New results on the convergence of the conjugate gradient method

R. Bouyouli[1], G. Meurant[2], L. Smoch[3] and H. Sadok[3,*]

[1] *Université Mohammed V, Faculté des sciences, département de Mathématiques, Rabat, Maroc.*
[2] *CEA/DIF, BP12, 91680 Bruyéres-le-Chatel, France.*E-mail : gerard.meurant@gmail.com
[3] *Laboratoire de Mathématiques Pures et Appliquées, Université du Littoral, zone universitaire de la Mi-voix, bâtiment H. Poincaré, 50 rue F. Buisson, BP 699, F-62228 Calais Cedex, France.* E-mail: smoch@lmpa.univ-littoral.fr and sadok@lmpa.univ-littoral.fr.

SUMMARY

This paper is concerned with proving theoretical results related to the convergence of the conjugate gradient method for solving positive definite symmetric linear systems. Considering the inverse of the projection of the inverse of the matrix, new relations for ratios of the A-norm of the error and the norm of the residual are provided starting from some earlier results of Sadok [16]. The proofs of our results rely on the well-known correspondence between the conjugate gradient method and the Lanczos algorithm. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: symmetric linear systems, conjugate gradient algorithm, Lanczos algorithm, Krylov subspace, Ritz values.

1. Introduction

The conjugate gradient (CG) method was originally developed in the early 1950s by Hestenes and Stiefel [5, 6] for solving a linear system of equations

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^n \tag{1}$$

where A is a symmetric positive definite (SPD) matrix.

Let x_0 be a given initial approximate solution of equation (1), the CG method first computes the initial residual $r_0 = b - Ax_0$ and then generates a sequence of approximate solutions $x_1,\ x_2,\dots$ such that the residual vectors $r_i = b - Ax_i$, $i = 0, 1, \dots$ can be written in the form

$$r_i = P_i(A)r_0,$$

where P_i belongs to the space π_i of i-th degree polynomials satisfying the relation $P_i(0) = 1$.

The CG polynomial P_i is implicitly computed by the algorithm and is such that the error $\epsilon_i = x - x_i$, which satisfies the relation $A\epsilon_i = r_i$, is minimized in the A-norm, defined for a vector $y \in \mathbb{R}^n$ as $\|y\|_A = (y^TAy)^{\frac{1}{2}}$, at each iteration. The CG A-norm of the error is therefore given by

$$\|\epsilon_i\|_A = \min_{P_i\in\pi_i} \|P_i(A)\epsilon_0\|_A. \tag{2}$$

Received 19 november 2007
Revised 26 march 2008

In exact arithmetic, the CG residual vectors r_i are mutually orthogonal, e.g. $(r_i, r_j) = 0$, $i \neq j$.

Various upper bounds for $\|\epsilon_i\|_A$ can be obtained by using other polynomials in π_i rather than the CG polynomial in equation (2), see [1], [4], [10]. In this paper we are concerned with finding exact expressions for the A-norm of the error as well as for ratios of the norm of the error to the norm of the residual. We will rely on the correspondence between CG and the Lanczos algorithm. In the course of the development of these theoretical results, we will also introduce new approximations to the eigenvalues of A. In some particular cases, they are similar to the harmonic Ritz values, see [20].

The remainder of the paper is organized as follows. In section 2, we describe in more details the Galerkin orthogonality conditions that define the CG iterates. Using the relationship between CG and the Lanczos algorithm (see [7], [8]) in section 3 leads to some new exact relations for the A-norm of error and the residual norm, depending essentially on Krylov matrices whose columns are the vectors of the natural basis of the Krylov subspace based on A and the initial residual. In section 4, using the QR factorizations of these Krylov matrices (which are closely linked to the Lanczos algorithm), some new expressions are derived for ratios of the norms of the error and the residual.

In this paper we assume exact arithmetic. For a summary of results about the behaviour of CG in finite precision arithmetic, see [11] or [12]. Throughout this paper, e_j stands for the j-th vector of the canonical basis.

2. The conjugate gradient method

In this section, we recall some theoretical properties of CG and give some expressions for $\|\epsilon_i\|_A$ involving Krylov matrices. Let us consider the linear system (1) with the SPD matrix A. Let $v \in \mathbb{R}^n$ and $\mathcal{K}_k(A, v) \equiv span\{v, Av, \dots, A^{k-1}v\}$ be the Krylov subspace constructed from A and v. According to equation (2), the CG iterates x_k are defined by

$$x_k - x_0 \in \mathcal{K}_k(A, r_0) \equiv \mathcal{K}_k, \tag{3}$$

and the so-called Galerkin orthogonality conditions

$$r_k = b - Ax_k = A\epsilon_k \perp \mathcal{K}_k. \tag{4}$$

It follows from property (3) that the residual vector r_k can be written as a linear combination of the Krylov vectors $A^i r_0$, $i = 0, 1, \dots, k$, which is written as

