

# NEW RESULTS ON THE CONVERGENCE OF THE CONJUGATE GRADIENT METHOD

R. BOUYOULI\*, G. MEURANT†, L. SMOCH‡, AND H. SADOK §

**Abstract.** This paper is concerned with proving theoretical results related to the convergence of the Conjugate Gradient method for solving positive definite symmetric linear systems. 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 [13]. These results use the well-known correspondence between the Conjugate Gradient method and the Lanczos algorithm.

**Key words:** symmetric linear systems, Conjugate Gradient method, Krylov subspace, Ritz values.

**1. Introduction.** The conjugate gradient (CG) method of Hestenes and Stiefel [3, 4] was originally developed in the early 1950s for solving a linear system of equations

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^n \quad (1.1)$$

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

Let  $x_0$  be an initial approximate solution of (1.1), the CG method first computes the initial residual  $r_0 = b - Ax_0$  and generates a sequence of approximate solutions  $x_1, x_2, \dots$  such that the residual vector  $r_i = b - Ax_i$  can be written in the form

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

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

The CG polynomial  $P_i$  is chosen 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 as  $\|y\|_A = (y^T A y)^{\frac{1}{2}}$ . The  $A$ -norm of the error is therefore given by

$$\|\epsilon_i\|_A = \min_{P_i \in \pi_i} \|P_i(A)\epsilon_0\|_A. \quad (1.2)$$

In exact arithmetic, the residuals  $r_i$  obtained by the CG method are orthogonal.

The remainder of this paper is organized as follows. In section 2, we describe the Petrov-Galerkin orthogonality conditions that define the CG iterates. Using the relationship between CG and the Lanczos algorithm [5], [6], this leads to obtaining some new relations for the error  $A$ -norm and the residual norm, depending essentially on Krylov matrices whose columns are the vectors of the natural basis of the Krylov

---

\*Université Mohammed V, Faculté des sciences, département de Mathématiques, Rabat, Maroc.

† CEA/DIF, BP12, 91680 Bruyères-le-Chatel, France. E-mail : [gerard.meurant@cea.fr](mailto:gerard.meurant@cea.fr).

‡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](mailto:smoch@lmpa.univ-littoral.fr).

§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: [sadok@lmpa.univ-littoral.fr](mailto:sadok@lmpa.univ-littoral.fr).

subspace based on  $A$  and the initial residual. In section 3, 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 will assume exact arithmetic. For a summary of results when using CG in finite precision arithmetic, see [8] or [9]. Throughout this paper,  $e_j$  will stand for the  $j$ -th vector of the canonical basis.

**2. The Conjugate Gradient method.** Let us consider the linear system (1.1) with the SPD matrix  $A$ . Let  $v \in \mathbb{R}^n$  and  $\mathcal{K}_k(A, v) \equiv \text{span}\{v, Av, \dots, A^{k-1}v\}$  be the Krylov subspace constructed from  $A$  and  $v$ . According to (1.2), the iterates  $x_k$  are defined by

$$x_k - x_0 \in \mathcal{K}_k(A, r_0) \equiv \mathcal{K}_k, \quad (2.1)$$

and the so-called Petrov-Galerkin orthogonality conditions

$$r_k = b - Ax_k = A\epsilon_k \perp \mathcal{K}_k. \quad (2.2)$$

It follows from (2.1) that

$$\epsilon_k = \epsilon_0 - \sum_{i=1}^k a_i A^{i-1} r_0. \quad (2.3)$$

where  $a_i \in \mathbb{R}$ ,  $1 \leq i \leq k$ . The orthogonality condition (2.2) can be written as

$$(A^i r_0)^T A \epsilon_k = 0, \quad 0 \leq i \leq k-1. \quad (2.4)$$

In an equivalent matrix form, using the Krylov matrix

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

the orthogonality condition writes as

$$(K_k^T A K_k) a = K_k^T r_0, \quad (2.5)$$

with  $a = (a_1, a_2, \dots, a_k)^T$ . Of course, in practice,  $x_k$  is not computed by solving the system giving  $a$  at each iteration  $k$ . The most usual form of the Conjugate Gradient algorithm (see for instance [2] or [8]) is given as:

#### Conjugate Gradient Algorithm

- **Init:**  $r_0 := b - Ax_0$ ,  $p_0 := r_0$ .
- **Iterate:** Until convergence do,
  1.  $\gamma_j := (r_j, r_j) / (Ap_j, p_j)$
  2.  $x_{j+1} := x_j + \gamma_j p_j$
  3.  $r_{j+1} := r_j - \gamma_j Ap_j$
  4.  $\beta_{j+1} := (r_{j+1}, r_{j+1}) / (r_j, r_j)$
  5.  $p_{j+1} := r_{j+1} + \beta_{j+1} p_j$

This algorithm needs only a matrix-vector product, vector additions and two inner products per iteration. In the following, we derive expressions for the  $A$ -norm of the error  $\|\epsilon_k\|_A$  and the norm of the residual  $\|r_k\|$ .

