A GENERALIZED CONJUGATE GRADIENT METHOD FOR NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS

by

Paul Concus
Gene H. Golub

STAN-CS-76-535
JANUARY 1976

Prepared for
The Energy Research and Development Administration
under Contract No. E(04-3) 326 PA No. 30

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

A GENERALIZED CONJUGATE GRADIENT METHOD FOR NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS

by

Paul Concus
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

and

Gene H. Golub
Computer Science Department
Stanford University
Stanford, CA 94305

Also issued as Lawrence Berkeley Laboratory Report 4636, University of California, Berkeley.

A GENERALIZED CONJUGATE GRADIENT METHOD FOR NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS

by

Paul Concus
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

and

Gene H. Golub
Computer Science Department
Stanford University
Stanford, CA 94305

We consider a generalized conjugate gradient method for solving systems of linear equations having nonsymmetric coefficient matrices with positive-definite symmetric part. The method is based on splitting the matrix into its symmetric and skew-symmetric parts, and then accelerating the associated iteration using conjugate gradients, which simplifies in this case, as only one of the two usual parameters is required. The method is most effective for cases in which the symmetric part of the matrix corresponds to an easily solvable system of equations. Convergence properties are discussed, as well as an application to the numerical solution of elliptic partial differential equations.

A GENERALIZED CONJUGATE GRADIENT METHOD FOR NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS

by

Paul Concus
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

and

Gene H. Golub
Computer Science Department
Stanford University
Stanford, CA 94305

0. Introduction

In a related paper [3] we discuss a generalized conjugate gradient (CG) iterative method for solving a system of real, linear, algebraic equations

$$Ax = b \ , \tag{1.1}$$

where A is symmetric and positive definite. The method is based on splitting off from A an approximating symmetric, positive-definite matrix M that corresponds to a system of equations more easily solvable than is (1.1), and then accelerating the associated iteration using CG. The method appears to be especially effective for sparse matrices A arising from the discretization of boundary-value problems for elliptic partial differential equations. For these cases, naturally arising selections for M often result in iteration matrices possessing eigenvalue distributions for which CG acceleration is effective.

The CG method has a number of attractive properties when used as an iterative procedure:

(i) It does not require an estimation of parameters.

(ii) It takes advantage of the distribution of the eigenvalues of the iteration operator.

(iii) It requires fewer restrictions on the matrix A for optimal behavior than do such methods and successive overrelaxation.

In this paper we remove the restriction that A be symmetric, and require only that its symmetric part $(A + A^T)/2$ be positive definite. We derive the generalized CG method for this case, taking for the approximating matrix M the symmetric part of A. We find that the method then simplifies, in that the computation of only one of the two CG parameters is required.

1. Derivation of the Method

We consider the system of linear equations

$$Ax = b \ , \tag{1.1}$$

where A is a given $n \times n$ real matrix and b is a given real n-vector. We rewrite (1.1) as the system

$$Mx = Nx + b \ , \tag{1.2}$$

where $M = M^T = (A + A^T)/2$ is the symmetric part of A, and $N = -N^T = -(A - A^T)/2$ is the negative of its skew-symmetric part. We assume that M is positive definite; In [3], we discuss the solution of equations of the form (1.2) by a generalized CG method, for the case in which M is symmetric and positive definite and N is symmetric. In this paper, we derive the corresponding algorithm for the case in which N is skew-symmetric.

Our interest is in those situations for which it is a simpler computational task to solve

$$Mz = d \tag{1.3}$$

than it is to solve (1.1), and for which, in a sense to be described later, $M^{-1}N$ is not too large.

Consider an iteration of the form

$$x^{(k+1)} = x^{(k-1)} + \omega_{k+1}(\alpha_k z^{(k)} + x^{(k)} - x^{(k-1)}) \ , \tag{1.4}$$

where

$$Mz^{(k)} = r^{(k)} \ , \tag{1.5}$$

with

$$r^{(k)} = b - (M-N)x^{(k)} = b - Ax^{(k)} \ ,$$

the residual at the kth step. The quantities α_k and ω_{k+1} are scalar parameters.

Many iterative methods can be described by (1.4), e.g., if N were symmetric, the Chebyshev semi-iterative method and Richardson second order method would be of this form (cf. [5]). The generalized conjugate gradient method described below, which is also of this form, has the advantage over those two methods that no apriori information about the spectral radius of $M^{-1}N$ is needed for estimating parameters. Furthermore, it takes advantage of the actual distribution of the eigenvalues of $M^{-1}N$.

