On the gradient based algorithm for solving the general coupled matrix equations

Davod Khojasteh Salkuyeh[1†] and Fatemeh Panjeh Ali Beik[‡]

[†]*Faculty of Mathematical Sciences, University of Guilan, Rasht, P.O. Box 1914, Iran*
email: khojasteh@guilan.ac.ir, salkuyeh@gmail.com

[‡]*Department of Mathematics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran*
email: f.beik@vru.ac.ir

Abstract

Recently, the gradient based iterative algorithms have been widely exploited for finding the (least-squares) solutions of the different kinds of the (coupled) linear matrix equations. Nevertheless, so far, the convergence of the propounded gradient based algorithms has been studied under the circumstance that the mentioned (coupled) linear matrix equations have a unique (least-squares) solution. In the present paper, we consider the consistent general coupled linear matrix equations which incorporate many of the recently investigated (coupled) linear matrix equations as their special instances. It is demonstrated that using a gradient based iterative algorithm for solving the referred coupled linear matrix equations is equivalent to extend the well-known Richardson method for solving the normal equations corresponding to the original coupled linear matrix equations. In addition, we prove the semi-convergence of the Richardson method when the coefficient matrix of the associated normal equations is singular. Finally, some numerical experiments are presented to illustrate the validity of our theoretical results.

Key words: Linear matrix equation; Iterative method; Richardson method; Semi-convergence.
2000 MSC: 15A24, 65F10.

1. Introduction

Consider the following consistent *general coupled matrix equations* (Ding and Chen, 2005, 2006)

$$\sum_{j=1}^{q} A_{ij}X_jB_{ij} = C_i, \qquad i = 1,\dots,p, \tag{1}$$

in which $A_{ij} \in \mathbb{R}^{r_i\times n_j}, B_{ij} \in \mathbb{R}^{m_j\times k_i}$ and $C_i \in \mathbb{R}^{r_i\times k_i}$ are given matrices and $X_j \in \mathbb{R}^{n_j\times m_j}$ are the unknown matrices where $i = 1,\dots,p$ and $j = 1,\dots,q$. Evidently, Equation (1) includes the well-known Sylvester equation $AX + XB = C$ and the Lyapunov equation $AX + XA^T = C$ as its special cases.

The linear matrix equations turn up in control theory, signal processing, filtering theory for continuous or discrete-time large-scale dynamical systems, model reduction, image restoration,

[1]Corresponding author
The work of the first author is supported by University of Guilan under Grant No. 904.

decoupling techniques for ordinary and partial differential equations, implementation of implicit numerical methods for ordinary differential equations, and block-diagonalization of matrices; for further details see Benner (2008), Bouhamidi and Jbilou (2008), Ding (2013), Hyland and Bernstein (1984), Jbilou et al. (1999), Jbilou and Riquet (2006), Salkuyeh and Toutounian (2006), Zhang (2011), Zhou and Duan (2008) and the references therein. In all of the above referred references, iterative algorithms have been employed to solve different kinds of linear matrix equations.

In the literature, several iterative methods have been examined for solving Equation (1) (or its particular cases) whereas their mentioned problem has a unique solution. Before presenting our main contribution, we briefly review some of the recently published works in which the considered main problem has a unique solution. For instance, Bouhamidi and Jbilou (2008) have proposed a Krylov subspace method for solving the generalized Sylvester matrix equation

$$\sum_{i=1}^{q} A_i X B_i = C, \tag{2}$$

where the matrices $A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{p\times p}$ $(i = 1,\ldots,q)$ and $C \in \mathbb{R}^{n\times p}$ are the given known matrices and the matrix $X \in \mathbb{R}^{n\times p}$ is the unknown matrix to be determined. Recently, Ding et al. (2010) have presented an iterative algorithm to solve the system of matrix equations $A_i X B_i = F_i$ for $i = 1,\ldots,p$. Ding, Liu and Ding (2008) have offered a gradient based iterative algorithm to solve $AXB + CXD = F$ where $A, C \in \mathbb{R}^{m\times m}$ and $B, D \in \mathbb{R}^{n\times n}$. Li and Wang (2010) have generalized the iterative method proposed in (Ding et al., 2008) for solving the following linear matrix equations

$$\sum_{i=1}^{r} A_i X B_i = C,$$

where $A_i \in \mathbb{R}^{p\times m}$, $B_i \in \mathbb{R}^{n\times q}$, $i = 1,\ldots,r$.

