Altman's methods revisited

C. Roland*, B. Beckermann* and C. Brezinski *

with a biography of M. Altman by T. Altman †

Abstract

In this paper we discuss two different methods of Altman for solving systems of linear equations. These methods can be considered as Krylov subspace type methods for solving a projected counterpart of the original system. We discuss the link to classical Krylov subspace methods, and give some theoretical and numerical results on their convergence behavior. Also, some historical remarks on the work of Altman are presented.

AMS subject classification: 65F10, 65E05.
Key words: Conjugate gradient method, Krylov subspace methods.

1 Introduction

Consider a system of linear equations

$$Ax = b, \tag{1}$$

where A is supposed to be hermitian positive definite of order p, and (without loss of generality) the right-hand side is supposed to be of euclidean norm $||b|| = 1$. In a series of papers [1, 2, 3, 4, 5, 6], Altman considers the associated problem

$$Ay = (Ay, b)b \tag{2}$$

which, with help of the orthogonal projector $P = I - bb^*$ onto the orthogononal complement of b, can be equivalently written as $PAy = 0$. Notice that the set of solutions of (2) is given by the set of scalar multiples of $A^{-1}b$, and hence, given any non-trivial solution y of (2), a solution of (1) is given by $x = y/(Ay, b)$. Corresponding to (2), Altman also considers the linear operator

$$r(y) := PAy = -P(b - Ay),$$

coinciding up to a sign with the projected residual of (1).

*Laboratoire de Mathématiques Appliquées (ANO), FRE CNRS 2222, UFR de Mathématiques Pures et Appliquées – M3, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq CEDEX, France, e-mail: {roland,bbecker,brezinsk}@ano.univ-lille1.fr

†Department of Computer Science and Engineering, University of Colorado at Denver, Campus Box 109, P.O. Box 173364, Denver, CO 80217-3364, USA.

In the above-mentioned papers, Altman proposes essentially two iterative methods giving approximate non-trivial solutions of (2) (and hence of (1) after normalization). Given an y_0 with $(y_0, b) \neq 0$, the first method presented in [1, Eqns (5) and (6)] and further analyzed and generalized in [6] minimizes the norm of the projected residual

$$y_{n+1} = y_n + \alpha_n r(y_n), \quad \alpha_n = \arg\min\{||r(y_n + \alpha r(y_n))|| : \alpha \in \mathbb{R}\}, \tag{3}$$

leading to a generalized method of minimal residuals. In the second method (see also [10, pp. 132-139]), Altman considers a quadratic form [2, Eqn. (7)] closely related to the quadratic form

$$G(y) := \left(\left(A - \frac{bb^*}{(A^{-1}b, b)} \right) y, y \right) \tag{4}$$

(Altman's quadratic form F involves r and its inverse, defined on the orthogonal complement of b, see the appendix A for further details). It is not difficult to check using formula (13) below that $G(y) \geq 0$, and $G(y) = 0$ if and only if y is a solution of (2). Then the so-called generalized steepest descent method introduced in [2, Eqns (10) and (11)] and further analyzed in [5] is given by

$$y_{n+1} = y_n + \alpha_n r(y_n), \quad \alpha_n = \arg\min\{G(y_n + \alpha r(y_n)) : \alpha \in \mathbb{R}\}. \tag{5}$$

For completeness, let us mention that relaxation approaches for these methods have been discussed in [3, 4], and additional projections on other subspaces have been investigated in [5, 6].

Both methods (3) and (5) may be considered as restarted versions (after one iteration) of generalized Krylov methods: here one minimizes on projected counterparts of the Krylov subspaces

$$K_n(A, c) := \mathrm{span}(c, Ac, ..., A^{n-1}c)$$

with the starting vector $c = b - Ay_0$. Indeed, (3) is the single-step version of the method

$$y_n = y_0 + \arg\min\{||r(y_0 + u)|| : u \in K_n(PA, P(b - Ay_0))\}, \tag{6}$$

referred in what follows as AMinRes, and (5) is the single-step version of the method

$$y_n = y_0 + \arg\min\{G(y_0 + u) : u \in K_n(PA, P(b - Ay_0))\}, \tag{7}$$

referred in what follows as ACG. To our knowledge, these latter methods have not yet been considered before.

The purpose of this note is to show that both approaches (6) and (7) (and thus also Altman's work) are mathematically equivalent to the classical algorithms of minimal residuals (MinRes) and conjugate gradients (CG), applied to the hermitian and semi positive definite matrix

$$\widetilde{A} := PAP. \tag{8}$$

As a consequence, we find coupled short term recurrences for the vectors y_n of (6) and (7), as well as error estimates involving the condition number of the matrix $\widetilde{A}$, derived already in a different manner by Altman [1, Eqn (16)], [2, Eqn (12)]. In particular, we establish interlacing properties of the corresponding eigenvalues, and conclude that both methods (6) and (7) behave always at least as well as the corresponding classical counterparts MinRes and CG, with an improvement of the speed of convergence only occurring for particular right-hand sides b. In this context it is interesting to observe that already Altman suggested to first transform the

original system (1) into $Ax' = b'$, $b' = b - Ax^*$, with $x' = x - x^*$, where x^* is some arbitrary vector, but he did not give any further device how to chose this vector x^*.

