Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Distributed Solving Sylvester Equations with an Explicit Exponential Convergence

Songsong Cheng * **Xianlin Zeng** ** **Yiguang Hong** *

* *Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China. (E-mail: sscheng@amss.ac.cn, yghong@iss.ac.cn)*
** *Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, 100081, Beijing, China. (xianlin.zeng@bit.edu.cn)*

Abstract: This paper addresses distributed achieving the least squares solution of Sylvester equations in the form of $AX + XB = C$. By decomposing the parameter matrices A, B and C, we formulate the problem of distributed solving Sylvester equations as a distributed optimization model and propose a continuous-time algorithm from the primal-dual viewpoint. Then, by constructing a Lyapunov function, we prove that the proposed algorithm can achieve a least squares solution of Sylvester equations with an explicit exponential convergence rate. Additionally, we illustrate the convergence performance by using a numerical example.

Keywords: Sylvester equations, Distributed optimization, Least squares solution, Exponential convergence.

1. INTRODUCTION

Because of their wide application in machine learning (Wang et al., 2016), control theory (Xu and Dubljevic, 2017), and robot manipulator (Zhang and Zheng, 2018), Sylvester equations have received intensive attentions in the fields of engineering and mathematics. In 1884, Sylvester developed the condition of the unique solution for the special case $AX + XB = \mathbf{0}$ (Sylvester, 1884), and then some scholars achieved several solvable conditions for the general case $AX + XB = C$ (Kučera, 1974; Gohberg and Lerer, 1988; Wimmer, 1996; Wang et al., 2019). Up to now, centralized algorithms have been proposed. For instance, by formulating the coefficient matrices into triangular form, Kleinman and Rao (1978) proposed an iterative algorithm, while by transforming the Sylvester equation as an optimization problem, Benner and Breiten (2014) achieved the low rank approximate solution.

However, in some fields, such as big data and complex systems, the dimensions of the Sylvester equation accumulate explosively. Therefore, the conventional centralized approaches are hard to rise to the challenges because of the limited abilities of computation and communication. In order to overcome these limitations, distributed optimization has attracted many research attentions (Yi and Hong, 2016). Because of the convenience of analysis and being implemented in hybrid physical systems, continuous-time based distributed optimization algorithms have became more and more popular (Liang et al., 2017; Kia et al., 2015; Shi et al., 2012). In recent years, some continuous-time algorithms have been proposed to distributedly solving the large scale linear matrix equations. Deng et al. (2019) proposed a continuous-time distributed algorithm to obtain the exact solution of Sylvester equations. However, this algorithm is invalid for a least squares solution. For Stein equations in the form of $X + AXB = C$ without exact solutions, Chen et al. (2019) developed a continuous-time distributed algorithm to achieve a least squares solution. However, this method is limited to the special Row-Column-Column structure of matrices A, B and C. By considering eight standard structures of A, B and C of the linear matrix equations $AXB = C$, Zeng et al. (2018) discussed four distributed algorithms to attain a least squares solution. All of these algorithms in Deng et al. (2019); Chen et al. (2019); Zeng et al. (2018) can solve the linear matrix equations in distributed manner with an exponential convergence, but it is hard to determine an explicit convergence rate.

In this work, we develop a distributed algorithm to solve $AX + XB = C$. By decomposing the matrices A, B and C with any row or column sub-blocks, we transform the problem as a universal distributed optimization model with one variable consensus constraint. Based on this optimization problem, we construct an augmented Lagrangian function and propose a distributed continuous-time algorithm. Moreover, by designing a Lyapunov function, we establish an explicit exponential convergence rate. The main contributions are listed as follows.

1) A universal distributed optimization model is established to handle any type of standard decomposition of the parameter matrices A, B and C.

⋆ This work was supported by the National Key Research and Development Program of China (2016YFB0901900), the National Natural Science Foundation of China under Grant 61733018, the National Postdoctoral Program for Innovative Talents (BX20180346), and the General Financial Grant from the China Postdoctoral Science Foundation (2019M660834).

2) In comparison with Deng et al. (2019), we remove the assumption on the existence of the exact solution and propose a more efficient continuous-time algorithm.
3) In comparison with Deng et al. (2019); Chen et al. (2019); Zeng et al. (2018), we prove that the proposed algorithm achieve the least squares solution with an explicit exponential convergence rate.

The rest of this paper is organized as follows. Section 2 gives some basic preliminaries, while Section 3 transforms the problem as a distributed optimization problem and designs a continuous-time algorithm. Then Section 4 shows that the proposed algorithm exponentially converges to the least squares solution with an explicit rate. Section 5 illustrates the proposed algorithm by showing an example and Section 6 concludes this paper.

2. PRELIMINARIES

2.1 Matrices

The real number set, n-dimensional real column vector set, and $n \times m$ real matrix set are denoted as $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$, respectively. $\mathbf{1}_n \in \mathbb{R}^n$ ($\mathbf{0}_n \in \mathbb{R}^n$, $\mathbf{0}_{n\times m} \in \mathbb{R}^{n\times m}$) is a vector (vector, matrix) with all of the elements are one (zero, zero). $I_n \in \mathbb{R}^{n\times n}$ is an identity matrix. For $\forall A \in \mathbb{R}^{n\times m}$, $A^\top$ (rank(A), ker(A)) means the transpose (rank, kernel) of A, a_{ij} denotes the i-th row and j-th column entry of A, $A_{i\cdot}$ ($A_{\cdot i}$) denotes the i-th row or row sub-block (column or column sub-block) of the matrix A with proper dimensions. $\|A\| = (\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2)^{\frac{1}{2}}$ is the Frobenius norm of A. Similarly, for any two real matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m}$, the Frobenius inner product can be calculated as $\langle A, B\rangle = \sum_{i=1}^n \sum_{j=1}^m a_{ij}b_{ij}$. $[m_i]$ and $[r_i]$ are two sequences with $\sum_{i=1}^n m_i = m$, $\sum_{i=1}^n r_i = r$, $m_{[i]} = \sum_{j=1}^i m_j$, and $r_{[i]} = \sum_{j=1}^i r_j$. Then two sub-block matrices $A_{i\cdot} \in \mathbb{R}^{m_i\times m}$ and $B_{\cdot i} \in \mathbb{R}^{r\times r_i}$ can be augmented as follows

$$\begin{cases} \bar{A}_i := \left[\mathbf{0}^\top_{m_{[i-1]}\times m}, A^\top_{i\cdot}, \mathbf{0}^\top_{(m-m_{[i]})\times m}\right]^\top, & \text{(1a)} \\ \bar{B}_i := \left[\mathbf{0}_{r\times r_{[i-1]}}, B_{\cdot i}, \mathbf{0}_{r\times(r-r_{[i]})}\right]. & \text{(1b)} \end{cases}$$

