Iterative least-squares solutions of coupled Sylvester matrix equations

Feng Ding 1, Tongwen Chen ∗
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2V4

Received 24 July 2003; received in revised form 9 June 2004; accepted 22 June 2004
Available online 26 August 2004

Abstract

In this paper, we present a general family of iterative methods to solve linear equations, which includes the well-known Jacobi and Gauss–Seidel iterations as its special cases. The methods are extended to solve coupled Sylvester matrix equations. In our approach, we regard the unknown matrices to be solved as the system parameters to be identified, and propose a least-squares iterative algorithm by applying a hierarchical identification principle and by introducing the block-matrix inner product (the star product for short). We prove that the iterative solution consistently converges to the exact solution for any initial value. The algorithms proposed require less storage capacity than the existing numerical ones. Finally, the algorithms are tested on computer and the results verify the theoretical findings.

Keywords: Sylvester matrix equation; Lyapunov matrix equation; Identification; Estimation; Least squares; Jacobi iteration; Gauss–Seidel iteration; Hadamard product; Star product; Hierarchical identification principle

1. Introduction

Lyapunov and Sylvester matrix equations play important roles in system theory [5,6,33–35]. Although exact solutions, which can be computed by using the Kronecker product, are important, the computational efforts rapidly increase with the dimensions of the matrices to be solved. For some applications such as stability analysis, it is often not necessary to compute exact solutions; approximate solutions or bounds of solutions are sufficient. Also, if the
parameters in system matrices are uncertain, it is not possible to obtain exact solutions for robust stability results [10,12,16,21–26,28–32,37].

Alternative ways exist which transform the matrix equations into forms for which solutions may be readily computed, for example, the Jordan canonical form [15], companion-type form [2,3], Hessenberg–Schur form [1,13]. In this area, Chu gave a numerical algorithm for solving the coupled Sylvester equations [7]; and Borno presented a parallel algorithm for solving the coupled Lyapunov equations [4]. But, these algorithms require computing some additional matrix transformation/decomposition; moreover, they are not suitable for more general coupled matrix equations of the form:

\[
\sum_{j=1}^{p} A_{ij} X_j B_{ij} = C_i, \quad i = 1, 2, \ldots, p,
\]

which includes the coupled Lyapunov and Sylvester equations as its special cases. In (1), \( X_j \in \mathbb{R}^{m \times n} \) are the unknown matrices to be solved; \( A_{ij}, B_{ij}, \) and \( C_{ij} \) represent constant (coefficient) matrices of appropriate dimensions. For such coupled matrix equations, the conventional methods require dealing with matrices whose dimensions are \( mn \times mn \). Such a dimensionality problem leads to computational difficulty in that excessive computer memory is required for computation and inversion of large matrices of size \( mn \times mn \). For instance, if \( m = n = p = 100 \), then \( mn \times mn = 10^6 \times 10^6 \).

In the field of matrix algebra and system identification, iterative algorithms have received much attention [8,10,14,27,32]. For example, Starke presented an iterative method for solutions of the Sylvester equations by using the SOR technique [36]; Jonsson and Kägström proposed recursive block algorithms for solving the coupled Sylvester matrix equations [18,19]; Kägström derived an approximate solution of the coupled Sylvester equation [20]. To our best knowledge, numerical algorithms for general matrix equations have not been fully investigated, especially the iterative solutions of the coupled Sylvester matrix equations, as well as the general coupled matrix equations in (1), and the convergence of the iterative solutions involved, which are the focus of this work.

In this paper, the problem will be tackled in a new way—we regard the unknown matrices \( X_j \) to be solved as the parameters (parameter matrices) of the system to be identified, and apply the so-called hierarchical identification principle to decompose the system into some subsystems, and derive iterative algorithms of the matrix equations involved. Our methods will generate solutions to the matrix equations which are arbitrarily close to the exact solutions.

The paper is organized as follows. In Section 2, we extend the well-known Jacobi and Gauss–Seidel iterations and present a large family of iterative methods. In Sections 3 and 4, we define the block-matrix inner product (the star product for short) and derive iterative algorithms for the coupled Sylvester matrix equations and general coupled matrix equations, respectively, and study the convergence properties of the algorithms. In Section 5, we give an example for illustrating the effectiveness of the algorithms proposed. Finally, we offer some concluding remarks in Section 6.

