the matrix A also have a simple zero eigenvalue.

We now describe our method for solving the system equations (3.1). It is solvable if and only if the right-hand side v is orthogonal to the left eigenvector of A which corresponds to the zero eigenvalue. It is shown in section 5 of [26] that the right hand side v is always consistent regardless of its values on $(C\Omega_h)$ if the data is already consistent on $\Omega_h \cup \partial\Omega_h$.

Guided by the classical potential theory, we make the Ansatz

$$u = B^{-1}v + B^{-1}UD\rho . \tag{3.3}$$

Here ρ is a m-vector to be determined. D is a nonsingular diagonal matrix containing certain scaling factors to be specified later. Computing the residual vector we obtain

$$\begin{aligned} Au - V &= (B - UV^T)(B^{-1}v + B^{-1}UD\rho) - v \\ &= (UD - UV^TB^{-1}UD)\rho - UV^TB^{-1}v . \end{aligned} \tag{3.4}$$

Because of the factor U the residuals are zero for all $x \in \Omega_h U(C\Omega)_h$. They must also vanish on $\partial\Omega_h$. We therefore multiply equation (3.4) by U^T and obtain

$$(D - V^T B^{-1} UD)\rho = V^T B^{-1} v . \tag{3.5}$$

Here we have used the relation $U^T U = I_m$, the $m \times m$ identity matrix. We shall refer to equation (3.5) as the capacitance matrix equation and the matrix on the left hand side of (3.5) as the capacitance matrix. It is shown in section 5 of [26] that the capacitance matrix which we shall denote by C has a simple zero eigenvalue and that the right hand side $V^T B^{-1} v$ of (3.5) is consistent if v is consistent for the original problem $Au = v$. For the special case when $v = UU^T v$, we can simply make the Ansatz $u = B^{-1} UD\rho$. The capacitance matrix equation now becomes

$$C\rho = U^T v . \tag{3.6}$$

Let ϕ^T satisfy $\phi^T C = 0$. Then $\phi^T = \phi^T V^T B^{-1} U$. Therefore $\phi^T V^T B^{-1} v = 0$ implies $\phi^T U^T v = 0$. Hence the right hand side $U^T v$ of equation (3.6) is again consistent if v is consistent for $Au = v$.

We now describe our choices of difference equations at the irregular mesh points.

Let $P \in \partial\Omega_h$. Let P^* be its closest point on $\partial\Omega$. Let W, E, N, S and NE be its western, eastern, northern, southern and northeastern neighbors on the mesh. Assume that the local orientation of the boundary is such that W is always in $(C\Omega)_h$ while N is either in $\partial\Omega_h$ or $(C\Omega)_h$ depending on whether P has one or two neighbors in $(C\Omega)_h$. Let $\alpha \leq \pi/4$ be the angle that the normal through P makes with the closest coordinate axis. We approximate the Neumann boundary conditions by the following first order scheme.

$$u(W) - (1 - \tan\alpha)u(P) - (\tan\alpha)u(S) = g(P^*)h \cos\alpha \tag{3.7}$$

$$u(N) - (1 - \tan\alpha)u(NE) - (\tan\alpha)u(E) = g(P^*)h \cos\alpha . \tag{3.8}$$

In our first scheme, we combine the equations (3.7) and (3.8) with the discrete Laplacian to form the following equation regardless of whether N is in $(C\Omega)_h$ or not

$$\cos\alpha[(3+\tan\alpha)u(P) - (1+\tan\alpha)u(S) - (1-\tan\alpha)u(NE) - (1+\tan\alpha)u(E)] \tag{3.9}$$
$$= \cos\alpha[h^2 f(P) + 2g(P^*)h\cos\alpha] .$$

We shall refer to this scheme as scheme I.N.a.

The second scheme is as follows. If $P \in \partial\Omega_h$ has two neighbors in $(C\Omega)_h$, we obtain the equation (3.9) as in scheme I.N.a. If only W is in $(C\Omega)_h$, we only use equation (3.7) to combine with the discrete Laplacian. We then multiply both sides of the combined equation by $2\cos\alpha$ instead of $\cos\alpha$ to obtain the following equation

$$2\cos\alpha[(3+\tan\alpha)u(P) - u(N) - u(E) - (1+\tan\alpha)u(S)] \tag{3.10}$$
$$= 2\cos\alpha[f(P) + g(P^*)h\cos\alpha] .$$

We shall refer to this scheme as scheme I.N.b. Both schemes I.N.a and I.N.b give rise to matrices A_{ll} that are positive semidefinite with null space of dimension one that consists only of constant functions. It is easily verified (see e.g. [3]) that the solutions of the discrete problems are $O(h \log h)$ approximations to the exact solutions.

The matrix D on the left hand side of equation (3.5) contains the scaling factors $d_P = \sec\alpha$. Here d_P is the diagonal element of D on the row corresponding to the point $P \in \partial\Omega_h$. The scaling factors $\cos\alpha$ and $2\cos\alpha$ in equation (3.9) and (3.10) respectively and the diagonal elements d_P of D are chosen so that the off diagonal part of C may be a formal approximation to the compact integral operator K defined by equations (2.3)-(2.4) with G^* replaced by the Green's function on a rectangle with Dirichlet boundary conditions. Because of the irregular patterns of points in $\partial\Omega_h$, the near diagonal part, the remaining part of C, will not in general be a formal approximation to the identity operator. It will be shown in section 5 that this near diagonal part is uniformly well conditioned in the spectral norm and that the singular values of C are distributed like that

of the sum of a positive definite symmetric operator and a compact symmetric operator. It is shown in [14] that the convergence of conjugate gradient method for solving operator equations with such operators depend asymptotically only on the spectral condition number of the positive definite symmetric parts of the operators. The method of proof in [14] does not apply in our case. We shall show, however, in section 6 with a different approach that the corresponding rates of convergence in our cases depend also asymptotically on the spectral condition numbers of B_h.

§ 4. Computational procedures and operation counts

We shall solve the matrix equation

$$C^T C\rho = C^T V^T B^{-1} v \tag{4.1}$$

by the conjugate gradient method. This is equivalent to solving the least square problem for the capacitance matrix equation. See section 6 for further details.