The remaining part of the paper is organized as follows: in Section 2 we explicitly state and prove the above equivalence claim, and discuss the behavior of MinRes/CG applied to consistent hermitian but singular systems. In Section 3 we study convergence properties of Altman's methods, both in the range of linear and super-linear convergence. Section 4 is devoted to the recursive computation of the iterates of Alman's methods. Since the analysis for (6) and (7) is quite similar, we will concentrate in this part only on (7). In Section 5 we present some numerical experiments confirming the theoretical observations of Section 3. Finally, in the appendix we discuss in Section A two quadratic forms, and add in Section B some historical remarks.

2 The equivalence

With the hermitian semi positive definite matrix $\widetilde{A} = PAP$ as in (8), consider the system of linear equations

$$\widetilde{A}\widetilde{x} = \widetilde{b} := -r(y_0) = -PAy_0. \tag{9}$$

In the following statement we give the exact link between Altman's algorithms (and their multi-step versions (6) and (7)) and classical algorithms like MinRes and CG, applied to (9).

Theorem 1. *The sequence $(y_n - y_0)_{n\geq 0}$ with y_n as in (6) is obtained by applying the MinRes algorithm with starting vector 0 to the system (9).*

Similarly, the sequence $(y_n - y_0)_{n\geq 0}$ with y_n as in (7) is obtained by applying the CG algorithm with starting vector 0 to (9).

Before presenting a proof of Theorem 1, let us have a closer look at the behavior of MinRes/CG applied to (9). Since $b^*\widetilde{b} = 0$, we obtain that

$$\widetilde{b} \in \mathrm{span}(b)^\perp = \mathrm{Ker}(\widetilde{A})^\perp = \mathrm{range}(\widetilde{A}),$$

and hence (9) is consistent, thought its matrix of coefficients is singular. The performance of Krylov subspace methods applied to inconsistent singular systems has been discussed by several authors, see for instance [11] and the citations therein. However, for consistent hermitian singular systems, the behavior is easily predictable: it is easily seen that

$$K_n(\widetilde{A}, \widetilde{b}) \subset \mathrm{Ker}(\widetilde{A})^\perp.$$

Thus, all iterates of MinRes/CG with starting vector 0 applied to (9) are elements of $\mathrm{Ker}(\widetilde{A})^\perp$. Notice also that (9) has a unique solution in $\mathrm{Ker}(\widetilde{A})^\perp$, namely $\widetilde{A}^\dagger\widetilde{b}$, where $\widetilde{A}^\dagger$ denotes the pseudo-inverse of $\widetilde{A}$. Thus, if one of these two algorithms terminates (after at most $\dim(K_n(\widetilde{A}, \widetilde{b})) \leq p-1$ iterations), then the corresponding iterate coincides with $\widetilde{A}^\dagger\widetilde{b}$. Moreover, the iterates of both algorithms converge to $\widetilde{A}^\dagger\widetilde{b}$, where the rate of convergence (either expressed in terms of the energy "norm" or the norm of the residual) can be bounded in the same way as for nonsingular hermitian systems.

In order to show the above theorem, we consider beside $P = I - bb^*$ the oblique projection operator

$$Q = I - \frac{A^{-1}bb^*}{(A^{-1}b, b)}.$$

The following properties are easily verified.

Lemma 2. *There holds*

$$PQ = Q, \tag{10}$$
$$QP = P, \tag{11}$$
$$PA = PAQ, \tag{12}$$
$$A - \frac{bb^*}{(A^{-1}b,b)} = AQ = Q^*AQ, \tag{13}$$
$$\widetilde{b} = -PAQy_0, \qquad \widetilde{A}^\dagger\widetilde{b} = -Qy_0, \tag{14}$$
$$K_n(PA, P(b - Ay_0)) = K_n(\widetilde{A}, \widetilde{b}). \tag{15}$$

Proof. By the definition of the matrix P and Q, and the fact that $||\,b\,|| = 1$, we obtain easily the four following properties

$$PQ = Q - bb^*Q = Q - bb^* + \frac{b(b^*A^{-1}b)b^*}{(A^{-1}b,b)} = Q,$$
$$QP = P - \frac{A^{-1}bb^*}{(A^{-1}b,b)}P = P - \frac{A^{-1}bb^*}{(A^{-1}b,b)} + \frac{A^{-1}b(b^*b)b^*}{(A^{-1}b,b)} = P,$$
$$PAQ = PA - \frac{Pbb^*}{(A^{-1}b,b)} = PA - \frac{(bb^* - b(b^*b)b^*)}{(A^{-1}b,b)} = PA,$$
$$Q^*AQ = AQ - \frac{bb^*Q}{(A^{-1}b,b)} = AQ - \left(\frac{bb^*}{(A^{-1}b,b)} - \frac{b(b^*A^{-1}b)b^*}{(A^{-1}b,b)}\right) = AQ.$$

Thus, the properties (10), (11), (12) and (13) are shown. By (12), we have $\widetilde{b} = -PAy_0 = -PAQy_0$. Thus, with (10), we deduce that

$$\widetilde{A}^\dagger\widetilde{b} = -\widetilde{A}^\dagger PAQy_0 = -\widetilde{A}^\dagger PAPQy_0 = -\widetilde{A}^\dagger\widetilde{A}Qy_0 = -PQy_0 = -Qy_0.$$

