

# ITERATIVE METHODS FOR COMPUTING GENERALIZED INVERSES RELATED WITH OPTIMIZATION METHODS

DRAGAN S. DJORDJEVIĆ and PREDRAG S. STANIMIROVIĆ

(Received 19 August 2002; revised 30 October 2003)

Communicated by A. J. Pryde

## Abstract

We develop several iterative methods for computing generalized inverses using both first and second order optimization methods in  $C^*$ -algebras. Known steepest descent iterative methods are generalized in  $C^*$ -algebras. We introduce second order methods based on the minimization of the norms  $\|Ax - b\|^2$  and  $\|x\|^2$  by means of the known second order unconstrained minimization methods. We give several examples which illustrate our theory.

2000 *Mathematics subject classification*: primary 46H05, 47A05, 90C30, 15A09.

*Keywords and phrases*: generalized inverses, Moore-Penrose inverse, iterative methods, nonlinear optimization.

## 1. Introduction

Let  $H$  and  $K$  be Hilbert spaces and let  $A$  be a bounded linear operator from  $H$  into  $K$ . In this paper we construct several iterative methods which are concerned in computing the Moore-Penrose generalized inverse of  $A$ . These iterative methods appear naturally in minimization methods of the first and the second order. Moreover, we investigate these methods in  $C^*$ -algebras.

Applications of the first order gradient minimization methods in computation of the Moore-Penrose inverse of  $A$  (when it exists) are well known. The steepest descent method in minimization of the functional  $Q(x) = \frac{1}{2}\|Ax - b\|^2$  ( $x \in H$ ,  $b \in K$ ) was introduced in [7] and [14] (see also [2, 5]). The problem of approximating the least squares solutions of  $Ax = b$  by means of the conjugate-gradient optimization technique was considered in [2, 5, 7, 9]. A conjugate-gradient method for computing

the Moore-Penrose generalized inverse was developed in [13]. A class of gradient methods for minimizing  $Q(x) = \frac{1}{2}\|Ax - b\|^2$ , defined by the real-valued step-size function  $s(x)$  defined on  $H$  and by a fixed real number  $0 < \alpha < 2$ :

$$(1.1) \quad x_{n+1} = x_n - \alpha s(x_n) A^*(Ax_n - b) = x_n - \alpha s(x_n) \nabla Q(x_n),$$

was analyzed in [10]. The weak steepest descent method, which is defined by means of the fixed step-size  $\alpha$  in all iterations, as well as its applications in computation of generalized inverses, were investigated in [1] and [14].

Let  $H$  be a Hilbert space and  $Q : H \mapsto \mathbb{R}$  be a twice differentiable functional. The second order minimization method (Newton's method) is defined by the following iterative procedure [3]:

$$(1.2) \quad x_{n+1} = x_n - [\nabla^2 Q(x_n)]^{-1} \nabla Q(x_n),$$

where  $\nabla Q(x_n)$  and  $\nabla^2 Q(x_n)$  denote the gradient and Hessian, respectively, of a given objective function  $Q$ , at the point  $x_n$ . A few quasi-Newton minimization methods for operators acting from  $\mathbb{R}^n$  into  $\mathbb{R}$  are defined in [3]. We use the following adaptation of a quasi-Newton minimization method:

$$(1.3) \quad x_{n+1} = x_n - (\lambda_n \mathbf{I} + \nabla^2 Q(x_n))^{-1} \nabla Q(x_n),$$

where  $Q : H \rightarrow \mathbb{R}$  is an arbitrary functional,  $\lambda_n > 0$  is a chosen scalar and  $\lambda_n \mathbf{I} + \nabla^2 Q(x_n)$  is positive and invertible.

The paper is organized as follows. In Section 2 we introduce and consider one iterative method for computing the Moore-Penrose inverse of a relatively regular element in a  $C^*$ -algebra. As a corollary we get that the Nashed steepest descent method is convergent in operator norm. We also prove that the gradient methods (1.1) are also convergent in operator norm. In Section 3 we construct and investigate some new methods related to the second order optimization. We also prove the convergence of these methods in  $C^*$ -algebras. Numerical examples are presented in Section 4.

Throughout this paper we always assume  $a \neq 0$  in a  $C^*$ -algebra.

## 2. Methods based on the first order optimization

We use  $\mathcal{L}(H, K)$  to denote the set of all bounded linear operators from  $H$  into  $K$ . For  $A \in \mathcal{L}(H, K)$ , we use  $\mathcal{R}(A)$  and  $\mathcal{N}(A)$ , respectively, to denote the range and the kernel of  $A$ . Also, by  $A^*$  we denote the adjoint operator of  $A$ . It is well known that  $A$  has the Moore-Penrose inverse  $A^\dagger \in \mathcal{L}(K, H)$  if and only if  $\mathcal{R}(A)$  is closed.

First we shall explain the steepest descent method, introduced in [11]. Suppose that  $A \in \mathcal{L}(H, K)$  has the Moore-Penrose inverse. For arbitrary  $b \in K$ , let  $r_n =$

$A^*Ax_n - A^*b$  and  $\alpha_n = \|r_n\|^2/\|Ar_n\|^2$ . In [11] it was proved that the iterative sequence  $x_{n+1} = x_n - \alpha_n r_n$  satisfies  $\lim x_n = A^\dagger b$  if and only if  $b \in \mathcal{R}(A^*)$ . Notice that every change of  $b \in K$  implies the construction of a new iterative method. The choice of  $\alpha_n$  in the  $n$ -th step of the steepest descent method ensures the most rapid decrease of  $\|Ax_{n+1} - b\|$ .

Therefore, taking  $X_0 \in \mathcal{L}(K, H)$ , the sequence  $(X_n)_n$  which is defined in the following way

$$(2.1) \quad X_{n+1} = X_n - \alpha_n A^*(AX_n - I)$$

obeys the property  $\lim X_n b = A^\dagger b$ .

We shall consider a more general method in  $C^*$ -algebras.

Let  $\mathcal{A}$  be a complex  $C^*$ -algebra with the unit 1. We say that  $a \in \mathcal{A}$  is *relatively regular*, provided that there exists some  $b \in \mathcal{A}$  satisfying  $aba = b$ . In this case  $b$  is called *an inner generalized inverse of  $a$* . It is well known that  $a$  is relatively regular if and only if there exists the Moore-Penrose inverse of  $a$ , denoted by  $a^\dagger$  ([6]). The element  $a^\dagger$  is the unique element of  $\mathcal{A}$  satisfying

$$aa^\dagger a = a, \quad a^\dagger aa^\dagger = a^\dagger, \quad (aa^\dagger)^* = aa^\dagger, \quad (a^\dagger a)^* = a^\dagger a.$$

If  $a \in \mathcal{A}$  is relatively regular and  $a^-$  denotes an arbitrary inner generalized inverse of  $a$ , then for any  $b, c \in \mathcal{A}$  we have  $(1 - aa^-)(ab - c) = (aa^-c - c)$ . Consequently,  $\|aa^-c - c\| \leq \|1 - aa^-\| \|ab - c\|$ . Taking  $(aa^-)^* = aa^-$ , or  $a^- = a^\dagger$ , we obtain the following minimization result:

$$\min_{b \in \mathcal{A}} \|ab - c\| = \|aa^\dagger c - c\|.$$

We begin with two auxiliary results. The first one can be found in [12, Theorem 5.7.8]. We use  $\sigma_{\mathcal{A}}(a)$  to denote the spectrum of  $a$  in the algebra  $\mathcal{A}$ .