In principle we can set up the matrix C by computing $V^T B^{-1} UD$. This takes at least m fast Poisson solvers and m^2 storage requirement. It is therefore much better to use the following algorithm. For any vector v we compute Cv as follows. Generate the mesh function UDv, use the fast solver to obtain $B^{-1}UDv$ and compute $V^T B^{-1} UDv$ at an expense on the order of m operations. The vector $C^T Cv$ can be obtained in this fashion at a cost of essentially two fast solvers. It is easily seen from (6.1) and (6.2) of section 6 that each iteration of the conjugate gradient method will therefore cost about two fast solvers. The theory presented in sections 5 and 6 does not preclude the possibility that the number of iterations to achieve a given accuracy grows like $\log m$ as we refine the mesh size. We have, however, consistently found in our experiments that the number of iterations stays constant as m increases and that we can achieve an accuracy of between two and three correct decimal digits for only four iterations. The operation counts for many discrete problems are therefore ten times that of a fast Poisson solver on a rectangle and the storage requirements are of the order n^2 where $n = 1/h$. We have used the generalized marching algorithm described in [2] for our fast solver on the rectangle. The operation counts of this fast solver is approximately $3n^2 \log_2(\frac{n+1}{k})$, where k is the size of each marching step, if $n = k2^{\ell} - 1$ for some positive integer ℓ. The marching algorithm is unstable for large k. We have, however, found that $k = 16$ is good enough for our purpose. The operation counts for many problems therefore do not exceed $120\ n^2$. It is possible that if fast Fourier transform methods are used to compute $B^{-1}UDv$ in the

computation of Cv, the algorithm will be even more efficient if we exploit the sparsity of the vector UDv in the Fourier analysis step. One big advantage of the capacitance matrix method is that it can be speeded up by the replacement of a subroutine, whenever a faster Poisson solver becomes available. Finally we mention that our algorithm can be used for the numerical solution of the Neumann problem for

$$-\Delta u + Cu = f \quad \text{on} \quad \Omega, \quad C > 0 ,$$

although the theoretical results in this work does not immediately apply.

§ 5. The distribution of singular values of C

We shall show that given $\varepsilon > 0$, then almost all the singular values of C lie in the interval $[d_1 - \varepsilon, d_2 + \varepsilon]$ where d_1 and d_2 are positive numbers independent of h. This is accomplished by first proving that B_h is uniformly well conditioned in the spectral norm and that almost all the singular values of K_h lie in the interval $[0, \varepsilon]$. Our main result then follows by a simple application of a well known result in matrix theory which we shall state below as lemma 5.8.

Definition. The matrix B_h which we refer to as the near diagonal part of the capacitance matrix is defined as follows. Each entry of B_h that corresponds to the irregular mesh points P and Q is zero if $d(P, Q) > \sqrt{h}$; otherwise $B_h(P, Q) \equiv C_h(P, Q)$. Here $d(\cdot, \cdot)$ denotes the Euclidean distance function.

Definition. The matrix K_h which we refer to as the off diagonal part of the capacitance matrix is defined to be the difference between M and B_h, i.e. $K_h \equiv C - B_h$.

Theorem 5.1. Given $\varepsilon > 0$, there exists a positive integer N such that for all $0 < h$, all except N singular values of K_h lie in $[0, \varepsilon]$.

This theorem will be a consequence of lemmas 5.1-5.6 below. First we need some basic results from the theory of collectively compact operators. Let $K : X \to X$ be a compact operator on a complex Banach space X.

Definition. A subset $S \subset X$ is sequentially compact if any sequence in S contains a convergent subsequence with limit in X.

Definition. A family of operators K_n on X is collectively compact if the set $\{K_n f : \|f\| \leq 1, \ f \in X, \ n = 1, 2, \ldots\}$ is sequentially compact in X. The following result is an immediate consequence of a theorem in [1]. See also Chapter 4 of [26].

Lemma 5.1. Let $\{K_n\}$ be a family of collectively compact operators on a complex Banach space X with K_n converging pointwise to a compact operator K. Given $\varepsilon > 0$, let μ_i,

with algebraic multiplicities m_i, be the $N(\varepsilon)$ eigenvalues of K with absolute values greater than ε. Then there exists a positive integer N^* and a $\varepsilon^* > 0$ such that for all $n \geq N^*$, each ε^* neighborhood of μ_i contains exactly m_i eigenvalues of K_n while all the other eigenvalues of K_n lie in an ε-neighborhood of zero.

We now construct a family of operators $\{K_S\}$ from $\{K_h\}$ as follows. Let ϕ, ψ both mappings from the unit interval $[0,1]$ to the real line, be a smooth parameterization of $\partial\Omega$. Let P_j, $j = 1, \ldots, m$ be the irregular mesh points. Let $P_j^* \equiv (\phi(t_j), \psi(t_j))$ be the points on $\partial\Omega$ which lie on the normals through P_j with $d(P_j, P_j^*) \leq h$. We require that

$$0 \leq t_1 < t_2 < \cdots < t_m < 1 .$$

Let L_m denote the space of m tuples with the sup norm. Let $C[0,1]$ denote the Banach space of continuous functions on the unit interval with the sup norm. Let

$$k(t, t_j) \equiv \gamma K_h(P_{i-1}, P_j) + (1 - \gamma) K_h(P_i, P_j) , \tag{5.1}$$

where

$$\gamma = (t - t_i)/(t_{i-1} - t_i), \quad t_{i-1} \leq t \leq t_i, \quad i = 2, \ldots, m ; \tag{5.2}$$

$$\gamma = (t - t_1 - 1)/(t_m - t_1 - 1), \quad t \geq t_m ; \quad P_{i-1} \equiv P_m, \quad P_i \equiv P_1 ; \tag{5.3}$$

$$\gamma = (t - t_1)/(t_m - t_1 - 1), \quad t \leq t_1 ; \quad P_{i-1} \equiv P_m, \quad P_i \equiv P_1 . \tag{5.4}$$

Define bounded linear operators $P_m : C[0,1] \to L_m$, $\bar{K}_m : L_m \to C[0,1]$, and $K_m : C[0,1] \to C[0,1]$ by

$$P_m f = v, \quad v_j = f(t_j) ; \tag{5.5}$$

$$(\bar{K}_m v)(t) = \sum_{j=1}^{m} k(t, t_j) v(t_j) ; \tag{5.6}$$

$$(K_m f)(t) = (\bar{K}_m P_m f)(t) = (\bar{K}_m V)(t) . \tag{5.7}$$

We then construct another family of operators $\{K'_m\}$ from $\{K_h^T\}$ by the above procedure.

Let $K_s \equiv K'_m K_m$. It is easily verified that $\{K_s\}$ is also the family of operators formed from $\{K_h^T K_h\}$ by the same procedure.

Lemma 5.2. Let $[X]$ denote the Banach space of bounded linear operators on a Banach space X. For any $\lambda \neq 0$,

$$(\lambda I - K_h^T K_h)^{-1} \varepsilon [L_m] \quad \text{iff} \quad (\lambda I - K_s)^{-1} \varepsilon [C[0,1]] .$$

Proof. For $\lambda = 1$, the lemma is proved in [1]. Exactly the same argument applies for any $\lambda \neq 0$. Therefore, the nonzero eigenvalues of $K_h^T K_h$ and K_s coincide.

