Brigham Young University
Department of Electrical and
Computer Engineering

459 Clyde Building
Provo, Utah 84602

Continued Fraction Error Bound for Conjugate Gradient Method

Karl Warnick

9 June 1998

MERS Technical Report # MERS 98-03
ECEN Department Report # TR-L100-98.3

Microwave Earth Remote Sensing (MERS)
Laboratory

Continued Fraction Error Bound for Conjugate Gradient Method

Short title: Conjugate Gradient Error Bound

Karl F. Warnick

Department of Electrical and Computer Engineering
459 Clyde Building
Brigham Young University
Provo, UT 84602

Abstract—An error bound for the conjugate gradient method for solution of a large linear system is developed which can be computed efficiently from quantities available at each iteration of the algorithm and requires only a lower bound on the eigenvalues of the linear operator.

Keywords: Conjugate Gradient Method, Continued Fractions

AMS(MOS) subject classification: 65F10

1. INTRODUCTION

The conjugate gradient method (CG) for solution of large linear systems enjoys widespread use as a numerical technique in computational electromagnetics, acoustics, fluid dynamics, signal processing, statistics, and many other fields. Despite the development of more advanced iterative algorithms which overcome breakdown problems and exhibit superior convergence properties, CG remains important due to ease of application and well understood behavior. A large body of literature quantifies the convergence rate of the algorithm in terms of properties of the spectrum of the linear operator, such as the condition number, eigenvalue clustering, isolated large or small eigenvalues, or gaps between eigenvalues [8, 6, 14, 1, 4, 7, 10, 12, 9]. The bounds developed in these and other papers allow the conjugate gradient algorithm to be terminated when the error grows acceptably small. In many applications, however, detailed knowledge of the structure of the eigenvalues is not readily available, and obtaining such information leads to additional computational burden. This

paper presents a new method for determining the error at each step of the conjugate gradient algorithm. An error bound is developed which can be computed efficiently from quantities available at each iteration of the algorithm and requires minimal information about the spectrum of the linear operator.

This error bound arises from the relationship between the conjugate gradient method for solution of a linear system $A\mathbf{x} = \mathbf{b}$, where A is symmetric and positive definite, and a corresponding continued fraction from the theory of the classical problem of moments [11]. The connection between CG and continued fractions involves the natural Hankel structure of the algorithm in terms of moments of the linear operator. Using this Hankel structure, the A–norm of the error $\mathbf{x} - \mathbf{x}_k$, where $\mathbf{x}_k$ is the approximate solution to the linear system at the kth step of the algorithm, can be expressed directly [2], but the exact expression depends on the unknown quantity $\mathbf{b} \cdot \mathbf{x}$. Most convergence theories are based on estimates of this error in terms of the spectrum of the linear operator. This paper develops a method for computing a bound on this error at each step of the algorithm using the approximants of a particular continued fraction. The even approximants of this continued fraction are equal to $\mathbf{b} \cdot \mathbf{x}_k$. These quantities are lower bounds for $\mathbf{b} \cdot \mathbf{x}$. The odd approximants are monotonically converging upper bounds for $\mathbf{b} \cdot \mathbf{x}_k$. The difference between the odd and even approximants therefore provides a nonincreasing bound for the A–norm $\|\mathbf{x} - \mathbf{x}_k\|_A$ of the error at each iteration of the algorithm. Evaluation of the odd approximants requires only a lower bound on the eigenvalues of the matrix of the linear system. The computational burden required to compute this error bound is trivial, since it can be determined without direct evaluation of high order Hankel determinants and only $O(1)$ additional scalar multiplications are required per iteration of the algorithm.

