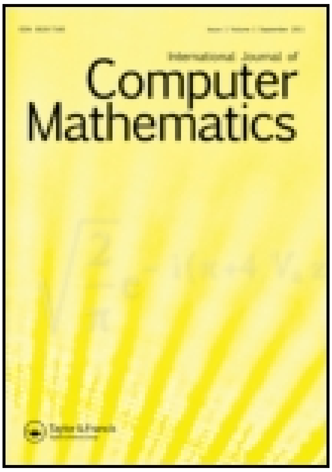

International Journal of Computer Mathematics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gcom20

Iterative algorithm for minimal norm least squares solution to general linear matrix equations

Zhao-Yan Li [a] & Yong Wang [a]
[a] Department of Mathematics, Harbin Institute of Technology, Harbin, China
Published online: 28 May 2010.

International Journal of Computer Mathematics
Vol. 87, No. 11, September 2010, 2552–2567

Iterative algorithm for minimal norm least squares solution to general linear matrix equations

Zhao-Yan Li* and Yong Wang

Department of Mathematics, Harbin Institute of Technology, Harbin, China

(*Received 18 April 2008; revised version received 29 November 2008; accepted 9 December 2008*)

This paper is concerned with minimal norm least squares solution to general linear matrix equations including the well-known Lyapunov matrix equation and Sylvester matrix equation as special cases. Two iterative algorithms are proposed to solve this problem. The first method is based on the gradient search principle for solving optimization problem and the second one can be regarded as its dual form. For both algorithms, necessary and sufficient conditions guaranteeing the convergence of the algorithms are presented. The optimal step sizes such that the convergence rates of the algorithms are maximized are established in terms of the singular values of some coefficient matrix. It is believed that the proposed methods can perform important functions in many analysis and design problems in systems theory.

Keywords: minimal norm least squares solution; iterative solutions; linear matrix equations; Lyapunov matrix equations; Sylvester matrix equations

2000 AMS Subject Classifications: 15A09; 15A12; 15A24

1. Introduction

The general linear matrix equation in the form of

$$A_1XB_1 + A_2XB_2 + \cdots + A_rXB_r = C, \tag{1}$$

including the well-known Lyapunov matrix equation and Sylvester matrix equation as special cases, plays an important role in control system theory [2,4,14–16,23]. For example, solutions to the Sylvester matrix equation $AX - XB = C$, where A, B, and C are known, can be used to parameterize the feedback gains in pole assignment problem for linear systems [2]. A more general case, $AX - EXF = C$, can also be used to achieve pole assignment, robust pole assignment, and observer design for descriptor linear systems [5,6,20,26]. Due to their wide applications, over the past several decades, the problem of searching for both analytical and numerical solutions to the Lyapunov and Sylvester matrix equations has been well investigated in the literature. For sample examples, see [1,7,8,10,22,24–26].

On the other hand, if solution to the linear matrix equation (1) is not unique or does not exist, only least squares and/or minimal norm solutions can be found. In fact, the minimal norm least

*Corresponding author. Email: zhaoyanlee@gmail.com; zhaoyanlee@yahoo.com.cn

ISSN 0020-7160 print/ISSN 1029-0265 online

DOI: 10.1080/00207160802684459
http://www.informaworld.com

squares solution to linear matrix equations has many applications in control system theory, for example, model reduction and system identification [8,18]. However, few results are available in the literature to deal with this problem. Only some special cases of Equation (1) are considered by some investigators. For instance, Ding *et al.* constructed an iterative algorithm to obtain the minimal norm least squares solution to the linear matrix equation $AX = C$ by using the hierarchical identification principle [7,8].

The linear matrix equation (1) can be converted to the simple vector form

$$\Upsilon x = c, \tag{2}$$

by using the Kronecker product [3] and then solved by some methods that can be applied on Equation (2) (for example [9,11,12,17]). However, such transformation is not recommended in practice for the reason that the dimensions of the associated coefficient matrix Υ are very high when the dimensions of A_i and B_i are large, which leads to computational difficulties as excessive computer memory is required to compute the inversion of a matrix of high dimensions. See [3,7] for detailed analysis. Moreover, when finding minimal norm least squares solutions to Equation (2) by using the methods mentioned in the above references, an augmentation on Equation (2) is generally required [18].

In this paper, we consider the problem of finding the minimal norm least squares solution to the linear matrix equation (1). Especially, we are interested in using the iterative algorithm to approximate the exact minimal norm least squares solution to such linear matrix equation. Using iteration to approximate exact solution to equations (linear or nonlinear) is popular in the literature (see, for example, [8,19,21] and the monograph [13]). In this paper, we present two methods. Our first method dealing with the case that Υ (see Equation (2) or (7) defined later) is of full column rank is based on the gradient search principle for solving optimization problem, as the gradient of the objective function defined in this paper is easy to compute. The second method dealing with the case that Υ is of full row rank can be regarded as the dual form of the first one. For both cases, we have analysed the convergence properties of the iterations, which allow one to obtain necessary and sufficient conditions. Furthermore, we have provided the optimal step size such that the convergence rate of the iteration, which is properly defined in this paper, is maximized. Both of these two methods are also able to produce the unique solution to the matrix equation (1) if Υ is non-singular which implies that our results are generalizations of those given in [28]. Numerical examples show that the proposed algorithms are very effective. It is believed that the proposed method can perform important functions in many analysis and design problems in systems theory.

The remainder of this paper is organized as follows. In Section 2, we give the problem formulation and some necessary preliminary results. Especially, the convergence rate of a general linear iteration is defined. In Section 3, we present two iterative algorithms to produce iterative minimal norm least squares solution to the linear matrix equation (1). Convergence properties are studied and the optimal step sizes such that the convergence rates of the iterations are maximized are presented. Examples are given in Section 4 to illustrate the effectiveness of the proposed algorithms. Section 5 concludes the paper.

Notations. Throughout this paper, we use $\mathrm{tr}(A)$, $\rho(A)$, A^{T}, $\mathrm{rank}(A)$, $\sigma_{\max}(A)$, and $\sigma_{\min}(A)$ to denote the trace, the spectral radius, the transpose, the rank, the maximal singular value, and the minimal singular value of matrix A, respectively. The notation $\|A\|_{\mathrm{F}}$ and $\|A\|_2$ refer to the Frobenius norm and 2-norm of the matrix A and $\mathbf{1}_{m\times n}$ refers to a matrix whose elements are 1. The Kronecker product of two matrices A and B are denoted by $A \otimes B$. The stretching function $\mathrm{vec}(A)$ where $A = [a_1 \ a_2 \ \cdots \ a_m]$ is defined as $\mathrm{vec}(A) = [a_1^{\mathrm{T}} \ a_2^{\mathrm{T}} \ \cdots \ a_m^{\mathrm{T}}]^{\mathrm{T}}$. Moreover, we use $\mathrm{cond}(A)$ to denote the condition number of a full row and/or column rank matrix A, i.e., $\mathrm{cond}(A) = \sigma_{\max}(A)/\sigma_{\min}(A)$.