Lemma 1. For the Frobenius inner product, we have

$$\begin{cases} \frac{\partial}{\partial X}\langle AX, B\rangle = A^\top B, \ \frac{\partial}{\partial X}\langle XA, B\rangle = BA^\top, \\ \frac{\partial}{\partial X}\langle A, BX\rangle = B^\top A, \ \frac{\partial}{\partial X}\langle A, XB\rangle = AB^\top. \end{cases} \tag{2}$$

Remark 1. According to Lemma 1, for a given quadratic function $H(X) = \frac{1}{2}\|\bar{H}(X)\|^2 = \frac{1}{2}\|AX + XB - C\|^2$, the corresponding gradient on X can be expressed as

$$\frac{\partial H(X)}{\partial X} = A^\top \bar{H}(X) + \bar{H}(X)B^\top. \tag{3}$$

2.2 Graph theory

For a network with nodes set $\mathcal{V} = \{1, 2, \cdots, n\}$ and edges set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, we denote the corresponding undirected communication graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. We say node j is the neighbor of node i if $\{j, i\} \in \mathcal{E}$. $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ is the adjacency matrix of the graph $\mathcal{G}$ with $a_{ij} = a_{ji} > 0$ if $\{j, i\} \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. Based on the adjacency matrix we calculate the degree matrix D and Laplacian matrix L as $D = \text{diag}\{\sum_{j=1}^n a_{1j}, \cdots, \sum_{j=1}^n a_{nj}\}$ and $L = D - A$, respectively. Specifically, if the undirected graph $\mathcal{G}$ is connected, the Laplacian matrix L is symmetric positive semi-definite, and the rank and kernel of L are $n-1$ and $k1_n$ with $k \in R$, respectively.

Assumption 1. The undirected graph $\mathcal{G}$ is connected.

2.3 Metric subregularity

For a map $H(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^n$, define gph$H = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n | \mathbf{y} = H(\mathbf{x})\}$ and $H^{-1}(\mathbf{y}) = \{\mathbf{y} \in \mathbb{R}^n | (\mathbf{x}, \mathbf{y}) \in \text{gph}H\}$. Then we introduce a lemma on κ-metric subregularity.

Lemma 2. Ioffe (2017) If gph$H$ is polyhedral, there exists a positive constant κ such that

$$d\big(\mathbf{x}, H^{-1}(\mathbf{y}^*)\big) \le \kappa\big\|H(\mathbf{x}) - \mathbf{y}^*\big\|, \ \forall \mathbf{x} \in \mathbb{R}^n, \tag{4}$$

where $d(\mathbf{x}, H^{-1}(\mathbf{y}^*)) = \inf_{\mathbf{z}\in H^{-1}(\mathbf{y}^*)}\|\mathbf{z} - \mathbf{x}\|$.

3. PROBLEM FORMULATION AND ALGORITHM

3.1 Problem formulation

Consider the problem of solving the Sylvester equation

$$AX + XB = C, \tag{5}$$

where $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{r\times r}$, and $C \in \mathbb{R}^{m\times r}$ are known matrices, $X \in \mathbb{R}^{m\times r}$ is an unknown matrix to be determined.

Definition 1. We call X^* is a least squares solution to (5), if X^* satisfies

$$X^* = \underset{X}{\text{argmin}}\|AX + XB - C\|^2. \tag{6}$$

In the scheme of distributedly solving (5) with the aid of multi-agent systems, the i-th agent only holds the i-th sub-blocks parameter matrices A, B, and C and exchanges information with their neighbors. For instance, matrices A, B, and C are decomposed as the Row-Column-Column structure, namely, $A = [A^\top_{1\cdot}, \cdots, A^\top_{n\cdot}]^\top \in \mathbb{R}^{m\times m}$, $A_{i\cdot} \in \mathbb{R}^{m_i\times m}$, $\sum_{i=1}^n m_i = m$, $B = [B_{\cdot 1}, \cdots, B_{\cdot n}] \in \mathbb{R}^{r\times r}$, $B_{\cdot i} \in \mathbb{R}^{r\times r_i}$, $\sum_{i=1}^n r_i = r$, $C = [C_{\cdot 1}, \cdots, C_{\cdot n}] \in \mathbb{R}^{m\times r}$, $C_{\cdot i} \in \mathbb{R}^{m\times r_i}$, and the i-th agent has access to $A_{i\cdot}$, $B_{\cdot i}$, and $C_{\cdot i}$. By referring (1a) and (1b), $A_{i\cdot}$, $B_{\cdot i}$, and $C_{\cdot i}$ can be augmented as $\bar{A}_i \in \mathbb{R}^{m\times m}$, $\bar{B}_i \in \mathbb{R}^{r\times r}$, and $\bar{C}_i \in \mathbb{R}^{m\times r}$, respectively. Then (5) can be transformed into

$$\sum_{i=1}^n \bar{A}_i X + X\sum_{i=1}^n \bar{B}_i - \sum_{i=1}^n \bar{C}_i = \mathbf{0}. \tag{7}$$

Remark 2. The aforementioned Row-Column-Column structure of matrices A, B, and C is just a case to show how to transform (5) into (7). Actually, for all of other 7 cases (including Column(C)-Column(C)-Column(C), RRC, CRC, RCR, CCR, RRR, and CRR) of the considered Sylvester equation in (5) can also be transformed into (7) equivalently.