From (1.4) and (1.5), we obtain

$$Mz^{(k+1)} = Mz^{(k-1)} - \omega_{k+1}(\alpha_k Az^{(k)} + M(z^{(k-1)} - z^{(k)})) \ . \tag{1.6}$$

For the generalized CG method, the parameters $\{\alpha_k, \omega_{k+1}\}$ are computed so that

$$z^{(p)^T} Mz^{(q)} = 0 \qquad \text{for } p \neq q \text{ and } p, q = 0,1,\ldots,n-1 \ . \tag{1.7}$$

Since M is an $n \times n$, symmetric, positive-definite matrix, (1.7) implies that for some $k<n$

$$z^{(k)} = \emptyset$$

and hence

$$x^{(k)} = x .$$

That is, the iteration converges in no more than n steps.

We derive the above result by induction. Assume

$$z^{(p)T} M z^{(q)} = 0 \qquad \text{for } p \neq q \text{ and } p, q = 0, 1, \ldots, k. \tag{1.8}$$

Since N is skew-symmetric, there holds that for any real n-vector w

$$w^T N w = 0 . \tag{1.9}$$

From (1.6), we have

$$z^{(k)T} M z^{(k+1)} = z^{(k)T} M z^{(k-1)} - \omega_{k+1}(\alpha_k z^{(k)T} A z^{(k)} + z^{(k)T} M(z^{(k-1)} - z^{(k)})),$$

and thus by (1.8) and (1.9),

$$z^{(k)T} M z^{(k+1)} = -\omega_{k+1}(\alpha_k z^{(k)T} M z^{(k)} - z^{(k)T} M z^{(k)}) .$$

Hence by choosing $\alpha_k \equiv 1$, we obtain $z^{(k)T} M z^{(k+1)} = 0$. Similarly, $z^{(k-1)T} M z^{(k+1)} = 0$ for the choice

$$\omega_{k+1} = \frac{z^{(k-1)T} M z^{(k-1)}}{z^{(k-1)T} M z^{(k-1)} - z^{(k-1)T} N z^{(k)}} . \tag{1.10}$$

We can simplify (1.10) by noting from (1.6), with (k+1) replaced with (k), that

$$z^{(k)T} M z^{(k)} = \omega_k z^{(k)T} N z^{(k-1)} ,$$

so that

$$-z^{(k-1)T} N z^{(k)} = z^{(k)T} M z^{(k)} / \omega_k .$$

We obtain

$$\omega_{k+1} = \left(1 + \frac{z^{(k)T} M z^{(k)}}{z^{(k-1)T} M z^{(k-1)}} \times \frac{1}{\omega_k} \right)^{-1} .$$

Then for $j \leq k-2$, we obtain from (1.6), (1.8), and (1.9) that

$$\begin{aligned} z^{(j)T} M z^{(k+1)} &= z^{(j)T} M z^{(k-1)} - \omega_{k+1}(z^{(j)T}(M-N)z^{(k)} - z^{(j)T} M(z^{(k-1)} - z^{(k)})) \\ &= \omega_{k+1} z^{(j)T} N z^{(k)} . \end{aligned} \tag{1.11}$$

But, since for $\alpha_j = 1$,

$$Mz^{(j+1)} = Mz^{(j-1)} - \omega_{j+1}(-Nz^{(j)} + Mz^{(j-1)}) ,$$

there holds

$$z^{(k)^T} M z^{(j+1)} = \omega_{j+1} z^{(k)^T} N z^{(j)} \qquad (1.12)$$

Thus, since from (1.8) the l.h.s. of (1.12) is zero, we have for $j \leq k-2$

$$z^{(j)^T} N z^{(k)} = 0 , \qquad (1.13)$$

which implies

$$z^{(j)^T} M z^{(k+1)} = 0 \quad \text{for } j \leq k-2 .$$

The desired result (1.7) then follows by induction.

The **generalized CG method** for the splitting $M = (A + A^T)/2$ is summarized as follows:

Algorithm

Let $x^{(0)}$ be a given vector and arbitrarily define $x^{(-1)}$. For k = 0,1,...

(1) Solve $Mz^{(k)} = r^{(k)}$, where $r^{(k)} = b - Ax^{(k)}$.