We shall now briefly describe the relations between various discrete and continuous Green's functions. Let the discrete analogue of the logarithmic Green's function be denoted momentarily by G. This discrete Green's function G has been studied in great detail in [4], [24], Chapter 3 of [27] and section 4 of [28]. It is translational invariant so that we may assume that the second parameter is fixed at the origin and define G as a function of one parameter by

$$G(a,b) = G(P;O) \quad \text{where} \quad P \equiv (a,b),\ O \equiv (0,0) .$$

Let $G^* : (a,b) \to R$ be defined by

$$G^*(a,b) = (1/2\pi) \log(a^2 + b^2) .$$

Let G_x, G_y be defined by

$$G_x(a,b) = G(a+h,b) - G(a,b) ;$$

$$G_y(a,b) = G(a,b+h) - G(a,b) .$$

Let G_x^*, G_y^* be similarly defined. It is shown in section 4 of [28] that for any nonnegative integers r and s, $r > s$, the following holds

$$(5.8) \qquad \max\{|G_x(sh,rh) - G_x^*(sh,rh)|,\ |G_y(sh,rh) - G_y^*(sh,rh|\} \leq (0.34)r^{-3} .$$

$$(5.9) \qquad G_x(sh,rh),\ G_y(sh,rh),\ -G_{xy}(sh,rh) \quad \text{are always positive} .$$

(5.10) $-G_{yy}(sh, rh - h)$ and $G_{yy}(sh, rh - h) - G_{yy}(sh - h, rh - h)$ are always nonnegative for $r \geq s$, $r \neq 0$ or $s \neq 0$.

The following property of G permits us to extend (5.8)-(5.10) to negative values of r and s

$$G(rh, sh) = G(sh, rh) = G(-sh, rh) = G(sh, -rh) . \tag{5.11}$$

Finally, G satisfies

$$h^2 \Delta_h G(rh, sh) = \begin{cases} 1, & r = 0, \ s = 0 ; \\ 0, & \text{otherwise} , \end{cases}$$

so that by (5.11), the following holds

$$G_x(0, 0) = G_y(0, 0) = 1/4 . \tag{5.12}$$

Let G' be the Green's function on the rectangle with zero Dirichlet boundary conditions. It is shown on p. 315-318 in [11] that G' and G^* differ only by a smooth function H and that B^{-1} and G differ by a mesh function H_h which is an $O(h)$ approximation to H. Using the same technique of proof used in [11], it is easy to see that if P, Q are both bounded away from ∂S uniformly in h, then $H_h(P, Q)$ is an $O(h^2)$ approximation to $H(P;Q)$. In what follows, we shall denote B^{-1} and G' by G and G^* unless stated otherwise.

Lemma 5.3. $\partial G^*/\partial \nu_P$ is uniformly continuous with respect to both parameters P and Q of G^* if both P and Q lie on a closed curve with continuously turning tangent and with continuous and bounded curvature.

Proof. This result is well known if G^* is the logarithmic potential. See e.g. [25]. Since G' and G^* differs only by a smooth function, the lemma clearly follows.

Lemma 5.4. Let P and Q be two points in $\partial\Omega_h$ with $d(P, Q) = h^\beta$, $\beta \leq 1/2$ and let P^* and Q^* be their corresponding points on $\partial\Omega$. Let α_P and α_Q be the angles that the normals through P and Q respectively make with the closest coordinate axis

$$K_h(P, Q) = 2[\partial G^*/\partial \nu_{P^*}](P^*; Q^*)\, h \sec \alpha_Q + O(h^{2-2\beta}) . \tag{5.13}$$

Proof. We shall only treat the case for scheme I.N.b when P has only one neighbor in $(C\Omega)_h$. The proof for all other cases is almost identical and will not be given. Let $W \in (C\Omega)_h$. Then

$$K_h(P, Q) = (\sec \alpha_Q / \sec \alpha_P)[\, 2G(W,Q) - 2(1 - \tan \alpha_P)G(P, Q) - 2 \tan \alpha_P G(S, Q)] \ .$$

By (5.8) and the discussion following (5.12),

$$G(W, Q) - G(P, Q) = G^*(W;Q) - G^*(P;Q) + O(h^\delta) \ ,$$

$$G(S, Q) - G(P, Q) = G^*(S;Q) - G^*(P;Q) + O(h^\delta) \ .$$

Here $\delta = \min\{2, 3(1 - \beta)\}$. Since the modulus of the second partial derivatives of $\log d(P, Q)$ is not greater than $[d(P, Q)]^{-2}$, it is easily verified that (5.13) holds.

Lemma 5.5. The family of operators $\{K_s\}$ is collectively compact on $C[0, 1]$.

Proof. We first show that $\{K_m\}$ is collectively compact on $C[0, 1]$. We construct a family of operators K_m^* on $[C[0, 1]]$ by the same procedure described above with $k(t, t_j)$ in (5.1) replaced by $k^*(t, t_j)$ defined as follows

$$k^*(t, t_j) = \gamma h \sec \alpha_{P_j} [\, \partial G^* / \partial \nu](P_{i-1}^*; P_j^*)$$

$$+ (1 - \gamma) h \sec \alpha_{P_j} [\, \partial G^* / \partial \nu](P_i^*; P_j^*) \ , \tag{5.9}$$

where the normal derivative is taken with respect to the first variable and γ is defined by (5.2)-(5.4). Let $\|k_t\| = \max_j |k(t, t_j)|$ and $\|k\| = \max_j \|k_{t_j}\|$. By lemma 5.3,

$$\|K_m^* f\| \le \text{constant } \|k\| \, \|f\| \ ,$$

$$|(K_m^* f)(t) - (K_m^* f)(t')| \le \text{constant } \|k_t - k_{t'}\| \, \|f\| \ .$$

Hence, $\{K_m^*\}$ is collectively compact on $C[0, 1]$ by the Ascoli-Arzela theorem. By lemma 6.3,

$$\|K_m - K_m^*\| \le \max_t \sum_{j=1}^{m} |k(t, t_j) - k^*(t, t_j)| = O(1) \quad \text{as} \quad m \to \infty \ .$$

Therefore $\{K_m\}$ and similarly $\{K'_m\}$ are collectively compact on $C[0,1]$. The theorem easily follows.

Lemma 5.6. $K_s f \to K^T K f$ for each $f \in C[0,1]$ where K is the compact integral operator defined by

$$(Kf)(t_P) = 2 \int_{\partial\Omega} [\partial G^*/\partial \nu_P] f \, ds \,, \tag{5.10}$$

where $P \equiv (\phi(t_P), \psi(t_P))$.

