SIAM J. NUMER. ANAL.
Vol. 15, No. 4, August 1978

0036-1429/78/1504-0011 $01.00/0

A LANCZOS METHOD FOR A CLASS OF NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS*

OLOF WIDLUND†

Abstract. Let L be a real linear operator with a positive definite symmetric part M. In certain applications a number of problems of the form $Mv = g$ can be solved with less human or computational effort than the original equation $Lu = f$. An iterative Lanczos method, which requires no a priori information on the spectrum of the operators, is derived for such problems. The convergence of the method is established assuming only that $M^{-1}L$ is bounded. If $M^{-1}L$ differs from the identity mapping by a compact operator the convergence is shown to be superlinear. The method is particularly well suited for large sparse systems arising from elliptic problems. Results from a series of numerical experiments are presented. They indicate that the method is numerically stable and that the number of iterations can be accurately predicted by our error estimate.

1. Introduction. It is the purpose of this paper to explore a Lanczos, conjugate gradient-like, method for the iterative solution of the equation

$$Lu = f. \tag{1.1}$$

We assume that the real linear operator L is semi-bounded, i.e.,

$$(u, Lu) \geqq c\|u\|^2, \tag{1.2}$$

for all u in a dense subset of a real Hilbert space H. Here c is a strictly positive constant and $(\cdot, \cdot)$ and $\|\cdot\|$ denote the inner product and the norm of the Hilbert space. We can derive an a priori estimate for any solution of the equation (1.1) because by the inequality (1.2) and the Schwarz inequality

$$c\|u\|^2 \leqq (u, Lu) \leqq \|Lu\| \cdot \|u\|,$$

or

$$\|u\| \leqq (1/c)\|Lu\|.$$

Borrowing some standard notations from the theory of iterative methods for elliptic finite difference equations (see Varga [28]), we introduce a splitting

$$L = M - N,$$

where we choose M as the symmetric and $-N$ as the skew-symmetric part of L. Thus

$$M = (L + L^T)/2 \quad \text{and} \quad N = (L^T - L)/2,$$

where L^T is the transpose of L induced by the Hilbert space inner product. The method considered is of practical interest for cases where a number of problems of the form

$$Mv = g$$

can be solved with less computational or human effort than (1.1). We find ourselves in such a situation if M is a finite difference approximation of a Helmholtz operator and the operator N represents a discretization of a transport term given by a partial differential operator with variable coefficients. We note that fast Helmholtz solvers now

* Received by the editors October 19, 1976, and in revised form February 22, 1977.

† Courant Institute of Mathematical Sciences, New York University, New York, New York 10012. The work presented in this paper was supported by the ERDA Mathematics and Computing Laboratory, Courant Institute of Mathematical Sciences, New York University, under Contract E(11-1)-3077 with the U.S. Energy Research and Development Administration.

are readily available for many of the regions which allow the separation of the variables; see Bank [2], Buneman [4], Buzbee, Golub and Nielson [7], Fischer, Golub, Hald, Leiva and Widlund [10], Hockney [14], [15] and Swarztrauber and Sweet [27]. The usefulness of these algorithms has been extended by the development of capacitance matrix methods which can be used for Helmholtz's equation on general regions; see Buzbee and Dorr [5], Buzbee, Dorr, George and Golub [6], George [11], Hockney [16] and Proskurowski and Widlund [23]. Our Lanczos method can also be considered for finite element Galerkin approximations for elliptic problems which are not self-adjoint. If we compare the use of a Gaussian elimination method for the resulting nonsymmetric matrix with the computation of the Cholesky factors of its symmetric part, we find that we can expect to save about half the storage and half the arithmetic operations by using the latter method. A wider repertoire of sparse matrix codes is also available for positive definite, symmetric problems, in which case pivoting for numerical stability ceases to be an issue. We note that the number of arithmetic operations in the forward elimination and backward substitution parts of these algorithms is typically much smaller than the effort required to generate the Gauss or Cholesky factors of the matrix.

In the development of the method, a central role is played by the operator

$$K = M^{-1}N$$

which we always assume to be bounded in the Hilbert space H. In § 3 we prove the convergence of our algorithm for an arbitrary bounded K and also its super-linear convergence for the case where K is a compact operator. The compactness of K can often be established for elliptic problems where N is a partial differential operator of lower order than M. The characteristic spectral distribution of a compact operator K, with a sole point of accumulation at the origin, is central in our proof of super-linear convergence. The spectrum of the corresponding finite dimensional operator, resulting from the discretization of M and N, frequently inherits the main features of its continuous counterpart. This leads to a convergence more rapid than that predicted by a standard comparison with a method based on Chebyshev polynomials. We note that results on super-linear convergence of the conjugate gradient method when applied to self-adjoint Fredholm integral equations of the second kind were obtained by Hayes [12] and further discussed in Proskurowski and Widlund [23].

An additional convergence result is also given. We show that the two sequences of error vectors corresponding to the approximate solutions, after an odd and even number of iterations respectively, have monotonically decreasing norms.