As the vector $\widetilde{b}$ verifies $P\widetilde{b} = \widetilde{b}$, we have $K_n(PA, P(b - Ay_0)) = K_n(PA, \widetilde{b}) = K_n(PAP, \widetilde{b}) = K_n(\widetilde{A}, \widetilde{b})$. Hence the lemma is proved. □

Proof of Theorem 1. Applying (12), (14), and (10), we see that, for any vector u,

$$||r(y_0 + Pu)|| = ||PAy_0 + PAPu|| = ||\widetilde{A}(Qy_0 + u)|| = ||\widetilde{b} - \widetilde{A}u||.$$

Taking into account (15) and the fact that $K_n(PA, P(b - Ay_0)) = PK_n(PA, P(b - Ay_0))$, we get for y_n as in (6)

$$y_n - y_0 = \arg\min\{||r(y_0 + Pu)|| : u \in K_n(PA, P(b - Ay_0))\} = \arg\min\{||\widetilde{b} - \widetilde{A}u|| : u \in K_n(\widetilde{A}, \widetilde{b})\},$$

the latter being the nth iterate of MinRes with starting vector 0 applied to (9).

Similarly, applying (13), (10), (11), and (14), we get for any vector u

$$G(y_0 + Pu) = \big((A - \frac{bb^*}{(A^{-1}b,b)})(y_0 + Pu), (y_0 + Pu)\big) = \big(AQ(y_0 + Pu), Q(y_0 + Pu)\big)$$
$$= \big(\widetilde{A}Q(y_0 + Pu), Q(y_0 + Pu)\big) = \big(\widetilde{A}(Qy_0 + u), (Qy_0 + u)\big)$$
$$= \big(\widetilde{A}(u - \widetilde{A}^\dagger\widetilde{b}), (u - \widetilde{A}^\dagger\widetilde{b})\big) =: \widetilde{G}(u),$$

and thus for y_n as in (7)

$$y_n - y_0 = \arg\min\{||G(y_0 + Pu)|| : u \in K_n(PA, P(b - Ay_0))\} = \arg\min\{\widetilde{G}(u) : u \in K_n(\widetilde{A}, \widetilde{b})\},$$

the latter being the nth iterate of CG with starting vector 0 applied to (9). Thus Theorem 1 is shown. □

3 Convergence

Before discussing bounds for the rate of convergence of Altman's methods, let us return to the termination property. As mentioned after Theorem 1, MinRes/CG with starting vector 0 applied to (9) is terminating (with value of the minimum being equal to 0) if and only if the corresponding iterate $y_n - y_0$ coincides with $\widetilde{A}^\dagger \widetilde{b}$, i.e. (compare with (14)),

$$y_n = y_0 + \widetilde{A}^\dagger \widetilde{b} = y_0 - Qy_0 = \frac{(y_0, b)}{(A^{-1}b, b)} A^{-1}b.$$

Thus, according to Theorem 1, the well-known terminating property for MinRes/CG yields a corresponding termination property for AMinRes/ACG.

However, as for other Krylov methods, one is more interested in convergence rates before reaching the stage of termination. If one plots the euclidean norm of the error as a function of the number of iterations on a semi-logarithmic scale, then following Nevanlinna [12] one may observe three different ranges which are more or less pronounced for particular examples: In general the curve will be first convex, then linear and finally concave, corresponding to the ranges of sublinear, of linear, and of superlinear convergence. In this description we do not take into account the effect of finite precision arithmetic, which of course in practical applications may lead to more complicated convergence curves. In case of a symmetric system (1) with general right-hand sides, the linear convergence behavior is quite well described in terms in the condition number of the underlying matrix of coefficients, see, e.g., [13, Theorem 6.6, Eqn. (6.105), and Corollary 6.1] for MinRes and CG. In contrast, the superlinear convergence behavior depends essentially on the eigenvalue distribution of the underlying matrix of coefficients, see [7, 8, 9] for a quantification of this statement. Roughly speaking, the superlinear convergence is pronounced if the eigenvalue distribution, especially for extremal eigenvalues, is far from the arcsine distribution on the convex hull of the spectrum.

Let us study here the behavior of ACG and AMinRes. Using the facts that, for $u \in \mathrm{Ker}(\widetilde{A})^\perp$,

$$||u - \widetilde{A}^\dagger \widetilde{b}||^2 \le ||\widetilde{A}^\dagger||\, \widetilde{G}(u), \quad ||u - \widetilde{A}^\dagger \widetilde{b}|| \le ||\widetilde{A}^\dagger||\, ||\widetilde{b} - \widetilde{A}u||,$$

we obtain as a consequence of our findings of the preceeding section

$$\left\| y_n - \frac{(y_0, b)}{(A^{-1}b, b)} A^{-1}b \right\|^2 = ||y_n - y_0 - \widetilde{A}^\dagger \widetilde{b}||^2 \le \begin{cases} ||\widetilde{A}^\dagger||^2\, ||r(y_n)||^2 & \text{for AMinRes,} \\ ||\widetilde{A}^\dagger||\, G(y_n) & \text{for ACG.} \end{cases}$$

The relative decrease of $||r(y_n)||$ or $G(y_n)$ is known from the corresponding classical decrease rates for MinRes/CG [13, Theorem 6.6, Eqn. (6.105), and Corollary 6.1], which are summarized in the following statement.