$$r_k = r_0 - \sum_{i=1}^{k} a_i \, A^i r_0. \tag{5}$$

with coefficients $a_i \in \mathbb{R}$, $1 \leq i \leq k$. The orthogonality condition (4) is nothing but

$$(A^i r_0)^T r_k = 0, \quad 0 \leq i \leq k-1. \tag{6}$$

In an equivalent matrix form, using the Krylov matrix K_k whose columns are the Krylov vectors

$$K_k = [r_0, Ar_0, \dots, A^{k-1}r_0],$$

and the representation

$$r_k = r_0 - AK_k a, \tag{7}$$

with $a = (a_1, a_2, \cdots, a_k)^T$, the orthogonality condition writes as $K_k^T r_k = 0$ or

$$(K_k^T A K_k) a = K_k^T r_0. \tag{8}$$

Of course, in practice, x_k is not computed by solving the linear system (8) for the coefficients a_i at each iteration k. This relation will be used only for theoretical purposes. The matrix $K_k^T A K_k$, known as a moment matrix, is dense, has Hankel structure and is badly conditioned. The most usual form of the conjugate gradient algorithm (see for instance [4], [1], [10] or [11]) is obtained by building an orthogonal basis of the Krylov subspace. The most used form of the conjugate gradient algorithm involves two coupled two-term recurrences. For later use we denote the two scalar coefficients of CG as

$$\gamma_j = (r_j, r_j)/(Ap_j, p_j), \quad \beta_{j+1} = (r_{j+1}, r_{j+1})/(r_j, r_j), \tag{9}$$

where r_j (resp. p_j) denotes the residual (resp. descent) vector and $(.\,,.)$ denotes the usual inner product.

The CG algorithm needs only a matrix-vector product, vector additions and multiplications by scalars and two inner products per iteration. In the following, we derive expressions for the A-norm of the error $\|\epsilon_k\|_A$. Similar results for the error norm of Krylov methods like GMRES or FOM were proved in [16]. They will be used in Section 4. We start by giving expressions of $\|\epsilon_k\|_A$ involving the Krylov matrix K_k.

Theorem 1. *Let $\epsilon_k = x - x_k$ be the conjugate gradient error at iteration k. Then, if the matrices $K_k^T A K_k$ and $K_{k+1}^T A^{-1} K_{k+1}$ are nonsingular, we have*

$$\|\epsilon_k\|_A^2 = (A\epsilon_k, \epsilon_k) = \frac{\det(K_{k+1}^T A^{-1} K_{k+1})}{\det(K_k^T A K_k)} = \frac{1}{e_1^T (K_{k+1}^T A^{-1} K_{k+1})^{-1} e_1}.$$

Proof. Since $A\epsilon_k = r_k$, it follows from relations (5) and (6), because of orthogonality conditions, that

$$\begin{aligned}(A\epsilon_k, \epsilon_k) &= (r_k, \epsilon_k)\\ &= (r_k, \epsilon_0 - \sum_{i=1}^{k} a_i\, A^{i-1} r_0)\\ &= (r_k, \epsilon_0).\end{aligned}$$

Using equations (8) and (7), we deduce, since $K_k^T A K_k$ is non singular, that

$$(A\epsilon_k, \epsilon_k) = (r_0, \epsilon_0) - (r_0, K_k (K_k^T A K_k)^{-1} K_k^T r_0). \tag{10}$$

We observe that the right hand side of equation (10) is the Schur complement of $K_k^T A K_k$ in the matrix $K_{k+1}^T A^{-1} K_{k+1}$, where

$$K_{k+1}^T A^{-1} K_{k+1} = \begin{pmatrix} r_0^T \epsilon_0 & r_0^T K_k \\ K_k^T r_0 & K_k^T A K_k \end{pmatrix}.$$

This is obtained since $K_{k+1} = [r_0 \; AK_k]$ and $A^{-1} K_{k+1} = [A^{-1} r_0 \; K_k]$. We can factor this matrix into a product of a block upper and a block lower triangular matrix (block UL factorization)

$$K_{k+1}^T A^{-1} K_{k+1} = \begin{pmatrix} 1 & r_0^T K_k (K_k^T A K_k)^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} (A\epsilon_k, \epsilon_k) & 0 \\ K_k^T r_0 & K_k^T A K_k \end{pmatrix}.$$

Taking determinants on both sides yields the formula that gives $\|\epsilon_k\|_A^2$ as the ratio of two determinants.

From the block UL factorization of the matrix $K_{k+1}^T A^{-1} K_{k+1}$, we deduce that

$$e_1^T (K_{k+1}^T A^{-1} K_{k+1})^{-1} e_1 = \frac{1}{(A\epsilon_k, \epsilon_k)}.$$

■

Corollary 1.