Results from some numerical experiments are reported in § 4. Additional experiments are planned for the near future for certain problems of continuum mechanics for which the operator N is nonlinear.

An equation of the form (1.1) with a complex linear operator L can also be treated by our method if it satisfies

$$\operatorname{Re}(u, Lu) \geqq c\|u\|^2,$$

c strictly positive, for all u in a dense subset of a complex Hilbert space. For example, if L is a partial differential operator with complex coefficients, a system of equations with real coefficients can be obtained by simply separating the real and imaginary parts of equation (1.1). It is easy to show that the resulting real operator satisfies condition (1.2) with respect to a very natural inner product of a real Hilbert space. The same argument is equally applicable for a linear, algebraic system of equations with a complex matrix.

The use of Lanczos and conjugate gradient methods for systems of linear equations with nonsymmetric coefficient matrices was discussed already in the early papers of

Lanczos [17], [18] and Hestenes and Stiefel [13]. For such problems they suggested that either two sets of so-called Lanczos vectors or a least squares formulation should be used. The special structure of our problems allows us to work with only one set of Lanczos vectors. In this respect our method parallels the interesting algorithm SYMMLQ developed by Paige and Saunders [22] for indefinite symmetric systems. The use of splitting and the conjugate gradient method for positive definite, symmetric problems has been considered by Axelsson [1], Bartels and Daniel [3], Concus, Golub, and O'Leary [9], Mejerink and van der Vorst [19], O'Leary [20], Wachspress [29] and others. The present paper is closely related to recent work by Concus and Golub [8].

2. Derivation of the algorithm. We will construct approximate solutions $u^{(k)}$, $k=1,2,\cdots$, of (1.1) by using the Krylov sequence

$$v^{(0)}, Kv^{(0)}, \cdots, K^{(k-1)}v^{(0)}, \cdots.$$

Here $v^{(0)}=M^{-1}r^{(0)}$ where $r^{(0)}\equiv f-Lu^{(0)}$ is the initial residual. Without loss of generality, we assume that the initial guess $u^{(0)}$ is zero and thus $r^{(0)}=f$. Following Lanczos [17], [18], we introduce vectors $v^{(i)}$, the Lanczos vectors, which serve as an orthogonal basis for the subspace spanned by the Krylov sequence. They satisfy a three term recurrence relationship

$$\alpha_j v^{(j+1)}=Kv^{(j)}-\gamma_j v^{(j)}-\beta_j v^{(j-1)}, \qquad j\geqq 0, \tag{2.1}$$

with

$$v^{(-1)}=0 \quad\text{and}\quad v^{(0)}=M^{-1}f.$$

The coefficients β_j and γ_j are chosen so that the $v^{(j)}$ are mutually orthogonal with respect to the real M-inner product defined by

$$(u,v)_M=(u,Mv).$$

We note that for an elliptic equation the quadratic form corresponding to this inner product is the Dirichlet form, or strain energy, of the problem. We note that the operator K is skew-symmetric with respect to the M-inner product because

$$(u,Kv)_M=(u,Nv)=-(Nu,v)=-(M^{-1}Nu,Mv)=-(Ku,v)_M. \tag{2.2}$$

We determine the coefficients in (2.1) as in the classical theory of orthogonal polynomials. The M-orthogonality of $v^{(j)}$, $v^{(j-1)}$ and $v^{(j+1)}$ and (2.2) makes $\gamma_j=0$. Similarly, we find that

$$\alpha_j\rho_{j+1}=(v^{(j+1)},Kv^{(j)})_M=(v^{(j+1)},Nv^{(j)}) \tag{2.3}$$

and

$$0=(v^{(j-1)},Kv^{(j)})_M-\beta_j\rho_{j-1}=(v^{(j-1)},Nv^{(j)})-\beta_j\rho_{j-1} \tag{2.4}$$

where

$$\rho_\ell=(v^{(\ell)},v^{(\ell)})_M=(v^{(\ell)},Mv^{(\ell)}).$$

By using the formulas (2.2), (2.3) and (2.4), we find that

$$\beta_{j+1}=-\alpha_j\rho_{j+1}/\rho_j.$$

We are free to choose the coefficients α_j arbitrarily and make

$$\begin{aligned}&\alpha_j+\beta_j=1, \qquad j>0,\\ &\alpha_0=1.\end{aligned} \tag{2.5}$$

We will see below that this choice is quite convenient. The coefficients α_j then satisfy

$$\alpha_{j+1} = 1 + \alpha_j(\rho_{j+1}/\rho_j), \qquad \alpha_0 = 1. \tag{2.6}$$

The M-orthogonality of $v^{(j+1)}$ and $v^{(0)}, \cdots, v^{(j-2)}$ is shown exactly as the corresponding result in the theory of orthogonal polynomials. We will see below that $v^{(j+1)} = M^{-1}r^{(j+1)}$, where $r^{(\ell)}$ is the ℓth residual. The vector $v^{(j+1)}$ will therefore vanish only if $u^{(j+1)}$ is the exact solution of (1.1).