Corollary 3 (Linear convergence). *With* $\kappa(\widetilde{A}) = ||\widetilde{A}||\, ||\widetilde{A}^\dagger||$ *and*

$$q := (\sqrt{\kappa(\widetilde{A})} - 1)/(\sqrt{\kappa(\widetilde{A})} + 1) < 1$$

we have for the iterates of AMinRes

$$\frac{||r(y_n)||}{||r(y_0)||} \le \frac{2}{q^n + q^{-n}} \le 2q^n,$$

and for the iterates of ACG

$$\sqrt{\frac{G(y_n)}{G(y_0)}} \le \frac{2}{q^n + q^{-n}} \le 2q^n.$$

In particular, for $n = 1$ we obtain the upper bound

$$\frac{2}{q^1 + q^{-1}} = \frac{\kappa(\widetilde{A}) - 1}{\kappa(\widetilde{A}) + 1} \leq \frac{\kappa(A) - 1}{\kappa(A) + 1},$$

the last inequality following from the next Lemma. In particular, we recover Altman's error estimates [1, Eqn (16)] and [2, Eqn (12)] for the single-step methods (3) and (5).

Finally, in the range of super-linear convergence, the convergence curve of both methods (6) and (7) (or equivalently MinRes/CG for (9)) is determined by the eigenvalue distribution of $\widetilde{A}$, which according to the following result is essentially the same as that for A. In what follows we write $\lambda_1(B) \geq \lambda_2(B) \geq \ldots \geq \lambda_p(B)$ for the eigenvalues of any hermitian matrix B of order p.

Lemma 4 (Superlinear convergence). *For the matrices A of (1) and $\widetilde{A}$ of (8) there holds*

$$\lambda_1(A) \geq \lambda_1(\widetilde{A}) \geq \lambda_2(A) \geq \lambda_2(\widetilde{A}) \geq \ldots \geq \lambda_p(A) > \lambda_p(\widetilde{A}) = 0.$$

Proof. It is clear that b is an eigenvector of $\widetilde{A}$ corresponding to the eigenvalue 0. Denote H_b the orthogonal complement of span(b), then we find an orthonormal basis of eigenvectors $v_1, \ldots, v_{p-1} \in H_b$, $v_p \in$ span(b) of the matrix $\widetilde{A}$ such that $\widetilde{A}v_i = \lambda_i(\widetilde{A})v_i$. Thus it remains to show that, for any $i \in \{1, \ldots, p-1\}$, there holds

$$\lambda_{i+1}(A) \leq \lambda_i(\widetilde{A}) \leq \lambda_i(A).$$

Define $S_i = \text{span}(v_1, \ldots, v_i)$. Using the Courant-Fischer Minmax Theorem, we get

$$\begin{aligned}
\lambda_i(A) &= \max_{dim(S)=i} \min_{\substack{y \in S \\ y \neq 0}} \frac{y^*Ay}{y^*y} \geq \min_{\substack{y \in S_i \\ y \neq 0}} \frac{y^*Ay}{y^*y} \\
&= \min_{\substack{y \in S_i \\ y \neq 0}} \frac{y^*P^*APy}{y^*y} \quad \text{since } S_i \subseteq H_b \\
&= \min_{\substack{y \in S_i \\ y \neq 0}} \frac{y^*\widetilde{A}y}{y^*y} = v_i^*\widetilde{A}v_i = \lambda_i(\widetilde{A}),
\end{aligned}$$

and hence $\lambda_i(A) \geq \lambda_i(\widetilde{A})$. Similarly, with $V_{p-i} = \text{span}(v_i, \ldots, v_{p-1})$ of dimension $p - i$

$$\begin{aligned}
\lambda_{i+1}(A) &= \min_{dim(S)=p-i} \max_{\substack{y \in S \\ y \neq 0}} \frac{y^*Ay}{y^*y} \leq \max_{\substack{y \in V_i \\ y \neq 0}} \frac{y^*Ay}{y^*y} \\
&= \max_{\substack{y \in V_i \\ y \neq 0}} \frac{y^*P^*APy}{y^*y} \quad \text{since } V_i \subseteq H_b \\
&= \max_{\substack{y \in V_i \\ y \neq 0}} \frac{y^*\widetilde{A}y}{y^*y} = v_i^*\widetilde{A}v_i = \lambda_i(\widetilde{A}).
\end{aligned}$$

□

The last two results enable us to compare ACG and CG for the system (1): according to Lemma 4, the convergence behavior in the super-linear range should be similar, but, according to Corollary 3, in the linear range there might be a different behavior, depending on whether $\kappa(A) = \lambda_1(A)/\lambda_p(A)$ is "much" larger as $\kappa(\widetilde{A}) = \lambda_1(\widetilde{A})/\lambda_{p-1}(\widetilde{A})$, this latter link depending on the choice of the right-hand side b. We will confirm these claims by some numerical experiments reported in Section 5.