Proof. Let $\{Q_j \equiv P_{i(j)},\ j = 1, \ldots, n\}$ be a subset of $\partial\Omega_h$ that is chosen as follows. The $t_{i(j)}$ are strictly increasing as j ranges from 1 to m and $d(Q_j, Q_{j+1})$ is between $\sqrt{h}$ and $2\sqrt{h}$. Let Q_j^* be the corresponding points on $\partial\Omega$. It is easily seen that as $h \to 0$,

$$(Kf)(t_P) = 2 \sum_{j=1}^{n} [\partial G^*/\partial \nu](P; Q_j^*) d(Q_j^*, Q_{j+1}^*) f(t_{i(j)}) + 0(1) \,. \tag{5.11}$$

The $d(Q_j^*, Q_{j+1}^*)$ on the right hand side of (5.11) can be replaced by $[i(j+1) - i(j)]\, h \sec \alpha_{Q_j}$ without affecting the $0(1)$ nature of the remaining term. It then easily follows from lemma 5.4 that $K_m f \to Kf$ and similarly $K'_m f \to K^T f$ for each $f \in C[0,1]$.

Proof of Theorem 5.1. By lemmas 5.1, 5.5 and 5.6, we see that Theorem 5.1 holds for the singular values of K_m. The theorem then follows because of lemma 5.2.

Theorem 5.2. $$0.25I \le B_h^T B_h \le 7.29I \quad \text{for scheme I.N.a ;}$$
$$0.25I \le B_h^T B_h \le 14I \quad \text{for scheme I.N.b .}$$

Proof. Below we give the proof for scheme I.N.a. Details of the proof for scheme I.N.b may be found in section 5 of [28], where it is shown that the B_h matrix for schemes I.N.a and I.N.b are essentially the same as that of schemes I.a and I.b considered in [28] respectively. We shall first prove that the following holds for scheme I.N.a

$$B_h^T B_h \ge I \,. \tag{5.14}$$

The following lemma is well known.

Lemma 5.7. Let the symmetric part of a matrix A satisfy

$$(A + A^T)/2 \geq \delta I, \quad \delta > 0 .$$

Then

$$A^T A \geq \delta^2 I .$$

Let B_s denote $B_h + B_h^T$. We shall show that

$$\min_{P \in \partial\Omega_h} \{B_s(P,P) - \sum_{Q \in \partial\Omega_h, Q \neq P} |B_s(P,Q)|\} \geq 1 \tag{5.15}$$

so that (5.14) holds because of lemma 5.7 and a well known Gerschgorin theorem.

Let $P \in \partial\Omega_h$. Assume that the local orientation of the boundary near P is such that for any point $P' \in \partial\Omega_h$ in a neighborhood of P, either W' and N', the western and northern neighbors of P', are both in $(C\Omega)_h$ or W' alone is in $(C\Omega)_h$. Let E and S denote the eastern and southern neighbors of P respectively. Then

$$B_h(P,Q) = (\sec \alpha_Q / \sec \alpha_P)[G_x(W;Q) - (\tan \alpha_P)\, G_y(S;Q) \tag{5.16}$$

$$+ G_x(N;Q) - (\tan \alpha_P)\, G_y(E;Q)] .$$

Here the subscripts x and y denote the forward differences in the x_1 and x_2 directions respectively taken with respect to the first parameter of G. Because of the band structure of B_h, we may assume without loss of generality that G is the discrete analogue of the logarithmic Green's function. The error resulting from this assumption is less than a constant times $\sqrt{h}$.

Assume that $\tan \alpha_P$ is bounded away from 0 and 1. Then for h sufficiently small, $(\partial\Omega_h)_{loc}$, a subset of $\partial\Omega_h$ which contains a $\sqrt{h}$ neighborhood of P can be partitioned into blocks as follows. Let

$$I_0 = \{(0,h), \ldots, (0, M_1 h)\} ,$$

$$I_k = \{(kh, M_k h + h), \ldots, (kh, M_{k+1} h)\}, \quad k = 1, \ldots, K_1 ,$$

$$I_{-k} = \{(-kh, -M_{-(k-1)} h), \ldots, (-kh, -M_{-k} h + h)\}, \quad k = 1, \ldots, K_2 .$$

Then

$$(\partial\Omega_h)_{\text{loc}} = \bigcup_{k=-K_2}^{K_1} I_k; \quad P \in I_0 .$$

Let m_k and m_{-k} be the number of points in the nonempty sets I_{k-1} and I_{-k} respectively. Then $M_k = m_1 + \cdots + m_k$; $M_{-k} = m_{-1} + \cdots + m_{-k}$; $M_0 = 0$. Let P_j denote the point with y-coordinate jh. By (5.8), (5.16) and the smoothness of $\partial\Omega$, it is easily verified that $\sum_{Q \neq P} |B_s(P, Q)|$ will remain essentially unchanged if $\tan \alpha_Q$ is replaced throughout by $\tan \alpha_P$. Let $a \equiv \tan \alpha_P$. Let $G(i, j) \equiv G(ih, jh)$. Let $P \equiv P_i$. Then

(5.17) $$B_s(P, P) = 2 + 2(1 - a)G_{xy}(0, 0) ;$$

$$B_s(P, P_j) = (1 - a)[G_{yy}(1, |j - i| - 1) - G_{yy}(0, |j - i| - 1)], \quad P_j \in I_0 .$$

Hence, if $P_j \in I_0$, then

(5.18) $$\sum_{j>i} B_s(P, P_j) = (1 - a)[G_x(0, 0) - G_x(0, M_1 - i) - G_x(0, 1) + G_x(0, M_1 + 1 - i)]$$

(5.19) $$\sum_{i>j} B_s(P, P_j) = (1 - a)[G_x(0, 0) - G_x(0, i - 1) - G_x(0, 1) + G_x(0, i)] .$$

Similarly, if $P_j \in I_k$, $k = 1, 2, \ldots, K_1$, then

$$B_s(P, P_j) = (1 - a)[G_{xy}(k, j - i) + G_{xy}(k - 1, j - i - 1)] ,$$

$$\sum_{P_j \in I_k}' B_s(P, P_j) = -(1 - a)[G_x(k, M_k + 1 - i) - G_x(k, M_{k+1} + 1 - i) + G_x(k - 1, M_k - i) - G_x(k - 1, M_{k+1} - i)] .$$

On the other hand, if $P_{-j} \in I_{-k}$, $k = 1, 2, \ldots, K_2$, then

$$B_s(P, P_{-j}) = (1 - a)[G_{xy}(k, i + j) + G_{xy}(k - 1, i + j - 1)] ,$$

$$\sum_{P_{-j} \in I_{-k}}' B_s(P, P_{-j}) = -(1 - a)[G_x(k, M_{-(k-1)} + i) - G_x(k, M_{-k} + i) + G_x(k - 1, M_{-(k-1)} + i - 1) - G_x(k - 1, M_{-k} + i - 1)] .$$