Denote by $J^{(k)}$ the tridiagonal matrix

$$J^{(k)} = \begin{bmatrix} 0 & \beta_1 & 0 & \cdots & 0 & 0 \\ \alpha_0 & 0 & \beta_2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & \alpha_{k-2} & 0 \end{bmatrix}.$$

The approximate solution $u^{(k)}$, $k \geqq 1$, is given the form

$$u^{(k)} = V^{(k)}y^{(k)},$$

where $y^{(k)}$ is a vector with k components and the columns of the matrix $V^{(k)}$ are $v^{(0)}, \cdots, v^{(k-1)}$. The vector $y^{(k)}$ is determined by the Galerkin condition that the residual

$$r^{(k)} \equiv f - Lu^{(k)} = f - (M - N)V^{(k)}y^{(k)}$$

be orthogonal to $v^{(0)}, \cdots, v^{(k-1)}$. The resulting linear system of equations is by formulas (2.3) and (2.4),

$$R^{(k)}(I - J^{(k)})y^{(k)} = \rho_0 e^{(0)},$$

where

$$e^{(0)} = (1, 0, \cdots, 0)^T.$$

The right-hand side has this form because

$$(v^{(\ell)}, f) = (v^{(\ell)}, Mv^{(0)}) = \rho_0\delta_{0\ell}.$$

The matrix $R^{(k)}$ is diagonal with the elements $\rho_0, \cdots, \rho_{k-1}$.

We now solve

$$(I - J^{(k)})y^{(k)} = e^{(0)}$$

by Gaussian elimination. By using the normalization (2.5) we find that the LU factorization of $I - J^{(k)}$ is very simple with

$$L^{(k)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & & & \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

and

$$U^{(k)} = \begin{bmatrix} \alpha_0 & -\beta_1 & 0 & \cdots & 0 & 0 \\ 0 & \alpha_1 & -\beta_2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & 0 & \alpha_{k-1} \end{bmatrix}.$$

By forward elimination, we then obtain

$$U^{(k)}y^{(k)}=\begin{bmatrix}1\\1\\\cdot\\\cdot\\\cdot\\1\end{bmatrix}.$$

It is now straightforward to verify that

$$\Delta u^{(\ell)}=(\omega_{\ell+1}-1)\,\Delta u^{(\ell-1)}+\omega_{\ell+1}v^{(\ell)},\qquad \Delta u^{(-1)}=0, \tag{2.7}$$

where $\omega_{\ell+1}=1/\alpha_\ell$, $\omega_1=1$ and $\Delta u^{(\ell)}=u^{(\ell+1)}-u^{(\ell)}$. We note that from formula (2.6) it follows that $\omega_\ell\leqq 1$. The formula (2.7) can be rewritten as

$$\begin{aligned}&u^{(\ell+1)}=u^{(\ell-1)}+\omega_{\ell+1}(v^{(\ell)}+u^{(\ell)}-u^{(\ell-1)}),\\&u^{(-1)}=u^{(0)}=0.\end{aligned} \tag{2.8}$$

From this formula it is easy to derive a corresponding relation for the residuals,

$$\begin{aligned}&r^{(\ell+1)}=r^{(\ell-1)}-\omega_{\ell+1}((M-N)v^{(\ell)}-r^{(\ell)}+r^{(\ell-1)}),\\&r^{(-1)}=0,\qquad r^{(0)}=f.\end{aligned} \tag{2.9}$$

If we compare formulas (2.9) and (2.1), we find that

$$v^{(\ell)}=M^{-1}r^{(\ell)}. \tag{2.10}$$

Formula (2.9) can therefore be rewritten as

$$\begin{aligned}r^{(\ell+1)}&=(1-\omega_{\ell+1})r^{(\ell-1)}+\omega_{\ell+1}Nv^{(\ell)}\\&=(1-\omega_{\ell+1})r^{(\ell-1)}+\omega_{\ell+1}NM^{-1}r^{(\ell)}.\end{aligned} \tag{2.11}$$

The following implementation of the method is due to Gene Golub.

ALGORITHM. Let $u^{(-1)}=u^{(0)}=0$ and let $\omega_1=1$. For $\ell=0,1,\cdots$:
Solve

$$Mv^{(\ell)}=r^{(\ell)}=f-Lu^{(\ell)}.$$

Compute

$$\rho_\ell=(v^{(\ell)},v^{(\ell)})_M$$

and for

$$\ell\geqq 1,\quad \omega_{\ell+1}=(1+(\rho_\ell/\rho_{\ell-1})/\omega_\ell)^{-1}.$$

Compute

$$u^{(\ell+1)}=u^{(\ell-1)}+\omega_{\ell+1}(v^{(\ell)}+u^{(\ell)}-u^{(\ell-1)}).$$