2. Extension of the Jacobi and Gauss–Seidel iterations

Consider the following linear equation:

\[
Ax = b.
\]

Here, \( A = [a_{ij}] \) \((i, j = 1, 2, \ldots, n)\) is a given full-rank \( n \times n \) matrix with non-zero diagonal elements, \( b \in \mathbb{R}^n \) is a constant vector, and \( x \in \mathbb{R}^n \) an unknown vector to be solved. Let \( D \) be the diagonal part of \( A \), and \( L \) and \( U \) be the
strictly lower and upper triangular parts of $A$:

$$D = \text{diag}[a_{11}, a_{22}, \ldots, a_{nn}] \in \mathbb{R}^{n \times n},$$

$$L = \begin{bmatrix}
0 & 0 & \cdots & \cdots & 0 \\
a_{21} & 0 & \cdots & \vdots & \vdots \\
a_{31} & a_{32} & 0 & \cdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{n,n-1} & 0
\end{bmatrix} \in \mathbb{R}^{n \times n},$$

$$U = \begin{bmatrix}
0 & a_{12} & a_{13} & \cdots & a_{1n} \\
0 & 0 & a_{23} & \cdots & a_{2n} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & a_{n-1,n}
\end{bmatrix} \in \mathbb{R}^{n \times n},$$

which satisfy $L + D + U = A$. Then both the Jacobi and Gauss–Seidel iterations can be expressed as \[8,14\]

$$Mx(k) = Nx(k-1) + b, \quad k = 1, 2, 3, \ldots,$$

where $x(k)$ is the iterative solution of $x$. For the Jacobi method, $M = D$ and $N = -(L + U)$; for the Gauss–Seidel method, $M = L + D$ and $N = -U$.

Unfortunately, the Jacobi and Gauss–Seidel iterations cannot guarantee that $x(k)$ converges to the exact solution $x = A^{-1}b$, and are not suitable for solving the non-square system: $Hx = g$ with $H \in \mathbb{R}^{m \times n}$. This motivates us to study new iterative methods.

Let $G \in \mathbb{R}^{n \times n}$ be a full-rank matrix to be determined and $\mu > 0$ be the step-size or convergence factor. We present a large family of iterative methods as follows:

$$x(k) = x(k-1) + \mu G [b - Ax(k-1)], \quad k = 1, 2, 3, \ldots,$$

which includes the Jacobi and Gauss–Seidel iterations as special cases. For example, when $G = D^{-1}$ and $\mu = 1$, we get the Jacobi method; when $G = (L + D)^{-1}$ and $\mu = 1$, we obtain the Gauss–Seidel method.

The following two lemmas are straightforward and their proofs are omitted here.

**Lemma 1.** If we take $G = A^T$, then the gradient iterative (or iterative gradient) algorithm,

$$x(k) = x(k-1) + \mu A^T [b - Ax(k-1)], \quad 0 < \mu < \frac{2}{\lambda_{\max}[A^TA]} \quad \text{or} \quad 0 < \mu < \frac{2}{\|A\|^2},$$

yields $\lim_{k \to \infty} x(k) = x$. Here, $\|X\|^2 = \text{tr}[XX^T]$.

**Lemma 2.** If we take $G = A^{-1}$, then the following iterative algorithm converges to $x$:

$$x(k) = x(k-1) + \mu A^{-1} [b - Ax(k-1)], \quad 0 < \mu < 2.$$

If $A$ is a non-square $m \times n$ full column-rank matrix, then we have $\lim_{k \to \infty} x(k) = x$ in the following:

$$x(k) = x(k-1) + \mu (A^T A)^{-1} A^T [b - Ax(k-1)], \quad 0 < \mu < 2.$$

It is easy to prove that the iterative solutions $x(k)$ in (4)–(6) all converge to the least-squares solution $(A^T A)^{-1} A^T b$ at a fast exponential rate, or are linearly convergent. When $\mu = 1$, the iteration in (6) gives $x(1) = (A^T A)^{-1} A^T b$. So (6) is also called the least-squares iterative algorithm.

The iterative algorithms in (4) and (6) are also suitable for solving non-square systems and are very useful for finding the iterative solutions of general matrix equations to be studied later; the convergence factors $\mu$ in (5) and (6) do not rely on the matrix $A$ and is easy to choose, although the algorithms in (5) and (6) require computing matrix inversion only at the first step.
3. Coupled Sylvester matrix equations

In this section, we study least squares iterative algorithms to solve the coupled Sylvester matrix equation

\[ AX + YB = C, \quad DX + YE = F. \] (7)

Here, \( A, D \in \mathbb{R}^{m \times m}, B, E \in \mathbb{R}^{n \times n} \) and \( C, F \in \mathbb{R}^{m \times n} \) are given constant matrices, \( X, Y \in \mathbb{R}^{m \times n} \) are the unknown matrices to be solved.