By introducing an intermediate matrix $M = [M_1^\top, M_2^\top, \cdots, M_n^\top]^\top \in \mathbb{R}^{nm\times r}$ with $M_i \in \mathbb{R}^{m\times r}$ for $\forall i \in \mathcal{V}$, we decouple (7) as

$$\begin{cases} \bar{A}_i X_i + X_i \bar{B}_i - \bar{C}_i + \sum_{j=1}^{n} a_{ij}(M_i - M_j) = \mathbf{0}, \\ X_i = X_j, \end{cases} \tag{8}$$

where the matrix $[a_{ij}]$ is the adjacent matrix of the undirected graph. Combining with the definition of the Laplacian matrix L, we rewrite (8) as

$$\begin{cases} \bar{A}_i X_i + X_i \bar{B}_i - \bar{C}_i + \bar{L}_{i\cdot} M = \mathbf{0}, \\ \bar{L} X = \mathbf{0}, \end{cases} \tag{9}$$

where $\bar{L} = L \otimes I_m$ and $\bar{L}_{i\cdot} = L_{i\cdot} \otimes I_m$. Therefore, we transform (9) into a following distributed optimization problem

$$\begin{aligned} \min_{X,M} \quad & \frac{1}{2}\sum_{i=1}^{n} \|\bar{A}_i X_i + X_i \bar{B}_i - \bar{C}_i + \bar{L}_{i\cdot} M\|^2, \\ \text{s.t.} \quad & \bar{L} X = \mathbf{0}. \end{aligned} \tag{10}$$

It can be verified the equivalence between an optimal solution of (10) and a least squares solution of (5).

3.2 Algorithm design

Based on (10), we construct an augmented Lagrangian function as

$$\mathcal{L}(X, M, \Lambda) = \frac{1}{2}\sum_{i=1}^{n} \|G_i\|^2 + \langle \Lambda, \bar{L}X \rangle + \frac{1}{2}\langle X, \bar{L}X \rangle, \tag{11}$$

where $\Lambda = [\Lambda_1^\top, \Lambda_2^\top, \cdots, \Lambda_n^\top]^\top \in \mathbb{R}^{nm\times r}$ with $\Lambda_i \in \mathbb{R}^{m\times r}$, for $\forall i \in \mathcal{V}$

According to (11), we design a continuous-time optimization algorithm from the primal-dual perspective, i.e., gradient descend for primal variables X and M and gradient ascent for the dual variable Λ

$$\begin{cases} \dot{X}_i^t = -\nabla_{X_i^t} \mathcal{L}(X^t, M^t, \Lambda^t), \\ \dot{M}_i^t = -\nabla_{M_i^t} \mathcal{L}(X^t, M^t, \Lambda^t), \\ \dot{\Lambda}_i^t = \nabla_{\Lambda_i^t} \mathcal{L}(X^t, M^t, \Lambda^t), \end{cases} \tag{12}$$

where $\nabla_{X_i^t}\mathcal{L}(X^t, M^t, \Lambda^t)$, $\nabla_{M_i^t}\mathcal{L}(X^t, M^t, \Lambda^t)$, and $\nabla_{\Lambda_i^t}\mathcal{L}$ (X^t, M^t, Λ^t) are the gradient of $\mathcal{L}(X^t, M^t, \Lambda^t)$ on variables X_i^t, M_i^t, and Λ_i^t, respectively. Based on the definition of $\mathcal{L}(X^t, M^t, \Lambda^t)$, we express the detailed update mechanism of X_i^t, M_i^t, and Λ_i^t in Algorithm 1.

Algorithm 1 Continuous-time Algorithm

Initialization: For each $i \in \mathcal{V}$
$X_i^0 \in \mathbb{R}^{m\times n}$, $M_i^0 \in \mathbb{R}^{m\times n}$, $\Lambda_i^0 \in \mathbb{R}^{m\times n}$.

Update flows: For each $i \in \mathcal{V}$,

$$G_i^t = \bar{A}_i X_i^t + X_i^t \bar{B}_i - \bar{C}_i + \bar{L}_{i\cdot} M^t, \tag{S1}$$

$$\dot{X}_i^t = -\big[\bar{A}_{vi}^\top G_i^t + G_i^t \bar{B}_i^\top + \bar{L}_{i\cdot}(\Lambda^t + X^t)\big], \tag{S2}$$

$$\dot{M}_i^t = -\bar{L}_{i\cdot}[G_1^{t\top}, \cdots, G_n^{t\top}]^\top, \tag{S3}$$

$$\dot{\Lambda}_i^t = \bar{L}_{i\cdot} X^t. \tag{S4}$$

4. CONVERGENCE PERFORMANCE

In this section, we analyze the convergence performance based on Lyapunov function. For the convenience of analysis, we formulate the algorithm in a compact form firstly.

Substituting $G_i^t = \bar{A}_i X_i^t + X_i^t \bar{B}_i - \bar{C}_i + \bar{L}_{i\cdot} M^t$ into the step (S2) of Algorithm 1 yields

$$\begin{aligned} \dot{X}_i^t = -\big[&\bar{A}_i^\top \bar{A}_i X_i^t + \bar{A}_i^\top X_i^t \bar{B}_i + \bar{A}_i X_i^t \bar{B}_i^\top + X_i^t \bar{B}_i \bar{B}_i^\top \\ &- \bar{A}_i^\top \bar{C}_i - \bar{C}_i \bar{B}_i^\top + \bar{A}_i^\top \bar{L}_{i\cdot} M^t + \bar{L}_{i\cdot} M^t \bar{B}_i^\top \\ &+ \bar{L}_{i\cdot}(\Lambda^t + X^t)\big]. \end{aligned} \tag{13}$$

Accumulate the column of X_i^t, M_i^t, Λ_i^t, and $\bar{C}_i$ as augmented column vectors $\mathbf{x}_i^t \in \mathbb{R}^{mr}$, $\mathbf{m}_i^t \in \mathbb{R}^{mr}$, $\boldsymbol{\lambda}_i^t \in \mathbb{R}^{mr}$, and $\mathbf{c}_i^t \in \mathbb{R}^{mr}$, respectively, and define $\mathbf{m}^t = \mathrm{col}\{\mathbf{m}_1^t, \cdots, \mathbf{m}_n^t\} \in \mathbb{R}^{nmr}$ and $\boldsymbol{\lambda}^t = \mathrm{col}\{\boldsymbol{\lambda}_1^t, \cdots, \boldsymbol{\lambda}_n^t\} \in \mathbb{R}^{nmr}$. Then we transform (13) into,