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_0\|_A^2} = \frac{1}{e_1^T (K_{k+1}^T\, A^{-1}\, K_{k+1}) e_1 \; e_1^T (K_{k+1}^T A^{-1} K_{k+1})^{-1} e_1}.$$

Proof. Since $\epsilon_0 = A^{-1}\, K_{k+1}\, e_1$, we have $(A\epsilon_0, \epsilon_0) = (\epsilon_0, r_0) = e_1^T\, K_{k+1}^T\, A^{-1} K_{k+1}\, e_1$ and this proves the result. ■

This leads to the following lower bound on the norm of the error.

Theorem 2. *We have*

$$1 > \frac{\|\epsilon_k\|_A}{\|\epsilon_0\|_A} \geq \frac{2\,\sqrt{\kappa(K_{k+1}^T A^{-1} K_{k+1})}}{\kappa(K_{k+1}^T A^{-1} K_{k+1}) + 1},$$

where κ denotes the condition number.

Proof. The Kantorovich inequality [21] says that for a vector y and a nonsingular matrix B, we have

$$\frac{(By, y)(B^{-1}y, y)}{(y, y)^2} \leq \frac{1}{4} \left(\sqrt{\kappa(B)} + \frac{1}{\sqrt{\kappa(B)}} \right)^2.$$

Using this with $y = e_1$ and $B = K_{k+1}^T A^{-1} K_{k+1}$, we obtain the lower bound. ■

The lower bound in the last result shows that there is no convergence of the CG algorithm as long as the matrix $K_{k+1}^T A^{-1} K_{k+1}$ is well-conditioned.

3. The Lanczos algorithm

In this section, we recall some basic facts about the Lanczos algorithm and its relation to CG, and we introduce new approximations to the eigenvalues of A.

3.1. The relationship between CG and Lanczos algorithms

It is well known (see, for instance [11]) that the CG algorithm is equivalent the Lanczos algorithm which generates a sequence of $n \times k$ matrices V_k whose columns are the Lanczos vectors v_i, $i = 1, \ldots, k$. These vectors are recursively constructed using a three-term recurrence

$$\eta_{j+1} v_{j+1} = A v_j - \alpha_j v_j - \eta_j v_{j-1}.$$

The coefficients α_j and η_j are defined to obtain mutually orthonormal basis vectors for the Krylov subspace $\mathcal{K}_k$. They define tridiagonal matrices T_k

$$T_k = \begin{pmatrix} \alpha_1 & \eta_2 & & & \\ \eta_2 & \alpha_2 & \eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{k-1} & \alpha_{k-1} & \eta_k \\ & & & \eta_k & \alpha_k \end{pmatrix}.$$

Using these notations we have the following well–known properties

$$V_k^T V_k = I_k, \quad \text{where} \quad V_k \equiv [v_1, \dots, v_k],$$

and we have the matrix relation

$$A\,V_k = V_k\,T_k + \eta_{k+1}\,v_{k+1}\,e_k^T. \tag{11}$$

Multiplying relation (11) by V_k^T implies that $T_k = V_k^T A V_k$. The Lanczos algorithm can be used to solve linear systems by defining iterates $x_k = x_0 + V_k y_k$. The coefficients in the vector $y_k \in \mathbb{R}^k$ are computed by requiring orthogonality of the corresponding residuals. They are obtained by solving a tridiagonal linear system

$$T_k y_k = \|r_0\| e_1.$$

The relationship between CG and the Lanczos algorithms is given in the following theorem, see for instance [11].

Theorem 3. *If x_0 and v_1 with $\|v_1\| = 1$ are such that $r_0 = b - Ax_0 = \|r_0\| v_1$ the Lanczos algorithm started from v_1 generates the same iterates as the CG algorithm started from x_0 when solving the linear system $Ax = b$ with A SPD and we have the following relations between the coefficients of the two algorithms*

$$\alpha_k = \frac{1}{\gamma_{k-1}} + \frac{\beta_{k-1}}{\gamma_{k-2}}, \quad \beta_0 = 0, \quad \gamma_{-1} = 1,$$

$$\eta_{k+1} = \frac{\sqrt{\beta_k}}{\gamma_{k-1}}.$$

The Lanczos vectors are related to the CG residual vectors by

$$v_{k+1} = (-1)^k \frac{r_k}{\|r_k\|}.$$

As seen previously, the A-norm of the CG error shows the important role played by the matrix $(K_k^T A^{-1} K_k)^{-1}$. Using the QR factorization of the Krylov matrix K_k, we obtain $K_k = V_k\,R_k$, where $V_k^T\,V_k = I_k$ and R_k is an upper triangular matrix. This orthonormal matrix V_k is the same as the one constructed by the Lanczos algorithm, see [11]. Consequently