2. Problem formulation and preliminaries

Consider a general linear matrix equation

$$A_1XB_1 + A_2XB_2 + \cdots + A_rXB_r = C, \tag{3}$$

where $A_i \in \mathbb{R}^{p\times m}$, $B_i \in \mathbb{R}^{n\times q}$, $i = 1, 2, \ldots, r$, are known matrices and $X \in \mathbb{R}^{m\times n}$ is a matrix to be determined. The problem studied in this paper is stated as follows.

Problem 1 Let

$$a = \min_{X\in\mathbb{R}^{m\times n}} \left\{ \left\| \sum_{i=1}^{r} A_iXB_i - C \right\|_{\mathrm{F}} \right\}.$$

Find a matrix $X \in \mathbb{R}^{m\times n}$ such that $\|X\|_{\mathrm{F}}$ is minimized and

$$\left\| \sum_{i=1}^{r} A_iXB_i - C \right\|_{\mathrm{F}} = a. \tag{4}$$

Note that for arbitrary $X \in \mathbb{R}^{m\times n}$,

$$\|X\|_{\mathrm{F}} = \|\mathrm{vec}(X)\|_{\mathrm{F}} = \|\mathrm{vec}(X)\|_2. \tag{5}$$

Let

$$f(X) = \| \sum_{i=1}^{r} A_iXB_i - C\|_{\mathrm{F}}.$$

Then by using the Kronecker product and the following well-known formulation [3]

$$\mathrm{vec}(AXB) = (B^{\mathrm{T}} \otimes A)\mathrm{vec}(X), \tag{6}$$

the function $f(X)$ can be converted to

$$\begin{aligned} f(X) &= \|\Upsilon\mathrm{vec}(X) - \mathrm{vec}(C)\|_{\mathrm{F}} \\ &= \|\Upsilon\mathrm{vec}(X) - \mathrm{vec}(C)\|_2, \end{aligned}$$

where Υ is defined as

$$\Upsilon = \sum_{i=1}^{r} (B_i^{\mathrm{T}} \otimes A_i) \in \mathbb{R}^{pq\times mn}. \tag{7}$$

Therefore, Problem 1 can be transformed into the following equivalent problem.

Problem 2 Let

$$\begin{aligned} b &= \min_{x\in\mathbb{R}^{mn}} \{\widetilde{f}(x)\} \\ &= \min_{x\in\mathbb{R}^{mn}} \{\|\Upsilon x - \mathrm{vec}(C)\|_2\}. \end{aligned}$$

Find a vector $x \in \mathbb{R}^{mn}$ such that $\|x\|_2$ is minimized and

$$\|\Upsilon x - \mathrm{vec}(C)\|_2 = b.$$

Recall that for arbitrary matrix $P \in \mathbb{R}^{m\times n}$, the Moore–Penrose inverse of P is a matrix X satisfying the following four equations

$$PXP = P, \quad XPX = X, \quad (PX)^{\mathrm{T}} = PX, \quad (XP)^{\mathrm{T}} = XP.$$

Denote $P^+ = X$. It is known that P^+ is unique and moreover,

$$P^+ = \begin{cases} (P^{\mathrm{T}}P)^{-1}P^{\mathrm{T}}, & P \text{ is of full column rank,} \\ P^{\mathrm{T}}(PP^{\mathrm{T}})^{-1}, & P \text{ is of full row rank.} \end{cases} \tag{8}$$

Then regarding solution to Problem 2, we have the following well-known result.

LEMMA 1 ([18]) *The unique solution x_* to Problem* 2 *is given by*

$$x_* = \Upsilon^+ \mathrm{vec}(C),$$

where Υ^+ is the Moore–Penrose inverse of matrix Υ.

In view of (8), we have the following corollary of Lemma 1.

COROLLARY 1 *The following two statements hold.*

(1) *If Υ is of full column rank, then the unique solution to Problem* 2 *is given by*

$$x_* = (\Upsilon^{\mathrm{T}}\Upsilon)^{-1}\Upsilon^{\mathrm{T}}\mathrm{vec}(C). \tag{9}$$

(2) *If Υ is of full row rank, then the unique solution to Problem* 2 *is given by*

$$x_* = \Upsilon^{\mathrm{T}}(\Upsilon\Upsilon^{\mathrm{T}})^{-1}\mathrm{vec}(C). \tag{10}$$

We next consider the convergence rate of a general linear iteration

$$X(k) = \sum_{i=1}^{p} \mathcal{A}_i X(k-1)\mathcal{B}_i + \mathcal{C}, \quad X(k) \in \mathbb{R}^{m\times n}, \tag{11}$$

where $\mathcal{A}_i$ and $\mathcal{B}_i$, $i = 1, 2, \ldots, p$, are constant matrices with appropriate dimensions. By means of the Kronecker product, Iteration (11) can be written as the following vector form

$$\mathrm{vec}(X(k)) = \Theta\mathrm{vec}(X(k-1)) + \mathrm{vec}(\mathcal{C}), \quad \Theta = \sum_{i=1}^{p}(\mathcal{B}_i^{\mathrm{T}} \otimes \mathcal{A}_i) \in \mathbb{R}^{mn\times mn}. \tag{12}$$

It is well known that Iteration (12) converges for arbitrary initial condition if and only if $\rho(\Theta) < 1$ [3,13,17]. Moreover, the smaller the $\rho(\Theta)$, the faster the iteration converges. For this reason, the number $-\log(\rho(\Theta))$ is usually used to denote the convergence rate of Iteration (12) [13]. For clarity, we first give the following definition of convergence rate for Iteration (11) (or 12).

DEFINITION 1 *Assume that Iteration* (11) *converges to the unique matrix X_∞ for arbitrary initial condition $X(0)$. The α-convergence rate for Iteration* (11) *is a scalar $\gamma = -\log\beta$ with $0 < \beta < 1$ such that*

$$\|X(k) - X_\infty\|_\alpha \le \kappa\beta^k\|X(0) - X_\infty\|_\alpha, \quad k \ge 0, \tag{13}$$

and there exists at least one $X(0) \neq X_\infty$ such that '$=$' hold in Equation (13). *In Equation* (13), *κ is a positive scalar independent of k and β, and α denotes a suitable matrix norm (e.g., $\alpha = 2$ or $\alpha = F$).*

Our next lemma shows that $-\log(\rho(\Theta))$ can indeed be used to denote the F-convergence rate of iteration (11) in the sense of Definition 1 in a special case.

LEMMA 2 *Assume that* $\Theta \in \mathbb{R}^{mn\times mn}$ *is a real symmetric matrix with* $\rho(\Theta) < 1$. *Then the* F-*convergence rate of Iteration* (11) *in the sense of Definition* 1 *is* $-\log(\rho(\Theta))$. *Moreover, let* $\lim_{k\to\infty} X(k) = X_\infty$, *then for arbitrary initial condition* $X(0)$, *there holds*

$$\|X(k) - X_\infty\|_{\mathrm{F}} \leq \rho^k(\Theta)\|X(0) - X_\infty\|_{\mathrm{F}}. \tag{14}$$

The proof of the above lemma is given in Appendix A.1. We further give another technical lemma that will be used later. The proof is simple and thus omitted.

