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 Duhljivic, 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 $AX + XB = 0$ (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 become 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 distribute 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, real matrix set, and matrix set are denoted as \( \mathbb{R} \), \( \mathbb{R}^{n} \), and \( \mathbb{R}^{n \times n} \), respectively. \( \mathbf{1}_{n} \in \mathbb{R}^{n} \) (\( \mathbf{0} \in \mathbb{R}^{n} \), \( \mathbf{0} \in \mathbb{R}^{n \times n} \)) is a vector (matrix) with all of the elements are one (zero, zero). \( \mathbf{I}_{n} \in \mathbb{R}^{n \times n} \) is an identity matrix. For any real matrix \( A \in \mathbb{R}^{m \times n} \), \( A^{\top} \) (\( \text{rank}(A) \), \( \ker(A) \)) means the transpose (rank, kernel) of \( A \). \( a_{ij} \) denotes the \( i \)-th row and \( j \)-th column entry of \( A \). \( A_{ij} \) (\( A_{*} \)) denotes the \( i \)-th row or row sub-block (column or column sub-block) of the matrix \( A \) with proper dimensions. \( ||A|| = (\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^{2})^{1/2} \) is the Frobenius norm of \( A \). Similarly, for any two real matrices \( A \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{p \times q} \), the Frobenius inner product can be calculated as \( \langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}b_{ij} \). [\( m \) and \( [r] \) are two sequences with \( \sum_{i=1}^{n} m_{i} = m \), \( \sum_{i=1}^{n} r_{i} = r \), \( m_{[i]} = \sum_{i=1}^{t} m_{i} \), and \( r_{[i]} = \sum_{i=1}^{t} r_{i} \). Then two sub-block matrices \( A_{[i]} \in \mathbb{R}^{m_{[i]} \times n_{[i]}} \) and \( B_{[i]} \in \mathbb{R}^{n_{[i]} \times r_{[i]}} \) can be augmented as follows:

\[
\begin{align*}
\tilde{A}_{i} & := [0_{m_{[i-1]} \times m_{i}}, A_{[i]}^{\top}, 0_{m - m_{[i-1]} \times m_{[i]}}]^{\top}, \\
\tilde{B}_{i} & := [0_{r \times r_{[i-1]}}, B_{[i]}, 0_{r 	imes (r - r_{[i-1]})}].
\end{align*}
\]

Lemma 1. For the Frobenius inner product, we have

\[
\begin{align*}
\frac{\partial}{\partial X} \langle AX, B \rangle &= A^{\top}B, \\
\frac{\partial}{\partial X} \langle AX, BX \rangle &= B^{\top}A, \\
\frac{\partial}{\partial X} \langle A, BX \rangle &= B^{\top}A, \\
\frac{\partial}{\partial X} \langle A, X \rangle &= A^{\top}B.
\end{align*}
\]

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

\[
\frac{\partial H(X)}{\partial X} = A^{\top}H(X) + H(X)B^{\top}.
\]

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 \( i \) is the neighbor of node \( j \) 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 \( a_{ij} = a_{ji} \geq 0 \) 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 \( D = \text{diag}(\sum_{j=1}^{n} a_{ij}, \cdots, \sum_{j=1}^{n} a_{ij}) \) 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 \( k_{L} \) with \( k \in \mathbb{R} \), respectively.

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

2.3 Metric subregularity

For a map \( H(x) : \mathbb{R}^{n} \rightarrow \mathbb{R}^{n} \), define \( gphH = \{ (x, y) \in \mathbb{R}^{n} \times \mathbb{R}^{n} | y = H(x) \} \) and \( H^{-1}(y) = \{ y \in \mathbb{R}^{n} | (x, y) \in gphH \} \). Then we introduce a lemma on \( \kappa \)-metric subregularity.

Lemma 2. Ioffe (2017) If \( gphH \) is polyhedral, there exists a positive constant \( \kappa \) such that

\[
d(x, H^{-1}(y^{*})) \leq \kappa \|H(x) - y^{*}\|, \forall x \in \mathbb{R}^{n},
\]

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

3. PROBLEM FORMULATION AND ALGORITHM

3.1 Problem formulation

Consider the problem of solving the Sylvester equation

\[
AX + XB = C,
\]

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

\[
X^{*} = \arg \min_{X} \|AX + XB - C\|^{2}.
\]