It is easily verified from (5.9) and (5.10) that

$$B_s(P, P_j) \le 0, \quad P_j \in I_k; \quad k = \pm 1, \pm 2, \cdots$$

$$B_s(P, P_j) \ge 0, \quad P_j \in I_0 . \tag{5.20}$$

Since we only want to obtain an upper bound for $\sum'_{P \neq P_j} |B_s(P, P_j)|$, we may assume without loss of generality that $B_s(P, P_j) \neq 0$ iff $P_j \in (\partial\Omega_h)_{loc}$. Then, by summing P_j over $j > M_1$

$$- \sum' B_s(P_i, P_j) = (1 - a)[G_x(1, M_1 + 1 - i) + G_x(0, M_1 - i)$$

$$+ \sum_{k=1}^{K-1}{}' \{G_{xx}(k, M_{k+1} + 1 - i) + G_{xx}(k - 1, M_{k+1} - i)\}$$

$$- G_x(K_1, M_{K_1+1} - i) - G_x(K - 1, M_{K_1+1} - i)] .$$

By (5.9)

$$\sum_{k=1}^{K_1}{}' [G_{yy}(k + 1; M_{k+1} - i) + G_{yy}(k, M_{k+1} - 1 - i)] < G_y(2, M_2 - i) + G_y(1, M_2 - i - 1) .$$

Hence, the following holds when P_j is summed over all $j > M_1$

$$\sum' |B_s(P, P_j)| < (1 - a)[G_x(1, M_1 + 1 - i) + G_x(0, M_1 - i) \tag{5.21}$$

$$+ G_y(2, M_2 - i) + G_y(1, M_2 - i - 1)] .$$

Similarly, the following holds when P_j is summed over all $j \le 0$

$$\sum' |B_s(P, P_j| < (1 - a)[G_x(1, i) + G_x(0, i - 1) + G_y(2, M_{-1} + i - 1) + G_y(1, M_1 + i - 2)] . \tag{5.22}$$

By (5.17)-(5.22),

$$B_s(P, P) - \sum_{P \neq P_j}{}' |B_s(P, P_j)| < 2 - 4(1 - a)[G_x(0, 0) - G_x(0, 1)] - (1 - a)\ H(i) , \tag{5.23}$$

where

$$H(i) = G_x(0, i) + G_x(1, i) + G_x(0, M_1 + 1 - i) + G_x(1, M_1 + 1 - i)$$

$$+ G_y(1, M_{-1} + i - 2) + G_y(2, M_{-1} + i - 1) + G\ (1, M_2 - i - 1) + G_y(2, M_2 - i) .$$

It is easily seen from the table on p. 292 of [24] that $H(i)$ attains its maximum at either $i = 1$ or $i = M_1$. It is then easily verified with the aid of the above mentioned table that

$$4aG_x(0,0) + 4(1-a)G_x(0,1) > (1-a)H(i), \quad i = 1, \ldots, M_1 . \tag{5.24}$$

By (5.23) and (5.24), we see that (5.15) holds for our choice of P.

The proof for other choices of P is almost identical and will not be completed. Thus we complete the proof for (5.6). We now proceed to prove

$$B_h^T B_h \leq (7.29)I . \tag{5.25}$$

We shall show that the following holds

$$\max\{\sum_Q{}' |B_h(P,Q)|, \sum_Q{}' |B_h^T(P,Q)|\} \leq 2.7 . \tag{5.26}$$

We assume that the local configuration of points in $\partial\Omega_h$ in a $\sqrt{h}$ neighborhood of P is similar to that in $(\partial\Omega_h)_{loc}$ described earlier in this section. Let $B_h^*(P,Q)$ be defined by equation (5.8) with G replaced by its continuous analogue G^*. Let $B_h^{T*}(P,Q)$ be similarly defined. We shall first show that

$$\sum_{|y_Q| \geq 3} |B_h^*(P,Q)| \leq 0.7 . \tag{5.27}$$

Let r, r^*, r_1, r', r_* and r_2 denote $d(P,Q)$, $d(w,Q)$, $d(S,Q)$, $d(NE,Q)$, $d(N,Q)$ and $d(E,Q)$ respectively. We again assume that $\tan\alpha_P \equiv \tan\alpha_Q \equiv a$ for all Q such that $B_h(P,Q) \neq 0$. Then

$$2\pi \cdot B_h^*(P,Q) = -\log(r^*/r) + a\log(r_1/r) - \log(r_*/r') + a\log(r_2/r') .$$

Let the coordinates of P, Q be $(0,0)$ and (x_Q, y_Q) respectively. We have

$$r^{*2} - r^2 = 2hx_Q + h^2;\ r_*^2 - r'^2 = 2hx_Q - h^2;\ r_1^2 - r^2 = 2hy_Q + h^2;\ r_2^2 - r'^2 = 2hy_Q - h^2 .$$

Hence,

$$4\pi B_h^*(P, Q) = (-2x_Q h + 2ay_Q h)(1/r^2 + 1/r'^2) + (a - 1)h^2(1/r^2 - 1/r'^2) + R_Q ,$$

where

$$|R_Q| \le [(2x_Q + h)^2 + a(2y_Q + h)^2]h^2/2r^4 + [(2x_Q - h)^2 + a(2y_Q - h)^2]h^2/2r'^4 .$$

It is easily verified that if P, Q are two points on $\partial\Omega$ with $d(P, Q) \le h^\gamma$, $0 < \gamma \le 1$, and ta is the tangent at Q to $\partial\Omega$, then

$$d(P, ta) \le (\kappa + O(1))h^{2\gamma} ,$$

where κ is the maximum absolute value of the curvatures of $\partial\Omega$. Hence, for all practical purposes we may assume $\kappa = 0$. Then

$$x_Q = ay_Q + \varepsilon h, \qquad |\varepsilon| \le 1 .$$

Hence,

$$\sum_{|y_Q| > 2} |4\pi \cdot B_h^*(P, Q)| \le F(r, r') ,$$

where

$$F(r, r') \equiv \sum_{|y_Q| > 2} \{2h^2(1/r^2 + 1/r'^2) + (1 - a)|1/r^2 - 1/r'^2|$$

$$+ (2x_Q^2 h^2 + 2ay_Q^2 h^2)(1/r^4 + 1/r'^4) + (1 + a)(h^4/2r^4 + h^4/2r'^4)$$

$$+ 2(x_Q + ay_Q)(h^3/2r^4 - h^3/2r'^4)\} .$$

Taking into account the local configuration of points in a $\sqrt{h}$ neighborhood of P for a certain $a \equiv \tan \alpha_P$, we easily see that for all $0 \le a \le 1$,

$$\sum_{|y_Q| > 2} (x_Q^2 + ay_Q^2)/r'^4 < \sum_{k=3} \{k^2/(k - 1)^4 + k^2/(k + 1)^4\} .$$