**LEMMA 2.1.** *If  $p \in \mathcal{A}$  is a nontrivial idempotent and  $a \in p\mathcal{A}p$ , then*

$$\sigma_{\mathcal{A}}(a) = \sigma_{p\mathcal{A}p}(a) \cup \{0\}.$$

The second result is a consequence of elementary calculations. For some facts we need properties of the group inverse in  $C^*$ -algebras. If 0 is not the accumulation point of the spectrum of  $a$  and  $p$  is the spectral idempotent of  $a$  corresponding to  $\{0\}$ , then  $a$  is invertible in the algebra  $(1 - p)\mathcal{A}(1 - p)$  and its inverse in this algebra is the generalized Drazin inverse of  $a$  (see [8]). It is well known that the generalized Drazin inverse of  $a$  commutes with every  $b$  satisfying  $ab = ba$ . If  $\lambda = 0$  is a pole of order  $k$  of the resolvent  $\lambda \mapsto (\lambda - a)^{-1}$ , then the generalized Drazin inverse becomes the (ordinary) Drazin inverse; in this case the Drazin index of  $a$  is equal to  $k$ . In particular,

if  $k \leq 1$ , then the Drazin inverse of  $a$  is known as the group inverse of  $a$ , denoted by  $a^\#$ . Naturally,  $a$  is invertible if and only if  $k = 0$  and in this case  $a^{-1} = a^\#$ . In the case when it exists,  $a^\#$  is the unique element of  $\mathcal{A}$  satisfying  $aa^\#a = a$ ,  $a^\#aa^\# = a^\#$  and  $aa^\# = a^\#a$ .

**LEMMA 2.2.** *Let  $a \in \mathcal{A}$  be relatively regular. Then the following holds:*

- (1)  $(a^*)^\dagger = (a^\dagger)^*$ ;
- (2)  $(a^*a)^\dagger = a^\dagger(a^*)^\dagger = (a^*a)^\#$ ;
- (3)  $a^\dagger = (a^*a)^\dagger a^*$ ;
- (4)  $(a^*a)^\dagger$  commutes with every element of  $\mathcal{A}$  which commutes with  $a^*$ ;
- (5)  $a^*a$  is invertible in the algebra  $a^\dagger a \mathcal{A} a^\dagger a = \mathcal{B}$  and  $(a^*a)^\dagger = (a^*a)_{\mathcal{B}}^{-1}$ ;
- (6)  $a^\dagger aa^* = a^* = a^*aa^\dagger$ .

It is important to notice that if  $a$  is relatively regular, then  $(a^*a)^\#$  exists. Hence, 0 is not the accumulation point of the spectrum  $\sigma_{\mathcal{A}}(a^*a)$ .

For a selfadjoint element  $d \in \mathcal{A}$ , we use  $\mathcal{M}_{\mathcal{A}}(d)$  and  $m_{\mathcal{A}}(d)$ , respectively, to denote the upper and the lower bound of the spectrum of  $d$  in the algebra  $\mathcal{A}$ .

Now we formulate the result inspired by the Nashed steepest descent method.

**THEOREM 2.3.** *Let  $a \in \mathcal{A}$  be relatively regular,  $\mathcal{B} = a^\dagger a \mathcal{A} a^\dagger a$  and let  $c, x_0 \in \mathcal{A}$  be arbitrary. Let  $(\lambda)_n$  be a sequence of positive numbers such that*

$$0 < \epsilon \leq \lambda_n \leq 2 \max \{ [\mathcal{M}_{\mathcal{B}}(a^*a)]^{-1}, [m_{\mathcal{B}}(a^*a)]^{-1} \} - \delta$$

*holds for some  $\epsilon, \delta > 0$  and for all  $n$ . Then the sequence  $(x_n)_n$  generated by the iterative method  $x_{n+1} = x_n - \lambda_n a^*(x_n - c)$ ,  $n \geq 0$ , converges to  $a^\dagger c + (1 - a^\dagger a)x_0$ . Consequently,  $\lim x_n = a^\dagger c$  if and only if  $a^\dagger ax_0 = x_0$ .*

**PROOF.** Let  $a^\dagger ax_0 = x_0$ . Using Lemma 2.2 (6) and the induction on  $n$  it follows that  $a^\dagger ax_n = x_n$  for all  $n$ . We compute

$$\begin{aligned} a^*ax_{n+1} - a^*c &= a^*ax_n - a^*c - \lambda_n a^*a(a^*ax_n - a^*c) \\ &= (1 - \lambda_n a^*a)(a^*ax_n - a^*c). \end{aligned}$$

Multiplying previous equality by  $(a^*a)^\dagger$  from the left side and using Lemma 2.2 (3) and (4), we get

$$x_{n+1} - a^\dagger c = (1 - \lambda_n a^*a)(x_n - a^\dagger c)$$

and

$$x_{n+1} - a^\dagger c = a^\dagger a(x_{n+1} - a^\dagger c) = a^\dagger a(1 - \lambda_n a^*a)a^\dagger a(x_n - a^\dagger c).$$

The norm estimate follows:  $\|x_{n+1} - a^\dagger c\| \leq \|a^\dagger a(1 - \lambda_n a^* a)a^\dagger a\| \|x_n - a^\dagger c\|$ . To ensure the convergence, we must have  $\|a^\dagger a(1 - \lambda_n a^* a)a^\dagger a\| \leq q < 1$ . We know the following

$$\begin{aligned} & \|a^\dagger a(1 - \lambda_n a^* a)a^\dagger a\| \\ &= \max \{ |\mathcal{M}_{\mathcal{B}}[a^\dagger a(1 - \lambda_n a^* a)a^\dagger a]|, |m_{\mathcal{B}}[a^\dagger a(1 - \lambda_n a^* a)a^\dagger a]| \}. \end{aligned}$$

Notice that  $\mathcal{M}_{\mathcal{B}}[a^\dagger a(1 - \lambda_n a^* a)a^\dagger a] = 1 - \lambda_n m_{\mathcal{B}}(a^* a)$  and  $m_{\mathcal{B}}[a^\dagger a(1 - \lambda_n a^* a)a^\dagger a] = 1 - \lambda_n \mathcal{M}_{\mathcal{B}}(a^* a)$ . Since  $a^* a$  is invertible in  $\mathcal{B}$ , we conclude that  $\|a^\dagger a(1 - \lambda_n a^* a)a^\dagger a\| \leq q < 1$  holds for all  $n$  if and only if

$$0 < \epsilon < \lambda_n < 2 \max \{ [\mathcal{M}_{\mathcal{B}}(a^* a)]^{-1}, [m_{\mathcal{B}}(a^* a)]^{-1} \} - \delta$$

holds for some  $\epsilon, \delta > 0$  and for all  $n$ . Obviously,  $\lim x_n = a^\dagger c$ .