2. NOTATION

Let A be an N by N symmetric, positive definite matrix, and $\mathbf{b}$ be a real vector of length N. We define the $n+1$ by $m+1$ Hankel matrix $H_p^{n,m}$ of moments of A to be

$$H_p^{n,m} = \begin{bmatrix} \mu_p & \mu_{p+1} & \cdots & \mu_{p+m} \\ \mu_{p+1} & \mu_{p+2} & \cdots & \mu_{p+m+1} \\ \vdots & \vdots & & \vdots \\ \mu_{p+n} & \mu_{p+n+1} & \cdots & \mu_{p+n+m} \end{bmatrix} \tag{1}$$

where $\mu_k = \mathbf{b} \cdot A^k \mathbf{b}$. A square Hankel matrix is denoted $H_p^n = H_p^{n,n}$. The Hankel determinants Δ_n, $\Delta_n^{(1)}$, and $\Delta_n^{(-1)}$ are defined by

$$\Delta_n = |H_0^n| \tag{2a}$$

$$\Delta_n^{(1)} = |H_1^n| \tag{2b}$$

$$\Delta_n^{(-1)} = \left| H_{-1}^n \right| \tag{2c}$$

where $\mu_{-1} = 0$ in the definition of $\Delta_n^{(-1)}$.

The conjugate gradient algorithm for solution of $A\mathbf{x} = \mathbf{b}$ is initialized by setting

$$\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0 \tag{3a}$$

$$\mathbf{d}_0 = \mathbf{r}_0 \tag{3b}$$

and proceeds via the iteration

$$\gamma_k = \frac{\mathbf{r}_k \cdot \mathbf{r}_k}{\mathbf{r}_k \cdot A\mathbf{d}_k} \tag{4a}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma_k \mathbf{d}_k \tag{4b}$$

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \gamma_k A\mathbf{d}_k \tag{4c}$$

$$\eta_k = \frac{\mathbf{r}_{k+1} \cdot \mathbf{r}_{k+1}}{\mathbf{r}_k \cdot \mathbf{r}_k} \tag{4d}$$

$$\mathbf{d}_{k+1} = \mathbf{r}_{k+1} + \eta_k \mathbf{d}_k \tag{4e}$$

where $\mathbf{r}_k$ is equal to the residual $\mathbf{b} - A\mathbf{x}_k$ and $\mathbf{d}_k$ represents the kth search direction along which the objective function $f(\mathbf{y}) = (1/2)\mathbf{y} \cdot A\mathbf{y} - \mathbf{b} \cdot \mathbf{y}$ is minimized. In the following, we assume that the initial guess $\mathbf{x}_0$ is equal to the zero vector.

From a result of Brezinski [2], we have that

$$\mathbf{b} \cdot \mathbf{x}_{k+1} = -\frac{\Delta_{k+1}^{(-1)}}{\Delta_k^{(1)}}. \tag{5}$$

By repeated application of the determinant identity

$$\left| \begin{array}{c|c} a & \mathbf{p}^{\mathrm{T}} \\ \hline \mathbf{q} & M \end{array} \right| = |M|(a - \mathbf{p}^{\mathrm{T}} M^{-1} \mathbf{q}) \tag{6}$$

and the block matrix inverse formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + E\Delta^{-1}F & -E\Delta^{-1} \\ -\Delta^{-1}F & \Delta^{-1} \end{bmatrix} \tag{7}$$

where $\Delta = B - CA^{-1}D$, $E = A^{-1}D$, and $F = CA^{-1}$, the recursive determinant identity

$$\left| \begin{array}{c|c|c} a & \mathbf{p} & b \\ \hline \mathbf{r}^{\mathrm{T}} & M & \mathbf{q}^{\mathrm{T}} \\ \hline c & \mathbf{t} & d \end{array} \right| = \frac{\left| \begin{array}{c|c} a & \mathbf{p} \\ \hline \mathbf{r}^{\mathrm{T}} & M \end{array} \right| \left| \begin{array}{c|c} M & \mathbf{q}^{\mathrm{T}} \\ \hline \mathbf{t} & d \end{array} \right| - \left| \begin{array}{c|c} \mathbf{r}^{\mathrm{T}} & M \\ \hline c & \mathbf{t} \end{array} \right| \left| \begin{array}{c|c} \mathbf{p} & b \\ \hline M & \mathbf{q}^{\mathrm{T}} \end{array} \right|}{|M|} \tag{8}$$

can be proved. In this expression, M is an n by n matrix, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$, and $\mathbf{t}$ are vectors of length n, and a, b, c, and d are scalars. By taking the left hand side of Eq. (8) to be $\Delta_{k+1}^{(-1)}$, we have that

$$\frac{\Delta_{k+1}^{(-1)}}{\Delta_k^{(1)}} = \frac{\Delta_k^{(-1)}}{\Delta_{k-1}^{(1)}} - \frac{\Delta_k^2}{\Delta_k^{(1)} \Delta_{k-1}^{(1)}}. \tag{9}$$

The first term on the right is equal to $-\mathbf{b} \cdot \mathbf{x}_k$, so that the iterative nature of the algorithm is explicit in this expression.