**2.1. The norms  $\|\epsilon_k\|_A$  and  $\|r_k\|$ .** We start by giving expressions of  $\|\epsilon_k\|_A$  involving the Krylov matrix  $K_k$ .

**THEOREM 2.1.** *Let  $\epsilon_k = x - x_k$  be the Conjugate Gradient error. Then, if the matrices  $K_k^T AK_k$  and  $K_{k+1}^T A^{-1} K_{k+1}$  are nonsingular,*

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

*Proof.* Since  $A\epsilon_k = r_k$ , it follows from (2.3) that

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

By using condition (2.4), we obtain

$$\begin{aligned} (\epsilon_k, A\epsilon_k) &= (r_k, \epsilon_0) \\ &= (r_0, \epsilon_0) - \sum_{i=1}^k a_i (A^{i-1} r_0, r_0) \\ &= (r_0, \epsilon_0) - (r_0, K_k a). \end{aligned} \tag{2.6}$$

where  $a$  is given by solving the system (2.5), that is  $a = (K_k^T AK_k)^{-1} K_k^T r_0$ . We observe that the right hand side of (2.6) is a Schur complement for the matrix

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

Cramer's rule gives that

$$\|\epsilon_k\|_A^2 = \det \begin{pmatrix} r_0^T \epsilon_0 & r_0^T K_k \\ K_k^T r_0 & K_k^T AK_k \end{pmatrix} / \det(K_k^T AK_k),$$

where  $\det(B)$  stands for the determinant of the matrix  $B$ .

Note that  $K_{k+1} = [ r_0, AK_k ]$  and  $A^{-1} K_{k+1} = [ \epsilon_0, K_k ]$ . Therefore

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

Using the hypothesis that  $K_{k+1}^T A^{-1} K_{k+1}$  is nonsingular and by using Cramer's rule for computing the (1, 1) element of the inverse of  $K_{k+1}^T A^{-1} K_{k+1}$ , the following result holds

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

□

Since  $\epsilon_0 = A^{-1} K_{k+1} e_1$  we have  $(\epsilon_0, A\epsilon_0) = (\epsilon_0, r_0) = e_1^T K_{k+1}^T A^{-1} K_{k+1} e_1$ . Consequently

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

By using the Kantorovich inequality [16] we obtain the following lower bound.

THEOREM 2.2.

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

where  $\kappa$  denotes the condition number.

This last result shows that there is no CG convergence as long as the matrix  $K_{k+1}^T A^{-1} K_{k+1}$  is well-conditioned.

It is well known that the CG algorithm has a close relationship with the Lanczos algorithm which is the following:

**Lanczos Algorithm**

- Set  $v_1 = \frac{r_0}{\|r_0\|}$ ,  $v_0 \equiv 0$
- For  $j = 1, 2, \dots, k$ :
  1.  $w_j := Av_j - \eta_j v_{j-1}$
  2.  $\alpha_j := (w_j, v_j)$
  3.  $w_j := w_j - \alpha_j v_j$
  4.  $\eta_{j+1} := \|w_j\|_2$ .
  5. If  $\eta_{j+1} = 0$  then Stop
  6.  $v_{j+1} := w_j / \eta_{j+1}$ .

The Lanczos algorithm generates matrices  $V_j$  whose columns are the Lanczos vectors  $v_i$ ,  $i = 1, \dots, j$  and a tridiagonal matrix  $T_k$ ,

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

It is easy to see that we have the following well known properties

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

that is  $V_k$  is an orthonormal matrix and

$$A V_k = V_k T_k + \eta_{k+1} v_{k+1} e_k^T. \quad (2.7)$$

Multiplying relation (2.7) 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  $y_k$  are computed by requiring orthogonality of the residuals. They are obtained by solving

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

The relationship between CG and the Lanczos algorithm is given in the following theorem, see for instance [8].

THEOREM 2.3. *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*

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

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

and the Lanczos vectors are related to the CG residuals by

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

As seen previously, the  $A$ -norm of the error shows the important role played by the matrix  $(K_k^T A^{-1} K_k)^{-1}$ . Using the  $QR$  decomposition 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 in the Lanczos algorithm, see [8]. Consequently

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

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

$$\hat{T}_k = (V_k^T A^{-1} V_k)^{-1}.$$

In the next section we will study the interesting properties of the matrix  $\hat{T}_k$ . We will prove that its eigenvalues, which in the sequel will be called the Petrov-Galerkin values (since (2.2) is a Petrov-Galerkin condition) are approximations to the eigenvalues of the matrix  $A$ , as those of the Lanczos matrix  $T_k$  which are known as Ritz values.