There are many variants. The parameter ρ_ℓ can thus be computed by the formula

$$\rho_\ell=(v^{(\ell)},r^{(\ell)}) \tag{2.12}$$

either in the same subroutine which produces $v^{(\ell)}=M^{-1}r^{(\ell)}$ or at the expense of temporarily storing $r^{(\ell)}$. An operator-vector multiplication is saved by using (2.12). We can also choose to store $u^{(\ell)}$ and $\Delta u^{(\ell-1)}$ instead of $u^{(\ell)}$ and $u^{(\ell-1)}$ and compute $\Delta u^{(\ell)}$ by using (2.7) rather than (2.8). As we have already noted the residual $r^{(\ell)}$ can be computed recursively by (2.11). We note that by doing so, we can design an algorithm which requires only two subroutines which produce Nx and $M^{-1}x$ respectively for a given vector x. If the initial guess $u^{(0)}$ differs from zero, we must replace f by the initial residual in the algorithm and the computation of this residual will of course require the multiplication of $u^{(0)}$ with the operator M. By using formulas (2.4) and (2.5) we can compute $\omega_{\ell+1} = 1/\alpha_\ell$ without using the recursive formula given above. At this time, we have no specific recommendations on the choice between these different variants. We have tried several in our numerical experiments and have noticed only marginal differences. We refer to Reid [24] for a careful discussion of related problems for the standard conjugate algorithm and to Paige [21] for an interesting error analysis of the Lanczos transform of symmetric matrices.

We note that the Krylov sequence is completely specified by its first element $v^{(0)}$. We have worked exclusively with $v^{(0)} = M^{-1}r^{(0)}$. This is a very natural choice especially when the norm of the operator K is small. In the symmetric case different methods of selecting this first element have been considered (see Paige and Saunders [22]) and an alternative choice might be worth our consideration even in the present context.

3. Results on convergence. We first establish that the method converges for any bounded operator K by using a standard argument for Galerkin methods, see Strang and Fix [26]. Let $S^{(k)}$ be the subspace spanned by $v^{(0)}, v^{(1)}, \cdots, v^{(k-1)}$ and denote by $e^{(k)} = u^{(k)} - u$ the error in the kth step. By the Galerkin condition

$$(z, Le^{(k)}) = 0, \tag{3.1}$$

for all $z \in S^{(k)}$. By the skewness of N, the equation (3.1) and the Schwarz inequality

$$\begin{aligned} \|e^{(k)}\|_M^2 &= (u^{(k)} - u, M(u^{(k)} - u)) \\ &= (u^{(k)} - u, L(u^{(k)} - u)) = (z - u, L(u^{(k)} - u)) \\ &\leqq \|z - u\|_M \|(I - K)\, e^{(k)}\|_M, \end{aligned} \tag{3.2}$$

for all $z \in S^{(k)}$. We now extend our real space H to a complex Hilbert space in the obvious way. The spectrum of K, $\sigma(K)$, is a subset of $\{\mu;\ |\mathrm{Im}\ \mu| \leqq \Lambda, \mathrm{Re}\ \mu = 0\}$ for some sufficiently large Λ. By applying a spectral decomposition of the bounded skew-Hermitian operator K and the spectral mapping theorem, see Riesz and Sz-Nagy [25, Chap. VII], we obtain from (3.2)

$$\|e^{(k)}\|_M \leqq \sqrt{1+\Lambda^2} \min_{z \in S^{(k)}} \|z - u\|_M. \tag{3.3}$$

Any element $z \in S^{(k)}$ can be represented as a polynomial in K, or $I - K$, with real coefficients and of degree $k-1$ operating on $v^{(0)} = M^{-1}f = (I-K)u$. It therefore has the form

$$z = P_k(I-K)u,$$

where the polynomial $P_k(\lambda)$ is of degree k and vanishes at the origin. We therefore get

$$\min_{z \in S^{(k)}} \|z - u\|_M = \min_{Q_k(0)=1} \|Q_k(I-K)u\|_M.$$

By the spectral mapping theorem, we find that

$$\min \|z - u\|_M \leqq \min_{Q_k(0)=1} \max_{\mu \in \sigma(K)} |Q_k(1-\mu)|\, \|u\|_M, \tag{3.4}$$

where Q_k is a polynomial of degree k with real coefficients. An upper bound is obtained by choosing

$$Q_k(1-\mu)=\frac{T_k(\mu/(i\Lambda))}{T_k(1/(i\Lambda))}$$

where $T_k(\lambda)$ is the Chebyshev polynomial defined by

$$T_k(\lambda)=\cosh\,(k\,\cosh^{-1}(\lambda)), \tag{3.5}$$

where the branch of $\cosh^{-1}(\lambda)$ with nonnegative real part is chosen. Since $S^{(k)}\subset S^{(k+1)}$, an estimate obtained by the inequality (3.3) for a certain value of k is also valid for $k+1$. We can therefore limit our considerations to even values of k. The Chebyshev polynomials are even functions for even k, and the polynomial Q_k chosen therefore has real coefficients. For $\mu\in\sigma(K)$, $-1\leqq\mu/(i\Lambda)\leqq 1$ and therefore $|T_k(\mu/(i\Lambda))|\leqq 1$.