First, let us introduce some notation. The notation \( I_n \) is the identity matrix of \( n \times n \). For two matrices \( M \) and \( N \), \( M \otimes N \) is their Kronecker product. For two \( m \times n \) matrices \( X \) and \( Y \) with \( X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m \times n} \), \( \text{col}[X] \) is an \( mn \)-dimensional vector formed by columns of \( X \)

\[ \text{col}[X] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{mn} \quad \text{and} \quad \text{col}[X, Y] = \begin{bmatrix} \text{col}[X] \\ \text{col}[Y] \end{bmatrix} \in \mathbb{R}^{2mn}. \]

The following result is well-known.

**Lemma 3.** Eq. (7) has a unique solution if and only if the matrix

\[ S_2 := \begin{bmatrix} I_n \otimes A & B^T \otimes I_n \\ I_n \otimes D & E^T \otimes I_m \end{bmatrix} \in \mathbb{R}^{(2mn) \times (2mn)} \]

is non-singular; in this case, the unique solution is given by

\[ \text{col}[X, Y] = S_2^{-1} \text{col}[C, F], \] (8)

and the corresponding homogeneous matrix equation \( AX + YB = 0, \quad DX + YE = 0 \) has a unique solution: \( X = Y = 0 \).

In order to derive the iterative solution to (7), we need to introduce the intermediate matrices \( b_1 \) and \( b_2 \) as follows:

\[ b_1 := \begin{bmatrix} C - YB \\ F - YE \end{bmatrix}, \]

\[ b_2 := [C - AX, \quad F - DX]. \] (10)

Then from (7), we obtain two fictitious subsystems

\[ S_1 : G_1 X = b_1, \quad S_2 : YH_1 = b_2. \]

Here, \( G_1 := \begin{bmatrix} A \\ D \end{bmatrix} \) and \( H_1 := [B, E] \).

Let \( X(k) \) and \( Y(k) \) be the iterative solutions of \( X \) and \( Y \). Referring to Lemma 2, it is not difficult to get the iterative solutions to \( S_1 \) and \( S_2 \) as follows:

\[ X(k) = X(k - 1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \left\{ b_1 - \begin{bmatrix} A \\ D \end{bmatrix} X(k - 1) \right\}, \] (11)

\[ Y(k) = Y(k - 1) + \mu (b_2 - Y(k - 1)B, E)[B, E]^T (H_1 H_1^T)^{-1}. \] (12)
Substituting (9) into (11) and (10) into (12) gives

\[
X(k) = X(k-1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \begin{bmatrix} C - YB \\ F - YE \end{bmatrix} - \begin{bmatrix} A \\ D \end{bmatrix} X(k-1)
\]

\[
= X(k-1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \begin{bmatrix} C - YB - AX(k-1) \\ F - YE - DX(k-1) \end{bmatrix}.
\]

(13)

\[
Y(k) = Y(k-1) + \mu \{ C - AX(k-1) - Y(k-1)B; \quad F - DX(k-1) - Y(k-1)E \}[B; \quad E]^T (H_1 H_1^T)^{-1},
\]

(14)

Here, a difficulty arises in that the expressions on the right-hand sides of (13) and (14) contain the unknown parameter matrix \( Y \), respectively, so it is impossible to realize the algorithm in (13) and (14). Our solution is based on the hierarchical identification principle: the unknown variables \( Y \) in (13) and \( X \) in (14) are replaced by their estimates \( Y(k-1) \) and \( X(k-1) \). Thus, we obtain the least-squares iterative solutions \( X(k) \) and \( Y(k) \) of the coupled Sylvester equation in (7)

\[
X(k) = X(k-1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \begin{bmatrix} C - AX(k-1) - Y(k-1)B \\ F - DX(k-1) - Y(k-1)E \end{bmatrix},
\]

(15)

\[
Y(k) = Y(k-1) + \mu \{ C - AX(k-1) - Y(k-1)B; \quad F - DX(k-1) - Y(k-1)E \}[B; \quad E]^T (H_1 H_1^T)^{-1},
\]

(16)

\[
\mu = \frac{1}{m+n} \quad \text{or} \quad \mu = \frac{1}{\lambda_{\text{max}}[G_1 (G_1^T G_1)^{-1} G_1^T] + \lambda_{\text{max}}[H_1^T (H_1 H_1^T)^{-1} H_1]}. \tag{17}
\]

The least-squares iterative algorithm in (15)–(17) requires computing the matrix inversions \( (G_1^T G_1)^{-1} \) and \( (H_1 H_1^T)^{-1} \) only once at the first step. To initialize the algorithm, we take \( X(0) = Y(0) = 0 \) or some small real matrix, e.g., \( X(0) = Y(0) = 10^{-6} I_{m \times n} \) with \( I_{m \times n} \) being an \( m \times n \) matrix whose elements are 1.