$$\begin{aligned} \dot{\mathbf{x}}_i^t = -\big[&(I_r \otimes \bar{A}_i^\top \bar{A}_i + \bar{B}_i^\top \otimes \bar{A}_i^\top + \bar{B}_i \otimes \bar{A}_i \\ &+ \bar{B}_i \bar{B}_i^\top \otimes I_m)\mathbf{x}_i^t - (I_r \otimes \bar{A}_i^\top + \bar{B}_i \otimes I_m)\mathbf{c}_i \\ &+ (I_r \otimes \bar{A}_i^\top + \bar{B}_i \otimes I_m)\hat{L}_{i\cdot}\mathbf{m}^t + \hat{L}_{i\cdot}(\boldsymbol{\lambda}^t + \mathbf{x}^t)\big], \end{aligned} \tag{14}$$

where $\hat{L}_{i\cdot} = L_{i\cdot} \otimes I_{mr} \in \mathbb{R}^{mr\times nmr}$. Defining $P_i := I_r \otimes \bar{A}_i^\top \bar{A}_i + \bar{B}_i^\top \otimes \bar{A}_i^\top + \bar{B}_i \otimes \bar{A}_i + \bar{B}_i \bar{B}_i^\top \otimes I_m \in \mathbb{R}^{mr\times mr}$ and $Q_i := I_r \otimes \bar{A}_i^\top + \bar{B}_i \otimes I_m \in \mathbb{R}^{mr\times mr}$, we abbreviate (14)

$$\dot{\mathbf{x}}_i^t = -\big[P_i \mathbf{x}_i^t + Q_i \hat{L}_{i\cdot}\mathbf{m}^t - Q_i \mathbf{c}_i + \hat{L}_{i\cdot}(\boldsymbol{\lambda}^t + \mathbf{x}^t)\big]. \tag{15}$$

With $\mathbf{x}^t = \mathrm{col}\{\mathbf{x}_1^t, \cdots, \mathbf{x}_n^t\} \in \mathbb{R}^{nmr}$, we rewrite (15) in a compact form

$$\dot{\mathbf{x}}^t = -\big[(P + \hat{L})\mathbf{x}^t - Q\mathbf{c} + Q\hat{L}\mathbf{m}^t + \hat{L}\boldsymbol{\lambda}^t\big], \tag{16}$$

where $\hat{L} = L \otimes I_{mr} \in \mathbb{R}^{nmr\times nmr}$, $P = \mathrm{diag}\{P_1, \cdots, P_n\} \in \mathbb{R}^{nmr\times nmr}$, $Q = \mathrm{diag}\{Q_1, \cdots, Q_n\} \in \mathbb{R}^{nmr\times nmr}$, and $\mathbf{c} = \mathrm{col}\{\mathbf{c}_1, \cdots, \mathbf{c}_n\} \in \mathbb{R}^{nmr}$.

Similarly, we rewrite steps (S2) and (S3) of Algorithm 1 as the following compact form

$$\begin{cases} \dot{\mathbf{m}}^t = -\hat{L}^\top\big[R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\big], \\ \dot{\boldsymbol{\lambda}}^t = \hat{L}\mathbf{x}^t, \end{cases} \tag{17}$$

where $R = \mathrm{diag}\{R_1, \cdots, R_n\} \in \mathbb{R}^{nmr\times nmr}$ with $R_i = I_r \otimes \bar{A}_i + \bar{B}_i^\top \otimes I_m \in \mathbb{R}^{mr\times mr}$.

Apparently, according to the definition of P, Q, and R, it is not hard to verify $P = R^\top R$ and $Q = R^\top$. Therefore, we transform the dynamics of Algorithm 1 into

$$\begin{cases} \dot{\mathbf{x}}^t = -\big[R^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) + \hat{L}(\mathbf{x}^t + \boldsymbol{\lambda}^t)\big], \\ \dot{\mathbf{m}}^t = -\hat{L}^\top\big[R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\big], \\ \dot{\boldsymbol{\lambda}}^t = \hat{L}\mathbf{x}^t. \end{cases} \tag{18}$$

The equilibria of (18) satisfy

$$R^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}) + \hat{L}(\mathbf{x}^* + \boldsymbol{\lambda}^*) = \mathbf{0}, \tag{19a}$$

$$\hat{L}^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}) = \mathbf{0}, \tag{19b}$$

$$\hat{L}\mathbf{x}^* = \mathbf{0}. \tag{19c}$$

Define $\delta_{R_m}^2$ and δ_{L_m} as the largest eigenvalues of $R^\top R$ and $\hat{L}$, respectively. Then, for the dynamics in (18), we construct a Lyapunov function as

$$V^t = 2\delta_{L_m} V_1^t + V_2^t, \tag{20}$$

where

$$\begin{aligned} V_1^t &= \frac{1}{2}\big(\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2\big), \\ V_2^t &= \frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2 + \frac{1}{2}\langle \mathbf{x}^t, \hat{L}\mathbf{x}^t\rangle \\ &\quad + \langle \boldsymbol{\lambda}^t, \hat{L}\mathbf{x}^t\rangle - \frac{1}{2}\|R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}\|^2, \end{aligned} \tag{21}$$

with $\tilde{\mathbf{x}}^t = \mathbf{x}^t - \mathbf{x}^*$, $\tilde{\mathbf{m}}^t = \mathbf{m}^t - \mathbf{m}^*$, and $\tilde{\boldsymbol{\lambda}}^t = \boldsymbol{\lambda}^t - \boldsymbol{\lambda}^*$.

Based on the Lyapunov function, we analyze the convergence performance of the proposed algorithm. Prior to present the convergence performance, we show two lemmas: Lemma 3 presents the upper and lower bound of the Lyapunov function and Lemma 4 shows that the Lyapunov function is non-increasing.

Lemma 3. Under Assumption 1, let (X^t, M^t, Λ^t) be generated by Algorithm 1. Then we bound V^t as follows

$$\delta_{L_m} V_1^t \le V^t \le 2(\delta_{R_m}^2 + 2\delta_{L_m}) V_1^t, \tag{22}$$

where $\delta_{R_m}^2$ and δ_{L_m} are the largest eigenvalues of matrices $R^\top R$ and $\hat{L}$, respectively.