3. CONTINUED FRACTION ERROR BOUND

Using the Hankel structure of the conjugate gradient method, we define a continued fraction which is closely related to CG through the values of its approximants. This continued fraction is similar in form to other continued fractions which arise in connection with CG and iterative solution algorithms of the same family [2, 3], but contains a parameter which is initially arbitrary and will later be specified such that the even and odd approximants become an error bound for the algorithm.

Consider the power series expansion

$$\mathbf{b}\cdot\mathbf{x} = \frac{\mathbf{b}\cdot\mathbf{b}}{s} - \frac{\mathbf{b}\cdot(A-sI)\mathbf{b}}{s^2} + \frac{\mathbf{b}\cdot(A-sI)^2\mathbf{b}}{s^3} - \cdots \tag{10}$$

Denoting by $\tilde{\mu}_k$ the moments $\mathbf{b}\cdot(A-sI)^k\mathbf{b}$, the series can be written in the form

$$I(s) = \frac{\tilde{\mu}_0}{s} - \frac{\tilde{\mu}_1}{s^2} + \frac{\tilde{\mu}_2}{s^3} - \cdots \tag{11}$$

Such a power series can be transformed into a corresponding continued fraction [5, 11]

$$C(s) = \cfrac{\tilde{\alpha}_1}{s + \cfrac{\tilde{\alpha}_2}{1 + \cfrac{\tilde{\alpha}_3}{s + \cfrac{\tilde{\alpha}_4}{1+\cdots}}}} \tag{12}$$

where the coefficients of the continued fraction are ratios of Hankel determinants of the coefficients of the power series,

$$\tilde{\alpha}_{2n} = \frac{\tilde{\Delta}^{(1)}_{n-1}\tilde{\Delta}_{n-2}}{\tilde{\Delta}^{(1)}_{n-2}\tilde{\Delta}_{n-1}} \tag{13a}$$

$$\tilde{\alpha}_{2n+1} = \frac{\tilde{\Delta}^{(1)}_{n-2}\tilde{\Delta}_{n}}{\tilde{\Delta}^{(1)}_{n-1}\tilde{\Delta}_{n-1}}. \tag{13b}$$

A tilde denote the appearance of the moments $\tilde{\mu}_k$ of $A - sI$ rather than moments of A in the Hankel determinates. The even and odd approximants to this continued fraction are Padé approximants for $\mathbf{b}\cdot\mathbf{x}$, since the power series expansion of $C(s)$ is precisely Eq. (10).

In the remainder of this section, we relate the coefficients α_n to the values γ_k and η_k produced by the continued fraction algorithm, show that the approximants to $C(s)$ are upper and lower bounds for $\mathbf{b}\cdot\mathbf{x}$, and develop a result on the dependence on s of the $\tilde{\alpha}_n$ which will be employed in the next section to show that the error bound can be computed efficiently without direct evaluation of the Hankel determinants or moments of $A - sI$.

Even approximants $C_{2n}(s)$ to the continued fraction $C(s)$ are defined by setting $\tilde{\alpha}_{2n+1}$, $\tilde{\alpha}_{2n+2}$, ... to zero in Eq. (12). The odd approximants $C_{2n+1}(s)$ are defined analogously. From Eq. (12), we have that

$$C_{2n}(0) = \frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_3}{\alpha_2\alpha_4} + \cdots + \frac{\alpha_1\alpha_3\cdots\alpha_{2n-1}}{\alpha_2\alpha_4\cdots\alpha_{2n}}. \tag{14}$$

Using the definition of the α_k, this can be rewritten as

$$C_{2n}(0) = \beta_0 + \beta_1 + \cdots + \beta_{n-1} \tag{15}$$

where

$$\beta_k = \frac{\Delta_k^2}{\Delta_k^{(1)}\Delta_{k-1}^{(1)}}. \tag{16}$$

By comparison of this expression with Eq. (9), we can see that $\beta_k = \mathbf{b}\cdot(\mathbf{x}_{k+1} - \mathbf{x}_k)$. This demonstrates that the even approximants $C_{2k}(0)$ are equal to $\mathbf{b}\cdot\mathbf{x}_k$.