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-Row-Column-Column structure, namely, \( A = [A_{1}, \cdots, A_{m}]^{\top} \in \mathbb{R}^{m \times m} \) where \( A_{i} \in \mathbb{R}^{m \times m} \), \( \sum_{i=1}^{m} m_{i} = m \), \( B = [B_{1}, \cdots, B_{r}] \in \mathbb{R}^{r \times r} \), \( B_{i} \in \mathbb{R}^{r \times r} \), \( \sum_{i=1}^{r} r_{i} = r \), \( C = [C_{1}, \cdots, C_{n}] \in \mathbb{R}^{m \times r} \), \( C_{i} \in \mathbb{R}^{m \times r} \), and the \( i \)-th agent has access to \( A_{i}, B_{i}, \) and \( C_{i} \). By referring (1a) and (1b), \( A_{i}, B_{i}, \) and \( C_{i} \) can be augmented as \( A_{i} \in \mathbb{R}^{m \times m} \), \( B_{i} \in \mathbb{R}^{r \times r} \), and \( C_{i} \in \mathbb{R}^{m \times r} \), respectively. Then (5) can be transformed into

\[
\sum_{i=1}^{n} A_{i}X + X \sum_{i=1}^{n} B_{i} - \sum_{i=1}^{n} C_{i} = 0.
\]

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-Column-Column-Column(C-C-C-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 \( M = [M_{1}^{\top}, M_{2}^{\top}, \cdots, M_{n}^{\top}]^{\top} \in \mathbb{R}^{m \times r} \) with \( M_{i} \in \mathbb{R}^{m \times r} \) for \( \forall i \in \mathcal{V} \), we decouple (7) as

\[
\sum_{i=1}^{n} A_{i}X + X \sum_{i=1}^{n} B_{i} - \sum_{i=1}^{n} C_{i} = 0.
\]
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

$$\begin{align*}
\dot{X}_i &= X_iB_i - C_i + L_i M, \\
X_i &= X_j,
\end{align*}$$

(9)

where $L = L \otimes I_n$ and $L_i = L_i \otimes I_m$. Therefore, we transform (9) into a following distributed optimization problem

$$\min_{X,M} \frac{1}{2} \sum_{i=1}^{n} \| A_iX_i + X_iB_i - C_i + \bar{L}_i M \|^2,$$

(10)

s.t. $LX = 0$.

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

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

(11)

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

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

$$\begin{align*}
\dot{X}_i &= -\nabla_{X_i} L(X^t, M^t, \Lambda^t), \\
\dot{M}_i &= -\nabla_{M_i} L(X^t, M^t, \Lambda^t), \\
\dot{\Lambda}_i &= \nabla_{\Lambda_i} L(X^t, M^t, \Lambda^t),
\end{align*}$$

(12)

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

**Algorithm 1  Continuous-time Algorithm**

**Initialization:** For each $i \in V$

$$X_i^0 \in \mathbb{R}^{m \times n}, \quad M_i^0 \in \mathbb{R}^{m \times n}, \quad \Lambda_i^0 \in \mathbb{R}^{m \times n}.$$  

**Update flows:** For each $i \in V$,

$$\begin{align*}
G_i^t &= \dot{A}_i X_i^t + X_i^t B_i - C_i + L_i M^t, \\
\dot{X}_i^t &= -[\dot{A}_i G_i^t + G_i^t \dot{B}_i^t + L_i (\Lambda^t + X^t_i)], \\
\dot{M}_i^t &= -L_i [G_i^t, \cdots, G_n^t]^T, \\
\dot{\Lambda}_i^t &= \bar{L}_i X_i^t.
\end{align*}$$

**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 $G_i^t = \dot{A}_i X_i^t + X_i^t B_i - C_i + \bar{L}_i M^t$ into the step (S2) of Algorithm 1 yields

$$\begin{align*}
\dot{X}_i^t &= -[A_i^T A_i X_i^t + A_i^T X_i^t B_i + A_i^T X_i^t B_i^T + X_i^t B_i L_i M^t + L_i M^t B_i^T - A_i^T C_i - C_i X_i^t + A_i^T L_i L_i M^t + L_i M^t B_i^T],
\end{align*}$$

(13)

where $L_i = L_i \otimes I_m$. Therefore, we transform (13) into,

$$\begin{align*}
\dot{X}_i^t &= -[L_i \otimes \dot{A}_i A_i + \dot{B}_i^T \otimes \Lambda_i - \dot{\Lambda}_i \otimes A_i + \dot{B}_i \otimes \Lambda_i + \dot{\Lambda}_i \otimes B_i, \bar{L}_i M^t + \bar{L}_i (\Lambda^t + X^t_i)],
\end{align*}$$

where $\bar{L}_i = L_i \otimes I_m, \bar{L}_i \otimes I_m, \bar{L}_i \otimes I_m$. Defining $P_i := L_i \otimes \Lambda_i A_i + \dot{B}_i^T \otimes \Lambda_i + \dot{\Lambda}_i \otimes B_i, \bar{L}_i \otimes I_m \in \mathbb{R}^{nm \times nm}$ and $Q_i := L_i \otimes \Lambda_i A_i + \dot{B}_i \otimes \Lambda_i + \dot{\Lambda}_i \otimes B_i, \bar{L}_i \otimes I_m \in \mathbb{R}^{nm \times nm}$.