LEMMA 3 *Assume that* $m_i, i = 1, 2, \ldots, n$, *are some given positive scalars. Denote* $m_{\max} = \max_{1\leq i\leq n}\{m_i\}$ *and* $m_{\min} = \min_{1\leq i\leq n}\{m_i\}$. *Then*

$$\min_{0<u<(2/m_{\max})} \max_{1\leq i\leq n}\{|1 - um_i|\} = \frac{m_{\max} - m_{\min}}{m_{\max} + m_{\min}}. \tag{15}$$

Moreover, *the unique* u_{opt} *such that the above relation holds is*

$$u_{\mathrm{opt}} = \frac{2}{m_{\max} + m_{\min}}.$$

3. Main results

3.1 *Iterative solution to Problem 1: Υ is of full column rank*

Denote a new objective function

$$J(X) = \frac{1}{2}\left\|\sum_{i=1}^{r} A_i X B_i - C\right\|_{\mathrm{F}}^2.$$

Note that $J(X) = f^2(X)$. Therefore, $f(X)$ is minimized if and only if $J(X)$ is minimized. The idea of our method is to use the gradient search method to find the optimal solution such that $J(X)$ is minimized. This can be done because of the fact that the gradient of the objective function $J(X)$ is easy to compute. The result is given as follows and the proof is provided in Appendix A.2.

LEMMA 4 *The gradient* $(\partial/\partial X)J(X)$ *is given by*

$$\frac{\partial J(X)}{\partial X} = \sum_{i=1}^{r} A_i^{\mathrm{T}}\left(\sum_{j=1}^{r} A_j X B_j - C\right) B_i^{\mathrm{T}}.$$

Therefore, the gradient-based iterative algorithm can be constructed as follows

$$X(k) = X(k-1) - \mu\sum_{i=1}^{r} A_i^{\mathrm{T}}\left(\sum_{j=1}^{r} A_j X(k-1) B_j - C\right) B_i^{\mathrm{T}}, \tag{16}$$

where μ is the step size specified later.

THEOREM 1 *Assume that Υ is of full column rank. Let X_* be the unique solution to Problem* 1.

(1) *Iteration* (16) *converges to a constant matrix X_∞ for arbitrary initial condition $X(0)$ if and only if*

$$0 < \mu < \frac{2}{\sigma_{\max}^2(\Upsilon)}. \tag{17}$$

Furthermore, if Equation (17) is satisfied, then the minimal norm least squares solution $X_* = X_\infty$.

(2) *For arbitrary μ satisfying Equation (17), the* F*-convergence rate of Iteration* (16) *in the sense of Definition* 1 *is given by*

$$\gamma = -\log(\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon)).$$

Moreover, there holds

$$\|X(k) - X_*\|_{\mathrm{F}} \leq \rho^k(I - \mu\Upsilon^{\mathrm{T}}\Upsilon)\|X(0) - X_*\|_{\mathrm{F}}. \tag{18}$$

(3) *The* F*-convergence rate of Iteration* (16) *is maximized when*

$$\mu = \mu_{\mathrm{opt}} = \frac{2}{\sigma_{\max}^2(\Upsilon) + \sigma_{\min}^2(\Upsilon)}. \tag{19}$$

In this case, the maximal F*-convergence rate is given by*

$$\gamma_{\mathrm{opt}} = -\log\left(\frac{\mathrm{cond}^2(\Upsilon) - 1}{\mathrm{cond}^2(\Upsilon) + 1}\right). \tag{20}$$

Proof (1) It follows from Equation (16) that

$$X(k) = X(k-1) - \mu \sum_{i=1}^{r}\sum_{j=1}^{r} A_i^{\mathrm{T}} A_j X(k-1) B_j B_i^{\mathrm{T}} - \mu \sum_{i=1}^{r} A_i^{\mathrm{T}} C B_i^{\mathrm{T}}. \tag{21}$$

Taking vec on both sides of Equation (21) and using Equation (6), we get

$$\mathrm{vec}(X(k)) = \Theta\mathrm{vec}(X(k-1)) + \mu\Upsilon^{\mathrm{T}}\mathrm{vec}(C), \tag{22}$$

where

$$\begin{aligned}
\Theta &= I - \mu \sum_{i=1}^{r}\sum_{j=1}^{r}(B_i B_j^{\mathrm{T}} \otimes A_i^{\mathrm{T}} A_j) \\
&= I - \mu \sum_{i=1}^{r}\sum_{j=1}^{r}(B_i \otimes A_i^{\mathrm{T}})(B_j^{\mathrm{T}} \otimes A_j) \\
&= I - \mu \sum_{i=1}^{r}(B_i \otimes A_i^{\mathrm{T}}) \sum_{j=1}^{r}(B_j^{\mathrm{T}} \otimes A_j) \\
&= I - \mu\Upsilon^{\mathrm{T}}\Upsilon.
\end{aligned} \tag{23}$$

Therefore, Iteration (16) converges to a finite matrix X_∞ for arbitrary initial condition $X(0)$ if and only if Θ is Schur stable, i.e., $\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon) < 1$. We note that $I - \mu\Upsilon^{\mathrm{T}}\Upsilon$ is a symmetric matrix and Υ is of full column rank. Then

$$\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon) = \max_{1\le i\le mn}\{|1 - \mu\sigma_i^2(\Upsilon)|\}.$$

Accordingly, $\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon) < 1$ if and only if $|1 - \mu\sigma_i^2(\Upsilon)| < 1$ which is equivalent to Equation (17). Now let Iteration (16) converge to X_∞ as k approaches to infinity, i.e., $\lim_{k\to\infty} X(k) = X_\infty$. Then it follows from Equation (22) that

$$\mathrm{vec}(X_\infty) = (I - \mu\Upsilon^{\mathrm{T}}\Upsilon)\mathrm{vec}(X_\infty) + \mu\Upsilon^{\mathrm{T}}\mathrm{vec}(C), \tag{24}$$

which in turn implies

$$\mathrm{vec}(X_\infty) = (\Upsilon^{\mathrm{T}}\Upsilon)^{-1}\Upsilon^{\mathrm{T}}\mathrm{vec}(C).$$

Therefore, it follows from Corollary 1 that $\mathrm{vec}(X_\infty)$ is the solution to Problem 2. Hence, X_∞ is the solution to Problem 1.

(2). Notice that Iterations (21) and (22) are, respectively, in the form of Equations (11) and (12). Moreover, the matrix $\Theta = I - \mu\Upsilon^{\mathrm{T}}\Upsilon$ is real symmetric. Then the F-convergence rate of Iteration (16) is $-\log(\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon))$ in accordance with Lemma 2. As a result, Inequality (18) follows directly from Inequality (13).