Now, suppose that  $a^\dagger a x_0 \neq x_0$ . For any  $x \in \mathcal{A}$  we denote  $x' = a^\dagger a x$  and  $x'' = (1 - a^\dagger a)x$ . Now we have  $x_0'' \neq 0$ . From Lemma 2.2 (6) we conclude that

$$x'_1 = x'_0 - \lambda_0 a^*(a x_0 - c) = x'_0 - \lambda_0 a^*(a x'_0 - c)$$

and  $x''_1 = x''_0$ . By induction on  $n$  we conclude that  $x'_{n+1} = x'_n - \lambda_n a^*(a x'_n - c)$  and  $x''_{n+1} = x''_0$ . From the first part of the proof we get  $\lim x_n = a^\dagger c + (1 - a^\dagger a)x_0$ .  $\square$

**REMARK 2.1.** In the case  $c = 1$ , the limit  $d = a^\dagger + (1 - a^\dagger a)x_0$  is an inner generalized inverse of  $a$  and  $ad = aa^\dagger$  is selfadjoint. Moreover, the condition  $a^\dagger a x_0 = x_0$  can be ensured taking  $x_0 = a^* e$  for an arbitrary  $e \in \mathcal{A}$ .

**REMARK 2.2.** In the case when  $A \in \mathcal{L}(H, K)$  is relatively regular, the same method as in Theorem 2.3 can be applied, taking  $C \in \mathcal{L}(K)$ ,  $X_0 \in \mathcal{L}(K, H)$ . Now the condition  $A^\dagger A X_0 = X_0$  can be replaced with  $\mathcal{R}(X_0) \subset \mathcal{R}(A^*)$ . We use  $\mathcal{B} = A^\dagger A \mathcal{L}(H) A^\dagger A = \mathcal{L}(\mathcal{R}(A^*))$  instead of  $\mathcal{B} = a^\dagger a \mathcal{A} a^\dagger a$ . It is useful to consider operator matrices. For example,

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix} : \begin{bmatrix} \mathcal{R}(A^*) \\ \mathcal{N}(A) \end{bmatrix} \rightarrow \begin{bmatrix} \mathcal{R}(A) \\ \mathcal{N}(A^*) \end{bmatrix},$$

where  $A_1$  is invertible. We also have

$$A^* = \begin{bmatrix} A_1^* & 0 \\ 0 & 0 \end{bmatrix} : \begin{bmatrix} \mathcal{R}(A) \\ \mathcal{N}(A^*) \end{bmatrix} \rightarrow \begin{bmatrix} \mathcal{R}(A^*) \\ \mathcal{N}(A) \end{bmatrix}$$

and

$$A^\dagger = \begin{bmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} : \begin{bmatrix} \mathcal{R}(A) \\ \mathcal{N}(A^*) \end{bmatrix} \rightarrow \begin{bmatrix} \mathcal{R}(A^*) \\ \mathcal{N}(A) \end{bmatrix}.$$

The critical point – use  $m_{\mathcal{B}}(a^* a)$  and  $\mathcal{M}_{\mathcal{B}}(a^* a)$  – is now changing: we have to use  $m_{\mathcal{B}}(A_1^* A_1) = m_{\mathcal{B}}(A^* A|_{\mathcal{R}(A^*)})$  and  $\mathcal{M}_{\mathcal{B}}(A_1^* A_1) = \mathcal{M}_{\mathcal{B}}(A^* A|_{\mathcal{N}(A^*)})$ .

**COROLLARY 2.4.** *The sequence  $(X_n)_n$ , constructed in the steepest descent method explained in (2.1), converges to  $A^\dagger + (I - A^\dagger A)X_0$  in the operator norm.*

**PROOF.** Using Theorem 2.3, we can prove that the steepest descent method produces the sequence  $(X_n)_n$  converging to  $A^\dagger + (I - A^\dagger A)X_0$  in the operator norm of  $\mathcal{L}(K, H)$ , if we can prove that

$$0 < \epsilon < \alpha_n < 2 \max \{ [\mathcal{M}_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}, [m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1} \} - \delta$$

holds for some  $\epsilon, \delta > 0$  and all  $n$ .

The restriction  $A|_{\mathcal{R}(A^*)} = A_1$  (see the notation in Remark 2.2) is invertible. Hence, we get

$$\begin{aligned} 0 < j(A_1)^2 &= \inf_{\|x\|=1} \|A_1x\|^2 = \inf_{\|x\|=1} (A_1x, A_1x) \\ &= \inf_{\|x\|=1} (A_1^*A_1x, x) = m_{\mathcal{B}}(A_1^*A_1) = m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)}), \end{aligned}$$

where  $\mathcal{B} = \mathcal{L}(\mathcal{R}(A^*))$ . Since  $r_n \in \mathcal{R}(A^*)$  for all  $n$ , we have  $\|Ar_n\| \geq j(A)\|r_n\|$  implying

$$\alpha_n = \frac{\|r_n\|^2}{\|Ar_n\|^2} \leq j(A)^{-2} = [m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}.$$

On the other hand, since  $\|Ar_n\| \leq \|A\|\|r_n\|$ , we get

$$\frac{1}{\|A\|^2} \leq \frac{\|r_n\|^2}{\|Ar_n\|^2} = \alpha_n.$$

Finally, using Lemma 2.1, we get the following

$$\begin{aligned} \|A\|^2 = \|A^*A\| &= \mathcal{M}_{\mathcal{L}(H)}(A^*A) = \mathcal{M}_{\mathcal{B}}(A_1^*A_1) \\ &= \mathcal{M}_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)}) \geq m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)}). \end{aligned}$$

Hence, the interval  $[\|A\|^{-2}, [m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}]$  is non-empty and it is contained in the interval  $[\epsilon, 2 \max \{ [\mathcal{M}_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}, [m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1} \} - \delta]$  for some  $\epsilon, \delta > 0$ . Hence, the steepest descent method introduced by Nashed in [11] is based on the construction of a certain sequence of operators converging (in the operator norm) to the Moore-Penrose inverse of a given operator.  $\square$

**REMARK 2.3.** Consider the class of gradient methods from [10]. These methods are related to the operator equation  $Au = b$ , where  $A \in \mathcal{L}(H, K)$  is relatively regular. The set of all least square solutions of this equation is denoted by  $S$ . Here  $x \in S$  if and only if  $\|Ax - b\| = \min_u \|Au - b\|$ . Such an  $x$  exists, since  $A$  is relatively regular. It

is well known that  $x \in S$  if and only if  $x = Gb$  for some  $G \in \mathcal{L}(K, H)$ , where  $G$  is an inner generalized inverse of  $A$  and  $AG$  is selfadjoint (compare this fact with the one in Remark 2.1). For an arbitrary  $x \in H$ ,  $P_S(x)$  is defined as  $P_S(x) = y$ , where

$$\|x - y\| = \min\{\|x - v\| : v \in S\}.$$

Such an  $y$  exists, since  $S$  is closed and convex set. Moreover, there exists the unique element  $x^* \in S$  of minimal norm. It is also well known that  $x^* = A^\dagger b$  holds. Let  $\Delta x = x - P_S(x)$ . The following is known:  $\Delta x \in \mathcal{R}(A^*)$  and  $P_S(x) = A^\dagger b + (I - A^\dagger A)b$ . Let  $0 < \alpha < 2$  and let  $x \mapsto s(x)$  be a real function defined on  $H \setminus S$ , satisfying

$$\|A\|^{-2} \leq s(x) \leq \frac{\|A\Delta x\|^2}{\|A^*A\Delta x\|^2}, \quad x \in H \setminus S.$$