$$(K_k^T A^{-1} K_k)^{-1} = R_k^{-1} (V_k^T\,A^{-1}\,V_k)^{-1}\,R_k^{-T} = R_k^{-1} \widehat{T_k}\,R_k^{-T},$$

where $\widehat{T}_k$ is defined as

$$\widehat{T_k} = (V_k^T\,A^{-1}\,V_k)^{-1}.$$

3.2. Properties of the matrix $\widehat{T}_k$

We will now study the interesting properties of the matrix $\widehat{T}_k$. We will first prove that this matrix is tridiagonal. We will also show that $\widehat{T}_k$ is nothing but the matrix T_k except for the (k, k) diagonal element, that is $\widehat{T}_k$ is a rank-one modification of T_k. We will prove that its eigenvalues, which in the sequel will be called the Galerkin values (since relation (2.2) is a Galerkin condition) are approximations to the eigenvalues of the matrix A. The eigenvalues of $\widehat{T}_k$ behave as those of the Lanczos matrix T_k known as the Ritz values. Some interlacing relations between both sets of approximations will be given.

The Galerkin values are also related to the harmonic Ritz values [13]. In fact, they are equal to the harmonic Ritz values when the first Lanczos vector v_1 is chosen as $v_1 = Av, v \in \mathbb{R}^n$, but otherwise they are different.

To prove the next theorem we need the following lemma which is proved in Zhang [21, p. 207].

Lemma 1. *Let U be an orthogonal matrix. If the eigenvalues of the SPD matrix A are ordered such that $\lambda_n \leq \ldots \leq \lambda_1$ then $\forall y \in \mathbb{R}^n$, we have*

1. $0 \leq y^T(U^T\, A\, U)\, y - y^T(U^T\, A^{-1}\, U)^{-1}\, y \leq (\sqrt{\lambda_1} - \sqrt{\lambda_n})^2,$
2. $0 \leq y^T(U^T\, A^2\, U)\, y - y^T(U^T\, A\, U)^2\, y \leq \dfrac{(\lambda_1 - \lambda_n)^2}{4}.$

In the following result we characterize the matrix $\widehat{T}_k$. Similar results were proved differently in [9] and also in [20] for the harmonic Ritz values.

Theorem 4. *Let $\lambda_n, \ldots, \lambda_1$ be the eigenvalues of the matrix A arranged as in Lemma 1, then*

$$\widehat{T}_k = T_k - \tau_k e_k\, e_k^T, \tag{12}$$

where τ_k is a positive real element such that

$$0 \leq \tau_k \leq (\sqrt{\lambda_1} - \sqrt{\lambda_n})^2.$$

Proof. Invoking relation (11) we deduce that

$$\begin{aligned}
I_k &= V_k^T A^{-1} V_k T_k + \eta_{k+1} V_k^T A^{-1} v_{k+1} e_k^T,\\
\widehat{T}_k &= T_k + \eta_{k+1} \widehat{T}_k V_k^T A^{-1} v_{k+1} e_k^T,\\
\widehat{T}_k &= T_k + u_k\, e_k^T,
\end{aligned}$$

where $u_k = \eta_{k+1} \widehat{T}_k V_k^T A^{-1} v_{k+1}$. Since $\widehat{T}_k$ and T_k are symmetric, it is obvious that $u_k\, e_k^T$ is also symmetric. Hence $u_k\, e_k^T = \tau_k e_k\, e_k^T$ and $\widehat{T}_k$ is tridiagonal.

In Lemma 1, we set, $U = V_k$ and $y = e_k$ to obtain the second part of the theorem. ■

Theorem 4 shows that the only unknown parameter of $\widehat{T}_k$ is τ_k. Indeed we have

$$\widehat{T}_k = (V_k^T\, A^{-1}\, V_k)^{-1} = \begin{pmatrix}
\alpha_1 & \eta_2 & & & \\
\eta_2 & \alpha_2 & \eta_3 & & \\
 & \ddots & \ddots & \ddots & \\
 & & \eta_{k-1} & \alpha_{k-1} & \eta_k \\
 & & & \eta_k & \alpha_k - \tau_k
\end{pmatrix}.$$

Let $\theta_i^{(k)}$ and $\widehat{\theta}_i^{(k)}$ be respectively the eigenvalues of T_k (Ritz values) and $\widehat{T}_k$ (Galerkin values). We arrange them as

$\theta_k^{(k)} \leq \ldots \leq \theta_2^{(k)} \leq \theta_1^{(k)}$ and $\widehat{\theta}_k^{(k)} \leq \ldots \leq \widehat{\theta}_2^{(k)} \leq \widehat{\theta}_1^{(k)}$.

In the following theorem we give some interlacing properties relating the eigenvalues of the three matrices T_k, $\widehat{T}_k$ and A.