In the case that A_{ij}'s and B_{ij}'s in (1) are square matrices, Beik and Salkuyeh (2011) have extended the global FOM (Gl-FOM) and the global GMRES (Gl-GMRES) methods to resolve (1). Ding and Chen (2005, 2006) have proposed two gradient based iterative algorithms for solving (1) and finding its unique least-squares solution, respectively. In (Zhou et al., 2009a), a gradient based iterative method has been suggested for solving Equation (1) and the presented method has been extended for its general form. An extension of the gradient based method for solving a class of linear matrix equations has been offered in (Xie et al., 2009). In an alternative research work, Xie et al. (2010) have developed the gradient based algorithm for solving the matrix equation $AXB + CX^TD = F$ and also for more general and complex coupled linear matrix equations.

Suppose that $A = [a_{ij}]_{m\times s}$ and $B = [b_{ij}]_{n\times q}$ defined over complex (real) number field, the Kronecker product of the matrices A and B is specified as the $mn \times sq$ matrix $A \otimes B = [a_{ij}B]$. The "vec" operator transforms a matrix A of size $m \times s$ to a vector $a = \mathrm{vec}(A)$ of size $ms \times 1$ by stacking the columns of A. In this paper, the following relation is utilized (Bernstein, 2009)

$$\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X).$$

Let us assume that

$$\mathcal{M} = \begin{pmatrix} B_{11}^T \otimes A_{11} & B_{12}^T \otimes A_{12} & \cdots & B_{1q}^T \otimes A_{1q} \\ B_{21}^T \otimes A_{21} & B_{22}^T \otimes A_{22} & \cdots & B_{2q}^T \otimes A_{2q} \\ \vdots & \vdots & & \vdots \\ B_{p1}^T \otimes A_{p1} & B_{p2}^T \otimes A_{p2} & \cdots & B_{pq}^T \otimes A_{pq} \end{pmatrix}.$$

Straightforward computations show that the general coupled matrix equations (1) is equivalent to the following linear system of equations

$$\mathcal{M}\mathcal{X} = \mathcal{C}, \tag{3}$$

where

$$\mathcal{X} = \begin{pmatrix} \mathrm{vec}(X_1) \\ \mathrm{vec}(X_2) \\ \vdots \\ \mathrm{vec}(X_q) \end{pmatrix}, \quad \mathcal{C} = \begin{pmatrix} \mathrm{vec}(C_1) \\ \mathrm{vec}(C_2) \\ \vdots \\ \mathrm{vec}(C_p) \end{pmatrix}.$$

Obviously, the size of the coefficient matrix of the linear system (3) would be huge even for moderate values of n_j's, m_j's r_i's and k_i's ($i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$). Consequently, it is more desirable to apply an iterative method for solving the original system (1) instead of the linear system (3).

Before ending this section, in order to reveal the novelty of our results, we give a brief survey on the recently published works in the subject of this paper. In (Li and Wang, 2010), to compute a minimal norm least squares solution to (2) the authors have focused on the following problem

$$\alpha = \min_{X \in \mathbb{R}^{m\times n}} \left\{ \left\| \sum_{i=1}^{r} A_i X B_i - C \right\|_F \right\}. \tag{4}$$

Using the vec operator, an equivalent problem is considered which consists of finding X^* such that

$$f(X^*) = \min_{X \in \mathbb{R}^{m\times n}} \left\| \Upsilon \mathrm{vec}(X) - \mathrm{vec}(C) \right\|_2,$$

where $\Upsilon = \sum_{i=1}^{r} (B_i^T \otimes A_i)$. In the case that (4) has a unique solution, a gradient based algorithm has been examined. The convergence of the algorithm has been studied under the assumption that Υ is a full row (column) rank matrix. The case that Υ is neither of full column rank nor of full row rank has been left as a project to be investigated. In (Li et al., (2010)), the following problem has been mentioned

$$\min_{X \in \mathbb{R}^{m\times n}} \left\{ \left\| \sum_{i=1}^{r} A_i X B_i + \sum_{j=1}^{s} C_j X^T D_j - E \right\|_F \right\}.$$

By an analogous strategy employed by Li and Wang (2010), a gradient based algorithm together with its convergence analysis have been given for the case that the coefficient matrix obtained after exploiting the "vec(.)" operator is full rank. In (Zhou et al., 2009a,b, 2010), the proposed gradient based algorithms have been investigated under the same assumption, i.e., the coefficient matrix appears after using the "vec(.)" operator is assumed to be full rank. We sight in the gradient based algorithm in an alternative point of view. In fact, we construct the gradient based algorithm in a different way. Our varied approach helps us to investigate the semi-convergence of the algorithm. That is, we omit the restriction that the mentioned problem must have a unique solution. Hence, the presented results are more general. More precisely, it is shown that the gradient based method presented in the literature for solving (1) is mathematically equivalent to applying the well-known Richardson method for the corresponding normal equations $\mathcal{M}^T\mathcal{M}\mathcal{X} = \mathcal{M}^T\mathcal{C}$. To the best of our knowledge, the gradient based iterative algorithm for solving (1) (and its special cases) has been presented under the assumption that the matrix $\mathcal{M}^T\mathcal{M}$ is nonsingular; for further details see Ding and Chen (2005,2006), Ding et al. (2008), Ding et al. (2010) and Zhou et al. (2009a,b) and the reference therein. We prove that under a mild condition the nonsingularity of $\mathcal{M}^T\mathcal{M}$ can be disregarded.