(2) Compute

$$\omega_{k+1} = \left(1 + \frac{z^{(k)^T} M z^{(k)}}{z^{(k-1)^T} M z^{(k-1)}} \frac{1}{\omega_k} \right)^{-1}, \qquad k \geq 1$$

$$\omega_1 = 1 .$$

(3) Compute

$$x^{(k+1)} = x^{(k-1)} + \omega_{k+1}(z^{(k)} + x^{(k)} - x^{(k-1)}) .$$

In the computation of ω_{k+1}, one need not recompute $Mz^{(k)}$ since $r^{(k)}$ can be saved from step (1).

A simple induction argument shows that for all k, there holds

$$0 < \omega_{k+1} \leq 1 ,$$

unlike the case $N = N^T$, for which $\omega_{k+1} \geq 1$.

Note that since $z^{(p)^T} Mz^{(q)} = 0$ for $p \neq q$ and since by (1.13), $z^{(p)^T} Nz^{(q)} = 0$ for $|p-q| \neq 1$, there holds

$$z^{(p)^T} Az^{(q)} = 0 \quad \text{for} \quad |p-q| > 1 .$$

Remarks concerning alternative forms of the generalized CG algorithm, which can be more efficient for actual computation, can be found in [3].

The calculated vectors $\{z^{(k)}\}_{k=0}^{n-1}$ will not generally be M-orthogonal in practice because of roundoff errors. One might consider forcing the newly calculated vectors to be M orthogonal by a procedure such as Gram-Schmidt. However, this would require the storage of all previously obtained vectors.

Cur basic approach is to permit the gradual loss of orthogonality and with it the finite termination property of CG. We consider primarily the iterative aspects of the algorithm. In fact for solving large sparse systems arising from the discretization of elliptic partial differential equations, the application of principal interest for us and for which the generalized CG method seems particularly effective, convergence to desired accuracy often occurs within a number of iterations small compared with n.

2. Some Properties of the Method

In [3], there are presented some optimality properties, convergence properties, and eigenvalue relationships for the case in which A is symmetric. We discuss in this section related matters for the case in which M is symmetric and positive-definite and N is skew-symmetric.

2.1. From (1.6) with $\alpha_k = 1$ we obtain

$$z^{(k+1)} = z^{(k-1)} - \omega_{k+1}(-M^{-1}Nz^{(k)} + z^{(k-1)}) = (1 - \omega_{k+1})\, z^{(k-1)} + \omega_{k+1}M^{-1}Nz^{(k)}, \quad (2.1)$$

which may be viewed as a relaxation of an iteration with iteration matrix

$$L = M^{-1}N .$$

We note that L is similar to a skew-symmetric matrix and hence that all the eigenvalues of L are either pure imaginary and occur in conjugate pairs, or are zero.

The eigenvalues of L can be determined directly from the generalized CG method in the same manner as for the symmetric case. We write (2.1) as

$$Lz^{(k)} = \left(1 - \frac{1}{\omega_{k+1}}\right) z^{(k-1)} + \frac{1}{\omega_{k+1}} z^{(k+1)}$$

or

$$L[z^{(0)}, z^{(1)}, \ldots, z^{(n-1)}] = [z^{(0)}, z^{(1)}, \ldots, z^{(n-1)}] \begin{bmatrix} 0 & 1 - \frac{1}{\omega_2} & & & \\ 1 & 0 & 1 - \frac{1}{\omega_3} & & \\ & \frac{1}{\omega_2} & 0 & \ddots & \\ & & \frac{1}{\omega_3} & \ddots & 1 - \frac{1}{\omega_n} \\ & & & \frac{1}{\omega_{n-1}} & 0 \end{bmatrix}$$

In matrix notation, the above equation can be written as

$$LZ = ZJ .$$

Assuming the columns of Z are linearily independent, it follows that

$$J = Z^{-1}LZ .$$

It can be shown that the kth principal minor of J yields very good estimates of the extreme eigenvalues of L, even in the presence of rounding errors. Note that although the matrix J is not skew-symmetricitis diagonally similar to such a matrix.