The gradient method for minimizing  $\|Ax - b\|$  is defined in the following way (see [10]):

$$(2.2) \quad x_{n+1} = \begin{cases} x_n - \alpha s(x_n)A^*(Ax_n - b); & x_n \in H \setminus S \\ x_n; & x_n \in S. \end{cases}$$

Finally, it is proved that  $\lim x_n = P_S(x_0) = A^\dagger b + (I - A^\dagger A)x_0$  holds.

This method can be considered as a construction of certain sequence of operators  $(X_n)_n$ , where  $\lim X_n b = A^\dagger b + (I - A^\dagger A)X_0 b$ . Notice that any change of  $b$  implies the change of the iterative method.

**COROLLARY 2.5.** *If  $A \in \mathcal{L}(H, K)$  is relatively regular, then the sequence  $(X_n)_n$  constructed in (2.2) converges to  $A^\dagger + (I - A^\dagger A)X_0$  in operator norm.*

**PROOF.** Since  $\Delta x \in \mathcal{R}(A^*)$ , it is easy to prove (as in Corollary 2.4) that the sequence  $\alpha s(x_n)$  is contained in the interval

$$[\epsilon, 2 \max \{[\mathcal{M}_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}, [m_{\mathcal{B}}(A^*A|_{\mathcal{R}(A^*)})]^{-1}\} - \delta]$$

for some  $\epsilon, \delta > 0$ . □

### 3. Methods based on the second order optimization

We minimize the functional  $Q(x) = \frac{1}{2}\|Ax - b\|^2$ . The gradient and Hessian of  $Q(x)$  are equal to  $\nabla Q(x_n) = A^*(Ax_n - b)$  and  $\nabla^2 Q(x_n) = A^*A$ , respectively. In view of the constant Hessian  $\nabla^2 Q(x) = A^*A$ , the Newton's optimization method (1.2) is not applicable in the minimization of  $Q(x)$  in the case when  $A^*A$  is singular. But,

it is possible to use the quasi-Newton method (1.3). This idea gives us the following iterative method of the type (1.3):

$$X_{n+1} = X_n - (\lambda_n I + A^*A)^{-1}A^*(AX_n - C).$$

This method, obviously, can be applied in  $C^*$ -algebras. We prove that this method converges to a certain generalized inverse of  $A$ . Taking some natural assumptions, we obtain that the method converges to the Moore-Penrose inverse of  $A$ .

We begin with the following result, in which we construct the method and prove the convergence.

**THEOREM 3.1.** *Suppose that  $a \in \mathcal{A}$  is relatively regular;  $x_0, c \in \mathcal{A}$  are arbitrary, and  $(\lambda_n)_n$  is a bounded sequence of positive numbers. Then the iterative method*

$$(3.1) \quad x_{n+1} = x_n - (\lambda_n + a^*a)^{-1}a^*(ax_n - c)$$

converges to  $a^\dagger c + (1 - a^\dagger a)x_0$ . Consequently,  $\lim x_n = a^\dagger c$  if and only if  $a^\dagger ax_0 = x_0$ .

**PROOF.** Let  $a^\dagger ax_0 = x_0$ . By induction on  $n$  we obtain  $a^\dagger ax_n = x_n$  for every  $n$ . Now we compute

$$\begin{aligned} a^*ax_{n+1} - a^*c &= a^*ax_n - a^*c - a^*a(\lambda_n + a^*a)^{-1}(a^*ax_n - a^*c) \\ &= \lambda_n(\lambda_n + a^*a)^{-1}(a^*ax_n - a^*c). \end{aligned}$$

Multiplying the previous equality by  $(a^*a)^\dagger$  from the left side, we get

$$x_{n+1} - a^\dagger c = \lambda_n(\lambda_n + a^*a)^{-1}(x_n - a^\dagger c).$$

Now we have the following:

$$\begin{aligned} x_{n+1} - a^\dagger c &= a^\dagger a(x_{n+1} - a^\dagger c) \\ &= \lambda_n a^\dagger a(\lambda_n + a^*a)^{-1}a^\dagger a(x_n - a^\dagger c). \end{aligned}$$

Let  $\mathcal{B} = a^\dagger a \mathcal{A} a^\dagger a$ . Since  $a^*a$  is invertible in  $\mathcal{B}$ , we know that  $m_{\mathcal{B}}(a^*a) > 0$ . Now the following holds

$$\begin{aligned} \|x_{n+1} - a^\dagger c\| &\leq \lambda_n \mathcal{M}_{\mathcal{B}}[a^\dagger a(\lambda_n + a^*a)^{-1}a^\dagger a] \|x_n - a^\dagger c\| \\ &= \frac{\lambda_n}{\lambda_n + m_{\mathcal{B}}(a^*a)} \|x_n - a^\dagger c\|. \end{aligned}$$

Since the function  $t \mapsto t[(t + m_{\mathcal{B}}(a^*a))]^{-1}$  is increasing and the sequence  $(\lambda_n)_n$  is bounded, we conclude that there exists some  $q, 0 < q < 1$ , such that for every  $n$

$$\|x_{n+1} - a^\dagger c\| \leq q \|x_n - a^\dagger c\|.$$



Hence,  $\lim x_n = a^\dagger c$ .

Suppose that  $a^\dagger ax_0 \neq x_0$ . Denote  $x' = a^\dagger ax$  and  $x'' = (1 - a^\dagger a)x$  for all  $x \in \mathcal{A}$ . We conclude that

$$x'_1 = x'_0 - (\lambda_n + a^*a)^{-1}a^*(ax_0 - c) = x'_0 - (\lambda_n + a^*a)^{-1}a^*(ax'_0 - c)$$

and  $x_1 = x'_1 + x''_0$ . By induction on  $n$  we get  $x_n = x'_n + x''_0$  for every  $n$ . From the first part we get  $\lim x'_n = a^\dagger c$  and consequently  $\lim x_n = a^\dagger c + (1 - a^\dagger a)x_0$ .  $\square$

To get the idea for the next method, consider a relatively regular operator  $A$  on Hilbert spaces. Let us consider the functional

$$Q_{\alpha_n}(x) = \|Ax - b\|^2/2 + \alpha_n \|x\|^2/2$$

where  $\alpha_n$  is a real quantity. If  $x_{\alpha_n}$  is the minimizer of  $Q_{\alpha_n}(x)$ , then  $\lim_{\alpha_n \rightarrow 0^+} x_{\alpha_n} = A^\dagger b$ . It is not difficult to verify the following:

$$\nabla Q_{\alpha_n}(x_n) = A^*(Ax_n - b) + \alpha_n x_n, \quad \nabla^2 Q_{\alpha_n}(x_n) = A^*A + \alpha_n \mathbf{I}.$$