(3). According to the above item, the F-convergence rate of Iteration (16) is maximized if and only if $-\log(\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon))$ is maximized, or equivalently, $\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon)$ is minimized. By definition,

$$\min_{0<\mu<\frac{2}{\sigma_{\max}^2(\Upsilon)}} \rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon) = \min_{0<\mu<\frac{2}{\sigma_{\max}^2(\Upsilon)}} \max_{1\le i\le mn}\{|1 - \mu\sigma_i^2(\Upsilon)|\}. \tag{25}$$

We notice that Equation (25) is in the form of Equation (15). Therefore, according to Lemma 3, $\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon)$ is minimized if μ is chosen as in Equation (19). Moreover,

$$\begin{aligned}\rho(I - \mu_{\mathrm{opt}}\Upsilon^{\mathrm{T}}\Upsilon) &= \frac{\sigma_{\max}^2(\Upsilon) - \sigma_{\min}^2(\Upsilon)}{\sigma_{\max}^2(\Upsilon) + \sigma_{\min}^2(\Upsilon)} \\ &= \frac{\mathrm{cond}^2(\Upsilon) - 1}{\mathrm{cond}^2(\Upsilon) + 1},\end{aligned} \tag{26}$$

which implies Equation (20). At last, we show that $\mu = \mu_{\mathrm{opt}}$ satisfies Condition (17). Since Υ is of full column rank, we have $\sigma_{\min}(\Upsilon) \neq 0$, that is $\mu_{\mathrm{opt}} < \mu_{\max}$. This proves Theorem 1. ■

The following corollary can be immediately obtained from Theorem 1.

COROLLARY 2 *Consider the following linear matrix equation*

$$AX = B. \tag{27}$$

If A is a non-square $p \times m$ full column rank matrix, then the iteration

$$X(k) = X(k-1) + \mu A^{\mathrm{T}}(B - AX) \tag{28}$$

converges to the minimal norm least squares solution $X_* = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}B$ to the linear matrix equation (27) *for arbitrary initial condition $X(0)$ if and only if*

$$0 < \mu < \frac{2}{\sigma_{\max}^2(A)}.$$

Moreover, if $\mu = 2/\sigma_{\max}^2(A) + \sigma_{\min}^2(A)$, then the F*-convergence rate of Iteration* (28) *is maximized.*

Remark 1 Under the condition of Corollary 2, the following iteration

$$X(k) = X(k-1) + \mu(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}(B - AX),\ 0 < \mu < 2, \tag{29}$$

is used in [8] to obtain the minimal norm least squares solution X_* to the linear matrix equation (27). Note that matrix inversion is required in Equation (29) which may lead to numerical problems. In fact, if $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ has been computed, we need not to use the iteration (29) but the formulation $X_* = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}B$ to compute the minimal norm least squares solution X_* directly.

Remark 2 We see from Theorem 1 that the algorithm converges exponentially and depends on the number $\rho(I - \mu\Upsilon^{\mathrm{T}}\Upsilon)$. Though Item 3 of Theorem 1 provides a method to select the optimal value μ such that the convergence of the algorithm is maximized, the convergence is generally fast (Example 2 in this paper) and sometimes can be rather slow (Example 1 in this paper). Our further study should focus on improving the convergence performances in general case. A possible way is to use the pre-conditioned technique [3].

To apply Theorem 1, we need to compute the singular value of matrix Υ which is difficult in practice. To overcome this shortcoming, by using the method in [28], we can provide the following corollary.

COROLLARY 3 *Algorithm* (16) *converges to the unique solution to Problem* 1 *provided*

$$0 < \mu < \frac{2}{r\sum_{i=1}^{r}\sigma_{\max}^2(A_i)\sigma_{\max}^2(B_i)}. \tag{30}$$

Proof By using the norm inequality $\|A+B\|_2 \le \|A\|_2 + \|B\|_2$, we obtain

$$\begin{aligned}\sigma_{\max}^2(\Upsilon) &= \left\|\sum_{k=1}^{r}(B_k^{\mathrm{T}} \otimes A_k)\right\|_2^2 \\ &\le \left(\sum_{k=1}^{r}\|B_k^{\mathrm{T}} \otimes A_k\|_2\right)^2.\end{aligned} \tag{31}$$

It is easy to verify that

$$\begin{aligned}\|A \otimes B\|_2 &= \sigma_{\max}(A \otimes B) \\ &= \sigma_{\max}(A)\sigma_{\max}(B).\end{aligned}$$

By using this fact and the Hölder inequality

$$\left(\sum_{i=1}^{r}a_i\right)^2 \le r\sum_{i=1}^{r}a_i^2,$$

Inequality (31) can be simplified as

$$\begin{aligned}\sigma_{\max}^2(\Upsilon) &\le \left(\sum_{i=1}^{r}(\sigma_{\max}(A_i)\sigma_{\max}(B_i^{\mathrm{T}}))\right)^2 \\ &\le r\sum_{i=1}^{r}\sigma_{\max}^2(A_i)\sigma_{\max}^2(B_i^{\mathrm{T}}).\end{aligned}$$

Then we have

$$\frac{2}{\sigma_{\max}^2(\Upsilon)} \geq \frac{2}{r\sum_{i=1}^{r}\sigma_{\max}^2(A_i)\sigma_{\max}^2(B_i^{\mathrm{T}})}.$$

The proof is completed by using Theorem 1. ∎

3.2 *Iterative solution to Problem 1: Υ is of full row rank*

We next consider the case that Υ is of full row rank. In this case, Iteration (16) fails to converge as $\rho(I-\mu\Upsilon^{\mathrm{T}}\Upsilon)\geq 1$ no mater what μ is. Alternatively, we present another iterative algorithm to solve Problem 1.

Construct the following iteration

$$Y(k) = Y(k-1) - \mu\sum_{i=1}^{r} A_i\left(\sum_{j=1}^{r} A_j^{\mathrm{T}} Y(k-1) B_j^{\mathrm{T}}\right) B_i + \mu C, \tag{32}$$

with initial condition $Y(0)$ and step size μ which is to be determined later. Iteration (32) can be regarded as the dual form of Iteration (16).

THEOREM 2 *Assume that Υ is of full row rank. Let X_* be the unique solution to Problem* 1.