By using the relations (3.4) and (3.5), we obtain, for k even,

$$\min_{z\in S^{(k)}}\|z-u\|_M\leqq(\cosh\{k\log(\sqrt{1+\Lambda^{-2}}+\Lambda^{-1})\})^{-1}\|u\|_M,$$

and by combining this formula with the inequality (3.3), we deduce

$$\begin{aligned}\|e^{(k)}\|_M&\leqq\sqrt{1+\Lambda^2}(\cosh\{k\log(\sqrt{1+\Lambda^{-2}}+\Lambda^{-1})\})^{-1}\|u\|_M\\&\leqq 2\sqrt{1+\Lambda^2}(\sqrt{1+\Lambda^{-2}}+\Lambda^{-1})^{-k}\|u\|_M.\end{aligned}\tag{3.6}$$

We note that, by assumption, $u^{(0)}=0$ and therefore $\|u\|_M=\|e^{(0)}\|_M$.

If K is compact the spectrum is discrete and the eigenvalues μ_i cluster at the origin. Since K is skew-Hermitian and real, the eigenvalues are purely imaginary and appear in complex conjugate pairs. Let us order them such that

$$\Lambda=|\mu_1|=|\mu_2|\geqq|\mu_3|=|\mu_4|\geqq\cdots\geqq|\mu_{2\ell+1}|=|\mu_{2\ell+2}|\geqq\cdots.$$

We now consider a family of polynomials

$$Q_{k,\ell}(1-\mu)=\left(\prod_{j=1}^{\ell}\frac{\mu^2-\mu_{2j-1}^2}{1-\mu_{2j-1}^2}\right)\left(\frac{T_{k-2\ell}(\mu/(i|\mu_{2\ell+1}|))}{T_{k-2\ell}(1/(i|\mu_{2\ell+1}|))}\right),$$

where $0\leqq 2\ell\leqq k$, and k even. We note that the first factor is less than one in absolute value for all $\mu\in\sigma(K)$. By repeating the arguments which lead to the inequality (3.6), we therefore obtain, for k even,

$$\begin{aligned}\|e^{(k)}\|_M&\leqq\sqrt{1+|\mu_1|^2}\min_{0\leqq\ell\leqq k/2}(\cosh\{(k-2\ell)\log(\sqrt{1+|\mu_{2\ell+1}|^{-2}}+|\mu_{2\ell+1}|^{-1})\})^{-1}\|u\|_M\\&\leqq 2\sqrt{1+|\mu_1|^2}\min_{0\leqq\ell\leqq k/2}(\sqrt{1+|\mu_{2\ell+1}|^{-2}}+|\mu_{2\ell+1}|^{-1})^{-(k-2\ell)}\|u\|_M.\end{aligned}\tag{3.7}$$

For any $\varepsilon>0$, we now choose a value of $\ell=\ell^*$ such that $|\mu_{2\ell^*+1}|\leqq\varepsilon$, and therefore by the inequality (3.7)

$$\|e^{(k)}\|_M\leqq 2\sqrt{1+|\mu_1|^2}(\varepsilon/2)^{(k-2\ell^*)}\|u\|_M,$$

which establishes the superlinear convergence of the method when K is a compact operator. We note that a more refined estimate can be obtained from the inequality

(3.7) if additional information about the asymptotic distribution of the eigenvalues is available.

Formulas (3.6) and (3.7) do not exclude an increase of the norm of the error in an early stage of the iteration. It is in fact easy to show that these estimates cannot be substantially improved for $k=1$. We can however establish the following descent property:

$$\|e^{(k+1)}\|_M^2 = \|e^{(k-1)}\|_M^2 - \|e^{(k+1)} - e^{(k-1)}\|_M^2, \qquad k \geqq 1, \tag{3.8}$$

which shows, among other things, that the two series $\|e^{(0)}\|_M$, $\|e^{(2)}\|_M, \cdots$ and $\|e^{(1)}\|_M$, $\|e^{(3)}\|_M, \cdots$ are monotonically decreasing. To establish this relation, we first note that it is equivalent to the M-orthogonality of the vectors $e^{(k+1)}$ and $e^{(k+1)}-e^{(k-1)}$. The Lanczos vector $v^{(k+1)}=(I-K)\,e^{(k+1)}$ is M-orthogonal to $e^{(k+1)}-e^{(k-1)}=u^{(k+1)}-u^{(k-1)} \in S^{(k+1)}$. Our proof can therefore be concluded by showing that $K\,e^{(k+1)}$ is M-orthogonal to $e^{(k+1)}-e^{(k-1)}$. It is easy to show that

$$e^{(k+1)} = \omega_{k+1} K\, e^{(k)} + (1-\omega_{k+1})\, e^{(k-1)},$$

with $e^{(1)} = K\,e^{(0)}$. Hence $e^{(k)} = \tilde{P}_k(K)\,e^{(0)}$ where the polynomial $\tilde{P}_k$ is an odd function for k odd and an even function for k even. By using (2.2), we find that

$$(K\,e^{(k+1)}, e^{(k+1)})_M = 0$$

and

$$(K\,e^{(k+1)}, e^{(k-1)})_M = (e^{(0)}, -K\tilde{P}_{k+1}(-K)\tilde{P}_{k-1}(K)\,e^{(0)})_M,$$

which also vanishes since the polynomial $\lambda\tilde{P}_{k+1}(-\lambda)\tilde{P}_{k-1}(\lambda)$ is an odd function.