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

$$\begin{align*}
\dot{x}^t &= -(P + \hat{L} x^t - c + \hat{Q} \dot{m}^t + \hat{L} \dot{x}^t),
\lambda^t &= \dot{L} \lambda^t,
\end{align*}$$

(16)

where $\hat{L} = L \otimes I_m \in \mathbb{R}^{nm \times nm}$, $P = \text{diag}(P_1, \cdots, P_n) \in \mathbb{R}^{nm \times nm}$, $Q = \text{diag}(Q_1, \cdots, Q_n) \in \mathbb{R}^{nm \times nm}$, and $c = \text{col}(c_1, \cdots, c_n) \in \mathbb{R}^{nm}$.

$$\begin{align*}
\delta_0^2 \text{ and } \delta_0^2 \text{ as the largest eigenvalues of } R^T R \text{ and } \bar{L}, \text{ respectively. Then, for the dynamics in (18), we construct a Lyapunov function as}
\end{align*}$$

$$\text{Lyapunov function as}$$
\[ V^t = 2 \delta_{Lm} V_1^t + V_2^t, \quad (20) \]

where
\[
V_1^t = \frac{1}{2} \left( \|\tilde{x}^t\|^2 + \|\tilde{m}^t\|^2 + \|\tilde{\lambda}^t\|^2 \right),
\]
\[
V_2^t = \frac{1}{2} \|R \tilde{x}^t + \tilde{L} \tilde{m}^t - c\|^2 + \frac{1}{2} \left( \langle x^t, \tilde{L} x^t \rangle - \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 \right),
\]
with \( \tilde{x}^t = x^t - x^*, \tilde{m}^t = m^t - m^*, \text{and} \tilde{\lambda}^t = \lambda^t - \lambda^*. \)

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 \( (X^t, M^t, \Lambda^t) \) be generated by Algorithm 1. Then we bound \( V^t \) as follows
\[
\delta_{Lm} V_1^t \leq V^t \leq 2(\delta_{Rm}^2 + 2\delta_{Lm}) V_1^t, \quad (22)
\]
where \( \delta_{Rm}^2 \) and \( \delta_{Lm} \) are the largest eigenvalues of matrices \( R^T R \) and \( L \), respectively.

**Proof.** According to (19c), we modify \( \langle x^t, \tilde{L} x^t \rangle \), \( \langle \Lambda^t, \tilde{L} \Lambda^t \rangle \), and (19a) as
\[
\left\{ \begin{array}{l}
\langle x^t, \tilde{L} x^t \rangle = \langle \tilde{x}^t, \tilde{L} \tilde{x}^t \rangle, \quad (23a) \\
\langle \Lambda^t, \tilde{L} \Lambda^t \rangle = \langle \tilde{\Lambda}^t, \tilde{L} \tilde{\Lambda}^t \rangle, \quad (23b) \\
R^T (R x^t + \tilde{L} m^t - c) + \tilde{L} \Lambda^t = 0 \quad (23c)
\end{array} \right.
\]

Based on (19b) and (23a)-(23c), we rewrite \( V_2^t \) as
\[
V_2^t = \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 - \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2
- \langle \tilde{m}^t - m^*, \tilde{L} (R x^t + \tilde{L} m^t - c) + \tilde{L} \Lambda^t \rangle
+ \frac{1}{2} \left( \langle x^t, \tilde{L} x^t \rangle - \langle x^t - x^*, \tilde{L} \Lambda^t \rangle \right)
\geq \frac{1}{2} \|\tilde{x}^t, \tilde{L} \tilde{x}^t\| + \langle \tilde{\Lambda}^t, \tilde{L} \tilde{\Lambda}^t \rangle - \langle x^t - x^*, \tilde{L} \Lambda^t \rangle
\geq \langle \tilde{\Lambda}^t, \tilde{L} \tilde{\Lambda}^t \rangle
\geq -\frac{\delta_{Lm}}{2} \left( \|\tilde{x}^t\|^2 + \|\tilde{m}^t\|^2 + \|\tilde{\lambda}^t\|^2 \right),
\]
where the first inequality follows from the convexity of function \( \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 \) on \( x^t \) and \( m^t \); the second inequality follows from the semi-positive definite of matrix \( \tilde{L} \). Combining (20) and (24) implies
\[
V^t \geq \frac{\delta_{Lm}}{2} \left( \|\tilde{x}^t\|^2 + \|\tilde{m}^t\|^2 + \|\tilde{\lambda}^t\|^2 \right) = \delta_{Lm} V_1^t. \quad (25)
\]
Reconsidering \( \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 \), we have
\[
\frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 \leq \frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2
= \frac{1}{2} \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle + \frac{1}{2} \langle \tilde{m}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle
= \frac{1}{2} \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
+ \frac{1}{2} \langle \tilde{m}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle
+ \frac{1}{2} \langle \tilde{\lambda}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle.
\]