, 2.2 As in §2 of [3], define

$$K = I - M^{-1}N = I - L .$$

Then we have, as for the symmetric case,

$$z^{(k)} = [I - KP_{k-1}(K)]z^{(0)},$$

where

$$P_{k-1}(K) = \sum_{j=0}^{k-1} \beta_j^{(k-1)} K^j$$

is a polynomial in K of degree k-1. Correspondingly, we have

$$x^{(k)} = x^{(0)} + P_{k-1}(K)\, z^{(0)} .$$

As for the symmetriccase, we define the weighted error function

$$E(x^{(k)}) = \frac{1}{2}\, e^{(k)^T}(M-N)\, e^{(k)} , \qquad (2.2)$$

where

$$e^{(k)} = x - x^{(k)}$$

For -the present case, (2.2) becomes

$$E(x^{(k)}) = \frac{1}{2}\, e^{(k)^T} Me^{(k)} .$$

Assuming that (M-N) is nonsingular, we obtain, using

$$z^{(0)} = Ke^{(0)}$$

and

$$e^{(k)} = [I - KP_{k-1}(K)]\, e(P) ,$$

the expression

$$E(x^{(k)}) = \frac{1}{2}\, e^{(0)^T} [I - KP_{k-1}(K)]^T M[I - KP_{k-1}(K)]\, e^{(0)} . \qquad (2.3)$$

The result for the symmetric case, that the polynomials $P_{k-1}(K)$ generated by CG minimize $E(x^{(k)})$ over the choice of all polynomials of degree k-1, does not hold here in general. Widlund [7] has shown, however, that there does hold

$$E(x^{(k)}) \leq \max_j (1+\lambda_j^2)\ E(y) \qquad (2.4$$

for any y of the form

$$y = x^{(0)} + S_{k-1}(K)\, z^{(0)} ,$$

where $S_{k-1}(K)$ is a polynomial in K of degree k-1. Here λ_j, j = 1,2,...,n, are the eigenvalues of L.

We remark that, as for the symmetric case, the generalized CG method converges in only p steps if K has only p < n distinct eigenvalues. This same result holds also if K has a larger number of distinct eigenvalues but $e^{(0)}$ lies in a subspace generated by the eigenvectors associated with only p of these eigenvalues.

2.3. Let us consider the polynomials $S_{k-1}(K)$ generated by the Richardson second order method, for which $\omega_1 = 1$ and $\omega_{k+1} \equiv \omega$, a fixed parameter, for $k \geq 1$. For this case, (1.4) with $\alpha_k \equiv 1$ becomes

$$x^{(k+1)} = x^{(k-1)} + \omega(z^{(k)} + x^{(k)} - x^{(k-1)}) , \qquad k \geq 1 ,$$

and we have

$$e^{(k)} = [I - KS_{k-1}(K)]\, e^{(0)} \equiv T_{k,\omega}(L)\, e^{(0)} .$$

We seek a value of ω for which the spectral radius of $T_{k,\omega}(L)$ is a minimum.

Denote by $\rho(X)$ the spectral radius of a matrix X. By using an argument similar to that given in [4, pp. 18-24], it can be shown that for

$$\hat{\omega} = \frac{2}{1 + \sqrt{1 + \rho^2(L)}}$$

there holds

$$\rho(T_{k,\omega}(L)) \geq \rho(T_{k,\hat{\omega}}(L)) ,$$

where

$$\rho(T_{k,\hat{\omega}}(L)) = \theta^k \left(1 + \frac{1-\theta^2}{1+\theta^2}\, k \right) \qquad (2.5)$$

and

$$\theta = \frac{\rho(L)}{1 + \sqrt{1 + \rho^2(L)}} = \sqrt{(1 - \hat{\omega})} \quad . \qquad (2.$$

To carry out the Richardson second order method we would need to have an estimate of P(L). It is interesting to note that here also $0 < \omega \leq 1$. As for CG, underrelaxation is preferred for the case of skew-symmetric N.

2.4. One can use for y in (2.4) the optimal kth Richardson second-order iterate to obtain an asymptotic error estimate for the generalized CG method. Doing so yields, with the use of (2.5) and (2.6),

$$E(x^{(k)}) \leq C\theta^{2k} \left[1 + \frac{1 - \theta^2}{1 + \theta^2} k \right] E(x^{(0)}) ,$$

where C is a constant independent of k.