It is well known that when the conjugate gradient method is used for a linear system of equations with a positive definite, symmetric matrix and n unknowns, then in the absence of roundoff, the exact solution is obtained in at most n steps; see Hestenes and Stiefel [13]. It is also easy to show that the calculation terminates in at most p steps if the initial error can be represented exactly as a linear combination of p eigenvectors of the operator. The same results are true for our algorithm. We need only note that K maps the p-dimensional subspace spanned by these p eigenvectors into itself. The $(p+1)$st Lanczos vector must be linearly dependent on, and orthogonal to, the previous Lanczos vectors and must therefore vanish.

4. Numerical results. A FORTRAN program was prepared for and run on the CDC 6600 at the ERDA Mathematics and Computing Laboratory of the Courant Institute. Single precision, between fourteen and fifteen decimal digits, was used throughout. Problems with known solutions and between 10 and 3969 variables were run and the M-norm of the error was computed in each step. The overall impression is that the method is quite robust and performs in accordance with the theory given in § 3. With our program the solution of a new problem requires only the preparation of subroutines which calculate M^{-1}, N and M times a given vector.

The iteration was stopped as soon as the ratio of the M-norms of the current and initial Lanczos vectors, $\sqrt{\rho_n/\rho_0}$, fell below a prescribed tolerance. We have found this stopping criterion quite satisfactory even for problems where the operator K has a large norm. The M-norm of the error for such a problem will, in general, fail to converge monotonically, and often changes by several powers of ten from one step to the next. In our approximately fifty experiments for which the assigned tolerance was met, we always found that the M-norm of the error was smallest at the last step. The computed

TABLE 1

Results from experiments solving equation (4.2). The number of mesh points is N and the number of iterations is I. The iteration was stopped when $\rho_I/\rho_0 \leqq 10^{-15}$ or when $I = 200$.

$a(x, y)$	N	I	Solution	ρ_I/ρ_0	$\log_{10}(\|e^{(I)}\|_M/\|e^{(0)}\|_M)$
1	49	7	random	.395E−16	−8.20
1	225	7	random	.484E−16	−8.16
1	961	6	random	.447E−15	−7.68
1	961	7	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.395E−15	−7.70
1	3969	7	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.416E−15	−7.69
10	49	16	random	.353E−16	−8.23
10	225	17	random	.117E−15	−7.99
10	961	16	random	.380E−15	−7.76
10	961	17	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.930E−15	−7.46
10	3969	17	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.932E−15	−7.49
100	49	50	random	.448E−17	−8.53
100	225	83	random	.487E−15	−7.63
100	961	90	random	.514E−15	−7.82
100	961	82	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.797E−15	−7.07
100	3969	82	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.881E−15	−6.97
1000	49	54	random	.283E−15	−8.00
1000	225	200	random	.122E−8	−4.41
1000	961	200	random	.988E−4	−1.85
1000	961	200	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.533E−7	−2.66
1000	3969	200	$\sin \pi x \sin \pi y$ $\cdot \exp((x/2+y)^3)$	.414E−8	−2.32
$2 \exp(3.5(x^2+y^2))$	49	31	random	.169E−15	−7.59
$2 \exp(3.5(x^2+y^2))$	225	60	random	.308E−15	−7.59
$20 \exp(3.5(x^2+y^2))$	49	76	random	.784E−17	−7.95
$20 \exp(3.5(x^2+y^2))$	225	200	random	.147E−8	−3.61
$200 \exp(3.5(x^2+y^2))$	49	86	random	.653E−16	−8.13

values of the series, $\|e^{(0)}\|_M, \|e^{(2)}\|_M, \cdots$ and $\|e^{(1)}\|_M, \|e^{(3)}\|_M \cdots$, were each monotonically decreasing, except for an occasional individual element of the series, experimentally confirming the descent property established in § 3. At times, the odd series showed a much higher asymptotic rate of convergence than the even one.

In all the experiments reported in our table and diagram a version of the algorithm was used which updated the residuals by (2.11) and the solution by (2.7). Certain experiments were also run using (2.1) and (2.8). The values of ω_ℓ and $\|e^{(\ell)}\|_M$ changed only in the eighth through twelfth significant decimal digits except for some large changes in the two final steps of the algorithm when it was applied to some difficult and rather atypical problems where the number of iterations equaled or exceeded the number of variables.