(1) *Iteration (*32*) converges to a finite matrix Y_∞ for arbitrary initial condition if and only if*

$$0 < \mu < \frac{2}{\sigma_{\max}^2(\Upsilon)}. \tag{33}$$

Furthermore, if Equation (33) *is satisfied and* $\lim_{k\to\infty} Y(k) = Y_\infty$, *then*

$$X_* = \sum_{i=1}^{r} A_i^{\mathrm{T}} Y_\infty B_i^{\mathrm{T}}. \tag{34}$$

(2) *For arbitrary μ satisfying Equation* (33), *the* F-*convergence rate of Iteration* (32) *is given by*

$$\gamma = -\log(\rho(I-\mu\Upsilon\Upsilon^{\mathrm{T}})).$$

Moreover, there holds

$$\|Y(k)-Y_\infty\|_{\mathrm{F}} \leq \rho^k(I-\mu\Upsilon\Upsilon^{\mathrm{T}})\|Y(0)-Y_\infty\|_{\mathrm{F}}.$$

(3) *The* F-*convergence rate of Iteration* (32) *is maximized when $\mu=\mu_{\mathrm{opt}}$ which is given by Equation* (19). *In this case, the maximal* F-*convergence rate of Iteration* (32) *is given by Equation* (20).

Proof *Proof of Item* 1. Taking vec on both sides of Equation (32) gives

$$\mathrm{vec}(Y(k)) = \mathrm{vec}(Y(k-1)) - \mu\Psi\mathrm{vec}(Y(k-1)) + \mu\mathrm{vec}(C), \tag{35}$$

where

$$\begin{aligned}\Psi &= \sum_{i=1}^{r}\sum_{j=1}^{r} B_i^{\mathrm{T}} B_j \otimes A_i A_j^{\mathrm{T}} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} (B_i^{\mathrm{T}} \otimes A_i)(B_j \otimes A_j^{\mathrm{T}}) \\ &= \Upsilon\Upsilon^{\mathrm{T}}.\end{aligned}$$

Then Iteration (35) can be simplified as

$$\mathrm{vec}(Y(k)) = (I - \mu\Upsilon\Upsilon^{\mathrm{T}})\mathrm{vec}(Y(k-1)) + \mu\mathrm{vec}(C). \tag{36}$$

Therefore, $\lim_{k\to\infty} Y(k) = Y_\infty$ for arbitrary initial condition $Y(0)$ if and only if $\rho(I - \mu\Upsilon\Upsilon^{\mathrm{T}}) < 1$ which is equivalent to Equation (33).

Taking limit on both sides of Equation (36) produces

$$\mathrm{vec}(Y_\infty) = (I - \mu\Upsilon\Upsilon^{\mathrm{T}})\mathrm{vec}(Y_\infty) + \mu\mathrm{vec}(C),$$

which in turn implies

$$\mathrm{vec}(Y_\infty) = (\Upsilon\Upsilon^{\mathrm{T}})^{-1}\mathrm{vec}(C).$$

It follows from Corollary 1 and from the above equation that

$$\mathrm{vec}(X_*) = \Upsilon^{\mathrm{T}}(\Upsilon\Upsilon^{\mathrm{T}})^{-1}\mathrm{vec}(C) = \Upsilon^{\mathrm{T}}\mathrm{vec}(Y_\infty),$$

which is just Equation (34).

Proofs of Items 2–3. We notice that Iterations (32) and (36) are, respectively, in the standard form of Equations (11) and (12). The proof is thus similar to the proof of Theorem 1 and omitted here. ■

Similar to Corollary 2, we can obtain the following corollary of Theorem 2.

COROLLARY 4 *Consider the linear matrix Equation* (27). *If A is a non-square $p \times m$ full row rank matrix, then the iteration*

$$Y(k) = Y(k-1) - \mu AA^{\mathrm{T}} Y(k-1) + \mu B \tag{37}$$

converges to a finite matrix Y_∞ as k approaches to infinity for arbitrary initial condition $Y(0)$ if and only if

$$0 < \mu < \frac{2}{\sigma_{\max}^2(A)}. \tag{38}$$

Moreover, let Equation (38) *be satisfied and* $\lim_{k\to\infty} Y(k) = Y_\infty$. *Then the minimal norm least squares solution to the linear matrix Equation* (27) *is given by $X_* = A^{\mathrm{T}} Y_\infty$. Furthermore, if $\mu = 2/(\sigma_{\max}^2(A) + \sigma_{\min}^2(A))$, then the* F-*convergence rate of Iteration* (37) *is maximized.*

Remark 3 If Υ is a square and non-singular matrix, then both Iterations (16) and (32) can be used to construct the unique solution to the linear matrix Equation (3). However, both Iterations (16) and (32) are invalid in the case that Υ is neither of full column rank nor of full row rank. How to deal with this case is still a project to be investigated.

Remark 4 Similar results for Algorithm (32) to Corollary 3 can be easily obtained.

4. Illustrative examples

Example 1 Consider the following linear matrix equation

$$A_1 X B_1 + A_2 X B_2 = C, \tag{39}$$

where A_1, A_2, B_1, B_2, C, and the exact minimal norm least squares solution X_* are, respectively, given by

$$A_1 = \begin{bmatrix} 1.0000 & 2.0000 \\ -1.0000 & 0.5000 \\ 0 & 1.0000 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1.0000 & -2.0000 \\ 0 & 1.0000 \\ 2.0000 & -1.0000 \end{bmatrix}, \quad C = \begin{bmatrix} -4.0000 & 2.0000 \\ 0 & 1.0000 \\ -3.0000 & 2.0000 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 1.0000 & -2.0000 \\ -1.0000 & 1.0000 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1.0000 & 0 \\ -1.0000 & 1.0000 \end{bmatrix}, \quad X_* = \begin{bmatrix} -0.5000 & 0.9000 \\ -0.2000 & 1.2667 \end{bmatrix}.$$

The exact solution X_* in the above is obtained by using Equation (9). In this case, the matrix $\Upsilon \in \mathbb{R}^{6\times 4}$ is of full column rank. Therefore, we should apply Iteration (16) to compute $X(k)$. Taking the initial condition $X(0) = 10^{-6}\mathbf{1}_{2\times 2}$. The results corresponding to $\mu = \mu_{\text{opt}}$ are shown in Table 1, where

$$\epsilon(k) = \frac{\|X(k) - X_*\|_{\text{F}}}{\|X_*\|_{\text{F}}} \tag{40}$$

is the relative iteration error. To demonstrate that $\mu = \mu_{\text{opt}}$ can indeed guarantee better convergence performance, several convergence curves are shown in Figure 1 where the y-axis denotes the relative iteration error $\epsilon(k)$. Except for μ_{opt}, all the other step sizes are computed according to Corollary 3. It is clear to see that the convergence performance associated with $\mu = \mu_{\text{opt}}$ is better than that associated with the other step sizes.

Example 2 Still consider the linear matrix Equation (39) but with the following parameters

$$A_1 = \begin{bmatrix} 1.0000 & 0 & -1.0000 \\ 0.50000 & 0 & -3.0000 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1.000 & -2.000 \\ -1.000 & 1.000 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} -2.000 & 2.000 & 0 \\ -1.000 & 1.000 & 1.000 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1.000 & -3.000 \\ 2.000 & 1.000 \end{bmatrix}$$

$$X_*^{\text{T}} = \begin{bmatrix} -0.0733 & -0.6260 & -0.2255 \\ 0.4734 & -0.2031 & 0.1327 \end{bmatrix}, \quad C = \begin{bmatrix} -4.000 & 2.000 \\ 1.000 & -3.000 \end{bmatrix}.$$

The exact solution X_* is obtained by using Equation (10) as the matrix $\Upsilon \in \mathbb{R}^{4\times 6}$ is of full row rank in this case. Hence, Iteration (32) should be used to construct iterative solutions to Problem 1.