**2.2. Properties of the matrix  $\hat{T}_k$ .** In this section we study the structure of the matrix  $\hat{T}_k$ . We will first prove that this matrix is tridiagonal. We will also show that  $\hat{T}_k$  is nothing but the matrix  $T_k$  except for the  $(k, k)$  diagonal element, that is  $\hat{T}_k$  is a rank-one modification of  $T_k$ . The eigenvalues of  $\hat{T}_k$  behave as the Ritz values. Some interlacing relations between both sets of approximations will be given.

To prove the next theorem we need the following lemma which is proved in Zhang [16].

LEMMA 2.4. *Let  $U$  be an orthogonal matrix. If the eigenvalues of the SPD matrix  $A$  are ordered such that  $\lambda_n \leq \dots \leq \lambda_1$  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.  $0 \leq y^T (U^T A^2 U) y - y^T (U^T A U)^2 y \leq \frac{(\lambda_1 - \lambda_n)^2}{4}$ .

In the following result we characterize the matrix  $\hat{T}_k$ .

THEOREM 2.5. *Let  $\lambda_n, \dots, \lambda_1$  be the eigenvalues of the matrix  $A$  arranged as in Lemma 2.4, then*

$$\hat{T}_k = T_k - \tau_k e_k e_k^T, \tag{2.8}$$

where  $\tau_k$  is a positive real element such that

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

*Proof.* Invoking relation (2.7) we deduce that

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

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 2.4, we set,  $U = V_k$  and  $y = e_k$  to obtain the second part of the theorem.  $\square$

Theorem 2.5 shows that the only unknown parameter of  $\widehat{T}_k$  is  $\tau_k$ . 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 the eigenvalues of  $T_k$  (Ritz values) and  $\widehat{T}_k$  (Petrov-Galerkin values) respectively. We arrange them as

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

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

**THEOREM 2.6.** *There exist nonnegative real numbers  $m_1, \dots, m_k$  such that*

$$\widehat{\theta}_i^{(k)} = \theta_i^{(k)} - \tau_k m_i$$

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

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

*Proof.*

1) and 2). The matrix  $\widehat{T}_k$  is obtained by perturbing the matrix  $T_k$  by a rank-one matrix. From theorem 8.1.5 of [2, p. 412], we obtain the first two statements.

3) The matrix  $T_{k-1}$  is a square submatrix of order  $(k-1)$  of  $\widehat{T}_k$  obtained by deleting the last row and the last column. Hence by using the interlacing Cauchy theorem for eigenvalues [2, p. 411], we get the inequalities given in 3).

4) Finally, for the last part, we use the relation  $\widehat{T}_k^{-1} = V_k A^{-1} V_k$  and Corollary 4.4 of [14, p. 198] to deduce that

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

which completes the proof.  $\square$

**3. Relationship between  $\|r_k\|$  and  $\|\epsilon_k\|_A$ .** 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. We have the following result which appears to be new.

**THEOREM 3.1.** *Let  $\epsilon_{k-1}$  and  $\epsilon_k$  be the errors obtained by the Conjugate Gradient method at steps  $k-1$  and  $k$  respectively. We have*

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{1}{\det(V_k^T A V_k) \det(V_k^T A^{-1} V_k)} = 1 - \frac{\det(\widehat{T}_k)}{\det(T_k)} = 1 - \prod_{i=1}^k \frac{\widehat{\theta}_i^{(k)}}{\theta_i^{(k)}} = \tau_k e_k^T T_k^{-1} e_k. \quad (3.1)$$

*Proof.* Since  $A^{-1}K_{k+1} = [A^{-1}r_0, K_{k-1}, A^{k-1}r_0]$ , we deduce that,

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

Applying Sylvester's identity [1] to the matrix  $K_{k+1}^T A^{-1} K_{k+1}$ , we obtain

$$\det(K_{k+1}^T A^{-1} K_{k+1}) \det(K_{k-1}^T A K_{k-1}) = \det(K_k^T A^{-1} K_k) \det(K_k^T A K_k) - \det(K_k^T K_k)^2. \quad (3.2)$$

By using Theorem 2.1, the following result holds

$$\frac{\|\epsilon_k\|_A^2}{\|\epsilon_{k-1}\|_A^2} = 1 - \frac{\det(K_k^T K_k)^2}{\det(K_k^T A^{-1} K_k) \det(K_k^T A K_k)}. \quad (3.3)$$

By using the QR-factorization of  $K_k$ , we obtain

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

Relation (2.8) also gives

$$\det(\widehat{T}_k) = (\alpha_k - \tau_k) \det(T_{k-1}) - \eta_k^2 \det(T_{k-2}). \quad (3.4)$$

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

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

The proof is completed by using (3.3).  $\square$

**Remarks:**

1. If  $k = 1$ , it is easy to show, 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, that an optimal bound is given by

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

2. Using Theorem 2.6, we deduce that

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

Now, we recall a result proved by Sadok in [13].