Difference approximations of the ordinary differential equation

$$(4.1)\qquad \begin{aligned} &-\partial_x^2 u + [\partial_x(a(x)u) + a(x)\partial_x u]/2 = f(x), \qquad 0 \leqq x \leqq 1, \\ &\partial_x u(0) \text{ and } u(1) \text{ given,} \end{aligned}$$

were used in our first series of experiments. By replacing the derivatives by centered differences, we approximate the second derivative by a symmetric, positive definite tridiagonal matrix and, for any choice of a real function $a(x)$, the first order term by a skew-symmetric matrix. The boundary conditions lead to a symmetric matrix which has a very simple LU factorization. By Rellich's theorem (see Yosida [30, Chap. X, 3]) the operator K which corresponds to the differential equation is compact. It is also easy to see that its eigenvalues $\mu_{2\ell-1}$, $\mu_{2\ell}$ behave asymptotically like $\pm i$ const.$/\ell+o(1/\ell)$. The eigenvalue distribution for the discrete problem is similar.

Let us consider the special case of $a(x)=A=$ const. It is easy to see that if formula (3.6) provided a sharp bound, the number of iterations to obtain a prescribed decrease of the M-norm of the error would grow linearly with A for large values of A. A series of experiments was carried out to assess the importance of the clustering of the eigenvalues. We found that the number of iterations required for problem (4.1) is approximately const. $A^{0.60}$.

The main series of experiments was carried out with a difference approximation of the partial differential equation

$$(4.2)\qquad -\partial_x^2 u-\partial_y^2 u+[\partial_x(a(x,y)u)+a(x,y)\partial_x u]/2=f(x,y),\qquad 0\leqq x\leqq 1,\quad 0\leqq y\leqq 1,$$

with Dirichlet data given on the boundary of the square. Centered differences were used and the symmetric, linear systems of equations corresponding to the Laplace operator were solved by Buneman's method; see Buzbee, Golub and Nielson [7]. In the special case when $a(x,y)=A=$ const., the eigenvalues of the operator K corresponding to the continuous problem (4.2) are $\pm iA/(2\pi\sqrt{k^2+\ell^2})$, $k, \ell=1,2,\cdots$. Thus the cluster of the eigenvalues at the origin is less pronounced than for problem (4.1). A series of

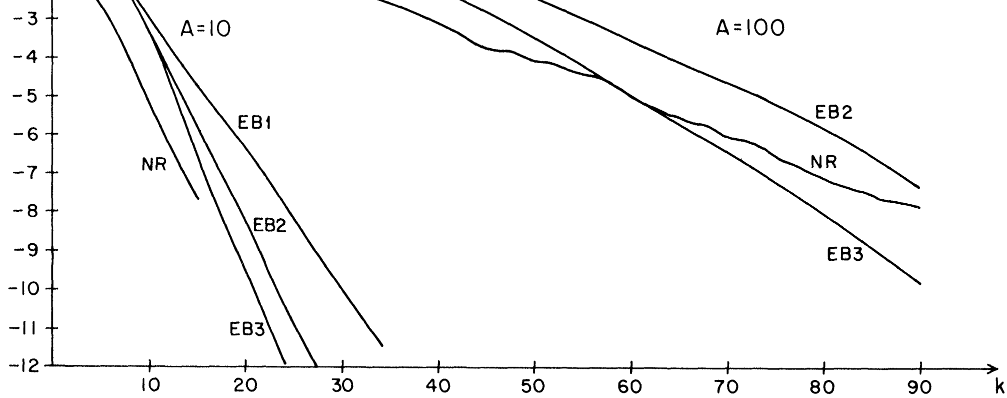

FIG. 1. NR = *results from numerical experiments solving* (4.2) *with* $a(x,y)=A=$ const. EB1 = *error bounds obtained by using formula* (3.6). EB2 = *error bounds obtained from* (3.7), *with the fact that this problem has double eigenvalues disregarded.* EB3 = *error bounds obtained from* (3.7) *by taking this multiplicity of the eigenvalues into account.*

experiments showed that the number of iterations required for a prescribed decrease of the M-norm of error is approximately const. $A^{0.69}$ for problem (4.2). The compactness of the operator K for a general choice of $a(x, y)$ can again by established by using Rellich's theorem.

Table 1 demonstrates that the rate of convergence of the algorithm is very insensitive to a change in the number of variables and that the regularity of the solution also plays a very minor role. In Fig. 1, we compare results from two experiments with the error bounds (3.6) and (3.7).

Acknowledgment. Early work on this paper was carried out while the author visited the Serra House of the Computer Science Department at Stanford University in August of 1975. Thanks are due foremost to Gene Golub for a number of illuminating discussions. A version of the algorithm (see § 2) was first derived by him (see Concus and Golub [8]). Conversations with Roland Glowinski, Dianne P. O'Leary, Thomas Manteuffel and Henk van der Vorst proved quite helpful.