Proof. According to (19c), we modify$\langle \mathbf{x}^t, \hat{L}\mathbf{x}^t\rangle$, $\langle \boldsymbol{\lambda}^t, \hat{L}\mathbf{x}^t\rangle$, and (19a) as

$$\begin{cases} \langle \mathbf{x}^t, \hat{L}\mathbf{x}^t\rangle = \langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\mathbf{x}}^t\rangle, & (23\text{a}) \\ \langle \boldsymbol{\lambda}^t, \hat{L}\mathbf{x}^t\rangle = \langle \boldsymbol{\lambda}^t, \hat{L}\tilde{\mathbf{x}}^t\rangle, & (23\text{b}) \\ R^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}) + \hat{L}\boldsymbol{\lambda}^* = \mathbf{0}. & (23\text{c}) \end{cases}$$

Based on (19b) and (23a)-(23c), we rewrite V_2^t as

$$\begin{aligned} V_2^t &= \frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2 - \frac{1}{2}\|R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}\|^2 \\ &\quad - \big\langle \mathbf{x}^t - \mathbf{x}^*, R^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}) + \hat{L}\boldsymbol{\lambda}^*\big\rangle \\ &\quad - \big\langle \mathbf{m}^t - \mathbf{m}^*, \hat{L}^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c})\big\rangle \\ &\quad + \frac{1}{2}\big\langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\mathbf{x}}^t\big\rangle + \langle \boldsymbol{\lambda}^t, \hat{L}\tilde{\mathbf{x}}^t\rangle \\ &\ge \frac{1}{2}\big\langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\mathbf{x}}^t\big\rangle + \big\langle \boldsymbol{\lambda}^t, \hat{L}\tilde{\mathbf{x}}^t\big\rangle - \big\langle \mathbf{x}^t - \mathbf{x}^*, \hat{L}\boldsymbol{\lambda}^*\big\rangle \\ &\ge \big\langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\tilde{\mathbf{x}}^t\big\rangle \\ &\ge -\frac{\delta_{L_m}}{2}\big(\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2\big), \end{aligned} \tag{24}$$

where the first inequality follows from the convexity of function $\frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2$ on $\mathbf{x}^t$ and $\mathbf{m}^t$; the second inequality follows from the semi-positive definite of matrix $\hat{L}$. Combining (20) and (24) implies

$$V^t \ge \frac{\delta_{L_m}}{2}\big(\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2\big) = \delta_{L_m} V_1^t. \tag{25}$$

Reconsidering $\frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2$, we have

$$\begin{aligned} &\frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2 - \frac{1}{2}\|R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}\|^2 \\ &= \frac{1}{2}\langle R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t, R(\mathbf{x}^t + \mathbf{x}^*) + \hat{L}(\mathbf{m}^t + \mathbf{m}^*) - 2\mathbf{c}\rangle \\ &= \frac{1}{2}\langle \tilde{\mathbf{x}}^t, R^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\rangle \\ &\quad + \frac{1}{2}\langle \tilde{\mathbf{m}}^t, \hat{L}^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\rangle. \\ &\quad + \frac{1}{2}\langle \tilde{\mathbf{x}}^t, R^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c})\rangle \\ &\quad + \frac{1}{2}\langle \tilde{\mathbf{m}}^t, \hat{L}^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c})\rangle. \end{aligned} \tag{26}$$

According to (19a) and (19b), $\hat{L}^\top\mathbf{c} = \hat{L}^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^*)$ and $R^\top\mathbf{c} = R^\top(R\mathbf{x}^* + \hat{L}\mathbf{m}^*) + \hat{L}\boldsymbol{\lambda}^* = \mathbf{0}$. Then we modify (26) as

$$\begin{aligned} &\frac{1}{2}\|R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\|^2 - \frac{1}{2}\|R\mathbf{x}^* + \hat{L}\mathbf{m}^* - \mathbf{c}\|^2 \\ &= \frac{1}{2}\langle R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t, R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t\rangle - \langle \tilde{\mathbf{x}}^t, \hat{L}\boldsymbol{\lambda}^*)\rangle. \end{aligned} \tag{27}$$

Consequently, we establish the upper bound of V_2^t as

$$\begin{aligned} V_2^t &= \frac{1}{2}\big\langle R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t, R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t\big\rangle + \big\langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\boldsymbol{\lambda}}^t)\big\rangle \\ &\quad + \frac{1}{2}\big\langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\mathbf{x}}^t\big\rangle \\ &\le (\delta_{R_m}^2 + \delta_{L_m})\|\tilde{\mathbf{x}}^t\|^2 + \delta_{L_m}^2\|\tilde{\mathbf{m}}^t\|^2 + \frac{\delta_{L_m}}{2}\|\tilde{\boldsymbol{\lambda}}^t\|^2 \\ &\le (\delta_{R_m}^2 + \delta_{L_m})\big(\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2\big). \end{aligned} \tag{28}$$

Combining (21) and (28) yields

$$\begin{aligned} V^t &\le (\delta_{R_m}^2 + 2\delta_{L_m})\big(\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2\big) \\ &= 2(\delta_{R_m}^2 + 2\delta_{L_m}) V_1^t. \end{aligned} \tag{29}$$

Therefore, according to (25) and (29), we achieve the conclusion. □

Lemma 4. Under Assumption 1, let (X^t, M^t, Λ^t) be generated by Algorithm 1. Then $\dot{V}^t$ is not positive.

Proof. Note that

$$\begin{aligned} \dot{V}_1^t &= \big\langle \dot{\mathbf{x}}^t, \nabla_{\mathbf{x}^t} V_1^t\big\rangle + \langle \dot{\mathbf{m}}^t, \nabla_{\mathbf{m}^t} V_1^t\rangle + \big\langle \dot{\boldsymbol{\lambda}}^t, \nabla_{\boldsymbol{\lambda}^t} V_1^t\big\rangle \\ &= -\big\langle \tilde{\mathbf{x}}^t, R^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) + \hat{L}(\mathbf{x}^t + \boldsymbol{\lambda}^t)\big\rangle \\ &\quad - \big\langle \tilde{\mathbf{m}}^t, \hat{L}^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\big\rangle + \big\langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\mathbf{x}^t\big\rangle \\ &= -\big\langle \tilde{\mathbf{x}}^t, R^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\big\rangle \\ &\quad - \big\langle \tilde{\mathbf{m}}^t, \hat{L}^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\big\rangle \\ &\quad - \big\langle \tilde{\mathbf{x}}^t, \hat{L}\mathbf{x}^t + \hat{L}\boldsymbol{\lambda}^t\big\rangle + \big\langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\mathbf{x}^t\big\rangle \\ &= \Omega_1^t + \Omega_2^t + \Omega_3^t + \Omega_4^t, \end{aligned} \tag{30}$$