We now relate the coefficients α_n of the continued fraction to γ_k and η_k. The coefficient of the vector $A^k\mathbf{b}$ in the Krylov subspace expansion for $\mathbf{x}_{k+1}$ implied by the algorithm (4) is equal to $(-1)^k\gamma_0\gamma_1\cdots\gamma_k$. This coefficient can also be written as [2]

$$c_k = (-1)^k \frac{\Delta_k}{\Delta_k^{(1)}}. \tag{17}$$

Since $(-1)^k\gamma_0\gamma_1\cdots\gamma_k = -c_{k-1}\gamma_k$, we have that

$$\gamma_k = \frac{\Delta_{k-1}^{(1)}\Delta_k}{\Delta_k^{(1)}\Delta_{k-1}}. \tag{18}$$

By comparison with Eq. (13a), we obtain the result

$$\alpha_{2k} = 1/\gamma_{k-1} \tag{19}$$

for the even coefficients of the continued fraction.

It can be shown by making use of properties of the conjugate gradient algorithm that $\mathbf{r}_k\cdot\mathbf{r}_k = \mathbf{b}\cdot\mathbf{d}_k$. From the CG algorithm and Eq. (9), we have that

$$\mathbf{b}\cdot\mathbf{d}_k = \beta_k/\gamma_k. \tag{20}$$

The definition of η_k together with Eq. (18) then yields

$$\eta_k = \frac{{\Delta_{k-1}^{(1)}}^2\Delta_{k+1}}{{\Delta_k^{(1)}}^2\Delta_{k-1}}. \tag{21}$$

Using Eq. (19) and the definition of the α_n, we obtain for $k > 0$,

$$\alpha_{2k+1} = \eta_{k-1}/\gamma_{k-1}. \tag{22}$$

For $k = 0$, $\alpha_1 = \mathbf{b}\cdot\mathbf{b}$. With this expression and Eq. (19), we have established relationships between the η_k and γ_k produced by the conjugate gradient method and the coefficients α_k of the corresponding continued fraction $C(s)$ evaluated at $s = 0$.

We now employ the relationship between the conjugate gradient method and the Lanczos algorithm to obtain a result on the dependence of the $\tilde{\alpha}_n$ on s, in order to show that the even

approximants of $C(s)$ are independent of s. This result will also allow us in the following section to obtain the $\tilde{\alpha}_n$ from the quantities η_k and γ_k. Each step in the iteration of the Lanczos algorithm produces a tridiagonalization of the form $Q_k^T A Q_k = T_k$, where Q_k is orthogonal and has as its first column the vector $\mathbf{b}$ normalized to unit magnitude. The matrix T_k has components equal to [4]

$$[T_k]_{i,i} = \frac{\eta_{i-2}}{\gamma_{i-2}} + \frac{1}{\gamma_{i-1}} \tag{23a}$$

$$[T_k]_{i,i+1} = [T_k]_{i+1,i} = \frac{\sqrt{\eta_{i-1}}}{\gamma_{i-1}} \tag{23b}$$

where $\eta_{-1} = 0$. Using the relationships (19) and (22), this can be transformed into

$$T_k = \begin{bmatrix} \alpha_2 & \sqrt{\alpha_2\alpha_3} & & & & \\ \sqrt{\alpha_2\alpha_3} & \alpha_3+\alpha_4 & \sqrt{\alpha_4\alpha_5} & & & \\ & \sqrt{\alpha_4\alpha_5} & \alpha_5+\alpha_6 & \sqrt{\alpha_4\alpha_5} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \sqrt{\alpha_{2k-2}\alpha_{2k-1}} & \alpha_{2k-1}+\alpha_{2k} & \end{bmatrix} \tag{24}$$