**THEOREM 3.2.** *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*

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

Using this theorem we can give a new bound for the (possible) increase of the residual.

**THEOREM 3.3.** *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 (2.7) can be rewritten as

$$A V_k = V_{k+1} \begin{pmatrix} T_k \\ \eta_{k+1} e_k^T \end{pmatrix}.$$

Multiplying this relation by its transpose we find

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

Using the second part of Lemma 2.4, we deduce that

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

then, by the Courant-Fischer Minimax Theorem, Theorem 2.5 and Theorem 3.2, we conclude that

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

□

By considering Theorem 3.1 and Theorem 3.2, we have

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

The following result relates the  $A$ -norm of the error to the norm of the residual.

**THEOREM 3.4.** *Let  $r_k$  be the residual obtained at the  $k$ -th step,  $\epsilon_k$  and  $\epsilon_{k-1}$  the CG errors obtained at  $k$ -th and  $(k-1)$ -th steps respectively. Then*

1.

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



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 [13] 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 2.1, we see that

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

Using the  $QR$  decomposition of the Krylov matrix  $K_k = V_k R_k$ , we obtain

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

Consequently

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

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

**THEOREM 3.5.** *Let  $r_k$  be the residual and  $\epsilon_k$  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 (3.4), we have

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

The assertion follows from Theorem 3.4.  $\square$

Finally, we relate our results using Krylov matrices to a formula for difference of the squares of the  $A$ -norm of the error in successive iterations proved in Hestenes and Stiefel [3].

**THEOREM 3.6.** *Let  $r_k$  be the residual obtained at the  $k$ -th step,  $\epsilon_k$  and  $\epsilon_{k-1}$  the CG errors obtained at  $k$ -th and  $(k-1)$ -th steps respectively. Then*

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

with  $\gamma_{k-1} = \frac{\det(T_{k-1})}{\det(T_k)}$ , the parameter computed in CG.

*Proof.* By using (3.2), we deduce

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

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

**4. 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. We have shown that  $\hat{T}_k = (V_k^T A^{-1} V_k)^{-1}$  is a tridiagonal matrix and a rank-one modification of the Lanczos matrix  $T_k$ . This modification is characterized by an important parameter  $\tau_k$  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 we can, at least, compute good approximations of the parameter  $\tau_k$  during CG iterations. This will be considered in a forthcoming paper. It could lead to complementing the bounds on the  $A$ -norm of the error that can be cheaply obtained using Gauss quadrature, see [8] for a summary of these techniques.

#### REFERENCES

- [1] F. R. GANTMACHER, *The theory of Matrices*, vol. 1, Chelsea, New York, 1959.
- [2] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Third Edition 1996, The Johns Hopkins University Press.
- [3] M. R. HESTENES AND E. STIEFEL, *Methods of Conjugate Gradients for solving linear systems*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409-436.
- [4] 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, pp. 83-102.
- [5] C. LANCZOS, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, v 45, 1950, pp. 255-282.
- [6] C. LANCZOS, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bur. Standards, v 49, 1952, pp. 33-53.
- [7] G. MEURANT, *Computer solution of large linear systems*, North-Holland, 1999.
- [8] G. MEURANT, *The Lanczos and Conjugate Gradient algorithms, from theory to finite precision computations*, SIAM, 2006.
- [9] G. MEURANT AND Z. STRAKÖS, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numerica v 15, Cambridge University Press, 2006, pp. 471-542.
- [10] C. C. PAIGE, B. N. PARLETT AND H. A. VAN DER VORST, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Num. Lin. Alg. with Appl., 2 1995, pp. 115-134.
- [11] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. 12, 1975, pp. 617-629.
- [12] Y. SAAD, *Krylov subspace methods for solving large unsymmetric linear systems*, Math. Comp. 37, 1981, pp. 105-126.
- [13] H. SADOK, *Analysis of the convergence of the minimal and the orthogonal residual methods*, Numer. Algorithms 40, 2005, pp. 101-115.
- [14] G. W. STEWART AND J. G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.
- [15] H. A. VAN DER VORST, *Iterative Krylov Methods for Large Linear Systems*, Cambridge university press, Cambridge, 2003.
- [16] F. ZHANG, *Matrix Theory*, Springer, N.Y. 1999.