Table 1. The iterative solutions to the matrix Equation (39) with $\mu = \mu_{\text{opt}}$.

k	x_{11}	x_{12}	x_{21}	x_{22}	$\epsilon \times 100\%$
5	−0.4004487709	0.9185200988	−0.7261052752	0.5705864483	53.41313089
10	−0.2012802428	0.8243088396	−0.1012826980	0.8448172543	32.32933255
15	−0.4420345949	0.9381598416	−0.3962905996	1.018250031	19.70940151
20	−0.3860262644	0.8762270342	−0.1633303245	1.111181219	12.02025139
25	−0.4780414671	0.9148018845	−0.2730197057	1.174463497	7.330981693
30	−0.4575509502	0.8912417636	−0.1863612783	1.208859576	4.471066798
35	−0.4918246063	0.9055174485	−0.2271606631	1.232374450	2.726843443
40	−0.4842095202	0.8967438990	−0.1949269385	1.245165180	1.663065107
45	−0.4969589199	0.9020525047	−0.2101027261	1.253911353	1.014281021
50	−0.4941265275	0.8987888868	−0.1981130163	1.258668950	0.618596341
55	−0.4988688322	0.9007634573	−0.2037578261	1.261922181	0.377273581
60	−0.4978152934	0.8995495130	−0.1992981146	1.263691823	0.230094078
65	−0.4995792490	0.9002839769	−0.2013977670	1.264901900	0.140331281
70	−0.4991873731	0.8998324362	−0.1997389256	1.265560139	0.085586159
75	−0.4998434968	0.9001056285	−0.2005199156	1.266010241	0.052197846
80	−0.4996977340	0.8999376727	−0.1999028903	1.266255081	0.031834764
X_*	−0.5000000000	0.9000000000	−0.2000000000	1.266666667	0.000000000

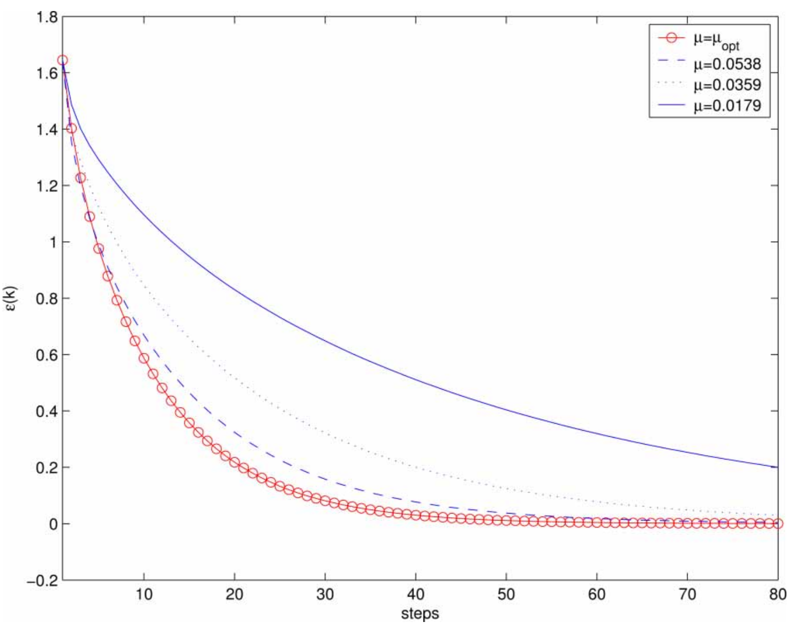

Figure 1. Convergence performances with different step sizes in Example 1.

Take $Y(0) = 10^{-6}\mathbf{1}_{3\times 2}$ and

$$X(k) = \sum_{i=1}^{2} A_i^{\mathrm{T}} Y(k) B_i^{\mathrm{T}}.$$

The relative iteration error is still defined as in Equation (40). When $\mu = \mu_{\text{opt}}$, the computing results are given in Table 2 where $X(k) = [x_{ij}]$, $i = 1, 2, 3$, $j = 1, 2$. Shown in Figure 2 are the convergence performances of the iteration with different step sizes. Again, except for μ_{opt}, all the other step sizes are computed according to Corollary 3. We also clearly see that $\mu = \mu_{\text{opt}}$ can indeed lead to better convergence performances.

Table 2. The iterative solutions to the matrix Equation (39) with $\mu = \mu_{\text{opt}}$.

k	x_{11}	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}	$\epsilon \times 100\%$
5	−0.03906904	0.46957129	−0.63848802	−0.21989548	−0.19779809	0.12798626	5.737290787
10	−0.07320991	0.47064177	−0.62292707	−0.20547563	−0.22538354	0.13555974	0.648608966
15	−0.07281730	0.47332417	−0.62619757	−0.20334086	−0.22510011	0.13265467	0.073331691
20	−0.07325281	0.47333393	−0.62600348	−0.20315421	−0.22545437	0.13275495	0.008290877
25	−0.07324780	0.47336823	−0.62604527	−0.20312693	−0.22545074	0.13271780	0.000937366
30	−0.07325336	0.47336835	−0.62604279	−0.20312454	−0.22545527	0.13271909	0.000105978
35	−0.07325330	0.47336879	−0.62604332	−0.20312419	−0.22545522	0.13271861	0.000011981
40	−0.07325337	0.47336879	−0.62604329	−0.20312416	−0.22545528	0.13271863	0.000001354
45	−0.07325337	0.47336880	−0.62604330	−0.20312416	−0.22545528	0.13271862	0.000000153
50	−0.07325337	0.47336880	−0.62604330	−0.20312416	−0.22545528	0.13271862	0.000000017
X_*	−0.07325337	0.47336880	−0.62604330	−0.20312416	−0.22545528	0.13271862	0

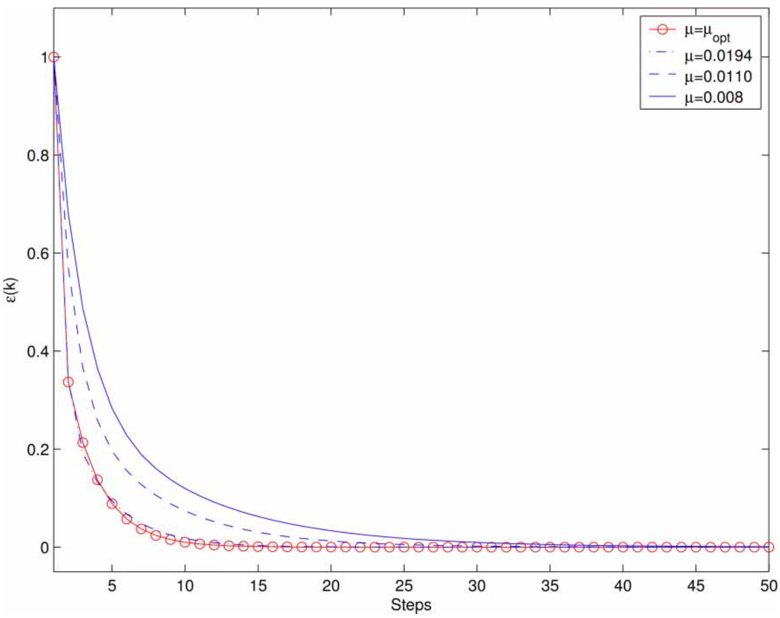

Figure 2. Convergence performances with different step sizes in Example 2.