**Theorem 1.** If the coupled Sylvester equation in (7) has a unique solution \( X \) and \( Y \), then the iterative solution \( X(k) \) and \( Y(k) \) given by the algorithm in (15)–(17) converges to \( X \) and \( Y \) for any finite initial values \( X(0) \) and \( Y(0) \), i.e.,

\[
\lim_{k \to \infty} X(k) = X \quad \text{and} \quad \lim_{k \to \infty} Y(k) = Y.
\]

**Proof.** Define two error matrices

\[
\tilde{X}(k) = X(k) - X, \quad \tilde{Y}(k) = Y(k) - Y.
\]

By using (7) and (15)–(16), it is not difficult to get

\[
\tilde{X}(k) = \tilde{X}(k-1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \begin{bmatrix} -AX(k-1) - \tilde{Y}(k-1)B \\ -DX(k-1) - \tilde{Y}(k-1)E \end{bmatrix},
\]

(18)

\[
\tilde{Y}(k) = \tilde{Y}(k-1) + \mu \{ -AX(k-1) - \tilde{Y}(k-1)B; \quad -DX(k-1) - \tilde{Y}(k-1)E \}[B; \quad E]^T (H_1 H_1^T)^{-1}, \tag{19}
\]

Taking the norm in (18) and using the formula

\[
\|G_1 [X + (G_1^T G_1)^{-1} Y]\|^2 = \text{tr}[(X + (G_1^T G_1)^{-1} Y)^T (G_1^T G_1)(X + (G_1^T G_1)^{-1} Y)]
\]

\[
= \text{tr}[X^T (G_1^T G_1) X + 2X^T Y + Y^T (G_1^T G_1)^{-1} Y]
\]

\[
= \|G_1 X\|^2 + 2\text{tr}[X^T Y] + \|(G_1^T G_1)^{-1/2} Y\|^2,
\]

Theorem 1 follows.
According to Lemma 3, we have
\[ \hat{X}(k) \rightarrow 0 \] then
[43x80]or
[43x206]and using (20) and (21), we have
\[ k \rightarrow \infty \]
It follows that as \( k \rightarrow \infty \),
\[ A \hat{X}(k) + \hat{Y}(k) B = 0, \quad D \hat{X}(k) + \hat{Y}(k) E = 0. \]
According to Lemma 3, we have \( \hat{X}(k) \rightarrow 0 \) and \( \hat{Y}(k) \rightarrow 0 \) as \( k \rightarrow \infty \). Thus Theorem 1 is proven. \( \square \)
Then the block-matrix star product is defined as
\[
\ast
\]
\[
A_{ij}X_{ij}B_{ij} + A_{12}X_{21}B_{12} + \cdots + A_{1p}X_{p1}B_{1p} = C_1,
\]
\[
A_{21}X_{21}B_{21} + A_{22}X_{22}B_{22} + \cdots + A_{2p}X_{p2}B_{2p} = C_2,
\]
\[
\cdots
\]
\[
A_{p1}X_{p1}B_{p1} + A_{p2}X_{p2}B_{p2} + \cdots + A_{pp}X_{pp}B_{pp} = C_p.
\]
(22)

Here, \( A_{ij} \in \mathbb{R}^{m \times m} \), \( B_{ij} \in \mathbb{R}^{n \times n} \) and \( C_i \in \mathbb{R}^{m \times n} \) are given constant matrices, \( X_i \in \mathbb{R}^{m \times n} \) are the unknown matrix to be solved.

The general coupled matrix equations (22) include the following matrix equations as the special cases: (i) the discrete-time Sylvester equation: \( AXB^T + X = C \) [18,19]; (ii) the discrete-time Lyapunov equation: \( AXA^T - X = C \) [18,19]; (iii) the generalized Sylvester (Lyapunov) equation: \( AXB + CXD = F \) [7]; (iv) the coupled Sylvester equations as discussed in the preceding section [7,20]; and (v) the general coupled Lyapunov matrix equations associated with linear jump parameter systems [4].

In order to more succinctly express the least-squares iterative algorithm to be presented later, we introduce the block-matrix star product—the star (\( \ast \)) product for short, denoted by notation \( \ast \), which differs from Hadamard (inner) product [9,11,17,38] and general matrix multiplication. Let
\[
X = \begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_p
\end{bmatrix} \in \mathbb{R}^{(mp) \times n},
Y = \begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_p
\end{bmatrix} \in \mathbb{R}^{(np) \times m},
A_i = \begin{bmatrix}
A_{i1} \\
A_{i2} \\
\vdots \\
A_{ip}
\end{bmatrix} \in \mathbb{R}^{(mp) \times m},
\]
\[
B_i = [B_{i1}, B_{i2}, \ldots, B_{pi}] \in \mathbb{R}^{n \times (np)},
S_A = [A_{ij}],
S_B = [B_{ij}],
S_B^T = [B_{ij}^T],
S_p = [B_{ij} \otimes A_{ij}], \quad i, j = 1, 2, \ldots, p.
\]