Similarly, according to the quasi-Newton method (1.3), we give the iterative method which produces the following approximations  $x_n$  of the minimizer  $x_{\alpha_n}$

$$x_{n+1} = x_n - ((\lambda_n + \alpha_n)\mathbf{I} + A^*A)^{-1} [A^*(Ax_n - b) + \alpha_n x_n], \quad n = 0, 1, \dots$$

For fixed values  $\lambda_n = \lambda$ ,  $n = 0, 1, \dots$ , the approximations  $x_n$  are of the form

$$x_{n+1} = x_n - ((\lambda + \alpha_n)\mathbf{I} + A^*A)^{-1} [A^*(Ax_n - b) + \alpha_n x_n], \quad n = 0, 1, \dots$$

This idea suggests to define the following iterative method in  $C^*$ -algebras

$$x_{n+1} = x_n - (\beta_n + a^*a)^{-1} [a^*(ax_n - 1) + \alpha_n x_n] \quad (n \geq 0),$$

where  $(\alpha_n)_n$  strongly decreases to 0 and  $\beta_n > \alpha_n$  for sufficiently large  $n$ .

In order to prove the convergence of the proposed method, we formulate an auxiliary result from [4, Theorem 3.3 (c)].

**LEMMA 3.2.** *If 0 is not the point of accumulation of  $\sigma_{\mathcal{A}}(a^*a)$ , then*

$$\lim_{\lambda \rightarrow 0} (\lambda + a^*a)^{-1} a^* = \lim_{\lambda \rightarrow 0} a^* (\lambda + aa^*)^{-1} = a^\dagger.$$

Actually, the result in [4, Theorem 3.3 (c)] is proved in the case when  $a^* = a^*a(a^*a)^\#a^* = a^\dagger aa^*$ , which always holds according to our Lemma 2.2 (6).

The convergence result follows.

**THEOREM 3.3.** *Let  $a \in \mathcal{A}$  be relatively regular, let  $(\alpha_n)_n$  be a sequence strongly decreasing to 0 and let  $(\beta_n)_n$  be a bounded sequence of positive numbers such that  $\beta_n - \alpha_n > 0$  for all  $n$ . Consider the iterative method*

$$(3.2) \quad x_{n+1} = x_n - (\beta_n + a^*a)^{-1}(a^*ax_n - a^* + \alpha_nx_n).$$

There are two possible cases.

- (a) *If  $a^\dagger ax_0 = x_0$ , then  $\lim x_n = a^\dagger$ .*
- (b) *If  $a^\dagger ax_0 \neq x_0$ , then  $\lim x_n = a^\dagger + e^{\sum_{n=0}^{\infty} \ln(1-\alpha_n\beta_n^{-1})}(1 - a^\dagger a)x_0$ . In this case  $\lim x_n = a^\dagger$  if and only if the series  $\sum(\alpha_n/\beta_n)$  is divergent.*

**PROOF.** Let  $y_n = (\alpha_n + a^*a)^{-1}a^*$ . From Lemma 3.2 we get that  $\lim y_n = a^\dagger$ . Notice that

$$\begin{aligned} x_{n+1} &= x_n - (\beta_n + a^*a)^{-1}(\alpha_n + a^*a)[x_n - (\alpha_n + a^*a)^{-1}a^*] \\ &= x_n - (\beta_n + a^*a)^{-1}(\alpha_n + a^*a)(x_n - y_n). \end{aligned}$$

Now we compute

$$\begin{aligned} x_{n+1} - y_n &= x_n - y_n - (\beta_n + a^*a)^{-1}(\alpha_n + a^*a)(x_n - y_n) \\ &= (\beta_n + a^*a)^{-1}[\beta_n + a^*a - (\alpha_n + a^*a)](x_n - y_n) \\ &= (\beta_n - \alpha_n)(\beta_n + a^*a)^{-1}(x_n - y_n). \end{aligned}$$

Consequently, we get

$$x_{n+1} - y_n = \left( \frac{\beta_n}{\beta_n - \alpha_n} + \frac{a^*a}{\beta_n - \alpha_n} \right)^{-1} (x_n - y_n).$$

(a) Suppose that  $a^\dagger ax_0 = x_0$ . Notice that  $a^\dagger ay_n = y_n$  holds for all  $n$ . By induction on  $n$  we get that  $a^\dagger ax_n = x_n$  is satisfied for all  $n$ . Hence we have

$$x_{n+1} - y_n = a^\dagger a(x_{n+1} - y_n) = a^\dagger a \left( \frac{\beta_n}{\beta_n - \alpha_n} + \frac{a^*a}{\beta_n - \alpha_n} \right)^{-1} a^\dagger a(x_n - y_n).$$

Again, let  $\mathcal{B} = a^\dagger a\mathcal{A}a^\dagger a$ ; whence  $a^*a$  is invertible in  $\mathcal{B}$  and  $m_{\mathcal{B}}(a^*a) > 0$ . Notice that

$$\begin{aligned} \left\| a^\dagger a \left( \frac{\beta_n}{\beta_n - \alpha_n} + \frac{a^*a}{\beta_n - \alpha_n} \right)^{-1} a^\dagger a \right\| &= \mathcal{M}_{\mathcal{B}} \left( a^\dagger a \left( \frac{\beta_n}{\beta_n - \alpha_n} + \frac{a^*a}{\beta_n - \alpha_n} \right)^{-1} a^\dagger a \right) \\ &= \frac{\beta_n - \alpha_n}{\beta_n + m_{\mathcal{B}}(a^*a)} \leq \frac{\beta_n}{\beta_n + m_{\mathcal{B}}(a^*a)}. \end{aligned}$$

The function  $t \mapsto t[t + m_{\mathcal{B}}(a^*a)]^{-1}$  is increasing and the sequence  $(\beta_n)_n$  is bounded. Hence,

$$\left\| a^\dagger a \left( \frac{\beta_n}{\beta_n - \alpha_n} + \frac{a^*a}{\beta_n - \alpha_n} \right)^{-1} a^\dagger a \right\| \leq q < 1$$

is satisfied for all  $n$ . For an arbitrary  $\epsilon > 0$ , there exists some  $n_0$  such that  $\|y_n - a^\dagger\| < \epsilon$  holds for all  $n \geq n_0$ . This implies

$$\begin{aligned} \|x_{n+1} - a^\dagger\| &\leq \|x_{n+1} - y_n\| + \|y_n - a^\dagger\| \\ &\leq q\|x_n - y_n\| + \|y_n - a^\dagger\| \leq q\|x_n - a^\dagger\| + (1 + q)\|y_n - a^\dagger\| \\ &\leq q\|x_n - a^\dagger\| + (1 + q)\epsilon \leq q^{n-n_0+1}\|x_{n_0} - a^\dagger\| + \epsilon(1 + q)/(1 - q). \end{aligned}$$

We see that  $\lim x_n = a^\dagger$ .