where $\Omega_1^t := -\big\langle \tilde{\mathbf{x}}^t, R^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\big\rangle$, $\Omega_2^t := -\big\langle \tilde{\mathbf{m}}^t, \hat{L}^\top(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c})\big\rangle$, $\Omega_3^t := -\big\langle \tilde{\mathbf{x}}^t, \hat{L}\mathbf{x}^t + \hat{L}\boldsymbol{\lambda}^t\big\rangle$, and $\Omega_4^t := \big\langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\mathbf{x}^t\big\rangle$. For $\Omega_1^t + \Omega_2^t$, we have

$$
\begin{aligned}
\Omega_1^t + \Omega_2^t =& -\langle \tilde{\mathbf{x}}^t, R^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) \rangle \\
& -\langle \tilde{\mathbf{m}}^t, \hat{L}^\top \big(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\big) \rangle \\
=& -\langle \tilde{\mathbf{x}}^t, R^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) \rangle \\
& +\langle \tilde{\mathbf{x}}^t, R^\top (R\mathbf{x}^* + \hat{L}\mathbf{m}^t - \mathbf{c}) \rangle \\
& -\langle \tilde{\mathbf{x}}^t, R^\top (R\mathbf{x}^* + \hat{L}\mathbf{m}^t - \mathbf{c}) \rangle \\
& -\langle \tilde{\mathbf{m}}^t, \hat{L}^\top \big(R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}\big) \rangle \\
& +\langle \tilde{\mathbf{m}}^t, \hat{L}^\top \big(R\mathbf{x}^t + \hat{L}\mathbf{m}^* - \mathbf{c}\big) \rangle \\
& -\langle \tilde{\mathbf{m}}^t, \hat{L}^\top \big(R\mathbf{x}^t + \hat{L}\mathbf{m}^* - \mathbf{c}\big) \rangle \\
=& -\langle \tilde{\mathbf{x}}^t, P\tilde{\mathbf{x}}^t \rangle - \langle \tilde{\mathbf{m}}^t, \hat{L}^\top \hat{L}\tilde{\mathbf{m}}^t \rangle \\
& -\langle \tilde{\mathbf{x}}^t, R^\top (R\mathbf{x}^* + \hat{L}\mathbf{m}^t - \mathbf{c}) \rangle \\
& -\langle \tilde{\mathbf{m}}^t, \hat{L}^\top \big(R\mathbf{x}^t + \hat{L}\mathbf{m}^* - \mathbf{c}\big) \rangle .
\end{aligned} \tag{31}
$$

Referring to the equilibria in (19a) and (19b), we obtain

$$
\begin{cases}
R^\top R\mathbf{x}^* - R^\top \mathbf{c} = -R^\top \hat{L}\mathbf{m}^* - \hat{L}\boldsymbol{\lambda}^*, & (32a)\\
\hat{L}^\top (\hat{L}\mathbf{m}^* - \mathbf{c}) = -\hat{L}^\top R\mathbf{x}^*. & (32b)
\end{cases}
$$

Substituting (32a) and (32b) into (31) yields

$$
\begin{aligned}
\Omega_1^t + \Omega_2^t =& -\langle \tilde{\mathbf{x}}^t, R^\top R\tilde{\mathbf{x}}^t \rangle - \langle \tilde{\mathbf{m}}^t, \hat{L}^\top \hat{L}\tilde{\mathbf{m}}^t \rangle \\
& -\langle \tilde{\mathbf{x}}^t, R^\top \hat{L}\tilde{\mathbf{m}}^t - \hat{L}\boldsymbol{\lambda}^* \rangle - \langle \tilde{\mathbf{m}}^t, \hat{L}^\top R\tilde{\mathbf{x}}^t \rangle \\
=& -\langle R\tilde{\mathbf{x}}^t, R\tilde{\mathbf{x}}^t \rangle - \langle \hat{L}\tilde{\mathbf{m}}^t, \hat{L}\tilde{\mathbf{m}}^t \rangle \\
& -2\langle \hat{L}\tilde{\mathbf{m}}^t, R\tilde{\mathbf{x}}^t \rangle + \langle \tilde{\mathbf{x}}^t, \hat{L}\boldsymbol{\lambda}^* \rangle \\
\le& \langle \tilde{\mathbf{x}}^t, \hat{L}\boldsymbol{\lambda}^* \rangle .
\end{aligned} \tag{33}
$$

Combining (30) and (33), we have

$$
\begin{aligned}
\dot{V}_1^t \le& \langle \tilde{\mathbf{x}}^t, \hat{L}\boldsymbol{\lambda}^* \rangle - \langle \tilde{\mathbf{x}}^t, \hat{L}\mathbf{x}^t + \hat{L}\boldsymbol{\lambda}^t \rangle + \langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\mathbf{x}^t \rangle \\
=& -\langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\mathbf{x}}^t \rangle - \langle \tilde{\mathbf{x}}^t, \hat{L}\tilde{\boldsymbol{\lambda}}^t \rangle + \langle \tilde{\boldsymbol{\lambda}}^t, \hat{L}\tilde{\mathbf{x}}^t \rangle \\
\le& -\delta_{L_m}^{-1} \big\| \hat{L}\tilde{\mathbf{x}}^t \big\|^2,
\end{aligned} \tag{34}
$$

where the equality follows from $\hat{L}\mathbf{x}^t = \hat{L}\tilde{\mathbf{x}}^t$ since $\hat{L}\mathbf{x}^* = 0$ and the last inequality follows from the fact that the communication graph is undirected.