According to (19a) and (19b), \( \tilde{L} c = \tilde{L} (R x^t + \tilde{L} m^t) \) and \( R^T c = R^T (R x^t + \tilde{L} m^t) + \tilde{L} \lambda^t = 0 \). Then we modify (26) as
\[
\frac{1}{2} \|R x^t + \tilde{L} m^t - c\|^2 \leq \frac{1}{2} \langle R x^t + \tilde{L} m^t, R x^t + \tilde{L} m^t \rangle - \langle \tilde{x}^t, \tilde{L} \tilde{\lambda}^t \rangle.
\]

Consequently, we establish the upper bound of \( V_2^t \) as
\[
V_2^t \leq \frac{1}{2} \langle R x^t + \tilde{L} m^t, R x^t + \tilde{L} m^t \rangle + \langle \tilde{x}^t, \tilde{L} \tilde{\lambda}^t \rangle
+ \frac{1}{2} \langle \tilde{x}^t, \tilde{L} \tilde{\lambda}^t \rangle
\leq (\delta_{Rm}^2 + \delta_{Lm}) \|\tilde{x}^t\|^2 + (\frac{\delta_{Lm}}{2} \|\tilde{m}^t\|^2 + \frac{\delta_{Lm}}{2} \|\tilde{\lambda}^t\|^2
\leq (\delta_{Rm}^2 + \delta_{Lm}) (\|\tilde{x}^t\|^2 + \|\tilde{m}^t\|^2 + \|\tilde{\lambda}^t\|^2).
\]

Combining (21) and (28) yields
\[
V^t \leq \delta_{Rm}^2 + 2\delta_{Lm}) \left( \|\tilde{x}^t\|^2 + \|\tilde{m}^t\|^2 + \|\tilde{\lambda}^t\|^2 \right) = \delta_{Lm} V_1^t. \quad (29)
\]

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

**Lemma 4.** Under Assumption 1, let \( (X^t, M^t, \Lambda^t) \) be generated by Algorithm 1. Then \( V^t \) is not positive.