(b) Now suppose that  $a^\dagger ax_0 \neq x_0$ . Define  $x' = a^\dagger ax$  and  $x'' = (1 - a^\dagger a)x$  for all  $x \in \mathcal{A}$ . Now for all  $n$ , we have  $x'_{n+1} = x'_n - (\beta_n + a^*a)^{-1}(\alpha_n + a^*a)(x'_n - y_n)$ . According to the first part we know that  $\lim x'_n = a^\dagger$ . Notice that

$$x''_{n+1} = (\beta_n - \alpha_n)(\beta_n + a^*a)^{-1}x''_n.$$

From

$$\begin{aligned} \beta_n + a^*a &= (\beta_n + a^*a)a^\dagger a + (\beta_n + a^*a)(1 - a^\dagger a) \\ &= (\beta_n + a^*a)a^\dagger a + \beta_n(1 - a^\dagger a), \end{aligned}$$

it is easy to verify  $(\beta_n + a^*a)^{-1} = [(\beta_n + a^*a)a^\dagger a]_{\mathcal{B}}^{-1} + (1 - a^\dagger a)/\beta_n$ . Here  $[(\beta_n + a^*a)]_{\mathcal{B}}^{-1}$  is the ordinary inverse of  $(\beta_n + a^*a)a^\dagger a$  in the algebra  $\mathcal{B}$ . Finally, we get

$$x''_{n+1} = \frac{\beta_n - \alpha_n}{\beta_n}x''_n = \left(\prod_{k=0}^n \left(1 - \frac{\alpha_k}{\beta_k}\right)\right)x''_0.$$

The product  $\prod_{k=0}^\infty (1 - \alpha_k/\beta_k)$  is convergent if and only if the series  $\sum_{k=0}^\infty \ln(1 - \alpha_k/\beta_k)$  is convergent, that is, if and only if the series  $\sum_{k=0}^\infty (\alpha_k/\beta_k)$  is convergent.

If the series  $\sum_{k=0}^\infty \ln(1 - \alpha_k/\beta_k)$  is divergent, then its sum is equal to  $-\infty$ . In this case  $\prod_{k=0}^\infty (1 - \alpha_k/\beta_k) = 0$  and the statement (b) follows.  $\square$

**REMARK 3.1.** Notice that the limit  $d = a^\dagger + e^{\sum_{n=0}^\infty \ln(1 - \alpha_n/\beta_n)}(1 - a^\dagger a)x_0$  satisfies (again)  $ada = a$  and  $ad = aa^\dagger$  is selfadjoint.

**REMARK 3.2.** If  $\alpha_{n_0} = \beta_{n_0}$  for some  $n_0$ , then  $x_{n_0+1} = y_{n_0}$ . In this case

$$\prod_{n=0}^\infty (1 - \alpha_n/\beta_n) = 0 \quad \text{and} \quad \lim x_n = a^\dagger.$$

If  $\alpha_n = \beta_n$  for all  $n$ , then  $x_n = y_n$  for all  $n$  and  $\lim x_n = a^\dagger$ .

We can obtain the following norm estimate.

**THEOREM 3.4.** *If the conditions from Theorem 3.3 are satisfied, then the following holds:*

$$\|x_n - a^\dagger\| < \beta_{n-1} \|(\beta_{n-1} + a^*a)^{-1}\| [\|a^\dagger\| + \|x_{n-1}\|].$$

**PROOF.** Using Lemma 2.2 (6) we get

$$\begin{aligned} x_n - a^\dagger &= x_{n-1} - (\beta_{n-1} + a^*a)^{-1} (a^*ax_{n-1} - a^* + \alpha_{n-1}x_{n-1}) - a^\dagger \\ &= (\beta_{n-1} + a^*a)^{-1} [(\beta_{n-1} + a^*a)(x_{n-1} - a^\dagger) + a^* - a^*ax_{n-1} - \alpha_{n-1}x_{n-1}] \\ &= (\beta_{n-1} + a^*a)^{-1} [(\beta_{n-1} - \alpha_{n-1})x_{n-1} - \beta_{n-1}a^\dagger]. \end{aligned}$$

This implies

$$\|x_n - a^\dagger\| \leq \|(\beta_{n-1} + a^*a)^{-1}\| [(\beta_{n-1} - \alpha_{n-1})\|x_{n-1}\| + \beta_{n-1}\|a^\dagger\|].$$

Since  $\alpha_n$  is a decreasing sequence converging to 0, we have

$$\|x_n - a^\dagger\| < \beta_{n-1} \|(\beta_{n-1} + a^*a)^{-1}\| [\|a^\dagger\| + \|x_{n-1}\|]$$

which completes the proof. □

### 4. Numerical results

All numerical examples are derived using the stopping criterion  $\|X_1 - X_0\| \leq \epsilon$ , where  $\epsilon$  is a given small real number and  $X_0, X_1$  are two successive approximations.

**EXAMPLE 1.** Consider the matrix

$$A = \begin{bmatrix} 3 & 1 & 4 & 9 \\ 1 & 2 & 3 & 4 \\ 0 & -2 & -2 & 0 \\ -1 & 0 & -1 & -4 \end{bmatrix}.$$

We use the starting approximation  $X_{S1} = A^T$  which satisfies  $A^\dagger AX_0 = X_0$  and the starting approximation

$$X_{S2} = \begin{bmatrix} 3 & 1 & 4 & 9 \\ 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 9 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

which does not satisfy  $A^\dagger AX_0 = X_0$ .

(A) Let us choose starting value  $\lambda_0 = 0.01$  and the rule  $\lambda_n = \lambda_{n-1}/2$ . Using  $X_0 = X_{s_1}$  after 7 iterations determined by (3.1) we get the following approximation of the Moore-Penrose inverse  $A^\dagger$  with the precision  $\|X_1 - X_0\| = 0.7784079078267431 \times 10^{-12}$ :

$$\begin{bmatrix} 0.888888888888325 & -0.870370370370752 & -0.259259259259112 & 1.12962962962986 \\ -0.444444444445008 & 0.518518518518137 & -0.0370370370368896 & -0.481481481481247 \\ 0.444444444445008 & -0.35185185185147 & -0.296296296296444 & 0.648148148147914 \\ -0.333333333333333 & 0.388888888888889 & 0.222222222222222 & -0.611111111111111 \end{bmatrix}.$$

Using  $X_0 = X_{s_2}$  after 20 iterations, (3.1) gives the following approximation of the  $\{1, 3\}$  inverse for  $A$  after 8 iterative steps and with the precision  $0.2151594247836613 \times 10^{-14}$ :

$$\begin{bmatrix} 1.22222222222194 & -0.203703703703917 & 0.740740740740222 & 2.46296296296203 \\ -0.111111111111389 & 1.18518518518497 & 0.962962962962444 & 0.851851851850917 \\ 0.111111111111389 & -1.0185185185183 & -1.29629629629578 & -0.68518518518425 \\ -0.333333333333333 & 0.388888888888889 & 0.222222222222222 & -0.611111111111111 \end{bmatrix}.$$