The rest of the paper is organized as follows. In Section 2, we prove the semi-convergence of the well-known Richardson iterative method for solving the consistent linear system of equations $Ax = b$ where the coefficient matrix is singular. In Section 3, we present an iterative method by developing the Richardson iterative method to resolve the coupled linear matrix equations (1). It reveals that the obtained iterative method is equivalent to the gradient based iterative algorithm for solving (1). Numerical simulations are presented in Section 4 which illustrate the validity of the established theoretical results. Finally, the conclusion is the subject of Section 5.

2. Semi-convergence of the Richardson method

Consider the consistent linear system of equations

$$Ax = b, \tag{5}$$

where $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$ are given and $x \in \mathbb{R}^n$ is the unknown vector to be determined.

Let us assume that $A = M - N$ with M nonsingular. Generally, a basic iterative method to solve (5) is applied in the following form:

$$x(k+1) = Hx(k) + c, \tag{6}$$

where $H = M^{-1}N$ and $c = M^{-1}N$. The matrix H is called *semiconvergent* if $\lim\limits_{k\to\infty} H^k$ exists. In the case that the limit point is equal to zero then the matrix H is said to be *convergent*. It is well-known that the matrix H is convergent if and only if $\rho(H) < 1$ and is semiconvergent if and only if (Berman and Plemmons, 1994)

1. $\rho(H) = 1$;
2. $\operatorname{index}(I - H) = 1$, which means that
   $\operatorname{rank}(I - H) = \operatorname{rank}(I - H)^2$;
3. If $\mu \in \sigma(H)$ with $|\mu| = 1$, then $\mu = 1$, i.e.,
   $\upsilon(H) = \{|\mu| : \mu \in \sigma(H), \mu \neq 1\} < 1$.

In the following, we recollect some useful theorems which are utilized for studying the convergence of the Richardson method for solving (5) in general situation where the coefficient matrix A is singular.

Theorem 2.1. (Berman and Plemmons, 1994) *Let $A = M - N \in \mathbb{R}^{n\times n}$ with M nonsingular. Then for $H = M^{-1}N$ and $c = M^{-1}b$, the basic iterative method (6) converges to some solution x^* of linear system $Ax = b$ for each $x(0)$ if and only if H is semiconvergent.*

Theorem 2.2. (Berman and Plemmons, 1994) *Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Furthermore, suppose that $A = M - N$ with M nonsingular and $H = M^{-1}N$. Assume that $M^T + N$ is positive definite. Then H is semiconvergent if and only if A is positive semidefinite.*

Theorem 2.3. (Laub, 2005) *Suppose that $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and*

$$\mathcal{X} = \{x \in \mathbb{R}^n : x = \operatorname*{argmin}_{y\in\mathbb{R}^n} \|Ay - b\|_2\}.$$

Then $x \in \mathcal{X}$ if and only if $A^TAx = A^Tb$. Moreover, $x^* = A^+b$ is the unique solution of the problem

$$\min_{x\in\mathcal{X}} \|x\|_2,$$

where A^+ is the pseudoinverse of A.

We would like to comment here that the system $A^TAx = A^Tb$ is known as the normal equations and the vector $x^* = A^+b$ in Theorem 2.3 is called the least Euclidean norm solution.

Theorem 2.4. (Peng and Peng, 2006) *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \text{Range}(A)$. Suppose that the system of linear equations $Ax = b$ has a solution $x^* \in \text{Range}(A^T)$. Then x^* is an unique least Euclidean norm solution of the system of linear equations.*

In the sequel, we investigate the convergence of the Richardson method to solve the normal equations. That is, for solving the linear system $A^TAx = A^Tb$, we consider the following iterative method (Saad, 1996)

$$x(k+1) = x(k) + \mu A^T(b - Ax(k)) = (I - \mu A^TA)x(k) + \mu A^Tb, \tag{7}$$

where μ is a given positive real number.

Now, we establish the following useful theorem which presents a sufficient condition for the convergence of the iterative method (7).