4 The algorithm ACG

In the CG algorithm applied to (9) one constructs recursively $\widetilde{A}$-conjugated bases $p_0, ..., p_{n-1}$ of $K_n(\widetilde{A}, \widetilde{b})$, leading to one-dimensional minimization problems and thus coupled short-term recurrences. Observing that

$$\widetilde{r}_n := \widetilde{b} - \widetilde{A}(y_n - y_0) = \widetilde{b} - PA(y_n - y_0) = -r(y_n),$$

we obtain from Theorem 1 the following coupled system of recurrence relations (we keep the notations of CG as stated, e.g., in [13, Algorithm 6.17]) for the iterates y_n of ACG

Initialize $\widetilde{r}_0 = p_0 = -r(y_0)$,

Compute for $n = 0, 1, ...$ until $||\widetilde{r}_n||$ is sufficiently small

$$\alpha_n = \frac{(\widetilde{r}_n, \widetilde{r}_n)}{(\widetilde{A}p_n, p_n)} = \frac{(\widetilde{r}_n, \widetilde{r}_n)}{(r(p_n), p_n)} = \frac{(\widetilde{r}_n, \widetilde{r}_n)}{(Ap_n, p_n)},$$

$$y_{n+1} = y_n + \alpha_n p_n, \quad \widetilde{r}_{n+1} = \widetilde{r}_n - \alpha_n \widetilde{A}p_n = \widetilde{r}_n - \alpha_n r(p_n),$$

$$\widetilde{\beta}_n = \frac{(\widetilde{r}_{n+1}, \widetilde{r}_{n+1})}{(\widetilde{r}_n, \widetilde{r}_n)}, \quad p_{n+1} = \widetilde{r}_{n+1} + \widetilde{\beta}_n p_n.$$

As a consequence, for the sequence $x_n = y_n/(Ay_n, b)$ approaching the solution of (1) we obtain the residual

$$r_n := b - Ax_n = \frac{1}{(Ay_n, b)}((Ay_n, b)b - Ay_n) = -r(y_n)/(Ay_n, b) = \widetilde{r}_n/(Ay_n, b),$$

and the substitution $z_n = p_n/(Ay_n, b)$ leads to

$$x_{n+1} = \frac{(Ay_n, b)}{(Ay_{n+1}, b)}(x_n + \alpha_n z_n) = \frac{x_n + \alpha_n z_n}{1 + \alpha_n(Az_n, b)}$$

and

$$z_{n+1} = r_{n+1} + \widetilde{\beta}_n \frac{(Ay_n, b)}{(Ay_{n+1}, b)} z_n = r_{n+1} + \beta_n(1 + \alpha_n(Az_n, b))z_n, \quad \beta_n = \frac{(r_{n+1}, r_{n+1})}{(r_n, r_n)}.$$

Hence we get following equivalent formulation of ACG

Initialize $r_0 = z_0 = b - Ax_0$

Compute for $n = 0, 1, ...$ until $||r_n||$ is sufficiently small

$$\alpha_n = \frac{(r_n, r_n)}{(Az_n, z_n)}, \quad \nu_n := 1 + \alpha_n(Az_n, b),$$

$$x_{n+1} = \frac{1}{\nu_n}(x_n + \alpha_n z_n),$$

$$r_{n+1} = \frac{1}{\nu_n}(r_n - \alpha_n r(z_n)) = \frac{1}{\nu_n}(r_n - \alpha_n PAz_n) = \frac{1}{\nu_n}(r_n - \alpha_n[Az_n - (Az_n, b)b]),$$

$$\beta_n = \frac{(r_{n+1}, r_{n+1})}{(r_n, r_n)}, \quad z_{n+1} = r_{n+1} + \nu_n \beta_n z_n.$$

Notice that this method, as MinRes and CG, requires one matrix-vector product by iteration.

We conclude this section with the observation that the last identity for r_{n+1} in the algorithm ACG can be written as

$$r_{n+1} = \frac{1}{\nu_n} r_n - \frac{\alpha_n}{\nu_n}(Az_n - \frac{\nu_n - 1}{\alpha_n} b) = \frac{1}{\nu_n}(r_n - \alpha_n Az_n) + (1 - \frac{1}{\nu_n})b.$$

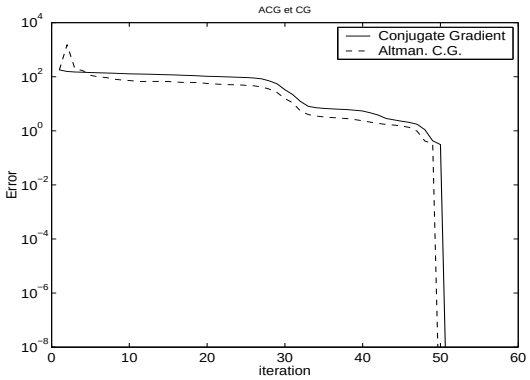

Figure 1: ACG (dashed line) versus CG (solid line) for $A = \text{tridiag}([-1, 2, -1])$ of dimension 50 resulting from the discretization of the one-dimensional Laplacian.

For $n = 0$ this observation gives raise to the following interpretation: the residual of the one-step version (5) of ACG is obtained by taking a convex combination of the residual of steepest descent (i.e., one iteration of CG applied to (1)) and the original residual. Such an relaxation interpretation was mentioned already by Altman in [4].

5 Numerical experiments

Let us now give some numerical results for illustrating the ACG method and compare it to the classical conjugate gradient algorithm for solving $Ax = b$ with two different kinds of test matrices A. All the computations were performed in MATLAB and all norms are the Euclidean ones.