5. Conclusion

This paper is concerned with an iterative method for finding the minimal norm least squares solution to linear matrix equation. We have achieved the following:

(1) Two iterative methods are proposed to solve this problem. These two methods deal with the case that Υ (see Equation (7)) is of full column rank and Υ is of full row rank, respectively.
(2) Necessary and sufficient conditions are provided to guarantee the convergence of the proposed algorithms.
(3) The optimal step size such that the convergence rates of the proposed algorithms are maximized is given explicitly in terms of the singular values of the matrix Υ.

Our results regarding numerical solutions to minimal norm least squares problem may have important applications in control system theory. Further study should focus on extending our methods to solve the problem of finding minimal norm least squares solutions to non-linear matrix equations.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 10771044) and the Natural Science Foundation of Heilongjiang Province (Grant No. 200605).

References

[1] R.H. Bartels and G.W. Stewart, *Solution of the equation* $AX + XB = C$, Comm. A. C. M. 15 (1972), pp. 820–826.
[2] S.P. Bhattacharyya and E. de Souza, *Pole assignment via Sylvester's equation*, Syst. Control Lett. 1(4) (1981), pp. 261–263.
[3] J.L. Chen and X.H. Chen, *Special Matrices*, Tsinghua University Press, Beijing, 2002 (in Chinese).
[4] T. Chen and L. Qiu, $\mathcal{H}_\infty$ *design of general multirate sampled-date control systems*, Automatica 30 (1994), pp. 1139–1152.
[5] D. Chu, H.C. Chan, and D.W.C. Ho, *Regularization of singular systems by derivative and proportional output feedback*, SIAM J. Matrix Anal. Appl. 19(1) (1998), pp. 21–38.
[6] D. Chu, V. Mehrmann, and N.K. Nichols, *Minimum norm regularization of descriptor systems by mixed output feedback*, Linear Algebra Appl. 296(1–3) (1999), pp. 39–77.
[7] F. Ding and T. Chen, *Iterative least squares solutions of coupled Sylvester matrix equations*, Syst. Control Lett. 54(2) (2005), pp. 95–107.
[8] F. Ding and T. Chen, *On iterative solutions of general coupled matrix equations*, SIAM J. Control Optim. 44(6) (2006), pp. 2269–2284.
[9] T. Gu, X. Liu, Z. Mo, and X. Chi, *Multiple search direction conjugate gradient method I: methods and their propositions*, Int. J. Comput. Math. 81(9) (2004), pp. 1133–1143.
[10] S.J. Hammarling, *Numerical solution of the stable, non-negative definite Lyapunov equation*, IMA J. Num. Anal. 2 (1982), pp. 303–323.
[11] T. Huang, G. Cheng, D.J. Evans, and X. Cheng, *AOR type iterations for solving preconditioned linear systems* Int. J. Comput. Math. 82(8) (2005), pp. 969–976.
[12] I.E. Kaporin, *Explicitly preconditioned conjugate gradient method for the solution of unsymmetric linear systems*, Int. J. Comput. Math. 44(1–4) (1992), pp. 169–187.
[13] C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.
[14] J. Lam and W.Y. Yan, *Pole assignment with optimal spectral conditioning*, Syst. Control Lett. 29(5) (1997), pp. 241–253.
[15] J. Lam, W. Yan, and T. Hu, *Pole assignment with eigenvalue and stability robustness*, Internat. J. Control 72(13) (1999), pp. 1165–1174.
[16] L. Qiu and T. Chen, *Unitary dilation approach to contractive matrix completion*, Linear Algebra Appl. 379 (2004), pp. 345–352.
[17] R.S. Varga, *Matrix Iterative Analysis*, 2nd ed., Series in Computational Mathematics, Springer-Verlag, New York, 2000.
[18] S.G. Wang and Z.H. Yang, *The Generalized Inverse of Matrix and its Applications*, Beijing University of Technology Press, Beijing, 1996 (in Chinese).
[19] Q. Wang, J. Lam, Y. Wei, and T. Chen, *Iterative solutions of coupled discrete Markovian jump Lyapunov equations*, Comput. Math. Appl. 55(4) (2008), pp. 843–850.
[20] G.F. Zhang, Q. Zhang, T. Chen, and Y. Lin, *On Lyapunov theorems for descriptor systems*, Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms 10(5) (2003), pp. 709–726.
[21] L. Zheng C. Li, and D. J. Evans, *Chebyshev acceleration for SOR-like method*, Int. J. Comput. Math. 82(5) (2005), pp. 583–593.
[22] B. Zhou and G. Duan, *An explicit solution to the matrix equation* $AX - XF = BY$, Linear Algebra Appl. 402 (2005), pp. 345–366.
[23] B. Zhou and G.R. Duan, *A new solution to the generalized Sylvester matrix equation* $AV - EVF = BW$, Syst. Control Lett. 55(3) (2006), pp. 193–198.
[24] B. Zhou and G. Duan, *Solutions to generalized Sylvester matrix equation by Schur decomposition*, Int. J. Syst. Sci. 38(5) (2007), pp. 369–375.
[25] B. Zhou and G. Duan, *Parametric solutions to the generalized Sylvester matrix equation* $AX - XF = BY$ *and the regulator equation* $AX - XF = BY + R$, Asian J. Control 9(4) (2007), pp. 475–483.
[26] B. Zhou and G.R. Duan, *On the generalized Sylvester mapping and matrix equations*, Syst. Control Lett. 57(3) (2008), pp. 200–208.
[27] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, New Jersey, 1996.
[28] B. Zhou, J. Lam, and G. Duan, *Gradient-based maximal convergence rate iterative method for solving linear matrix equations*, Int. J. Comput. Math. 87(3) (2010), pp. 515–527.