Theorem 2.5. *Let $A \in \mathbb{R}^{m\times n}$ be a nonzero matrix and $b \in \text{Range}(A)$. Moreover, assume that*

$$0 < \mu < \frac{2}{\sigma^2_{\max}(A)}, \tag{8}$$

where $\sigma_{\max}$ is the largest singular value of A. Then, the iterative method (7) converges to a solution of the normal equations $A^TAx = A^Tb$ for each initial guess $x(0)$. Moreover, in the case that $x(0) \in \text{Range}(A^T)$, the iterative method (7) converges to $x^* = A^+b$. Furthermore, the optimal value of μ is given by

$$\mu_{opt} = \frac{2}{\bar{\sigma}^2_{\min}(A) + \sigma^2_{\max}(A)},$$

where $\bar{\sigma}_{\min}$ is the smallest nonzero singular value of A.

Proof. Let us assume that $A^TA = M - N$, such that

$$M = \frac{1}{\mu}I \quad \text{and} \quad N = \frac{1}{\mu}I - A^TA.$$

Evidently, the matrix $M^T + N$ is symmetric and has the following form

$$M^T + N = \frac{2}{\mu}I - A^TA.$$

The eigenvalues of $M^T + N$ are given by

$$\lambda_i = \frac{2}{\mu} - \sigma_i^2(A),$$

where $\sigma_i(A)$'s are the singular values of A. By the assumption (8), we get

$$\lambda_i = \frac{2}{\mu} - \sigma_i(A)^2 \geq \frac{2}{\mu} - \sigma^2_{\max}(A) > 0.$$

Invoking the facts that $M^T + N$ is a symmetric matrix and all of its eigenvalues are positive, we can instantly conclude that $M^T + N$ is symmetric positive definite. As A^TA is a positive semidefinite matrix, by Theorem 2.2, we deduce that the matrix $H = I - \mu A^T A$ is semiconvergent. Thus, the first part of the theorem is fulfilled by using Theorem 2.1.

To prove the second part of the theorem, suppose that $x(0) \in \text{Range}(A^T)$. Using the mathematical induction, we may show that for $k = 1, 2, \ldots$, $x(k) \in \text{Range}(A^T)$. Assume that $x(j) \in \text{Range}(A^T)$, hence $x(j) = A^T w$ for some $w \in \mathbb{R}^m$. Therefore,

$$\begin{aligned} x(j+1) &= (I - \mu A^T A)x(j) + \mu A^T b = (I - \mu A^T A)A^T w + \mu A^T b \\ &= A^T((I - \mu A A^T)w + \mu b) \in \text{Range}(A^T). \end{aligned}$$

Consequently, $x^* = \lim_{k\to\infty} x(k) \in \text{Range}(A^T)$. Now, Theorem 2.4 implies that the $x^* = A^+ b$.

To prove the last part of the theorem, we must determine the value of μ in such a way that $\nu(H)$ is minimized where $\nu(H) = \{|1 - \mu\sigma_i| : \sigma_i \neq 0\}$. From Lemma 3 in (Li and Wang, 2010), it can be concluded that

$$\mu_{\text{opt}} = \frac{2}{\bar{\sigma}^2_{\min}(A) + \sigma^2_{\max}(A)},$$

which completes the proof. $\square$

3. Richardson method for the general coupled matrix equations

Consider the coupled linear matrix equations (1). As seen, Equation (1) is equivalent to the linear system of equations (3). Using the results described in the previous section, we propose the Richardson method for solving the normal equations corresponding to the linear system of equations (3).

The normal equations associated with (3) can be written as follows:

$$\mathcal{M}^T \mathcal{M} \mathcal{X} = \mathcal{M}^T \mathcal{C}. \tag{9}$$

Therefore, the Richardson iterative method to solve the above system takes the following form

$$\mathcal{X}(k+1) = \mathcal{X}(k) + \mu \mathcal{M}^T(C - \mathcal{M}\mathcal{X}(k)), \qquad k = 0, 1, \ldots, \tag{10}$$

where $\mathcal{X}(0)$ is a given initial guess. According to Theorem 2.5, the above iterative method is convergent to $\mathcal{X}^* = \mathcal{M}^+ \mathcal{C}$, if

$$\mathcal{X}(0) \in \text{Range}(\mathcal{M}^T) \quad \text{and} \quad 0 < \mu < \frac{2}{\sigma^2_{\max}(\mathcal{M})}.$$

In addition, the optimum value of μ is given by

$$\mu_{opt} = \frac{2}{\bar{\sigma}^2_{\min}(\mathcal{M}) + \sigma^2_{\max}(\mathcal{M})}.$$

Let us consider the iterative method (10). The (i, j)th block of $\mathcal{M}^T \mathcal{M}$ can be simplified as follows:

$$\begin{aligned} (\mathcal{M}^T \mathcal{M})_{ij} &= \begin{pmatrix} B_{1i} \otimes A^T_{1i} & \cdots & B_{pi} \otimes A^T_{pi} \end{pmatrix} \begin{pmatrix} B^T_{1j} \otimes A_{1j} \\ \vdots \\ B^T_{pj} \otimes A_{pj} \end{pmatrix} = \sum_{r=1}^{p} (B_{ri} \otimes A^T_{ri})(B^T_{rj} \otimes A_{rj}) \\ &= \sum_{r=1}^{p} (B_{ri} B^T_{rj} \otimes A^T_{ri} A_{rj}), \qquad i, j = 1, \ldots, q. \end{aligned}$$