We first present an example of a system (1) with a matrix A of dimension $p = 50$ resulting from the discretization of the one-dimensional Laplacian on $[-1, 1]$, the right-hand side b and the starting vector y_0 being chosen randomly. Here the eigenvalue distribution of A approaches the worst-case of an arcsine distribution, and hence superlinear convergence should only occur for very special right-hand sides b, compare with [9]. As expected, we observe in Figure 1 that the convergence rate is essentially linear, up to the stage where the termination property of CG/ACG gives suddenly convergence. We also notice that, by Lemma 4, the quantity $\kappa(\widetilde{A})$ lies between $\kappa(A) = \lambda_1(A)/\lambda_p(A)$ and $\lambda_2(A)/\lambda_{p-1}(A)$, which for our example give the numerical values $1.05 \cdot 10^3$, and $0.26 \cdot 10^3$. Thus the condition number of $\widetilde{A}$ is not essentially smaller than that of A, and the convergence behavior of CG and ACG should be similar. This is clearly confirmed by the results presented in Figure 1: for this example we obtain that ACG, compared to CG, allows to gain only one iteration.

The second group of examples were performed using the matrix $A = QDQ^*$, where

$$Q = (I - 2w_3w_3^*)(I - 2w_2w_2^*)(I - 2w_1w_1^*),$$

w_1,w_2 and w_3 are unitary random vectors, $D = diag(\lambda_1, \cdots, \lambda_p)$ is a diagonal matrix whose all components are $\lambda_i = \varepsilon + (i-1)$ for $i = 1, \cdots, p$, and $\varepsilon > 0$ is a scalar which will vary below. In our tests reported below we use $p = 1000$, the starting point $y_0 = (1, 1, ..., 1)^* \in \mathbb{R}^p$ for ACG, and the starting point $x_0 = (1, 1, ..., 1)^* \in \mathbb{R}^p$ for CG. The solution x of system (1) (and hence the right-hand side $b = Ax$) will be chosen either randomly, or in terms of the eigenvector v_i corresponding to the eigenvalue $\lambda_i(A)$ of A.

name of example	ε	solution x	cond(A)	cond$(\widetilde{A})$	CG	ACG
I	10^{-6}	v_p	9.9e+08	9.9e+02	243	194
II		$v_p + 10^{-8}v_{p-1}$	9.9e+08	9.9e+02	237	188
III		$v_p + 10^{-3}v_{p-1}$	9.9e+08	4.9e+08	245	235
IV		random	9.9e+08	9.9e+08	274	274
V	10^{-3}	v_p	9.9e+05	9.9e+02	240	187
VI		random	9.9e+05	9.9e+05	238	235
VII	1	v_p	1e+03	5e+02	180	180

Table 1: Number of iterations for CG versus ACG.

In Table 1, we report the results for seven different choices of parameters, namely, the choice of the variable ε and/or the solution x, the resulting condition numbers of the matrices A and $\widetilde{A}$, and the resulting number of iterations used to obtain an error norm $|| x_k - x || \leq 10^{-8}$ by the CG and the ACG method.

We first provide some explanation for the CG convergence behavior for these seven examples. The convergence behavior for the case of equidistant eigenvalues and $\varepsilon = 1$ and general right-hand sides was considered in [7, Corollary 3.2], here one essentially has only superlinear convergence for CG, and no range of linear convergence occurs, see also the plot of Example VII in Figure 4 below. For ε approaching zero, the condition number of A becomes worse. Here the superlinear convergence is delayed by some non-trivial range of linear convergence, corresponding to the quasi horizontal part on the left-hand side of the CG curves for examples I, II, III (which are essentially the same) and V and VI (again essentially the same).

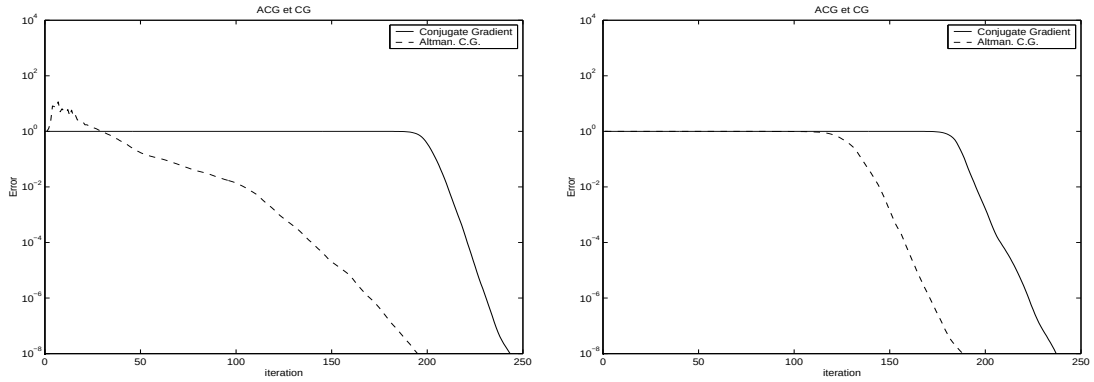

Figure 2: ACG (dashed line) versus CG (solid line), examples I (left) and II (right).

Concerning the convergence behavior of ACG, we note that, for small ϵ, there is an important gap between $\kappa(A) = \lambda_1(A)/\lambda_p(A)$ and $\lambda_2(A)/\lambda_{p-1}(A)$, and thus a potential improvement of ACG on CG in the range of linear convergence. Indeed, for the examples I, II and V we get an essential improvement for the number of iterations, see Table 1, Figure 2 and Figure 3. However, for general right-hand sides as in examples III, IV, and VI, no essential improvement is found even in the range of linear convergence (compare with Table 1, Figure 3 and Figure 4), confirming our theoretical observations of the end of Section 3. Moreover, for $\varepsilon = 1$ and hence well-conditioned A as in example VII (see Figure 4), the two methods CG and ACG behave identically.