Appendix

A.1 Proof of Lemma 2

Since $\rho(\Theta) < 1$, Iteration (12) converges to a finite vector as k approaches to infinity with arbitrary initial condition. Equivalently, Iteration (11) converges to X_∞ as k approaches to infinity with arbitrary initial condition $X(0)$, i.e.,

$$X_\infty = \sum_{i=1}^{p} \mathcal{A}_i X_\infty \mathcal{B}_i + \mathcal{C}.$$

Substituting this equation into Equation (11) gives

$$X(k) - X_\infty = \sum_{i=1}^{p} \mathcal{A}_i (X(k-1) - X_\infty)\mathcal{B}_i,$$

which, by using the Kronecker product, is equivalent to

$$\mathrm{vec}(X(k) - X_\infty) = \Theta \mathrm{vec}(X(k-1) - X_\infty), \tag{A1}$$

where Θ is defined in Equation (12). Since Θ is real and symmetric, we have $\|\Theta\|_2 = \rho(\Theta)$. Therefore, it follows from Equation (A1) that

$$\begin{aligned} \|\mathrm{vec}(X(k) - X_\infty)\|_2 &\leq \|\Theta\|_2 \|\mathrm{vec}(X(k-1) - X_\infty)\|_2 \\ &= \rho(\Theta)\|\mathrm{vec}(X(k-1) - X_\infty)\|_2 \\ &\leq \rho^k(\Theta)\|\mathrm{vec}(X(0) - X_\infty)\|_2, \end{aligned}$$

which in turn implies Equation (14) in view of Equation (5). To complete the proof, we need only to show that there exists at least one initial condition $X(0)$ such that the '$=$' holds in Equation (14). Since Θ is real and symmetric, there exists a unitary matrix U such that

$$U^{\mathrm{T}} \Theta U = \begin{bmatrix} \sigma_1 I_{v_1} & 0 & \cdots & 0 \\ 0 & \sigma_2 I_{v_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_h I_{v_h} \end{bmatrix} := \Xi \in \mathbb{R}^{mn \times mn}, \tag{A2}$$

where $\sigma_i, i = 1, 2, \ldots, h$, are the real eigenvalues of Θ and $\Sigma_{i=1}^{h} v_i = mn$. Assume that $|\sigma_1| > |\sigma_2| > \cdots > |\sigma_h|$. Then we clearly have $\rho(\Theta) = |\sigma_1|$. It follows from Equation (A1) and (A2) that

$$\begin{aligned} \mathrm{vec}(X(k) - X_\infty) &= \Theta^k \mathrm{vec}(X(0) - X_\infty) \\ &= U \Xi^k U^{\mathrm{T}} \mathrm{vec}(X(0) - X_\infty), \end{aligned}$$

or equivalently,

$$U^{\mathrm{T}} \mathrm{vec}(X(k) - X_\infty) = \Xi^k U^{\mathrm{T}} \mathrm{vec}(X(0) - X_\infty). \tag{A3}$$

Now we choose the initial condition $X_\#(0)$ such that

$$\mathrm{vec}(X_\#(0)) = U \begin{bmatrix} x_\# \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \mathrm{vec}(X_\infty), \tag{A4}$$

where $x_\#$ is a non-zero scalar. Then we clearly have $X_\#(0) \neq X_\infty$. By using Equation (A3), we can obtain

$$\begin{aligned} U^{\mathrm{T}} \mathrm{vec}(X(k) - X_\infty) &= \begin{bmatrix} \sigma_1^k I_{v_1} & 0 & \cdots & 0 \\ 0 & \sigma_2^k I_{v_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_h^k I_{v_h} \end{bmatrix} \begin{bmatrix} x_\# \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^k x_\# \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \end{aligned}$$

Equation (A4) clearly implies that $\|\mathrm{vec}(X_{\#}(0) - X_{\infty})\|_2 = |x_{\#}|$. Therefore,

$$\begin{aligned}\|\mathrm{vec}(X(k) - X_{\infty})\|_2 &= \|U^{\mathrm{T}}\mathrm{vec}(X(k) - X_{\infty})\|_2 \\ &= |\sigma_1^k||x_{\#}| = \rho^k(\Theta)|x_{\#}| \\ &= \rho^k(\Theta)\|\mathrm{vec}(X_{\#}(0) - X_{\infty})\|_2.\end{aligned}$$

That is to say

$$\|X(k) - X_{\infty}\|_{\mathrm{F}} = \rho^k(\Theta)\|X_{\#}(0) - X_{\infty}\|_{\mathrm{F}}.$$

The above equation and Inequality (14) complete the proof. ■

A.2 Proof of Lemma 4

We first introduce the following lemma.

LEMMA 5 ([27]) *Let A, B, and X be some matrices with appropriate dimensions. Then*

$$\frac{\partial \mathrm{tr}(AXB)}{\partial X} = A^{\mathrm{T}}B^{\mathrm{T}}, \quad \frac{\partial \mathrm{tr}(AX^{\mathrm{T}}B)}{\partial X} = BA \tag{A5}$$

$$\frac{\partial \mathrm{tr}(AXBX^{\mathrm{T}})}{\partial X} = A^{\mathrm{T}}XB^{\mathrm{T}} + AXB. \tag{A6}$$

Now we start to prove the lemma. Note that

$$\begin{aligned}J(X) &= \frac{1}{2}\left\|\sum_{i=1}^{r} A_i X B_i - C\right\|_{\mathrm{F}}^2 \\ &= \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{r} B_i^{\mathrm{T}} X^{\mathrm{T}} A_i^{\mathrm{T}} - C^{\mathrm{T}}\right)\left(\sum_{j=1}^{r} A_j X B_j - C\right) \\ &= \frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{r}\sum_{j=1}^{r} B_i^{\mathrm{T}} X^{\mathrm{T}} A_i^{\mathrm{T}} A_j X B_j\right) - \mathrm{tr}\left(C^{\mathrm{T}}\sum_{j=1}^{r} A_j X B_j\right) + \mathrm{tr}(C^{\mathrm{T}}C).\end{aligned}$$

Therefore, in view of Equation (A5), (A6), and the formulation $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, we have

$$\begin{aligned}\frac{\partial J(X)}{\partial X} &= \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{\partial}{\partial X}\mathrm{tr}(B_i^{\mathrm{T}} X^{\mathrm{T}} A_i^{\mathrm{T}} A_j X B_j) - \sum_{j=1}^{r}\frac{\partial}{\partial X}\mathrm{tr}(C^{\mathrm{T}} A_j X B_j) \\ &= \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{\partial}{\partial X}\mathrm{tr}(A_i^{\mathrm{T}} A_j X B_j B_i^{\mathrm{T}} X^{\mathrm{T}}) - \sum_{j=1}^{r}\frac{\partial}{\partial X}\mathrm{tr}(C^{\mathrm{T}} A_j X B_j) \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} A_i^{\mathrm{T}} A_j X B_j B_i^{\mathrm{T}} - \sum_{j=1}^{r} A_j^{\mathrm{T}} C B_j^{\mathrm{T}} \\ &= \sum_{j=1}^{r} A_j^{\mathrm{T}}\left(\sum_{i=1}^{r} A_i X B_i - C\right) B_j^{\mathrm{T}}.\end{aligned}$$

This completes the proof of Lemma 4. ■