For $i = 1, \ldots, q$, we represent the ith block row of $\mathcal{M}^T\mathcal{M}$ by $(\mathcal{M}^T\mathcal{M})_{i,:}$. The ith block of $\mathcal{M}^T\mathcal{M}\mathcal{X}$ is given by

$$\begin{aligned}
(\mathcal{M}^T\mathcal{M}\mathcal{X})_i &= (\mathcal{M}^T\mathcal{M})_{i,:}\mathcal{X} = \sum_{j=1}^{q}\left(\sum_{r=1}^{p}\left(B_{ri}B_{rj}^T \otimes A_{ri}^T A_{rj}\right)\mathrm{vec}(X_j)\right) \\
&= \sum_{j=1}^{q}\left(\sum_{r=1}^{p}\mathrm{vec}\left(\left(A_{ri}^T A_{rj}\right) X_j \left(B_{rj}B_{ri}^T\right)\right)\right) = \mathrm{vec}\left(\sum_{j=1}^{q}\sum_{r=1}^{p}(A_{ri}^T A_{rj})X_j(B_{rj}B_{ri}^T)\right) \\
&= \mathrm{vec}\left(\sum_{r=1}^{p}A_{ri}^T\left(\sum_{j=1}^{q}A_{rj}X_jB_{rj}\right)B_{ri}^T\right). \qquad (11)
\end{aligned}$$

On the other hand, the ith block of $\mathcal{M}^T C$ can be written as

$$(\mathcal{M}^T C)_i = \sum_{r=1}^{p}\left(B_{ri} \otimes A_{ri}^T\right)\mathrm{vec}(C_r) = \mathrm{vec}\left(\sum_{r=1}^{p}A_{ri}^T C_r B_{ri}^T\right). \qquad (12)$$

Assume that $(\mathcal{X})_i$ stands for the ith block of $\mathcal{X}$. Substituting Equations (11) and (12) in Equation (10), it can be verified that

$$\begin{aligned}
(\mathcal{X}(k+1))_i &= (\mathcal{X}(k))_i + \mu\left((\mathcal{M}^T C)_i - (\mathcal{M}^T\mathcal{M}\mathcal{X}(k))_i\right) \\
&= (\mathcal{X}(k))_i + \mu\left(\mathrm{vec}\left(\sum_{r=1}^{p}A_{ri}^T C_r B_{ri}^T - \sum_{r=1}^{p}A_{ri}^T\left(\sum_{j=1}^{q}A_{rj}X_j(k)B_{rj}\right)B_{ri}^T\right)\right).
\end{aligned}$$

Or equivalently,

$$X_i(k+1) = X_i(k) + \mu\sum_{r=1}^{p}A_{ri}^T\left(C_r - \sum_{j=1}^{q}A_{rj}X_j(k)B_{rj}\right)B_{ri}^T, \quad i = 1, \ldots, q. \qquad (13)$$

For arbitrary given matrices $W_j \in \mathbb{R}^{r_j \times k_j}$, $j = 1, \ldots, p$, we set

$$w = \begin{pmatrix} \mathrm{vec}(W_1) \\ \vdots \\ \mathrm{vec}(W_p) \end{pmatrix}.$$

It is not difficult to see that

$$\left(\mathcal{M}^T w\right)_i = \sum_{j=1}^{p}\mathrm{vec}\left(A_{ji}^T W_j B_{ji}^T\right) = \mathrm{vec}\left(\sum_{j=1}^{p}A_{ji}^T W_j B_{ji}^T\right), \qquad i = 1, 2, \ldots, q.$$

Therefore, in order to choose $\mathcal{X}(0) = \left((\mathrm{vec}(X_1(0)))^T, \ldots, (\mathrm{vec}(X_q(0)))^T\right)^T$ such that $\mathcal{X}(0) \in \mathrm{Range}(\mathcal{M}^T)$, it is sufficient to set

$$X_i(0) = \sum_{j=1}^{p}A_{ji}^T W_j B_{ji}^T, \qquad i = 1, \ldots, q. \qquad (14)$$

Consequently, the iterative method (13) with the initial guess of the form (14) and under the following condition

$$0 < \mu < \frac{2}{\sigma^2_{\max}(\mathcal{M})}, \tag{15}$$

converges to the solution $\mathcal{X}^* = \mathcal{M}^+\mathcal{C}$ of the linear system of equations (3).