We terminate our discussion of this second group of exemples by some concluding remarks. The work in [7, 8, 9] is based only on the (asymptotic) eigenvalue distribution, and gives some weak asymptotics for the error of Krylov subspace methods. For many examples, the bounds

obtained in [7] do not only describe the error asymptotically, but provide also an upper bound without passing to the limit, but this latter experimental observation has no theoretical justification. In our case, it seems that, in the superlinear range, the slopes of all twelve convergence curves (CG and ACG) do only depend on the ratio of the iteration index and of p, and coincide with the slope pointed out in [7]. But for the error there seems to be an additional multiplicative factor depending on the condition number, and, more importantly, due to this factor, the beginning of superlinear convergence seems to be delayed. In this context we should mention that the effect of ill-conditioning of the matrix of coefficients (which is related to [7, Condition (iii)]) was essentially neglected in [7], and its exact role for (non asymptotic) upper bounds for the CG in the superlinear range remains open.

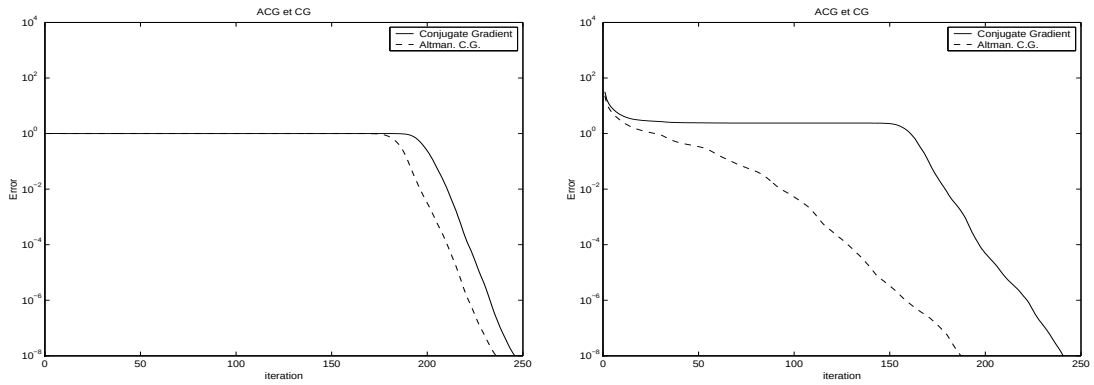

Figure 3: ACG (dashed line) versus CG (solid line), examples III (left) and V (right).

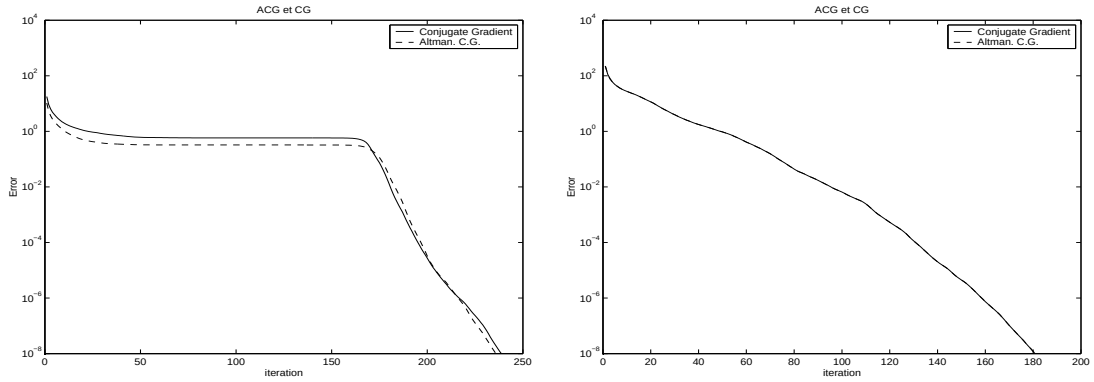

Figure 4: ACG (dashed line) versus CG (solid line), examples VI (left) and VII (right).

Appendix A: Representation of the quadratic form

Denote by H_b the the orthogonal complement of span(b). In [2], Altman uses the quadratic form

$$z \in H_b : \qquad F(z) := (r^{-1}(z), z). \tag{16}$$

In order to justify that this formula makes sense, Altman first shows that the restriction $r : H_b \mapsto H_b$ is hermitian, self-adjoint and positive definite, and thus the inverse of r exists on H_b. In fact, for $u \in H_b$ there holds $r(u) = PAu = PAPu = \widetilde{A}u$, and thus $r^{-1}(u) = \widetilde{A}^{\dagger}u$.

This observation allows us relate the quadratic form F of (16) to the quadratic form G of

(4): for any vector y there holds according to (12) and (10)

$$\begin{aligned} F(r(y)) &= (\widetilde{A}^{\dagger} r(y), r(y)) = (\widetilde{A}^{\dagger} PAy, PAy) = ((PA)^* \widetilde{A}^{\dagger} PAy, y) \\ &= ((PAQ)^* \widetilde{A}^{\dagger} PAQy, y) = (Q^* \widetilde{A} \widetilde{A}^{\dagger} \widetilde{A} Qy, y) = (Q^* \widetilde{A} Qy, y) = (Q^* AQy, y), \end{aligned}$$

and thus (13) implies that $F(r(y)) = G(y)$.