Hence,

$$F(r,r') \le \sum_{k=3} \{8/k^2 + (3-a)/(k-1)^2 + (1+a)/(k+1)^2$$
$$+ 2k^2(1/(k-1)^4 + 1/(k+1)^4)$$
$$+ 1/k^4 + 1/2(k-1)^4 + 1/2(k+1)^4$$
$$+ 2\sqrt{2}\,(1/k^3 + 1/(k+1)^3)\}$$
$$< 8.78.$$

This completes the proof of inequality (5.27). By (5.17),

$$|B_h(P,P)| \le 1 . \tag{5.28}$$

It is easily verified that $\Sigma|B_h(P,Q)|$, $P \neq Q$, $|y_Q| \le 2$ attains its maximum when $a = 0$ and P has two neighbors in $(C\Omega)_h$. A simple calculation with the aid of table I in [4] or table II in [24] shows that when summed over all Q with $0 < |y_Q| \le 2$,

$$\Sigma|B_h(P,Q)| \le 0.96 . \tag{5.29}$$

By theorem 4.3 of [28] and table I of [4], we see that

$$\sum_{|y_Q| \ge 3} |B_h(P,Q) - B_h^*(P,Q)| \le 0.04 . \tag{5.30}$$

By (5.27)-(5.30),

$$\sum_{Q \in \partial\Omega_h} |B_h(P,Q)| \le 2.7 . \tag{5.31}$$

Since the same inequality holds by a similar argument when B_h in (5.31) is replaced by B_h^T, we easily see that (5.26) and hence (5.25) holds.

Lemma 5.8. If $D = A + B$, where A and B are arbitrary matrices with singular values $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n \ge 0$ and $\beta_1 \ge \beta_2 \ge \cdots \ge \beta_n \ge 0$ respectively and $\delta_1 \ge \delta_2 \ge \cdots \ge \delta_n \ge 0$ are the singular values of D, then

$$\delta_{i+j+1} \le \alpha_{i+1} + \beta_{j+1}, \quad i, j \text{ positive integers} .$$

Proof. See e.g. ex. 28 on p. 89 of [20].

Theorem 5.3. Let d_1 and d_2 be the spectral bounds of B_h. Then given $\varepsilon > 0$, there exists a positive integer N independent of h such that all except N singular values of C lie in $[d_1 - \varepsilon, d_2 + \varepsilon]$.

Proof. Let C, K_h and B_h be matrices A, B and D respectively in lemma 5.8. By theorem 5.1, given $\varepsilon > 0$, there exists a positive integer p such that for all $h > 0$,

$$\beta_{p+j+1} \leq \varepsilon, \qquad j = 0, 1, 2, \ldots .$$

Since

$$d_1 \leq \delta_m \leq \alpha_{m-p-j} + \beta_{p+j+1}, \qquad j = 0, 1, 2, \ldots ,$$

only the last $p + 1$ singular values of C may lie to the left of $d_1 - \varepsilon$. Similarly, by letting B_h, $-K_h$ and C be the matrices A, B and D respectively in lemma 5.8, we see that only the first $p + 1$ singular values of C can lie to the right of $d_2 + \varepsilon$.

Theorem 5.4. Let $\|M\|$ and $R(M)$ denote the spectral norm and range of a matrix M respectively. Let A_{11} be the same as in (3.2). Let $A_s = A_{11}^T A_{11} | R(A_{11}^T)$ and $C_s = C^T C | R(C^T)$ be the restrictions of $A_{11}^T A_{11}$ and $C^T C$ to $R(A_{11}^T)$ and $R(C^T)$ respectively. Then

$$\|C_s^{-1}\| \leq \|A_s^{-1}\| \, \|B\|^2 \, \|D^{-1}\|^2 ,$$

where B and D are the same as in equation (3.5).

Proof. Let $v = UU^T v \in R(A)$. It is shown in section 3 that $U^T v \in R(C)$ so that we may make the Ansatz $u = B^{-1} U D \rho$ for the solution of $Au = v$ and solve the alternative form of capacitance matrix equation $C\rho = U^T v$ for ρ. Let ρ_1 be the eigenvector corresponding to the smallest eigenvalue of the positive definite matrix C_s. Let $C\rho_1 = U^T v_1$, where $v_1 = UU^T v_1$. Then

$$\|\rho_1\|^2 = \|C_s^{-1}\| \, \|U^T v_1\|^2 . \tag{5.32}$$

Here $\|v\|$ denotes the Euclidean norm of a vector v. Let $u_1 = B^{-1} U D \rho_1$. Let U^* and

U' be the extension operators from $\Omega_h \cup \partial\Omega_h$ to all mesh points and from Ω_h to $\Omega_h \cup \partial\Omega_h$ respectively that are defined the same way as U. Let $u_1^* = U^{*T}u_1 = u_2^* + u_3^*$ where $u_2^* \in N(A_{11})$, the null space of A_{11}; $u_3^* \in R(A_{11}^T)$. Because of the reducible structure of A, $U^{*T}v \in N(A_{11})$ if $v \in N(A)$. Hence we may write $u_1 = u_2 + u_3$, where $u_2 \in N(A)$, $U^{*T}u_2 = u_2^*$ and $U^{*T}u_3 = u_3^*$. Let $\varphi \neq 0 \in N(C)$. It is readily verified that $B^{-1}UD\varphi \neq 0 \in N(A)$. Since both $N(A)$ and $N(C)$ are of dimension one, we can choose φ_1 such that $B^{-1}UD\varphi_1 = u_2$. Then $u_3^* = U^{*T}B^{-1}U^*U'D(\rho_1 - \varphi_1)$. Clearly, $\|\rho_1 - \varphi_1\| \geq \|\rho_1\|$. Hence,

$$\|D^{-1}\|\|B\|\|u_3^*\| \geq \|\rho_1\| . \tag{5.33}$$

On the other hand, $A_{11}u_3^* = U^{*T}v_1$. Hence,

$$\|u_3^*\|^2 \leq \|A_s^{-1}\|\|v_1\|^2 . \tag{5.34}$$

The theorem then easily follows from (5.32)-(5.34).

§ 6. Rate of convergence of the conjugate gradient iterations

Let b denote the right hand side of the capacitance matrix equation. We are concerned with solving $C^TC\rho = C^Tb$ by the conjugate gradient method. It is shown in [21] that the conjugate gradient method gives the solution

$$\rho = C^+b + (I - P)\rho_0 .$$

Here C^+ is the generalized inverse of C. For any m-vector b, C^+b is the unique least square solution of $C\rho = b$ that is of the minimum Euclidean norm. P is the orhtogonal projection of L_m onto $R(C^T)$ and ρ_0 is the initial guess. We shall assume throughout this section that $\rho_0 \in R(C^T)$ so that $\rho = C^+b$. Let $v = C^Tb$. Then $\rho = C_s^{-1}v$. Moreover, from (6.1) and (6.2) below we see that all the relevant vectors generated in the conjugate gradient process are in $R(C^T)$. Hence, the original problem is reduced to solving $Q\rho = v$ by the conjugate gradient method where $Q \equiv C_s$ is a positive definite symmetric matrix.