**Proof.** Note that
\[
\dot{V}^t = \langle \dot{x}^t, \nabla_x V_1^t \rangle + \langle \dot{m}^t, \nabla_m V_1^t \rangle + \langle \dot{\lambda}^t, \nabla_{\lambda} V_1^t \rangle
= -\langle \dot{x}^t, R^T (R x^t + \tilde{L} m^t - c) + \tilde{L} (x^t + \lambda^t) \rangle
\]
\[
- \langle \dot{m}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle + \langle \dot{\lambda}^t, \tilde{L} \dot{\lambda}^t \rangle
= -\langle \dot{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
- \langle \dot{m}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle
- \langle \dot{\lambda}^t, \tilde{L} \dot{\lambda}^t \rangle + \langle \dot{\lambda}^t, \Lambda^t \rangle
= \Omega_1^t + \Omega_2^t + \Omega_3^t + \Omega_4^t,
\]
where \( \Omega_1^t := -\langle \dot{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle, \Omega_2^t := -\langle \dot{m}^t, \tilde{L} (R x^t + \tilde{L} m^t - c) \rangle, \Omega_3^t := -\langle \dot{\lambda}^t, \Lambda^t \rangle, \text{and} \Omega_4^t := \langle \dot{\lambda}^t, \dot{\lambda}^t \rangle. \) For \( \Omega_1^t + \Omega_2^t \), we have
\[
\Omega_1^t + \Omega_2^t = \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
- \langle \tilde{m}^t, \tilde{L}^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
= \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
+ \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
- \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
+ \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
- \langle \tilde{m}^t, \tilde{L}^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
+ \langle \tilde{m}^t, \tilde{L}^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
= \langle \tilde{x}^t, P \tilde{x}^t \rangle - \langle \tilde{m}^t, \tilde{L}^T \tilde{L} m^t \rangle
\]
\[
- \langle \tilde{x}^t, R^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
- \langle \tilde{m}^t, \tilde{L}^T (R x^t + \tilde{L} m^t - c) \rangle
\]
\[
\tag{31}
\]
\[
\text{Referring to the equilibria in (19a) and (19b), we obtain}
\]
\[
\begin{cases}
R^T R x^* - R^t c = - R^T \tilde{L} m^* - \tilde{L} x^* , \\
\tilde{L}^T (\tilde{L} m^* - c) = -\tilde{L}^T R x^* .
\end{cases}
\tag{32a,b}
\]
\[
\text{Substituting (32a) and (32b) into (31) yields}
\]
\[
\Omega_1^t + \Omega_2^t = \langle \tilde{x}^t, R^T R x^t \rangle - \langle \tilde{m}^t, \tilde{L}^T \tilde{L} m^t \rangle
\]
\[
- \langle \tilde{x}^t, R^T \tilde{L} m^t - \tilde{L} x^t \rangle - \langle \tilde{m}^t, \tilde{L}^T R x^t \rangle
\]
\[
= \langle R \tilde{x}^t, \tilde{R} \tilde{x}^t \rangle - \langle \tilde{L} \tilde{m}^t, \tilde{L} \tilde{m}^t \rangle
\]
\[
-\tilde{L}(\tilde{L} m^t, \tilde{R} x^t) + \langle \tilde{x}^t, \tilde{L} x^t \rangle
\]
\[
\tag{33}
\leq \langle \tilde{x}^t, \tilde{L} x^t \rangle .
\]
\[
\text{Combining (30) and (33), we have}
\]
\[
\tilde{V}_1^t \leq \langle \tilde{x}^t, \tilde{L} x^t \rangle - \langle \tilde{x}^t, \tilde{L} x^t + \tilde{L} x^* \rangle + \langle \tilde{x}^t, \tilde{L} x^t \rangle
\]
\[
= \langle \tilde{x}^t, \tilde{L} x^t \rangle - \langle \tilde{x}^t, \tilde{L} x^* \rangle + \langle \tilde{x}^t, \tilde{L} x^* \rangle
\]
\[
\leq -\delta_{\tilde{L} m}^{-1} \| \tilde{L} x^t \|^2 .
\]
\text{where the equality follows from \( \tilde{L} x^t = \tilde{L} x^* \) since \( \tilde{L} x^* = 0 \) and the last inequality follows from the fact that the communication graph is undirected.}
\[
\text{Furthermore, note that}
\]
\[
\tilde{V}_2^t = \langle \tilde{x}^t, \nabla_{\tilde{L} x^t} V_2^t \rangle + \langle \tilde{m}^t, \nabla_{\tilde{L} m^t} V_2^t \rangle + \langle \tilde{L} x^t, \nabla_{\tilde{L} x^t} V_2^t \rangle
\]
\[
= -\| R^T (R x^t + \tilde{L} m^t - c) + \tilde{L} (\tilde{x}^t + \tilde{L} x^t) \|^2
\]
\[
-\| \tilde{L}^T (R x^t + \tilde{L} m^t - c) \|^2 + \| \tilde{L} x^t \|^2 .
\]
\text{Combining (34) and (35), we obtain}
\[
\tilde{V}_t = \tilde{V}_1^t + \tilde{V}_2^t
\]
\[
= -\| R^T (R x^t + \tilde{L} m^t - c) + \tilde{L} (\tilde{x}^t + \tilde{L} x^t) \|^2
\]
\[
-\| \tilde{L}^T (R x^t + \tilde{L} m^t - c) \|^2 + \| \tilde{L} x^t \|^2
\]
\[
\leq 0 ,
\]
\text{which leads to the conclusion.}
\]
\text{Based on Lemmas 3 and 4, we show a convergence result of the proposed algorithm as follows.}
\text{\textit{Theorem 1.} Under Assumption 1, let \( (X^t, M^t, \Lambda^t) \) be generated by Algorithm 1. Then \( (X^t, M^t, \Lambda^t) \) exponentially converges to the optimal solution of (10) with the following exponential convergence rate,
\[
\| \tilde{X}^t \|^2 + \| \tilde{M}^t \|^2 + \| \tilde{\Lambda}^t \|^2 \leq \frac{2 V_0}{\delta_{\tilde{L} m}^{-2} + 2 \delta_{\tilde{L} m}^{-1}} ,
\]
Fig. 1. The trajectories of the estimation of $X^*$.

Fig. 2. The trajectories of $\|A^T(AX+XB-C)+(AX+XB-C)B^T\|_2$.


REFERENCES