Theorem 5. *There exist nonnegative real numbers $m_1, \ldots, m_k$ such that*

$$\widehat{\theta}_i^{(k)} = \theta_i^{(k)} - \tau_k\, m_i, \quad \textit{for} \quad i = 1, \ldots, k, \tag{13}$$

with $m_i \geq 0$ and $\displaystyle\sum_{i=1}^{k} m_i = 1$. Moreover

$$\begin{array}{l}
1)\, \theta_{i+1}^{(k)} \leq \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)}, \quad i \in \{1, \ldots, k-1\}, \\
2)\, \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)} \leq \widehat{\theta}_{i-1}^{(k)}, \quad i \in \{2, \ldots, k\}, \\
3)\, \widehat{\theta}_i^{(k)} \leq \theta_{i-1}^{(k-1)} \leq \widehat{\theta}_{i-1}^{(k)}, \quad i \in \{2, \ldots, k\}, \\
4)\, \lambda_{i+n-k} \leq \widehat{\theta}_i^{(k)} \leq \theta_i^{(k)} \leq \lambda_i, \quad i \in \{1, \ldots, k\}.
\end{array}$$

Proof.

Since the matrix $\widehat{T}_k$ is obtained by perturbing the matrix T_k by a rank-one matrix $\tau_k e_k e_k^T$, with τ_k nonnegative, then from Theorem 8.1.5 of [4, p. 412], we deduce that

$$\widehat{\theta}_i^{(k)} \in \left[\theta_{i+1}^{(k)}, \theta_i^{(k)}\right], \quad i \in \{1, \ldots, k-1\}.$$

We have also $T_k = \widehat{T}_k + \tau_k e_k\, e_k^T$, then

$$\theta_i^{(k)} \in \left[\widehat{\theta}_i^{(k)}, \widehat{\theta}_{i-1}^{(k)}\right], \quad i \in \{2, \ldots, k\}.$$

and there exist nonnegative real numbers m_i, for $i = 1, \ldots, k$, such that

$$\widehat{\theta}_i^{(k)} = \theta_i^{(k)} - \tau_k\, m_i, \quad \text{for} \quad i = 1, \ldots, k, \tag{14}$$

The matrix T_{k-1} is the leading submatrix of order $(k-1)$ of $\widehat{T}_k$ obtained by deleting the last row and the last column. Hence by using the Cauchy interlacing theorem for eigenvalues [4, p. 411], we get

$$\widehat{\theta}_k^{(k)} \leq \theta_{k-1}^{(k-1)} \leq \widehat{\theta}_{k-1}^{(k)} \leq \cdots \leq \widehat{\theta}_2^{(k)} \leq \theta_1^{(k-1)} \leq \widehat{\theta}_1^{(k)}$$

Finally, for the last part, we use the relation $\widehat{T}_k^{-1} = V_k^T\, A^{-1}\, V_k$, which shows that the matrix $\widehat{T}_k^{-1}$ is a section of the matrix A^{-1} (see the last section of [13]). Then using [13] or Corollary 4.4 of [19, p. 198] we deduce that

$$\widehat{\theta}_i^{(k)} \in [\lambda_{n-k+i}, \lambda_i] \qquad \text{for} \qquad i = 1, \ldots, k,$$

which ends the proof. ■

4. New expressions for $\|\epsilon_k\|_A$ and $\|r_k\|$

When using CG, we are concerned with the A-norm of the error because it corresponds to the energy norm occurring in some problems arising from partial differential equations and also because this norm is minimized at each CG iteration. In this section, using results from sections 2 and 3, we will derive formulas for ratios of norms of errors and residuals. For the A-norm of the error we have the following result which appears to be new. It shows the importance of the parameter τ_k in CG convergence.

Theorem 6. *Let ϵ_{k-1} and ϵ_k be the errors obtained by the conjugate gradient method at steps $k-1$ and k respectively. We have*

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{1}{\det(V_k^T A V_k)\det(V_k^T A^{-1} V_k)} = 1 - \frac{\det(\widehat{T}_k)}{\det(T_k)} = 1 - \prod_{i=1}^{k} \frac{\widehat{\theta}_i^{(k)}}{\theta_i^{(k)}} = \tau_k\, e_k^T T_k^{-1} e_k. \quad (15)$$

Proof. Since $A^{-1}K_{k+1} = \left[\begin{array}{ccc} A^{-1}r_0, & K_{k-1}, & A^{k-1}r_0 \end{array}\right]$, we have,

$$K_{k+1}^T A^{-1} K_{k+1} = \begin{pmatrix} r_0^T A^{-1} r_0 & r_0^T K_{k-1} & r_0^T A^{k-1} r_0 \\ K_{k-1}^T r_0 & K_{k-1}^T A K_{k-1} & K_{k-1}^T A^k r_0 \\ r_0^T A^{k-1} r_0 & r_0^T A^k K_{k-1} & r_0^T A^{2k-1} r_0 \end{pmatrix}.$$