(B) We apply iterative method (3.2) with  $\alpha_0 = 0.5$ ,  $\beta_0 = 1.5$ ,  $\alpha_1 = 0.5$ ,  $\beta_1 = 1.5$  and  $\alpha_n = 1/2^n$ ,  $\beta_n = n^2 * \alpha_n$ ,  $n > 1$ . In the case  $X_0 = X_{s_1}$  we get the following approximation for  $A^\dagger$  with the precision  $0.8145858162803369 \times 10^{-10}$  after 38 iterations:

$$\begin{bmatrix} 0.8888888888574836 & -0.8703703703390785 & -0.2592592592486855 & 1.1296296295877641 \\ -0.4444444442951403 & 0.5185185185035897 & -0.0370370370419484 & -0.481481481461641 \\ 0.44444444427987 & -0.35185185183547824 & -0.29629629629063636 & 0.6481481481261144 \\ -0.3333333331825743 & 0.3888888887388284 & 0.2222222221706717 & -0.6111111110909501 \end{bmatrix}.$$

In the case  $X_0 = X_{s_2}$  the method produces the following approximation for  $A^{(1,3)}$  with the precision  $0.9201250315032129 \times 10^{-10}$  after 87 iterative steps:

$$\begin{bmatrix} 1.002140322865836 & -0.6438675024175026 & 0.0804950426702385 & 1.5826353655360357 \\ -0.33119301046752425 & 0.7450213864713654 & 0.3027172648924681 & -0.028475745575009752 \\ 0.33119301046752425 & -0.5783547198046588 & -0.6360505982258156 & 0.19514241224155002 \\ -0.3333333333333337 & 0.3888888888888889 & 0.2222222222222222 & -0.6111111111111112 \end{bmatrix}.$$

(C) We now consider the case  $\alpha_0 = 1$ ,  $\beta_0 = 2$ ,  $\alpha_1 = 0.5$ ,  $\beta_1 = 1.5$ ,  $\alpha_2 = 0.25$ ,  $\beta_2 = 1.25$ ,  $\alpha_n = 1/2^n$ ,  $\beta_n = 1/n^2$ ,  $n > 2$ . In the case  $X_0 = X_{s_1}$ , iterative method (3.2) produces the following approximation for  $A^\dagger$  with the precision  $0.8179146910995648 \times 10^{-10}$  after 38 iterations:

$$\begin{bmatrix} 0.8888888888573743 & -0.87037037033896 & -0.2592592592486474 & 1.1296296295876076 \\ -0.444444444294535 & 0.5185185185035387 & -0.037037037041967674 & -0.48148148146157094 \\ 0.4444444442792064 & -0.3518518518354216 & -0.2962962962906151 & 0.6481481481260366 \\ -0.3333333331820203 & 0.3888888887382767 & 0.2222222221704846 & -0.6111111110908761 \end{bmatrix}.$$

In the case  $X_0 = X_{s_2}$  the method produces the following approximation for  $A^{(1,3)}$  with the precision  $0.667309429022048 \times 10^{-8}$  after 38 iterations:

$$\begin{bmatrix} 0.8885296579803937 & -0.8710888320929212 & -0.26033695187958905 & 1.1281927060796855 \\ -0.444803675306434 & 0.5178000567495776 & -0.03811472967290933 & -0.4829184049694931 \\ 0.44480367530490117 & -0.3511339008146044 & -0.2952186036596735 & 0.6495850716339587 \\ -0.3333333331820203 & 0.3888888887382767 & 0.2222222221704846 & -0.6111111110908761 \end{bmatrix}.$$

(D) Let us assume  $\alpha_0 = \alpha_1 = 1/5$ ,  $\beta_0 = \beta_1 = 1/2$ ,  $\alpha_n = 1/5^n$ ,  $\beta_n = 1/2^n$ ,  $n > 1$ . In the case  $X_0 = X_{S_1}$ , iterative method (3.2) produces the following approximation for  $A^\dagger$  with the precision  $0.5871735621397724 \times 10^{-10}$  after 18 iterations:

$$\begin{bmatrix} 0.888888888832127 & -0.8703703703647123 & -0.25925925925735943 & 1.1296296296220825 \\ -0.4444444444177217 & 0.5185185185158488 & -0.03703703703792724 & -0.4814814814779111 \\ 0.4444444444149706 & -0.3518518518489219 & -0.2962962962952712 & 0.6481481481441825 \\ -0.33333333333061704 & 0.3888888888861852 & 0.2222222222129343 & -0.6111111111074786 \end{bmatrix}.$$

In the case  $X_0 = X_{S_2}$  the method produces the following approximation for  $A^{(1,3)}$  with the precision  $0.6435232820789964 \times 10^{-10}$  after 27 iterations:

$$\begin{bmatrix} 1.039509072508886 & -0.5691300031303874 & 0.19260129160073503 & 1.732110364109587 \\ -0.29382426082444757 & 0.8197588857585015 & 0.4148235138229573 & 0.1209992529984758 \\ 0.29382426082444757 & -0.6530922190918348 & -0.7481568471562906 & 0.04566741366819088 \\ -0.3333333333333337 & 0.3888888888888889 & 0.2222222222222224 & -0.6111111111111112 \end{bmatrix}.$$

(E) Now consider the choice  $\alpha_0 = 1/2$ ,  $\beta_0 = 1/2 + 1/5$ ,  $\alpha_n = 1/5^n$ ,  $\beta_n = 1/2^n + 1/5^n$ ,  $n \geq 1$ . In the case  $X_0 = X_{S_1}$ , iterative method (3.2) produces the following approximation for  $A^\dagger$  with the precision  $0.8145866914197384 \times 10^{-10}$  after 38 iterations:

$$\begin{bmatrix} 0.8888888888574894 & -0.8703703703390749 & -0.2592592592486863 & 1.1296296295877613 \\ -0.44444444442950826 & 0.5185185185035932 & -0.03703703704194922 & -0.48148148146164393 \\ 0.44444444442798114 & -0.3518518518354818 & -0.29629629629063553 & 0.6481481481261173 \\ -0.33333333331825743 & 0.3888888888738829 & 0.22222222221706717 & -0.61111111110909501 \end{bmatrix}.$$

Since the series  $\alpha_n/\beta_n$  is convergent, in the case  $X_0 = X_{S_2}$  we again get an approximation for  $A^\dagger$ . Numerical precision  $0.8145866914197384 \times 10^{-10}$  is achieved after 38 iterations, and the result is

$$\begin{bmatrix} 0.8888888888574894 & -0.8703703703390749 & -0.2592592592486863 & 1.1296296295877613 \\ -0.44444444442950826 & 0.5185185185035932 & -0.03703703704194922 & -0.48148148146164393 \\ 0.44444444442798114 & -0.3518518518354818 & -0.29629629629063553 & 0.6481481481261173 \\ -0.33333333331825743 & 0.3888888888738829 & 0.22222222221706717 & -0.61111111110909501 \end{bmatrix}.$$