The columns of Q_k are an orthonormal basis for the Krylov subspace span$\{\mathbf{b}, A\mathbf{b}, \ldots, A^{k-1}\mathbf{b}\}$. This is equal to the jth Krylov subspace span$\{\mathbf{b}, (A-sI)\mathbf{b}, \ldots, (A-sI)^{k-1}\mathbf{b}\}$. Since Q_k is constructed iteratively, the Lanczos algorithm produces the same orthogonal matrix when applied to the matrix $A - sI$. Let $\tilde{T}_k$ denote the tridiagonal matrix produced for $A - sI$, so that $\tilde{T}_k$ is related to (24) by replacement of the coefficients α_n with $\tilde{\alpha}_n$. We then have that

$$Q_k^T (A - sI) Q_k = \tilde{T}_k. \tag{25}$$

Since Q_k is orthogonal, this becomes

$$T_k = \tilde{T}_k + sI. \tag{26}$$

From the well known continued fraction form for components of the inverse of a tridiagonal matrix [5], the even approximants of $C(s)$ are equal to

$$C_{2k}(s) = \mathbf{e}_1^{\mathrm{T}} (sI + \tilde{T}_k)^{-1} \mathbf{e}_1 \tag{27}$$

where $\mathbf{e}_1 = [1\ 0 \ldots 0]^{\mathrm{T}}$. This expression, together with Eq. (26), shows that the even approximants are independent of s. Note that Eq. (26) also implies that the odd approximants can be viewed as generalized approximants [11] for $C(0)$, since only the final denominator of each odd approximant depends on the parameter s.

The series (11) can be recast as a classical moment problem involving an unknown distribution function which must be determined from its moments. The weights of the discrete

distribution function are given by the projections of $\mathbf{b}$ onto the eigenvectors of A and the locations are given by the eigenvalues of $A - sI$. As long as $A - sI$ is positive definite and $\mathbf{b}$ is real, the corresponding distribution function is positive, nondecreasing, and confined to the positive real axis. A result from the theory of the classical problem of moments [5, 11] then shows that if s is positive and smaller than the minimal eigenvalue of A, the even appoximants are nondecreasing lower bounds and the odd approximants nonincreasing upper bounds for $\mathbf{b}\cdot\mathbf{x}$. We have established above that $C_{2k}(s) = \mathbf{b}\cdot\mathbf{x}_k$. Since the error $\|\mathbf{x} - \mathbf{x}_k\|_A^2$ is equal to $\mathbf{b}\cdot(\mathbf{x} - \mathbf{x}_k)$, we thus have that

$$\|\mathbf{x} - \mathbf{x}_k\|_A^2 \leq C_{2k+1}(s) - C_{2k}(s) \tag{28}$$

so that the approximants of $C(s)$ provide a bound for the error at each step of the conjugate gradient algorithm. It is straightforward to demonstrate that the bound $C_{2k+1}(s) - C_{2k}(s)$ satisfies the same finite termination property as the CG algorithm, so that if the solution $\mathbf{x}$ lies in an invariant subspace of A of dimension M, the bound vanishes for $k \geq M$.

4. IMPLEMENTATION

Direct computation of the approximants of $C(s)$ would require additional matrix–vector multiplications to obtain the moments $\tilde{\mu}_k = \mathbf{b}\cdot(A - sI)^k\mathbf{b}$ and the use of a numerically unstable algorithm [5] for evaluating the $\tilde{\alpha}_k$. Equation (26) provides a much more efficient method, since we must have that

$$\alpha_{2k}\alpha_{2k+1} = \tilde{\alpha}_{2k}\tilde{\alpha}_{2k+1} \tag{29a}$$

$$\alpha_{2k+1} + \alpha_{2k+2} = \tilde{\alpha}_{2k+1} + \tilde{\alpha}_{2k+2} + s \tag{29b}$$

for $k > 0$. For $k = 0$, $\tilde{\alpha}_1 = \mathbf{b}\cdot\mathbf{b}$ and $\alpha_2 = \tilde{\alpha}_2 + s$. Together with Eqs. (19) and (22), these relationships allow the coefficients $\tilde{\alpha}_k$ for arbitrary s to be computed from the values of γ_i and η_i produced by the conjugate gradient algorithm. The even and odd approximants C_n are then given by Eq. (12) with $\tilde{\alpha}_{n+1}$, $\tilde{\alpha}_{n+2}$, ... set to zero. The approximants can be evaluated using a well known recursive procedure [13] so that only a few scalar multiplications are required per iteration.