We now briefly describe the conjugate gradient method. See e.g. [9], [14], [15], [16] and [23] for details. Let $p_0 = -g_0 = v - Q\rho_0$. The conjugate gradient process generates a sequence of vectors ρ_k approximating the solution ρ by

$$\rho_{k+1} = \rho_k - [g_k^Tp_k/p_k^TQp_k]p_k , \tag{6.1}$$

$$p_{k+1} = -g_{k+1} + [g_{k+1}^TQp_k/p_k^TQp_k]p_k , \tag{6.2}$$

where $g_k = Qx_k - v$. The p_k are Q-conjugate, i.e. $p_i^TQp_j = 0$, $i \neq j$. The ρ_k minimizes the quadratic form $(1/2)w^TQw - v^Tw$ on the linear variety $\rho_0 + Y_k$ where Y_k is the subspace spanned by $\{p_0, p_1, \ldots, p_{k-1}\}$. The iterates ρ_k satisfy

$$\rho_k = \rho_0 + P_{k-1}(Q)g_0 \tag{6.3}$$

where P_{k-1} is a polynomial of degree $k - 1$. It is shown in [23] that among all iterative methods that satisfy (6.3), the conjugate gradient method is optimal in the sense that

$$E(\rho_k) = (1/2)(\rho_k - \rho)^T Q(\rho_k - \rho)$$

is minimal. It then easily follows that if λ_i are the eigenvalues of Q, then

$$E(\rho_k) \le \max_{\lambda_i}(1 + \lambda_i P_{k-1}(\lambda_i))^2 E(\rho_0) , \tag{6.4}$$

for any choice of a polynomial P_{k-1} of degree $k-1$. Let $Z : (a,b) \to R$, where (a,b) are ordered pairs of positive numbers a and b be defined by

$$Z(a,b) = [(1 - \sqrt{a})/(1 + \sqrt{a})]^b .$$

It is known (see e.g. [9] or [23]) that we can select $P_{k-1}(\lambda)$ such that

$$\max_{\lambda_i} |1 + \lambda_i P_{k-1}(\lambda_i)| \le 2Z(\kappa, k) , \tag{6.5}$$

where κ is the spectral condition number of Q. On the other hand, suppose all except N eigenvalues of Q lie in the interval $[c_1, c_2]$. Let λ_i, $i = 1, \ldots, N$ be the exceptional eigenvalues. Let $\kappa_1 = c_1/c_2$. Then as before we can choose $P_{k-N}(\lambda)$ such that

$$\max_{c_1 \le \lambda_i \le c_2} |1 + \lambda_i P_{k-N}(\lambda_i)| \le 2[(1 - \sqrt{\kappa_1})/(1 + \sqrt{\kappa_1})]^{k-N} .$$

Choose $P_{k-1}(\lambda)$ such that

$$1 + \lambda P_{k-1}(\lambda) = [1 + \lambda P_{k-N}(\lambda)](\lambda - \lambda_1) \cdots (\lambda - \lambda_N)(-1)^N/\lambda_1 \cdots \lambda_N .$$

Then,

$$\max_{\lambda_i} |1 + \lambda_i P_{k-1}(\lambda_i)| \le 2Z(\kappa_1, k - N) \max_{c_1 \le \lambda \le c_2} \{ \prod_{i=1}^{N} |1 - \lambda/\lambda_i| \} . \tag{6.6}$$

By (6.4)-(6.6) and theorem 5.3, we easily have the following theorem.

Theorem 6.1. Let κ and κ_1 be the spectral condition numbers of C_s and $B_h^T B_h$ respectively. Then given $\varepsilon > 0$, there exists a positive integer N independent of k and h such that

$$E(\rho_k)/E(\rho_0) \leq \min\{4Z(\kappa, 2k), 4Z(\kappa_1 - \varepsilon, 2k - 2N)\chi(\lambda)\} .$$

Here $\chi(\lambda) = \max_{c_1 \leq \lambda \leq c_2} \prod_{i=1}^{N} |1 - \lambda/\lambda_i|$ where λ_i, $i = 1, \ldots, N$ are the N eigenvalues of C_s that lie outside $[c_1 - \varepsilon, c_2 + \varepsilon]$. Here c_1 and c_2 are the spectral bounds of $B_h^T B_h$.

Corollary 6.1. The number of iterations needed to reduce $E(\rho_k)/E(\rho_0)$ to a given accuracy can grow no faster than constant $\log m$ as $h \to 0$.

Proof. By theorem 5.4, the smallest eigenvalue of C_s is larger than constant $\cdot\, m^{-4}$. The corollary is therefore an immediate consequence of theorems 5.2, 5.3 and 6.1.

§7. Survey of previous work on capacitance matrix methods

We give here only a brief survey of previous work on capacitance matrix methods. See also section 8 of [26] for more references and more details on some of the references mentioned here. C. W. Hockney gave a brief description of a method of this type in [18]. He credited Oscar Buneman for the idea.

The papers [7] and [13] by Buzbee, Dorr, George, and Golub and George respectively used the same Ansatz

$$u = B^{-1}v + B^{-1}UD\rho$$

as is used in our algorithm to treat the Dirichlet problem. It is then shown experimentally in [26] and theoretically in [27] that the resulting capacitance matrices C are ill-conditioned and that the singular values of C cluster around zero. The conjugate gradient method was used in [13] to solve the capacitance matrix equations using an iterative imbedding technique similar to that mentioned in section 4 of this work. The number of iterations used to achieve a given accuracy are proportional to the square root of m, the order of C. The regions considered in [7] are of a rather simple type. The matrices C are positive definite symmetric and the Cholesky method is used to factorize C. The numerical results are obtained on a CDC6600 and a gain in speed of a factor three is reported in [5] for runs on CDC 7600.

The paper [26] by Proskurowski and Widlund is probably the first one that exploits the similarity between the classical potential theory and the capacitance matrix method. It is shown experimentally there that by making the correct Ansatz guided by the classical potential theory the capacitance matrix method becomes a well posed problem. The matrices C for many test regions are uniformly well conditions in the spectral norm and the convergence of the conjugate gradient iterations for these regions appears to be independent of the mesh sizes. It is then shown theoretically in [26] that for a large class of domains

and some special schemes of approximating boundary conditions, the above observation is always valid. This work is mainly an outgrowth of [26]. It extends some of the theoretical estimates in [26] to all bounded domains with smooth boundaries.