Thanks are also due to Alexandra Banegas for her very valuable assistance in carrying out the numerical experiments. Three colleagues at the Courant Institute also provided assistance. Anthony Jameson provided a fine fast Poisson solver and Eugene Isaacson and Charles Peskin read an early version of this paper and made a number of useful suggestions.

REFERENCES

[1] O. Axelsson, *A class of iterative methods for finite element equations*, Comput. Methods Appl. Mech. Engrg., 9 (1976), pp. 123–137.

[2] R. E. Bank, *Marching algorithms and Gaussian elimination*, Proc. Symp. on Sparse Matrix Computations (Argonne National Lab., Sept. 1975), J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1975.

[3] R. Bartels and J. W. Daniel, *A conjugate gradient approach to nonlinear elliptic boundary value problems in irregular regions*, Conference on the Numerical Solution of Differential Equations (Dundee, Scotland, July 1973), Lecture Notes in Mathematics, vol. 363, Springer-Verlag, Berlin, pp. 1–11, 1973.

[4] O. Buneman, *A compact non-iterative Poisson solver*, Rep. SUIPR-294, Inst. for Plasma Research, Stanford Univ., 1969.

[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*, this Journal, 11 (1974), pp. 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*, this Journal, 8 (1971), pp. 722–736.

[7] B. L. Buzbee, G. H. Golub and C. W. Nielson, *On direct methods for solving Poisson's equation*, this Journal, 7 (1970), pp. 627–656.

[8] P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, Proc. Second Internat. Symp. on Computing Methods in Applied Sciences and Engineering, IRIA (Paris, Dec. 1975), Lecture Notes in Economics and Mathematical Systems, vol. 134, R. Glowinski and J. L. Lions, eds., Springer-Verlag, Berlin, 1976.

[9] P. Concus, G. H. Golub and D. P. O'Leary, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, Proc. Symp. on Sparse Matrix Computations (Argonne National Lab., Sept. 1975), J. R. Bunch and D. J. Rose, eds., Academic Press, New York, 1975.

[10] D. Fischer, G. H. Golub, O. Hald, C. Leiva and O. Widlund, *On Fourier–Toeplitz methods for separable elliptic problems*, Math. Comput., 28 (1974), pp. 349–368.

[11] J. A. George, *The use of direct methods for the solution of the discrete Poisson equation on non-rectangular regions*, Computer Science Department Rep. 159, Stanford Univ., 1970.

[12] R. M. Hayes, *Iterative methods of solving linear problems on Hilbert space*, Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues, O. Taussky, ed., Nat. Bur. of Standards Applied Math. Series, vol. 39, U.S. Govt. Printing Office, Washington, DC, 1954, pp. 71–103.

[13] M. R. HESTENES AND E. STIEFEL, *Method of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.

[14] R. W. HOCKNEY, *A fast direct solution of Poisson's equation using Fourier analysis*, J. Assoc. Comput. Math., 12 (1965), pp. 95–113.

[15] ———, *The Potential Calculation and Some Applications*, Methods in Computational Physics, vol. 9, Academic Press, New York, 1970.

[16] ———, *POT 4—A fast direct Poisson solver for the rectangle allowing some mixed boundary conditions and internal electrodes*, IBM Research, R.C. 2870, 1970.

[17] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.

[18] ———, *Solution of systems of linear equations by minimized iterations*, Ibid., 49 (1952), pp. 33–53.

[19] J. A. MEJERINK AND H. A. VAN DER VORST, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, report, Academisch Computer Centrum, Utrecht, Holland, 1975.

[20] D. P. O'LEARY, *Hybrid conjugate gradient algorithms for elliptic systems*, Computer Science Dept. Rep. 548, Stanford Univ., Stanford, CA, 1976.

[21] C. C. PAIGE, *Error analysis of the Lanczos algorithm for tridiagonalizing a symmetric matrix*, J. Inst. Math. Appl., 18 (1976), pp. 341–349.

[22] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, this Journal, 12 (1975), pp. 617–629.

[23] W. PROSKUROWSKI AND O. WIDLUND, *On the numerical solution of Helmholtz's equation by the capacitance matrix method*, Math. Comput., 30 (1976), pp. 433–468.

[24] J. K. REID, *On the method of conjugate gradients for the solution of large sparse systems of linear equations*, Large Sparse Sets of Linear Equations, J. K. Reid, ed., Academic Press, New York, 1971.

[25] F. RIESZ AND B. SZ-NAGY, *Functional Analysis*, Frederick Ungar, New York, 1955.

[26] W. G. STRANG AND G. J. FIX, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[27] P. SWARZTRAUBER AND R. SWEET, *Efficient FORTRAN subprograms for the solution of elliptic partial differential equations*, Rep. NCAR-TN/1A-109, National Center for Atmospheric Research, Boulder, CO, 1975.

[28] R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[29] E. L. WACHSPRESS, *Iterative Solution of Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics*, Prentice-Hall, Englewood Cliffs, NJ, 1966.

[30] K. YOSIDA, *Functional Analysis*, Springer-Verlag, Berlin, 1965.