Then the block-matrix star product is defined as
\[
X \ast Y = \begin{bmatrix}
X_1 & Y_1 \\
X_2 & Y_2 \\
\vdots & \vdots \\
X_p & Y_p
\end{bmatrix},
S_A \ast X = \begin{bmatrix}
A_{11}X_1 & A_{12}X_2 & \cdots & A_{1p}X_p \\
A_{21}X_1 & A_{22}X_2 & \cdots & A_{2p}X_p \\
\vdots & \vdots & \ddots & \vdots \\
A_{p1}X_1 & A_{p2}X_2 & \cdots & A_{pp}X_p
\end{bmatrix},
\]
\[
X \ast S_B = \begin{bmatrix}
X_1B_{11} & X_1B_{12} & \cdots & X_1B_{1p} \\
X_2B_{21} & X_2B_{22} & \cdots & X_2B_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
X_pB_{p1} & X_pB_{p2} & \cdots & X_pB_{pp}
\end{bmatrix},
S_A \ast S_B = \begin{bmatrix}
A_{11}B_{11} & A_{12}B_{12} & \cdots & A_{1p}B_{1p} \\
A_{21}B_{21} & A_{22}B_{22} & \cdots & A_{2p}B_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
A_{p1}B_{p1} & A_{p2}B_{p2} & \cdots & A_{pp}B_{pp}
\end{bmatrix}.
\]

In the above definitions, we assume that the dimensions of multiplier matrix and multiplicand matrix are compatible.

The block-matrix star Kronecker product, denoted by notation \( \otimes \), is defined by
\[
S_B^T \otimes S_A = S_p.
\]
Taking into account the dimension compatibility, the star product is superior to matrix multiplication. Note that \( AB \ast C = A(B \ast C) \neq (AB) \ast C \).
Lemma 4. Eq. (22) has a unique solution if and only if the matrix $S_p \in \mathbb{R}^{(mnp) \times (mnp)}$ is nonsingular; in this case, the solution is

$$\text{col}[X_1, X_2, \ldots, X_p] = S_p^{-1} \text{col}[C_1, C_2, \ldots, C_p];$$

and if $C_i = 0$ ($i = 1, 2, \ldots, p$), then the matrix equation in (22) has unique solutions $X_i = 0$ ($i = 1, 2, \ldots, p$).

In order to derive the iterative algorithm for solving the general coupled matrix equation in (22), we first consider the coupled Sylvester equation in (7) to a more general form

$$AXIB + IA YB = C, \quad DXIE + IDYE = F,$$

whose iterative solution can be expressed as

$$X(k) = X(k - 1) + \mu (G_1^T G_1)^{-1} \begin{bmatrix} A \\ D \end{bmatrix}^T \begin{bmatrix} [C - AX(k - 1)IB - IA Y(k - 1)B] \\ [F - DX(k - 1)IE - ID Y(k - 1)E] \end{bmatrix} * [I_B, I_E]^T,$$  \hspace{1cm} (23)

$$Y(k) = Y(k - 1) + \mu \begin{bmatrix} I_A \\ I_D \end{bmatrix}^T \begin{bmatrix} [C - AX(k - 1)IE - ID Y(k - 1)B] \\ [F - DX(k - 1)IE - ID Y(k - 1)E] \end{bmatrix} * [B, E]^T (H_1 H_1^T)^{-1}. \hspace{1cm} (24)$$

If $I_A$, $I_B$, $I_D$ and $I_E$ are identity matrices of appropriate dimensions, then the algorithm in (23) and (24) is equivalent to the one in (15) and (16).

Let $X_i(k)$ be the estimates or iterative solutions of $X_i$. We present the least squares iterative algorithm of computing the solutions $X_i(k)$ ($i = 1, 2, \ldots, p$) of the matrix equations (22) as follows:

$$X_i(k) = X_i(k - 1) + \mu (A_i^T A_i)^{-1} \begin{bmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ A_{ip} \end{bmatrix}^T \begin{bmatrix} C_1 - \sum_{j=1}^{p} A_{1j} X_j(k - 1) B_{1j} \\ C_2 - \sum_{j=1}^{p} A_{2j} X_j(k - 1) B_{2j} \\ \vdots \\ C_p - \sum_{j=1}^{p} A_{pj} X_j(k - 1) B_{pj} \end{bmatrix} * \begin{bmatrix} B_{1i}^T \\ B_{2i}^T \\ \vdots \\ B_{pi}^T \end{bmatrix} (B_i B_i^T)^{-1}, \hspace{1cm} (25)$$
Theorem 2. If the coupled matrix equation in (22) has unique solutions $X_i$, $i = 1, 2, \ldots, p$, then the iterative solutions $X_i(k)$ given by the algorithm in (25)–(26) converge to the solutions $X_i$ for any finite initial values $X_i(0)$, i.e.,