Appendix B: Biography of Mieczysław Altman by Tom Altman

Mieczysław Altman (1916-1997) was born in Kutno, Poland, 50km north of Łódź. He studied mathematics at the Warsaw University from 1937 until the outbreak of World War II in 1939. In 1940 he enrolled at Lwów University and worked directly under the tutelage of Stefan Banach for two years. He was his last student. After the Nazi invasion of USSR in 1941, he was forced to flee again, eventually settling at the University of Sverdlovsk (now Yekaterinburg), 1700km east of Moscow, in 1941-1942, and then in Tashkent, USSR (now Uzbekistan). There, he first obtained his Master in Mathematics in 1944 and he finished his Ph.D. in Mathematics in 1948. After his return to Poland in 1949, he learned that his entire family, including seven brothers and sisters, perished in the Łódź Jewish ghetto during the war. He held a position at the Institute of Mathematics of the Polish Academy of Sciences in Warsaw from 1949 to 1969, first as an Assistant Professor, 1949-1957, then as an Associate Professor, 1957-1958, and finally as a Full Professor and Director of the Numerical Analysis Department from 1958 to 1969.

In 1953, he married Wanda Kusal, M.D., and they had two children, Barbara (in 1956) and Tom (in 1958). For two years (1959-1960) Professor Altman took visiting positions at the California Institute of Technology in Pasadena and the Courant Institute in New York. He received the Poland's Banach Prize in Functional Analysis in 1958 and was Vice President of the Polish Mathematical Society in 1962-1963. Due to political pressures, in 1969 the Altman family left Poland and eventually settled in Baton Rouge, Louisiana, where M. Altman worked as a Professor of Mathematics, Louisiana State University, from 1970 until his retirement in 1987.

His 1977 book on *Contractors and Contractor Directions - Theory and Applications* (Marcel Dekker, New York, 1977) received international acclaim and recognition among mathematicians as the most encompassing theory for solving equations by analytical means. To honor the memory of his relatives, friends, and countrymen, he had dedicated his 1986 book on the *A Unified Theory of Nonlinear Operator and Evolution Equations with Applications - A New Approach to Nonlinear Partial Differential Equations* (Marcel Dekker, New York, 1986) to the victims of the Holocaust.

Even after his retirement, Professor Altman remained professionally active, publishing several journal papers and a book *A Theory of Optimization and Optimal Control for Nonlinear Evolution and Singular Equations with Applications to Nonlinear Partial Differential Equations* (World Scientific, Singapore, 1990).

Prof. Altman is the author of over 200 research papers in pure and applied mathematics, including functional and numerical analysis, mathematical programming, general optimization and optimal control theory, nonlinear differential and integral equations, and Banach algebras. A man of many talents, Professor Altman published his mathematical papers in a number of languages, including Russian, Polish, French, German, and English. He even coined the french word "contracteurs" (Contracteurs dans les algèbres de Banach, C.R. Acad. Sci. Paris, 274 (1972), 399-400).

He was a visiting professor at the Instituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Rome, Italy, 1969-1970, and Newcastle University, Australia, in 1973.

Acknowledgements: The authors gratefully acknowledge the help of Prof. Stefan Paszkowski and Prof. Andrzej Wakulicz for finding out more about the life of M. Altman, and the help of Prof. Tom Altman who provided the biography of M. Altman in Appendix B.

References

[1] M. Altman, A linear iterative method with a vector parameter, Bull. Pol. Acad. Sci. Math., 9 (1961) 169-174.

[2] M. Altman, A method of steepest ortho-descent, Bull. Pol. Acad. Sci. Math., 9 (1961) 575-580.

[3] M. Altman, A general method of steespest ortho-descent, Bull. Pol. Acad. Sci. Math., 9 (1961) 645-651.

[4] M. Altman, A gradient relaxation method for linear algebraic equations, Bull. Pol. Acad. Sci. Math., 9 (1961) 653-657.

[5] M. Altman, A generalized method of steepest descent, Bull. Pol. Acad. Sci. Math., 9 (1961) 733-737.

[6] M. Altman, A general method with minimum ortho-residua, Bull. Pol. Acad. Sci. Math., 9 (1961) 739-744.

[7] B. Beckermann, A.B.J. Kuijlaars, Superlinear convergence of conjugate gradients, *SIAM J. Num. Anal.* **39** (2001) 300-329.

[8] B. Beckermann, A.B.J. Kuijlaars, On the sharpness of an asymptotic error estimate for Conjugate Gradients, *BIT* **41** (2001), 856–867.

[9] B. Beckermann, A.B.J. Kuijlaars, Superlinear CG convergence for special right-hand sides, *Electr. Trans. Num. Anal.* **14** (2002) 1-19.

[10] C. Brezinski, *Projection methods for systems of equations*, North-Holland, Amsterdam (1997).

[11] B. Fischer, M. Hanke, M. Hochbruck, A note on conjugate-gradient type methods for indefinite and/or inconsistent linear systems, *Numer. Algo.* **11** (1996) 181-189.

[12] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[13] Y. Saad, *Iterative methods for sparse linear systems*, PWS Publishing, Boston, MA, 1996.