In the case that the matrix $\mathcal{M}$ is of full rank we can simplify our results in the next corollary which have already been presented in some papers; See for example (Zhou et al. 2008, 2009a, 2009b, 2010).

Corollary 3.1. *Assume that Equation (1) has a unique solution. Moreover, let*

$$0 < \mu < \frac{2}{\sigma^2_{\max}(\mathcal{M})}, \tag{16}$$

where $\sigma_{\max}$ is the largest singular value of $\mathcal{M}$. Then, the iterative method (7) converges to the unique solution for every initial guess $x(0)$. Furthermore, the optimal value of μ is given by

$$\mu_{opt} = \frac{2}{\sigma_1^2 + \sigma_r^2},$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero singular values of $\mathcal{M}$.

Remark 3.1. *(Zhou et al. 2008, 2009a, 2009b, 2010) Commonly, the upper bound in (15) can be too expensive to compute. Therefore, since*

$$\sigma_{\max}(\mathcal{M})^2 \leq \|\mathcal{M}\|_F^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \|B_{ij}^T \otimes A_{ij}\|_F^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \|B_{ij}\|_F^2 \|A_{ij}\|_F^2 =: R,$$

one may use

$$0 < \mu < \frac{2}{R},$$

instead of Equation (15) which is easy to compute.

4. Numerical examples

In this section, we present two numerical examples to show the effectiveness of the proposed algorithm. All the numerical experiments presented in this section were computed in double precision with some MATLAB codes.

Example 4.1. In this example, we consider the following matrix equation

$$A_{11}X_1B_{11} + A_{12}X_2B_{12} = C_1, \tag{17}$$

where

$$A_{11} = \begin{pmatrix} 1 & 7 & 3 \\ -3 & -5 & 1 \\ 4 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \quad B_{11} = \begin{pmatrix} -1 & -1 \\ 2 & 1 \\ -2 & 2 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 1 & 2 \\ -3 & 2 \end{pmatrix},$$

$$B_{12} = \begin{pmatrix} 2 & 4 \\ -1 & -2 \\ 1 & 2 \end{pmatrix}, \quad C_1 = \begin{pmatrix} -16 & 22 \\ 4 & 16 \\ 8 & -12 \\ -19 & -9 \end{pmatrix}.$$

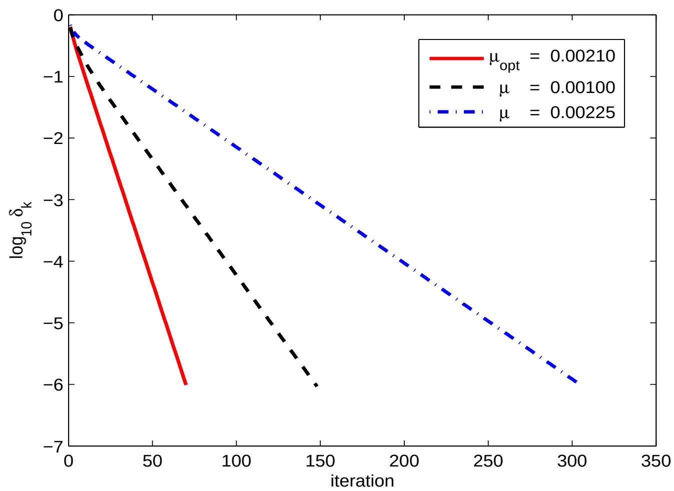

Figure 1: Convergence history for Example 1.

It is not difficult to inspect that the corresponding matrix $\mathcal{M}$ is of order 8×15 with $\mathrm{rank}(\mathcal{M}) = 7$. On the other hand, (X_1, X_2) with

$$X_1 = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 2 & 1 & -1 \\ -1 & 1 & 0 \end{pmatrix},$$

is a solution of Equation (17). Hence, this equation has an infinitely number of solutions. We apply the gradient based algorithm to solve Equation (17). The initial guess was $(X_1, X_2) = (0, 0)$ and the stopping criterion

$$\delta_k = \frac{\|R_1(k)\|_F}{\|R_1(0)\|_F} < 10^{-6},$$

was used where $R_1(k) = C_1 - (A_{11}X_1(k)B_{11} + A_{12}X_2(k)B_{12})$. The optimum value of μ is $\mu_{\text{opt}} = 0.00210$ and according to the equation (15) the upper bound of μ for the convergence is $\mu_u = 0.00230$. Hence, the method is convergent for $0 < \mu < 0.00230$. The method converges in 70, 148 and 305 iterates for μ_{opt}, $\mu = 0.00100$ and $\mu = 0.00225$, respectively. The convergence history is displayed in Figure 1.