§8. Numerical experiments

To illustrate the effectiveness of our algorithm, we have used linear polynomials as test functions. Truncation errors are not present and the right hand side of equation (8.1) is always consistent

$$A_{11}u = U^{*T}v \,. \tag{8.1}$$

Let u and u^* be respectively the exact and numerical solutions of (8.1). Let $E_{max} \equiv \|\delta_{x_1}(u - u^*)\|_\infty + \|\delta_{x_2}(u - u^*)\|_\infty$, where δ_{x_1} and δ_{x_2} denote the undivided forward differences in the x_1 and x_2 directions respectively. The domain Ω is an ellipse with the ratio of half axes equal to γ and the test function u satisfies $u(x) \equiv x_1$. The following is a table of numerical results obtained by test runs on the Univac 1110 at MACC, University of Wisconsin, Madison.

Table I

No. of iterations	γ	m	Norm of C.G. Residual	E_{max} (approx.)
4	1	36	.5172056-03	-04
4	1	76	.5409905-03	-04
4	1	108	.7234433-03	-04
4	0.7	32	.3189510-02	.2-03
4	0.7	64	.8658407-03	.2-03
4	0.7	92	.1266753-02	.2-03
4	0.5	60	.5768389-02	.2-03
4	0.5	84	.3497684-02	.2-03
7	1	108	.1820406-04	-06
7	0.7	92	.1372015-04	-06
7	0.5	84	.3218270-04	-06

Scheme I.N.a is used to obtain results listed in Table I. Typically it will take one or two more iterations to achieve similar accuracies if scheme I.N.b is used. The norm of C.G. residual given in the fourth column of Table I is the L_2 norm of the conjugate gradient residuals divided by the square root of the number of mesh points inside Ω.

§9. Conclusions

Since it takes two fast Poisson solvers to complete each conjugate gradient iteration, the total operation counts of the algorithm are approximately ten or eleven times that of a fast Poisson solver for the Laplace or Poisson equation respectively. It is reported in [2] that the operation counts of a fast Poisson solver can be reduced to $O(n^2)$ if the fast Fourier transform methods are combined with k cyclic reduction methods if k is proportional to $\log_2 n$. It is, however, more realistic to say that the operation counts of our algorithm are proportionally to $n^2 \log_2 n$ in the experiments carried out so far. Our theoretical estimate of a constant times $n^2(\log_2 n)^2$ is perhaps too conservative. It is shown in [27] and [28] that theoretical estimates of constant $\cdot\ n^2(\log_2 n)$ can be obtained for a special class of domains in some cases although it is still an open question whether this is true in general.

Acknowledgements

The author would like to thank Dr. Paul Concus of the Lawrence Berkeley Laboratory for many interesting and fruitful discussions when he first began work on this project. He is also greatly indebted to Professor Olof Widlund for suggesting the topic and to Professors A. Chorin and G. Golub for their encouraging suggestions.

REFERENCES

1. P. M. Anselone, Collectively compact operator, New Jersey, 1971.
2. R. E. Bank, Marching algorithms for elliptic boundary value problems, Ph. D. Thesis, 1975, Harvard University.
3. J. H. Bramble and B. E. Hubbard, The Neumann problem for Poisson's equation, SIAM J. Numer. Anal, Ser. B, 2 (1965), 1-14.
4. O. Buneman, Analytic inversion of the five-point Poisson operator, J. of Comp. Phys. 8, (1971), 500-505.
5. B. L. Buzbee and F. W. Dorr, The direct solution of the biharmonic equation on rectangular regions and the Poisson equation on irregular regions, SIAM J. Numer. Anal., 11 (1974), 753-763.
6. B. L. Buzbee, F. W. Dorr, J. A. George and G. H. Golub, The direct solution of the discrete Poisson equation on irregular regions, SIAM J. Numer. Anal., 8 (1971), 722-736.
7. B. L. Buzbee, G. H. Golub and C. W. Nielson, On direct methods for solving Poisson's equation, SIAM J. Numer. Anal., 7 (1970), 627-656.
8. R. Courant and D. Hilbert, Methods of mathematical physics, Interscience, 1953.
9. J. W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer, Anal.,
10. D. Fischer, G. Golub, O. Hald, C. Leiva and O. Widlund, On Fourier-Toeplitz methods for separable elliptic problems, Math. Comp., 28 (1974), 349-368.
11. G. E. Forsythe and W. R. Wasow, Finite difference methods for partial differential equations, Wiley, 1960.
12. P. R. Garabedian, Partial differential equations, Wiley, 1964.
13. J. A. George, The use of direct methods for the solution of the discrete Poisson equation on nonrectangular regions, Comp. Sci. Dept. Report 159, Stanford University, 1970

14. R. M. Hayes, Iterative methods of solving linear problems on Hilbert space, Nat. Bur. of Standards, Applied Math. Series, 39 (1954), 71-103.

15. M. R. Hestenes, The conjugate gradient method for solving linear systems, Proc. Symp. Appl. Math., 6 (1956), 83-102.

16. M. R. Hestenes and E. Stiefel, Method of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, 49 (1952), 409-436.

17. R. W. Hockney, A fast direct solution of Poisson's equation using Fourier analysis: J. Assoc. Comp. Mach., 12 (1965), 95-113.

18. R. W. Hockney, Formation and stability of virtual electrodes in a cylinder, J. Appl. Phys., 39 (1968), 4166-4170.

19. R. W. Hockney, The potential calculation and some applications, Methods in computational physics, 9 (1970), Academic Press.

20. A. S. Householder, The theory of matrices in numerical analysis, Blaisdell, 1964.

21. W. J. Kammarer and M. Z. Nashed, On the convergence of the conjugate methods for singular operator equations, SIAM J. Numer. Anal., 9 (1972), 165-181.

22. O. D. Kellog, Foundations of potential theory, Dover, 1953.

23. D. G. Luenberger, Introduction to linear and nonlinear programming, Addison-Wesley, 1973.

24. W. H. McCrea and F. J. W. Whipple, Random paths in two or three dimensions, Proc. Roy. Soc. Edinburg, Sect. A, 60, 281-298.

25. I. G. Petrovsky, Partial differential equations, Interscience, 1954.

26. W. Proskurowski and O. Widlund, On the numerical solution of Helmoltz equation by the capacitance matrix method, Math. Comp., 30 (1976), 433-468.

27. A. Shieh, Fast Poisson solvers on nonrectangular regions, Ph.D. Thesis, N.Y.U., 1976.

28. A. Shieh, Fast Poisson solvers on general regions, I. The Dirichlet problem, to appear.

29. O. Widlund, On the use of fast methods for separable finite difference equations for the solution of general elliptic problems, Sparse matrices and their applications, Ed. by D. J. Rose and R. A. Willoughby, Plenum Press, 1972.

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720