Furthermore, note that

$$
\begin{aligned}
\dot{V}_2^t =& \langle \dot{\mathbf{x}}^t, \nabla_{\mathbf{x}^t} V_2^t \rangle + \langle \dot{\mathbf{m}}^t, \nabla_{\mathbf{m}^t} V_2^t \rangle + \langle \dot{\boldsymbol{\lambda}}^t, \nabla_{\boldsymbol{\lambda}^t} V_2^t \rangle \\
=& -\big\| R^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) + \hat{L}(\mathbf{x}^t + \boldsymbol{\lambda}^t) \big\|^2 \\
& -\big\| \hat{L}^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) \big\|^2 + \big\| \hat{L}\mathbf{x}^t \big\|^2 .
\end{aligned} \tag{35}
$$

Combining (34) and (35), we obtain

$$
\begin{aligned}
\dot{V}^t =& \dot{V}_1^t + \dot{V}_2^t \\
=& -\big\| R^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) + \hat{L}(\mathbf{x}^t + \boldsymbol{\lambda}^t) \big\|^2 \\
& -\big\| \hat{L}^\top (R\mathbf{x}^t + \hat{L}\mathbf{m}^t - \mathbf{c}) \big\|^2 - \big\| \hat{L}\mathbf{x}^t \big\|^2 \\
\le& 0,
\end{aligned} \tag{36}
$$

which leads to the conclusion. □

Based on Lemmas 3 and 4, we show a convergence result of the proposed algorithm as follows.

Theorem 1. Under Assumption 1, let (X^t, M^t, Λ^t) be generated by Algorithm 1. Then (X^t, M^t, Λ^t) exponentially converges to the optimal solution of (10) with the following exponential convergence rate,

$$
\big\|\tilde{X}^t\big\|^2 + \big\|\tilde{M}^t\big\|^2 + \big\|\tilde{\Lambda}^t\big\|^2 \le \frac{2V^0}{\delta_{L_m}} \mathrm{e}^{-\frac{\kappa}{\delta_{R_m}^2 + 2\delta_{L_m}} t}, \tag{37}
$$

where $\tilde{X}^t = X^t - X^*$, $\tilde{M}^t = M^t - M^*$, $\tilde{\Lambda}^t = \Lambda^t - \Lambda^*$, and κ is a parameter defined hereinafter.

Proof. Considering the case of $\dot{V}^t = 0$ at time $t = t_1$ in (36), we have

$$
\begin{cases}
R^\top (R\mathbf{x}^{t_1} + \hat{L}\mathbf{m}^{t_1} - \mathbf{c}) + \hat{L}(\mathbf{x}^{t_1} + \boldsymbol{\lambda}^{t_1}) = \mathbf{0}, & (38a)\\
\hat{L}^\top (R\mathbf{x}^{t_1} + \hat{L}\mathbf{m}^{t_1} - \mathbf{c}) = \mathbf{0}, & (38b)\\
\hat{L}\mathbf{x}^{t_1} = \mathbf{0}, & (38c)
\end{cases}
$$

which means that the proposed algorithm access to the equilibria. From this point of view, before the dynamics of (16) and (17) arriving at the equilibria, $\dot{V} < 0$ holds.

According to the equilibria of (19a) and (19b), substituting $R^\top \mathbf{c} = R^\top (R\mathbf{x}^* + \hat{L}\mathbf{m}^*) + \hat{L}(\mathbf{x}^* + \boldsymbol{\lambda}^*)$ and $\hat{L}^\top \mathbf{c} = \hat{L}^\top (R\mathbf{x}^* + \hat{L}\mathbf{m}^*)$ into (36) yields

$$
\begin{aligned}
\dot{V}^t =& -\big\| (R^\top R + \hat{L})\tilde{\mathbf{x}}^t + R^\top \hat{L}\tilde{\mathbf{m}}^t + \hat{L}\tilde{\boldsymbol{\lambda}}^t \big\|^2 \\
& -\big\| \hat{L}^\top (R\tilde{\mathbf{x}}^t + \hat{L}\tilde{\mathbf{m}}^t) \big\|^2 - \big\| \hat{L}\tilde{\mathbf{x}}^t \big\|^2 \\
\le& -\big\| (R^\top R + \hat{L})\tilde{\mathbf{x}}^t + R^\top \hat{L}\tilde{\mathbf{m}}^t + \hat{L}\tilde{\boldsymbol{\lambda}}^t \big\|^2 \\
\le& -\kappa \big[\|\tilde{\mathbf{x}}^t\|^2 + \|\tilde{\mathbf{m}}^t\|^2 + \|\tilde{\boldsymbol{\lambda}}^t\|^2 \big],
\end{aligned} \tag{39}
$$

where the last inequality follows from Lemma 2 and κ is the smallest nonzero eigenvalue of $\mathrm{diag}\{(R^\top R + \hat{L})^\top (R^\top R + \hat{L}), \hat{L}^\top R R^\top \hat{L}, \hat{L}^\top \hat{L}\}$. Substituting (25) into (39), yields

$$
\dot{V}^t \le -\frac{\kappa}{\delta_{R_m}^2 + 2\delta_{L_m}} V^t. \tag{40}
$$

Then for $\forall t \in [0, \infty)$, we have

$$
V^t \le V^0 \mathrm{e}^{-\frac{\kappa}{\delta_{R_m}^2 + 2\delta_{L_m}} t}. \tag{41}
$$

According to V_1^t in (21), V^t in (22), and (41) , we obtain the conclusion. □

5. NUMERICAL SIMULATION

Consider the Sylvester equation with Row-Column-Column structure, whose parameter matrices are chosen as

$$
\begin{aligned}
& A_{1\cdot} = [2,1,5,3], B_{\cdot 1} = [1,2,1,3]^\top, C_{\cdot 1} = [3,4,1,2]^\top, \\
& A_{2\cdot} = [6,2,1,3], B_{\cdot 2} = [2,5,2,1]^\top, C_{\cdot 2} = [2,3,3,1]^\top, \\
& A_{3\cdot} = [1,2,1,4], B_{\cdot 3} = [4,4,1,3]^\top, C_{\cdot 3} = [1,2,1,2]^\top, \\
& A_{4\cdot} = [3,3,6,7], B_{\cdot 4} = [5,6,2,6]^\top, C_{\cdot 4} = [2,1,3,4]^\top.
\end{aligned}
$$

It is not hard to verify that there exists no exact solution for the Sylvester equation and one of the least squares solution can be calculated as

$$
X^* = \begin{bmatrix} 0.5535 & -0.6068 & 1.1924 & 0.0602 \\ -3.6657 & 2.5313 & -1.0092 & 0.7709 \\ -0.1440 & 0.6659 & -0.4360 & -0.0709 \\ 1.8179 & -1.0551 & 0.1781 & -0.0996 \end{bmatrix}.
$$