Example 4.2. Consider the following coupled linear matrix equations

$$\begin{cases} A_{11}X_1B_{11} + A_{12}X_2B_{12} = C_1, \\ A_{21}X_1B_{21} + A_{22}X_2B_{22} = C_2, \end{cases} \tag{18}$$

where

$$A_{11} = \begin{pmatrix} 1 & 2 \\ -3 & -6 \end{pmatrix}, \quad B_{11} = \begin{pmatrix} -1 & -1 \\ 2 & 1 \\ -5 & 1 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 2 & 1 & 3 \\ 1 & -1 & 0 \end{pmatrix},$$

$$B_{12} = \begin{pmatrix} 2 & 9 \\ 0 & -3 \end{pmatrix}, \quad A_{21} = \begin{pmatrix} 1 & 2 \\ -3 & -6 \\ 1 & 2 \end{pmatrix}, \quad B_{21} = \begin{pmatrix} -1 & -1 & -2 \\ 3 & 1 & -1 \\ 2 & -1 & 1 \end{pmatrix},$$

$$A_{22} = \begin{pmatrix} 2 & 1 & 3 \\ 5 & -1 & 0 \\ 2 & 3 & -1 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} 2 & -1 & 3 \\ -1 & -4 & 2 \end{pmatrix},$$

and

$$C_1 = \begin{pmatrix} 2 & 83 \\ 54 & 57 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 9 & -6 & 15 \\ 65 & 44 & 37 \\ -19 & -28 & 1 \end{pmatrix}.$$

One can verify that the corresponding matrix $\mathcal{M}$ is of order 13×12 with $\text{rank}(\mathcal{M}) = 9$ and (X_1, X_2) with

$$X_1 = \begin{pmatrix} 1 & -1 & -1 \\ 2 & -1 & 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 2 & -1 \end{pmatrix},$$

is a solution of Equation (18). Consequently, Equation (18) has an infinitely number of solutions. We apply the gradient based algorithm to solve Equation (18). The initial guess was $(X_1, X_2) = (0, 0)$ and the following stopping criterion is utilized

$$\delta_k = \max\{\frac{\|R_1(k)\|_F}{\|R_1(0)\|_F}, \frac{\|R_1(k)\|_F}{\|R_1(0)\|_F}\} < 10^{-6},$$

in which $R_i(k) = C_i - (A_{i1}X_1(k)B_{i1} + A_{i2}X_2(k)B_{i2})$, $i = 1, 2$. The optimum value of μ is $\mu_{\text{opt}} = 0.00109$ and the upper bound of μ for the convergence, referring Equation (15), is $\mu_u = 0.00112$. For μ_{opt}, $\mu = 0.00030$ and $\mu = 0.00111$, the method converges in 208, 762 and 483 iterates, respectively. The convergence history is depicted in Figure 2.

5. Conclusion

The semi-convergence of the well-known Richardson iterative method for the consistent linear system of equations has been established. We have developed the Richardson iterative for solving the normal equations corresponding to the general consistent coupled linear matrix equations. It has been demonstrated that, for solving the general coupled linear matrix equations, the extension of the Richardson iterative method is mathematically equivalent to apply the gradient based iterative method. The gradient based iterative algorithms for solving the general coupled linear matrix equations have been the subject of interest in the various of the recently published research works. However, the gradient based iterative algorithms together with their convergence analysis have been investigated only for the (coupled) linear matrix equations with unique solution. It is well-known that the (coupled) linear matrix equations have unique solution if and only if the coefficient matrix of their associated normal equations is nonsingular. We show that under a mild condition the nonsingularity can be disregarded. In the numerical simulations, we have illustrated the validity of our presented theoretical results. It seems that the generalization of the presented method to the coupled linear matrix equations containing X_j^H, $\bar{X}_j$ and etc. is straightforward.

Acknowledgments

The authors would like to express their sincere gratitude to the five anonymous referees for their valuable suggestions and constructive comments which have ameliorated the quality of the paper.

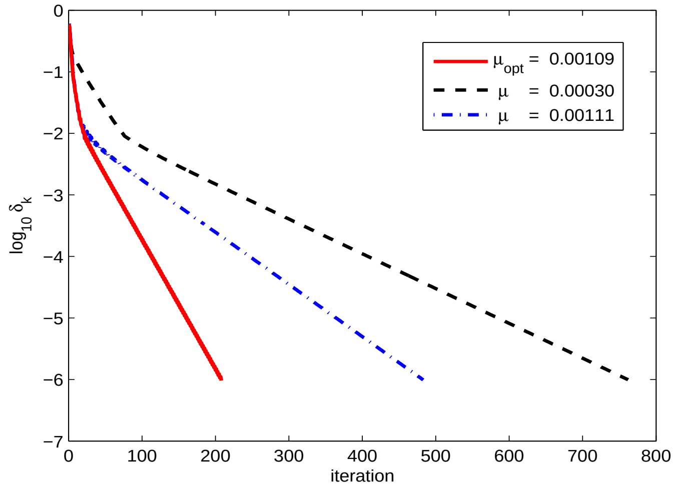

Figure 2: Convergence history for Example 2.