Applying Sylvester's identity (see [3] or [2]) to this matrix, we obtain

$$\begin{aligned} \det(K_{k+1}^T A^{-1} K_{k+1})\det(K_{k-1}^T A K_{k-1}) &= \begin{vmatrix} r_0^T A^{-1} r_0 & r_0^T K_{k-1} \\ K_{k-1}^T r_0 & K_{k-1}^T A K_{k-1} \end{vmatrix} \begin{vmatrix} K_{k-1}^T A K_{k-1} & K_{k-1}^T A^k r_0 \\ r_0^T A^k K_{k-1} & r_0^T A^{2k-1} r_0 \end{vmatrix} \\ &- \begin{vmatrix} r_0^T K_{k-1} & r_0^T A^{k-1} r_0 \\ K_{k-1}^T A K_{k-1} & K_{k-1}^T A^k r_0 \end{vmatrix} \begin{vmatrix} K_{k-1}^T r_0 & K_{k-1}^T A K_{k-1} \\ r_0^T A^{k-1} r_0 & r_0^T A^k K_{k-1} \end{vmatrix} \end{aligned}$$

We notice that the first matrix on the right hand side is $K_k^T A^{-1} K_k$ and the second one is $K_k^T A K_k$. Moreover, the third and fourth matrices are the transpose of each other and equal to $K_k^T K_k$. Therefore

$$\det(K_{k+1}^T A^{-1} K_{k+1})\ \det(K_{k-1}^T A K_{k-1}) = \det(K_k^T\ A^{-1}\ K_k)\ \det(K_k^T\ A\, K_k) - \det(K_k^T\ K_k)^2. \quad (16)$$

By using Theorem 1, the following result holds

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{\det(K_k^T K_k)^2}{\det(K_k^T A^{-1} K_k)\det(K_k^T A K_k)}. \quad (17)$$

By using the QR factorization of K_k which is $K_k = V_k R_k$, we obtain

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{1}{\det(V_k^T A V_k)\det(V_k^T A^{-1} V_k)} = 1 - \frac{\det(\widehat{T}_k)}{\det(T_k)} = 1 - \prod_{i=1}^{k} \frac{\widehat{\theta}_i^{(k)}}{\theta_i^{(k)}}. \quad (18)$$

Relation (12) also gives

$$\det(\widehat{T}_k) = (\alpha_k - \tau_k)\det(T_{k-1}) - \eta_k^2 \det(T_{k-2}).$$

By using the fact that $\det(T_k) = \alpha_k \det(T_{k-1}) - \eta_k^2 \det(T_{k-2})$, we deduce that

$$\det(\widehat{T}_k) = \det(T_k) - \tau_k \det(T_{k-1}),$$
$$\frac{\det(\widehat{T}_k)}{\det(T_k)} = 1 - \tau_k \frac{\det(T_{k-1})}{\det(T_k)}.$$

Using equation (18), we get

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = \tau_k \frac{\det(T_{k-1})}{\det(T_k)}.$$

On the other hand, since the matrix T_k has the following form

$$T_k = \begin{pmatrix} T_{k-1} & \eta_k e_{k-1} \\ \eta_k e_{k-1}^T & \alpha_k \end{pmatrix},$$

it follows from the Cholesky-like factorizations of T_{k-1} and T_k that

$$\frac{\det(T_{k-1})}{\det(T_k)} = e_k^T T_k^{-1} e_k,$$

which completes the proof. ■

Theorem 6provides an exact expression for the ratio of norms $\|\epsilon_k\|_A / \|\epsilon_{k-1}\|_A$. From this, bounds can be obtained which, unfortunately, are not optimal when $k > 1$.

Remarks:

1. If $k = 1$, by using the fact that

$$\frac{\|\epsilon_1\|_A^2}{\|\epsilon_0\|_A^2} = 1 - \frac{1}{(v_1^T A v_1)(v_1^T A^{-1} v_1)}$$

and the Kantorovich inequality, it is easy to show that an optimal bound is given by

$$\frac{\|\epsilon_1\|_A}{\|\epsilon_0\|_A} \leq \frac{1}{1 + 2\,\dfrac{\lambda_n}{\lambda_1 - \lambda_n}}.$$

2. Using Theorem 5, we deduce that

$$\frac{\|\epsilon_k\|_A}{\|\epsilon_{k-1}\|_A} \leq \sqrt{1 - \frac{\hat{\theta}_k^{(k)}}{\theta_1^{(k)}}} \leq \sqrt{1 - \frac{\lambda_n}{\lambda_1}}.$$

Notice that this last bound is weaker than the one obtained by using Chebyshev polynomials, see [1], [10]. We now consider the norm of the residual vector and the relationships between the error and residual norms. First, we recall a result proved by Sadok in [16].