$$
\lim_{k \to \infty} X_i(k) = X_i, \quad i = 1, 2, \ldots, p.
$$

Proof. Define the estimation error matrix

$$
\tilde{X}_i(k) = X_i(k) - X_i.
$$

Let

$$
\begin{bmatrix}
\tilde{C}_1(k) \\
\tilde{C}_2(k) \\
\vdots \\
\tilde{C}_p(k)
\end{bmatrix} = \begin{bmatrix}
\sum_{j=1}^{p} A_{1j} \tilde{X}_j(k-1) B_{1j} \\
\sum_{j=1}^{p} A_{2j} \tilde{X}_j(k-1) B_{2j} \\
\vdots \\
\sum_{j=1}^{p} A_{pj} \tilde{X}_j(k-1) B_{pj}
\end{bmatrix}.
$$

By using (22) and (25), it is not difficult to get

$$
\tilde{X}_i(k) = \tilde{X}_i(k-1) - \mu (A_i^T A_i)^{-1} \begin{bmatrix}
A_{1i}^T \\
A_{2i}^T \\
\vdots \\
A_{pi}^T
\end{bmatrix}^T \begin{bmatrix}
\tilde{C}_1(k) \\
\tilde{C}_2(k) \\
\vdots \\
\tilde{C}_p(k)
\end{bmatrix} + \begin{bmatrix}
B_{1i}^T \\
B_{2i}^T \\
\vdots \\
B_{pi}^T
\end{bmatrix} (B_i B_i^T)^{-1}.
$$

Defining a non-negative definite function

$$
V_i(k) = \| A_i \tilde{X}(k) B_i \|^2,
$$

using the above equation, the star product properties and formula

$$
\text{tr}[\{X + (A_i^T A_i)^{-1} Y (B_i B_i^T)^{-1} (A_i^T A_i)^{-1} (X + (A_i^T A_i)^{-1} Y (B_i B_i^T)^{-1} (X + (A_i^T A_i)^{-1} Y (B_i B_i^T)^{-1})\} (B_i B_i^T)]
$$

$$
= \text{tr}[(A_i^T A_i)^{-1} Y (B_i B_i^T)^{-1} Y (A_i^T A_i)^{-1} Y]
$$

$$
= \|A_i X B_i\|^2 + 2\text{tr}[X^T Y] + \text{tr}[(B_i B_i^T)^{-1} Y (A_i^T A_i)^{-1} Y]
$$

$$
\leq \|A_i X B_i\|^2 + 2\text{tr}[X^T Y] + \|A_i^T A_i\|^{-1/2} \|Y (B_i B_i^T)^{-1}\|^{-1/2}^2,
$$

we have

$$
V_i(k) = \text{tr}[\tilde{X}(k)^T (A_i^T A_i) \tilde{X}(k) (B_i B_i^T)]
$$

$$
\leq V_i(k-1) - 2\mu \text{tr} \begin{bmatrix}
A_{1i} \tilde{X}_i(k-1) B_{1i} \\
A_{2i} \tilde{X}_i(k-1) B_{2i} \\
\vdots \\
A_{pi} \tilde{X}_i(k-1) B_{pi}
\end{bmatrix}^T \begin{bmatrix}
\tilde{C}_1(k) \\
\tilde{C}_2(k) \\
\vdots \\
\tilde{C}_p(k)
\end{bmatrix} + \mu^2 \| \begin{bmatrix}
\tilde{C}_1(k) \\
\tilde{C}_2(k) \\
\vdots \\
\tilde{C}_p(k)
\end{bmatrix} \|^2.
$$

(27)
Summing for $i$ from 1 to $p$ yields

$$V(k) := \sum_{i=1}^{p} V_i(k)$$

$$\leq V(k - 1) - 2\mu \left\| \begin{bmatrix} \tilde{C}_1(k) \\ \vdots \\ \tilde{C}_p(k) \end{bmatrix} \right\|^2 + \mu^2 mnp \left\| \begin{bmatrix} \tilde{C}_1(k) \\ \vdots \\ \tilde{C}_p(k) \end{bmatrix} \right\|^2$$

$$= V(k - 1) - \mu (2 - \mu mnp) \sum_{i=1}^{p} \| \tilde{C}_i(k) \|^2$$

$$= V(k - 1) - \mu (2 - \mu mnp) \sum_{i=1}^{k} \sum_{l=1}^{p} \| \tilde{C}_i(l) \|^2.$$