References

[1] Beik FPA and Salkuyeh DK (2011) On the global Krylov subspace methods for solving general coupled matrix equations. *Computers & Mathematics with Applications* 62 (12): 4605-4613.

[2] Benner P (2008) Large-Scale matrix equations of special type. *Numerical Linear Algebra with Applications* 15 (9): 747–754.

[3] Berman A and Plemmons RJ (1994) Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York, third edition, Reprinted by SIAM, Philadelphia.

[4] Bernstein DS (2009) Matrix mathematics: theory, facts, and formulas. Princeton University Press, New Jersey.

[5] Bouhamidi A and Jbilou K (2008) A note on the numerical approximate solutions for generalized Sylvester matrix equations with applications. *Applied Mathematics and Computation* 206 (2): 687–694.

[6] Ding F and Chen T (2005) Iterative least-squares solutions of coupled Sylvester matrix equations. *Systems & Control Letters* 54 (2): 95–107.

[7] Ding F and Chen T (2006) On iterative solutions of general coupled matrix equations. *SIAM Journal on Control and Optimization* 44 (6): 2269-2284.

[8] Ding F (2013) Coupled-least-squares identification for multivariable systems. *IET Control Theory and Applications* 7 (1): 68–79.

[9] Ding F, Liu PX and Ding J (2008) Iterative solutions of the generalized Sylvester matrix equations by using the hierarchical identification principle. *Applied Mathematics and Computation* 197 (1): 41–50.

[10] Ding J, Liu Y and Ding F (2010) Iterative solutions to matrix equations of the form $A_iXB_i = F_i$. *Computers & Mathematics with Applications* 59 (11): 3500–3507.

[11] Hyland D and Bernstein D (1984) The optimal projection equations for fixed-order dynamic compensation. *IEEE Transactions on Automatic Control* 29 (11): 1034–1037.

[12] Jbilou K, Messaudi A and Sadok H (1999) Global FOM and GMRES algorithms for matrix equations. *Applied Numerical Mathematics* 31 (1): 49–63.

[13] Jbilou K and Riquet AJ (2006) Projection methods for large Lyapunov matrix equations. *Linear Algebra and its Applications* 415 (2): 344–358.

[14] Laub AJ (2005) Matrix Analysis for Scientists and Engineers, SIAM, Philadelphia.

[15] Li ZY and Wang Y (2010) Iterative algorithm for minimal norm least squares solution to general linear matrix equations. *International Journal of Computer Mathematics* 87 (11): 2552–2567.

[16] Li ZY, Wang Y, Zhou B and Duan GR (2010) Least-squares solution with the minimum-norm to general matrix equations via iteration. *Applied Mathematics and Computation* 215 (10): 3547–3562.

[17] Saad Y (1996) Iterative Methods for Sparse Linear Systems. PWS press, New York.

[18] Salkuyeh DK and Toutounian F (2006) New approaches for solving large Sylvester equations. *Applied Mathematics and Computation* 173 (1): 9–18.

[19] Peng ZY and Peng YX (2006) An efficient iterative method for solving the matrix equation $AXB + CYD = E$. *Numerical Linear Algebra with Applications* 13 (6): 473–485.

[20] Xie L, Ding J and Ding F (2009) Gradient based iterative solutions for general linear matrix equations. *Computers & Mathematics with Applications* 58 (7): 1441–1448.

[21] Xie L, Liu YJ and Yang HZ (2010) Gradient based and least squares based iterative algorithms for matrix equations $AXB + CX^TD = F$. *Applied Mathematics and Computation* 217 (5): 2191–2199.

[22] Zhang JJ (2011) A note on the iterative solutions of general coupled matrix equation. *Applied Mathematics and Computation* 217 (22): 9380–9386.

[23] Zhou B and Duan GR (2008) On the generalized Sylvester mapping and matrix equations. *Systems & Control Letters* 57 (3): 200–208.

[24] Zhou B, Duan GR and Li ZY (2009a) Gradient based iterative algorithm for solving coupled matrix equations. *Systems & Control Letters* 58 (5) : 327–333.

[25] Zhou B, Li ZY, Duan GR and Wang Y (2009b) Weighted least squares solutions to general coupled Sylvester matrix equations. *Journal of Computational and Applied Mathematics* 224 (2): 759–776.

[26] Zhou B, Lam J and Duan GR (2010) Gradient-based maximal convergence rate iterative method for solving linear matrix equations. *International Journal of Computer Mathematics* 87 (3): 515–527.