Theorem 7. *Let r_{k-1} and r_k be the residuals obtained by the conjugate gradient method at steps $k-1$ and k respectively, we obtain*

$$\frac{\|r_k\|}{\|r_{k-1}\|} = \eta_{k+1} \frac{\det(T_{k-1})}{\det(T_k)} = \eta_{k+1}\, e_k^T T_k^{-1} e_k.$$

It is known that when the A-norm of the error decreases monotonically, the norms of the CG residuals may oscillate. Using the previous theorem we can give a bound for the (possible) increase of the residual.

Theorem 8. *Let r_{k-1} and r_k be the residuals obtained by the conjugate gradient method at steps $k-1$ and k respectively, we obtain*

$$\frac{\|r_k\|}{\|r_{k-1}\|} \leq \frac{\kappa(A)-1}{2},$$

where $\kappa(A) = \lambda_1/\lambda_n$ is the condition number of A.

Proof. Formula (11) can be rewritten as

$$A\,V_k = V_{k+1}\left(\begin{array}{c} T_k \\ \eta_{k+1}e_k^T \end{array}\right).$$

Multiplying this relation by its transpose we find that

$$V_k^T\,A^2\,V_k = (V_k^T A V_k)^2 + \eta_{k+1}^2 e_k e_k^T.$$

From the second part of Lemma 1, we deduce that

$$\eta_{k+1} \leq \frac{\lambda_1 - \lambda_n}{2}.$$

Then, by using the Courant-Fischer Minimax Theorem, Theorem 4 and Theorem 7, we conclude that

$$\frac{\|r_k\|}{\|r_{k-1}\|} \leq \frac{\lambda_1 - \lambda_n}{2\,\theta_k^{(k)}} \leq \frac{\lambda_1 - \lambda_n}{2\,\lambda_n}.$$

■

The following result relates the A-norm of the error to the norm of the residual. We will see that the inverse of the matrix $\widehat{T}_{k+1}$ plays a key role here.

Theorem 9. *Let r_k be the residual obtained at the k-th step, ϵ_k and ϵ_{k-1} be the CG errors obtained at the k-th and $(k-1)$-th steps respectively. Then*

1.

$$\|r_k\|^2 = \frac{\det(\widehat{T}_{k+1})}{\det(T_k)}\|\epsilon_k\|_A^2 = \frac{\|\epsilon_k\|_A^2}{e_{k+1}^T\,\widehat{T}_{k+1}^{-1}\,e_{k+1}}.$$

2.

$$\hat{\theta}_{k+1}^{(k+1)} \leq \frac{\|r_k\|^2}{\|\epsilon_k\|_A^2} \leq \hat{\theta}_1^{(k+1)}.$$

Proof. It was proved by Sadok in [16] that

$$\|r_k\|^2 = \frac{\det(K_k^T K_k)\det(K_{k+1}^T K_{k+1})}{\det(K_k^T A K_k)^2}.$$

From Theorem 1, we see that

$$\|r_k\|^2 = \frac{\det(K_k^T K_k)\det(K_{k+1}^T K_{k+1})}{\det(K_k^T A K_k)\det(K_{k+1}^T A^{-1} K_{k+1})}(A\epsilon_k, \epsilon_k).$$

Using once again the QR factorization of the Krylov matrix $K_k = V_k R_k$, we obtain

$$\|r_k\|^2 = \frac{\det((V_{k+1}^T A^{-1} V_{k+1})^{-1})}{\det(V_k^T A V_k)}(A\epsilon_k, \epsilon_k).$$

Consequently

$$\|r_k\|^2 = \frac{\det(\widehat{T}_{k+1})}{\det(T_k)}(A\epsilon_k, \epsilon_k).$$

The second part of the theorem is obtained by bounding $\det(\widehat{T}_{k+1})/\det(T_k)$. ■

The following theorem gives the ratio of norms of the residual and the error as a function of CG and Lanczos parameters and τ_{k+1}.

Theorem 10. *Let r_k be the residual and ϵ_k be the CG error obtained at the k-th step. Then*

$$\frac{\|r_k\|^2}{\|\epsilon_k\|_A^2} = \alpha_{k+1} - \eta_{k+1}\frac{\|r_k\|}{\|r_{k-1}\|} - \tau_{k+1} = \frac{1}{\gamma_k} - \tau_{k+1}.$$

Proof. Using formula (18), we have

$$\frac{\det(\widehat{T}_{k+1})}{\det(T_k)} = \alpha_{k+1} - \tau_{k+1} - \eta_{k+1}^2\frac{\det(T_{k-1})}{\det(T_k)}.$$

The assertion follows from Theorem 9. ■

Finally, we relate our results which have been obtained using Krylov matrices to a formula for the difference of the squares of the A-norm of the error in successive iterations first proved in the seminal paper of Hestenes and Stiefel [5].