(F) For the choice  $\alpha_0 = 1/2^n$  and  $\beta_n = 5$ ,  $n \geq 0$ , in the case  $X_0 = X_{S_1}$  method (3.2) gives the approximation for  $A^\dagger$  with the precision  $0.975862514147117 \times 10^{-8}$  after 406 iterations:

$$\begin{bmatrix} 0.88888794783812 & -0.870370276596845 & -0.259259227496198 & 1.12962950409304 \\ -0.44444399761835 & 0.518518473993339 & -0.0370370521186507 & -0.481481421874688 \\ 0.44444395021972 & -0.351851802603508 & -0.296296279614848 & 0.648148082218357 \\ -0.33333288084808 & 0.38888843799784 & 0.22222206949596 & -0.61111105074938 \end{bmatrix}.$$

The series  $\alpha_n/\beta_n$  is convergent, so that in the case  $X_0 = X_{S_2}$  we also get an approximation for  $A^\dagger$ . Numerical precision  $0.9736262652199987 \times 10^{-8}$  is achieved after 402 iterations, and the result is

$$\begin{bmatrix} 0.913523867929914 & -0.821100111622034 & -0.185353982552945 & 1.22816982525257 \\ -0.419809325997662 & 0.56778861265089 & 0.03686815706372 & -0.382941140385912 \\ 0.41980932123669 & -0.401121940362275 & -0.370201487575917 & 0.549607802084412 \\ -0.333332878833 & 0.38888835219648 & 0.22222195290585 & -0.611111063683115 \end{bmatrix}.$$

**REMARK 4.1.** (i) In (3.2) the convergence is ensured for arbitrary positive and bounded values of the parameters  $\beta_n > \alpha_n \rightarrow 0, n = 0, 1, \dots$ . For larger values of the parameter  $\beta_n$ , iterations (3.2) behave like the first order gradient method, because of  $(\beta_n \mathbf{I} + A^*A)^{-1} \approx (1/\beta_n) \nabla Q(x_n)$ . In the case  $\beta_n \rightarrow 0_+$  we have  $\beta_n \mathbf{I} + A^*A \approx A^*A$ , so these iterative processes converge according to the Newton's method. In view of the global convergence of the steepest descent and the fast local convergence of the Newton's method, we suggest successive decreasing values for  $\beta_n, n = 0, 1, \dots$ , during the iterations.

(ii) Since  $\|x_{n+1} - x_n\| = \|(\beta_n \mathbf{I} + A^*A)^{-1} A^*(Ax_n - b)\|$ , the condition  $\|x_{n+1} - x_n\| \leq \epsilon$ , where  $\epsilon$  is a small real number, can be used as the stopping criterion for the iterative process (3.2).

(iii) Using the known result about the convergence of the quasi-Newton method from [3], we conclude the following: As long as  $\{\lambda_n\}$  converge to zero, iterations (3.2) are superlinearly convergent. Moreover, if  $\lambda_n \leq \mu \|\nabla Q(x_n)\|$  for some constant  $\mu$  for all sufficiently large  $n$ , then (3.2) converge quadratically.

**REMARK 4.2.** Effectiveness of methods based on the second order optimization with respect to the steepest descent method (2.1) is evident. For the chosen matrix  $A$ , in case when  $X_0 = X_{S_1}$  we get the following approximation of the Moore-Penrose inverse after 2850 iterative steps with the precision  $0.9984792937219064 \times 10^{-8}$ :

$$\begin{bmatrix} 0.88888887107898 & -0.870370352568972 & -0.259259253239727 & 1.1296296058087 \\ -0.444444435950256 & 0.518518510085312 & -0.0370370398993667 & -0.481481470185946 \\ 0.444444435128724 & -0.351851842483659 & -0.296296293139094 & 0.648148135622755 \\ -0.333333324635957 & 0.388888880397453 & 0.222222219313109 & -0.611111099710559 \end{bmatrix}.$$

Using  $X_0 = X_{S_2}$  we get the following approximation of  $\{1, 3\}$ -inverse after 5650 iterative steps with the precision  $0.9985799594416307 \times 10^{-8}$ :

$$\begin{bmatrix} 1.2222222979886 & -0.203703699773892 & 0.74074075406113 & 2.46296299115256 \\ -0.111111114702461 & 1.18518518332474 & 0.962962956650109 & 0.851851838488027 \\ 0.111111115096401 & -1.01851851644915 & -1.29629628928876 & -0.685185170359436 \\ -0.333333336954566 & 0.388888887018803 & 0.222222215859508 & -0.611111124590927 \end{bmatrix}.$$

## References

- [1] A. Ben-Israel and D. Cohen, 'On iterative computation of generalized inverses and associated projections', *SIAM J. Numer. Anal.* **3** (1966), 410–419.
- [2] S. R. Caradus, *Generalized inverses and operator theory*, Queen's Papers in Pure and Appl. Math. (Queen's University, Kingston, ON, 1978).
- [3] J. E. Dennis and J. J. Moré, 'Quasi-Newton methods, motivation and theory', *SIAM Rev.* (1) **19** (1977), 46–89.

- [4] D. S. Djordjević and P. S. Stanimirović, 'On the generalized Drazin inverse and generalized resolvent', *Czech. Math. J.* **51** (2001), 617–634.
- [5] C. W. Groetsch, *Generalized inverses of linear operators* (Marcel Dekker, New York, 1977).
- [6] R. E. Harte and M. Mbekhta, 'On generalized inverses in  $C^*$ -algebras', *Studia Math.* **103** (1992), 71–77.
- [7] W. J. Kammerer and M. Z. Nashed, 'On the convergence of the conjugate gradient method for singular linear operator equations', *SIAM Rev. (I)* **9** (1972), 165–181.
- [8] J. J. Koliha, 'A generalized Drazin inverse', *Glasgow Math. J.* **38** (1996), 367–381.
- [9] L. J. Lardy, 'A class of iterative methods of conjugate gradient type', *Numer. Funct. Anal. Optim.* **11** (1990), 283–302.
- [10] S. F. McCormick and G. H. Rodrigue, 'A uniform approach to gradient methods for linear operator equations', *J. Math. Anal. Appl.* **49** (1975), 275–285.
- [11] M. Z. Nashed, 'Steepest descent for singular linear operators equations', *SIAM J. Numer. Anal.* **7** (1970), 358–362.
- [12] V. Rakočević, *Functional analysis* (Naučna Knjiga, Belgrade, 1994) (in Serbian).
- [13] K. Tanabe, 'Conjugate-gradient method for computing the Moore-Penrose inverse and rank of a matrix', *J. Optimization Theory Appl.* **22** (1977), 1–23.
- [14] T. M. Whitney and R. H. Meany, 'Two methods related to the method of steepest descent', *SIAM J. Numer. Anal.* **4** (1967), 109–118.

Department of Mathematics

University of Niš

Faculty of Science and Mathematics

P.O. Box 224

Višegradska 33

18000 Niš

Serbia

e-mail: [dragan@pmf.ni.ac.yu](mailto:dragan@pmf.ni.ac.yu), [pecko@pmf.ni.ac.yu](mailto:pecko@pmf.ni.ac.yu)