If the convergence factor $\mu$ is chosen to satisfy

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

then

$$\sum_{k=1}^{\infty} \sum_{i=1}^{p} \| \tilde{C}_i(k) \|^2 < \infty.$$

It follows that as $k \to \infty$,

$$\sum_{i=1}^{p} \| \tilde{C}_i(k) \|^2 = \sum_{j=1}^{p} \| A_{ij} \tilde{X}_j(k - 1) B_{ij} \|^2 = 0,$$

or

$$\sum_{j=1}^{p} A_{ij} \tilde{X}_j(k - 1) B_{ij} = 0, \quad i = 1, 2, \ldots, p.$$

According to Lemma 4, we prove Theorem 2. □

From the proofs of Theorems 1 and 2, we can see that the iterative solutions in (15)–(17) and (25)–(26) are linearly convergent.

Let

$$X(k) = \begin{bmatrix} X_1(k) \\ X_2(k) \\ \vdots \\ X_p(k) \end{bmatrix} \in \mathbb{R}^{(mp) \times n}, \quad C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_p \end{bmatrix} \in \mathbb{R}^{(mp) \times n},$$

$$D_A = \text{diag}[(A_1^T A_1), (A_2^T A_2), \ldots, (A_p^T A_p)], \quad D_B = \text{diag}[(B_1 B_1^T), (B_2 B_2^T), \ldots, (B_p B_p^T)].$$

Then (22) can simply be expressed as

$$S_A * X * S_B I_{mp \times n} = C.$$
By using the star product properties, (25) can be written as the following more compact form:

\[
X(k) = X(k-1) + \mu D_A^{-1} S^T_A \begin{bmatrix} 
C_1 - \sum_{j=1}^p A_{1j} X_j(k-1) B_{1j} \\
C_2 - \sum_{j=1}^p A_{2j} X_j(k-1) B_{2j} \\
\vdots \\
C_p - \sum_{j=1}^p A_{pj} X_j(k-1) B_{pj} 
\end{bmatrix} * S_B^T D_B^{-1} 
= X(k-1) + \mu D_A^{-1} S^T_A \{ C - S_A * X(k-1) * S_B I_{n \times n} \} * S_B^T D_B^{-1}.
\]

Referring to Lemma 1, we also establish the gradient iterative algorithm for the solution of the general coupled matrix equation (22) as follows:

\[
X(k) = X(k-1) + \mu S^T_A \begin{bmatrix} 
C_1 - \sum_{j=1}^p A_{1j} X_j(k-1) B_{1j} \\
C_2 - \sum_{j=1}^p A_{2j} X_j(k-1) B_{2j} \\
\vdots \\
C_p - \sum_{j=1}^p A_{pj} X_j(k-1) B_{pj} 
\end{bmatrix} * S_B^T 
= X(k-1) + \mu S^T_A \{ C - S_A * X(k-1) * S_B I_{n \times n} \} * S_B^T,
\]

\[
\mu = \frac{1}{\sum_{i=1}^p \sum_{j=1}^p \|A_{ij} B_{ij}\|^2} 
\text{ or } 
\mu = \frac{1}{\sum_{i=1}^p \sum_{j=1}^p \max_{1 \leq i,j \leq p} |A_{ij} A^T_{ij}| \max_{1 \leq i,j \leq p} |B_{ij} B^T_{ij}|}.
\]

5. Example

In this section, we give an example to illustrate the performance of the proposed algorithms.

Suppose that the coupled matrix equations are \(AX + BY = C\), \(DX + YE = F\) with

\[
A = \begin{bmatrix} 2.00 & 1.00 \\ -1.00 & 2.00 \end{bmatrix}, \quad B = \begin{bmatrix} 1.00 & -0.20 \\ 0.20 & 1.00 \end{bmatrix}, \quad D = \begin{bmatrix} -2.00 & -0.50 \\ 0.50 & 2.00 \end{bmatrix},
\]

\[
E = \begin{bmatrix} -1.00 & -3.00 \\ 2.00 & -4.00 \end{bmatrix}, \quad C = \begin{bmatrix} 13.20 & 10.60 \\ 0.60 & 8.40 \end{bmatrix}, \quad F = \begin{bmatrix} -9.50 & -18.00 \\ 16.00 & 3.50 \end{bmatrix}.
\]

Then the solutions of \(X\) and \(Y\) from (8) are

\[
X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} 4.00 & 3.00 \\ 3.00 & 4.00 \end{bmatrix}, \quad Y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} 2.00 & 1.00 \\ -2.00 & 3.00 \end{bmatrix}.
\]

Taking \(X(0) = Y(0) = 10^{-6} \mathbf{1}_{2 \times 2}\), we apply the algorithm in (15) and (16) to compute \(X(k)\) and \(Y(k)\). The iterative solutions \(X(k)\) and \(Y(k)\) is shown in Table 1, where

\[
\delta = \sqrt{\frac{\|X(k) - X\|^2 + \|Y(k) - Y\|^2}{\|X\|^2 + \|Y\|^2}}
\]

is the relative error. The errors \(\delta\) with different convergence factors are shown in Fig. 1. From Table 1 and Fig. 1, it is clear that \(\delta\) are becoming smaller and smaller and goes to zero as \(k\) increases. This indicates that the proposed algorithm is effective.