Theorem 11. *Let r_k be the residual obtained at the k-th step, ϵ_k and ϵ_{k-1} the CG errors obtained at k-th and $(k-1)$-th steps respectively. Then*

$$\|\epsilon_{k-1}\|_A^2 - \|\epsilon_k\|_A^2 = \gamma_{k-1}\|r_{k-1}\|^2 = \frac{\|r_{k-1}\|.\|r_k\|}{\eta_{k+1}},$$

with $\gamma_{k-1} = \dfrac{\det(T_{k-1})}{\det(T_k)}$, one of the parameters computed in CG.

Proof. By using equation (16), we have

$$\|\epsilon_{k-1}\|_A^2 - \|\epsilon_k\|_A^2 = \frac{\det(K_k^T K_k)^2}{\det(K_{k-1}^T A K_{k-1})\det(K_k^T A K_k)} = \gamma_{k-1}\|r_{k-1}\|^2,$$

where $\gamma_{k-1} = \dfrac{\det(K_k^T K_k)\det(K_{k-1}^T A K_{k-1})}{\det(K_k^T A K_k)\det(K_{k-1}^T K_{k-1})} = \dfrac{\det(T_{k-1})}{\det(T_k)}$. ■

5. Conclusion

In this paper we have established new expressions for the A-norm of the error and the norm of the residual for the CG algorithm by using the relationships between CG and the Lanczos algorithm. We have shown that $\widehat{T}_k = (V_k^T A^{-1} V_k)^{-1}$ is a tridiagonal matrix and a rank-one modification of the Lanczos matrix T_k. This modification is characterized by an important parameter τ_k which is involved in the ratio of A-norms of the error at successive CG iterations and in the ratio of the norm of the residual to the A-norm of the error.

It remains to be seen if one can compute τ_k or, at least, compute good approximations of it, during CG iterations. This will be considered in a forthcoming paper. It could lead to complementing the bounds of the A-norm of the error that can be cheaply computed using Gauss quadrature, see [17, 18, 11] for a summary of these techniques.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referees for their valuable comments and suggestions.

REFERENCES

1. O. Axelsson, *Iterative solution methods*, Cambridge University Press, 1994.
2. C. Brezinski, *Biorthogonality and its applications to numerical analysis*, Marcel Dekker, 1992.
3. F.R. Gantmacher, *The theory of matrices*, vol. 1, Chelsea, New York, 1959.
4. G.H. Golub and C. F. Van Loan, *Matrix computations*, Third Edition 1996, The Johns Hopkins University Press.
5. M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, 1952; 49:409–436.
6. M.R. Hestenes, *The conjugate gradient method for solving linear systems*, Proc. Symposia in Appl. Math., vol VI, Numerical Analysis, Mc Graw-Hill, New-York, 1956; 83–102.
7. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, Journal of Research the National Bureau of Standards, 1950,; 45:255–282.
8. C. Lanczos, *Solution of systems of linear equations by minimized iterations*, Journal of Research the National Bureau of Standards, 1952; 49:33–53.
9. C.T. Lenard, *Rank-1 perturbations and the Lanczos method, inverse iteration, and Krylov subspaces*, Journal of the Australian Mathematical Society, Series B, 1995; 36(4):381–388
10. G. Meurant, *Computer solution of large linear systems*, North-Holland, 1999.
11. G. Meurant, *The Lanczos and conjugate gradient algorithms, from theory to finite precision computations*, SIAM, 2006.
12. G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numerica v 15, Cambridge University Press, 2006; 471–542.
13. C.C. Paige, B.N. Parlett and H.A. Van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numerical Linear Algebra with Applications, 1995; 2:115–134.
14. C.C. Paige and M.A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM Journal of Numerical Analysis, 1975; 12:617–629.
15. Y. Saad, *Krylov subspace methods for solving large unsymmetric linear systems*, Mathematics of Computation, 1981; 37:105–126.
16. H. Sadok, *Analysis of the convergence of the minimal and the orthogonal residual methods*, Numerical Algorithms, 2005; 40:201–216.
17. Z. Strakoš and P. Tichý, *On Error estimation in the conjugate gradient method and why it works in finite precision computations*, Electronic Transactions on Numerical Analysis, 2002; 13:56–80.
18. Z. Strakoš and P. Tichý, *Error estimation in preconditioned conjugate gradients*, BIT Numerical Mathematics, 2005; 45:789–817.
19. G.W. Stewart and J.G. Sun, *Matrix perturbation theory*, Academic Press, New York, 1990.
20. H.A. Van Der Vorst, *Iterative Krylov methods for large linear systems*, Cambridge University Press, Cambridge, 2003.
21. F. Zhang, *Matrix theory*, Springer, N.Y. 1999.