We solve the Sylvester equation with the aid of a multi-agent system, where four agents are connected by an undirected circular graph and the i-th agent has access to sub-blocks $A_{i\cdot}$, $B_{\cdot i}$, and $C_{\cdot i}$. We present the trajectories of elements of X_i^t and the trajectories of least squares error

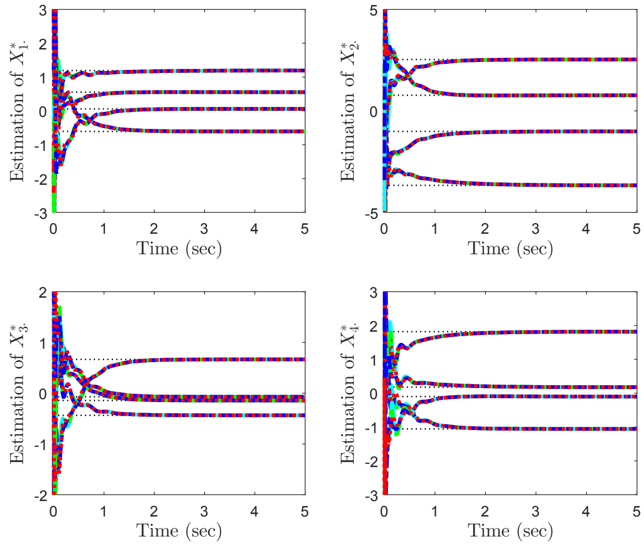

Fig. 1. The trajectories of the estimation of $X^*_{i\cdot}$.

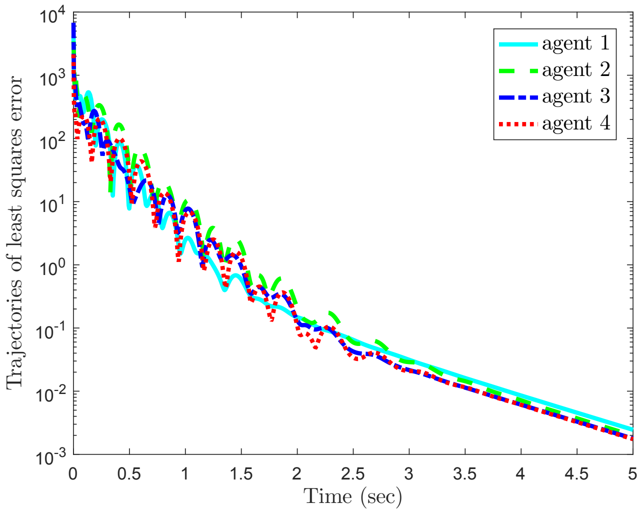

Fig. 2. The trajectories of $\|A^\top(AX + XB - C) + (AX + XB - C)B^\top\|$.

$\|A^\top(AX+XB-C)+(AX+XB-C)B^\top\|$ in Figures 1 and 2, respectively. The simulation results illuminate that the proposed algorithm can achieve the least squares solution of the Sylvester equation with an exponential convergence rate.

6. CONCLUSION

In this paper, we have formulated a distributed optimization method to solve Sylvester equations. From the primal-dual viewpoint, we have proposed a continuous-time distributed algorithm to achieve the least squares solution. By constructing a Lyapunov function, we have proved the exponential convergence of the proposed algorithm with an explicit convergence rate.

REFERENCES

Benner, P. and Breiten, T. (2014). On optimality of approximate low rank solutions of large-scale matrix equations. *Systems & Control Letters*, 67, 55–64.

Chen, G., Zeng, X., and Hong, Y. (2019). Distributed optimisation design for solving the Stein equation with constraints. *IET Control Theory & Applications*, 13(15), 2492–2499.

Deng, W., Zeng, X., and Hong, Y. (2019). Distributed computation for solving the Sylvester equation based on optimization. *IEEE Systems Control Letters*, 414–419.

Gohberg, I. and Lerer, L. (1988). Matrix generalizations of MG Krein theorems on orthogonal polynomials. In *Orthogonal Matrix-valued Polynomials and Applications*, 137–202. Springer.

Ioffe, A.D. (2017). *Variational Analysis of Regular Mappings: Theory and Applications.* Cham, Switzerland: Springer International Publishing AG.

Kia, S.S., Cortés, J., and Martínez, S. (2015). Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 55, 254–264.

Kleinman, D. and Rao, P. (1978). Extensions to the Bartels-Stewart algorithm for linear matrix equations. *IEEE Transactions on Automatic Control*, 23(1), 85–87.

Kučera, V. (1974). The matrix equation AX+XB=C. *SIAM Journal on Applied Mathematics*, 26(1), 15–25.

Liang, S., Zeng, X., and Hong, Y. (2017). Distributed nonsmooth optimization with coupled inequality constraints via modified lagrangian function. *IEEE Transactions on Automatic Control*, 63(6), 1753–1759.

Shi, G., Johansson, K.H., and Hong, Y. (2012). Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets. *IEEE Transactions on Automatic Control*, 58(3), 610–622.

Sylvester, J.J. (1884). Sur l'équation en matrices px = xq. *CR Acad. Sci. Paris*, 99(2), 67–71.

Wang, L., Li, D., He, T., and Xue, Z. (2016). Manifold regularized multi-view subspace clustering for image representation. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, 283–288. IEEE.

Wang, Q.W., He, Z.H., and Zhang, Y. (2019). Constrained two-sided coupled Sylvester-type quaternion matrix equations. *Automatica*, 101, 207–213.

Wimmer, H.K. (1996). The generalized Sylvester equation in polynomial matrices. *IEEE Transactions on Automatic Control*, 41(9), 1372–1376.

Xu, X. and Dubljevic, S. (2017). Output and error feedback regulator designs for linear infinite-dimensional systems. *Automatica*, 83, 170–178.

Yi, P. and Hong, Y. (2016). Distributed cooperative optimization and its applications. *Scientia Sinica Mathematica*, 46(10), 1547–1564.

Zeng, X., Liang, S., Hong, Y., and Chen, J. (2018). Distributed computation of linear matrix equations: an optimization perspective. *IEEE Transactions on Automatic Control*, 64(5), 1858–1873.

Zhang, Z. and Zheng, L. (2018). A complex varying-parameter convergent-differential neural-network for solving online time-varying complex Sylvester equation. *IEEE Transactions on Cybernetics*, 49(10), 3627–3639.