The remaining problem is the specification of the parameter s. As demonstrated in the previous section, the odd approximants yield upper bounds for $\mathbf{b}\cdot\mathbf{x}$ if s is smaller than the minimal eigenvalue of A. Regardless of the value of s, the odd approximants must converge to $\mathbf{b}\cdot\mathbf{x}$ in at most N iterations of the algorithm. The further s is from the minimal eigenvalue, however, the slower the convergence. Depending on the particular application of the conjugate gradient algorithm, one of several techniques might be employed to choose the parameter s. Gershgorin–type theorems could be employed to estimate the minimal

eigenvalue of A; this would be suited to the diagonally dominant case. Since the eigenvalues of the tridiagonal Lanczos matrices T_k approximate those of A, s could be taken to be smaller than the minimal eigenvalue of T_k. Finally, the small computational cost involved in finding the odd approximants allows their determination for many values of s and implementation of a search strategy to identify the best monotonically decreasing bound.

5. CONCLUSION

We have exploited the connection between the conjugate gradient method for solution of the linear system $A\mathbf{x} = \mathbf{b}$ and a corresponding continued fraction constructed from moments of the matrix A to develop a method for computing the error at each iteration of the algorithm. Using the approximants of the continued fraction, a nonincreasing bound on the A–norm of the error $\mathbf{x} - \mathbf{x}_k$ can be obtained if a lower bound on the minimal eigenvalue of the matrix of the linear system is available. Computation of the bound requires only a few scalar multiplications at each interation of the algorithm.

ACKNOWLEGEMENTS

This material is based in part upon work supported under a National Science Foundation Graduate Fellowship. The author would also like to acknowledge Stephen Wandzura for suggesting the methods of Gordon [5] in connection with the conjugate gradient method.

References

[1] Owe Axelsson and Gunhild Lindskog, On the rate of convergence of the preconditioned conjugate gradient algorithm, Numer. Math. 48 (1986), 499–523.

[2] Claude Brezinski, *Padé-type Approximation and General Orthogonal Polynomials*, Birkhäuser, 1980.

[3] Adhemar Bultheel and Marc Van Barel, Formal orthogonal polynomials and Hankel/Toeplitz duality, Numer. Alg. 10 (1995), 289–335.

[4] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 2 edition, 1993.

[5] Roy G. Gordon, Error bounds in equilibrium statistical mechanics, J. Math. Phys. 9 (1968), 655–663.

[6] A. Greenbaum, Comparison of splittings used with the conjugate gradient algorithm, Numer. Math. 33 (1979), 181–194.

[7] A. Greenbaum and Z. Strakos, Predicting the behavior of finite precision Lanczos and conjugate gradient computations, SIAM J. Matrix Anal. Apps. 13 (1992), 121–137.

[8] Shmuel Kaniel, Estimates for some computational techniques in linear algebra, Math. of Comp. 95 (1966), 369–378.

[9] Yvan Notay, On the convergence rate of the conjugate gradients in presence of rounding errors, Numer. Math. 65 (1993), 301–317.

[10] A. Ramage, Eigenvalue clustering and conjugate gradient convergence for elliptic partial differential equations, in: *Iterative Methods in Linear Algebra*, eds. R. Beauwens and P. de Groen, Elsevier, 1992, pp. 593–601.

[11] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.

[12] A. Van Der Sluis, The convergence behavior of conjugate gradients and Ritz values in various circumstances, in: *Iterative Methods in Linear Algebra*, eds. R. Beauwens and P. de Groen, Elsevier, 1992, pp. 49–66.

[13] H. S. Wall, *Continued Fractions*, D. Van Nostrand, 1948.

[14] Ragnar Winther, Some superlinear convergence results for the conjugate gradient method, SIAM J. Numer. Anal. 17 (1980), 14–17.