The effect of changing the convergence factor \(\mu\) is illustrated in Fig. 1. We see that the larger the convergence factor \(\mu\) is, the faster the convergence the algorithm (or, the smaller the estimation error). However, if \(\mu\) is too large, the algorithm may diverge. How to choose a best convergence factor is still a project to be studied.
Table 1
The iterative solutions ($\mu = 1/1.10$)

<table>
<thead>
<tr>
<th>$k$</th>
<th>$x_{11}$</th>
<th>$x_{12}$</th>
<th>$x_{21}$</th>
<th>$x_{22}$</th>
<th>$y_{11}$</th>
<th>$y_{12}$</th>
<th>$y_{21}$</th>
<th>$y_{22}$</th>
<th>$\delta$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.61430</td>
<td>2.99005</td>
<td>2.94096</td>
<td>3.69706</td>
<td>0.38948</td>
<td>-2.97539</td>
<td>3.27086</td>
<td>22.3259974</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.58609</td>
<td>3.05453</td>
<td>2.97031</td>
<td>3.27086</td>
<td>0.305453</td>
<td>-2.21107</td>
<td>3.09466</td>
<td>7.84857813</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3.94308</td>
<td>3.03387</td>
<td>2.90272</td>
<td>3.27086</td>
<td>0.234456</td>
<td>-2.27222</td>
<td>3.02519</td>
<td>1.42914360</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.96448</td>
<td>3.02170</td>
<td>2.95326</td>
<td>3.97523</td>
<td>0.85256</td>
<td>-2.01531</td>
<td>3.01515</td>
<td>0.85256301</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3.97879</td>
<td>3.01341</td>
<td>3.06025</td>
<td>3.97523</td>
<td>0.83128</td>
<td>-2.00897</td>
<td>3.00919</td>
<td>0.51331998</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.98723</td>
<td>3.00821</td>
<td>2.99615</td>
<td>3.87639</td>
<td>0.78180</td>
<td>-2.00533</td>
<td>3.00556</td>
<td>0.30979089</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>3.99229</td>
<td>3.00500</td>
<td>2.99767</td>
<td>3.87639</td>
<td>0.76180</td>
<td>-2.00200</td>
<td>3.00337</td>
<td>0.18728213</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.99534</td>
<td>3.00303</td>
<td>2.99859</td>
<td>3.87639</td>
<td>0.74180</td>
<td>-2.00071</td>
<td>3.00204</td>
<td>0.11329119</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>3.99718</td>
<td>3.00184</td>
<td>2.99915</td>
<td>3.87639</td>
<td>0.72180</td>
<td>-2.00071</td>
<td>3.00123</td>
<td>0.06855766</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.99829</td>
<td>3.00111</td>
<td>2.99948</td>
<td>3.87639</td>
<td>0.70180</td>
<td>-2.00071</td>
<td>3.00075</td>
<td>0.04149393</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>3.99879</td>
<td>3.00033</td>
<td>2.99948</td>
<td>3.87639</td>
<td>0.68180</td>
<td>-2.00071</td>
<td>3.00033</td>
<td>0.02429906</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>3.99929</td>
<td>3.00000</td>
<td>2.99948</td>
<td>3.87639</td>
<td>0.66180</td>
<td>-2.00071</td>
<td>3.00000</td>
<td>0.00719881</td>
<td></td>
</tr>
</tbody>
</table>


Fig. 1. The relative errors $\delta$ of Example 1 versus $k$ (dots) $\mu = \frac{1}{m+n} = \frac{1}{4}$, $\mu = \frac{1}{m+n}$, $\mu = \frac{1}{m+n}$, $\mu = \frac{1}{m+n}$, $\mu = \frac{1}{m+n}$.

6. Conclusions

A family of iterative methods for linear systems is presented and a least-squares iterative solution to coupled matrix equations are studied by using the hierarchical identification principle and the star product. The analysis indicates that the algorithms proposed can achieve a good convergence property for any initial values. How to use the conjugate gradient method to solve the coupled matrix equation requires further research. Although the algorithms are presented for linear coupled matrix equations, the idea adopted can be easily extended to study iterative solutions of more complex matrix equations and nonlinear matrix equations, e.g., the Riccati equation.

References