3. An Example

To illustrate the method, we give here a simple example for which one can easily estimate the spectral radius of L. Consider the problem

$$\begin{aligned} -\Delta u + \sigma u_x &= f(x,y) && (x,y) \in R \\ u &= g(x,y) && (x,y) \in \partial R , \end{aligned}$$

where σ is a constant and R is the unit square $0 < x,y < 1$. We discretize on a uniform mesh of width h, using for Δ the standard five-point, approximation Δ_h and for u_x at the point i,j the approximation $(U_{i+1,j} - U_{i-1,j})/(2h)$, where U_{ij} corresponds to $u(x,y)$ at $x = ih$, $y = jh$.

We consider solving the discrete problem by the algorithm of §1, for which

$$M = -\Delta_h$$

and

$$N = \begin{bmatrix} D & & & 0 \\ & D & & \\ & & \ddots & \\ 0 & & & D \end{bmatrix} , \text{ where } D = \frac{-\sigma}{2h} \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & 0 \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & 1 \\ 0 & & & -1 & 0 \end{bmatrix}$$

A fast direct method (cf. [1]) can be used in this case for the solution of the system of equations $Mz^{(k)} = r^{(k)}$. (Of course, a fast direct method could be used, without iteration, to solve the entire problem for this simple example.)

To estimate the rate of convergence, we wish to determine the extremal eigenvalues of $L = M^{-1}N$, that is

$$N\varphi = i\lambda M\varphi \ . \tag{3.1}$$

For the corresponding differential operators, the equivalent eigenproblem is

$$\begin{aligned} \sigma\varphi_x &= i\lambda(\varphi_{xx} + \varphi_{yy}) && (x,y) \in R \\ \varphi &= 0 && (x,y) \in \partial R \ , \end{aligned} \tag{3.2}$$

for which one readily finds, by separation of variables, the eigenvalues to be

$$\lambda_{j,l} = \frac{\sigma}{2\pi\sqrt{j^2 + l^2}} \ , \qquad j = 1, 2, \ldots . \quad l = 1, 2, \ldots .$$

The first eigenvalues $\lambda_{1,1}$ provide the uniform estimate for the spectral radius $\rho(L)$,

$$\rho(L) = |\lambda|_{max} \tilde{=} \frac{\sqrt{2}}{4\pi}|\sigma|, \tag{3.3}$$

for which

$$\theta \tilde{=} \frac{\sqrt{2}|\sigma|}{4\pi}\left[1 + \sqrt{1 + \frac{\sigma^2}{8\pi^2}}\right]^{-1} \ .$$

Direct computation of the eigenvalues of (3.1), which is somewhat more cumbersome than for (3.2), shows (3.3) to be good asymptotically to within $O(h^2)$ as $h \to 0$.

We remark that for the symmetric problem with σu_x replaced by Uu, and the splitting $M = -\Delta_h$ and $N = -\sigma I$, the estimate corresponding to (3.3) is [2] $|\lambda|_{max} \tilde{=} |\sigma|/(2\pi^2)$. Numerical experiments illustrating the behavior of the modified CG method on related examples can be found in [7].

The possibility of using CG on nonsymmetric matrices in the manner presented here first occurred to us while listening to a presentation by T. Manteuffel of his dissertation research [6]. We wish to thank O. Widlund for making available to us his results to appear in [7] and to thank both O. Widlund and T. Karasalo for their helpful comments. This work was supported in part by the Energy Research and Development Administration and by the National Science Foundation.

REFERENCES

[1] B. L. Buzbee, G. H. Golub, and C.W. Nielson, "On direct methods for solving Poisson's equation," SIAM J. Numer. Anal. 7 (1970), pp. 627-656.

[2] P. Concus and G.H. Golub, "Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations," SIAM J. Numer. Anal. 10 (1973), pp. 1103-1120.

[3] 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. Symposium on Sparse Matrix Computations, Argonne National Lab., Sept. 1975, Academic Press (to appear).

[4] G. H. Golub, "The use of Chebyshev matrix polynomials in the iterative solution of linear equations compared with the method of successive over-relaxation," Ph.D. Thesis, University of Illinois. 1959.

[5] G.H. Golub and R.S. Varga, "Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods," Numer. Math. 3 (1961), pp. 147-168.

[6] T. A. Manteuffel, "An iterative method for solving nonsymmetric linear systems with dynamic estimation of parameters," Ph.D. Thesis, University of Illinois, 1975.

[7] O. Widlund, Tech. Rept., Courant Institute, New York University (to appear).