

Distributed Computation for Solving the Sylvester Equation Based on Optimization

Wen Deng, Xianlin Zeng, and Yiguang Hong, *Fellow, IEEE*

Abstract

This paper solves the Sylvester equation in the form of $AX + XB = C$ in a distributed way, and proposes three distributed continuous-time algorithms for three cases. We start with the basic algorithm for solving a least squares solution of the equation, and then give a simplified algorithm for the case when there is an exact solution to the equation, followed by an algorithm with regularization case. Based on local information and appropriate communication among neighbor agents, we solve the distributed computation problem of the Sylvester equation from the optimization viewpoint, and we prove the convergence of proposed algorithms to an optimal solution in three different cases, with help of the convex optimization and semi-stability.

Index Terms

Sylvester equation, distributed algorithm, convex optimization, least squares solution, regularization, semi-stability.

I. INTRODUCTION

Under the influence of big data and large-scale systems, distributed optimization and computation have attracted more and more research attention. Both discrete-time algorithms [1]–[3] and continuous-time algorithms [4]–[6] have been given for various distributed optimization problems. The basic idea is that many interconnected agents in a network, having local information

W. Deng and Y. Hong are with the School of Mathematical Sciences, University of Chinese Academy of Sciences; Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China (dengwena@amss.ac.cn, yghong@iss.ac.cn).

X. Zeng is with Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing, 100081, China (xianlin.zeng@bit.edu.cn).

separately, cooperate with their neighbors to exchange information and achieve global goals eventually.

With the rapid development of distributed optimization algorithms, the idea of using distributed methods to solve matrix equations has attracted much interest. The Sylvester equation is an important class of matrix equations, which has wide application in control theory, systems theory and many other fields [7]–[10]. For instance, the Sylvester equation plays a significant role in computing invariant subspaces [11], achieving pole assignment [12] and model reduction [13]. There have been many centralized algorithms for solving matrix equations, such as Schur decomposition methods, Krylov-subspace methods, and iterative methods [9], [14]–[16]. However, those centralized algorithms for solving the Sylvester equation $AX + XB = C$ mainly need to deal with the whole two coefficient matrices A and B , which could not be applied directly to many distributed scenarios, including the one we consider in this paper. Besides, the parallel distributed computation for the Sylvester equation [17] could not work only for such local information either, because it needs to transform A and B to real Schur form at first.

Recently, there have been several works about solving the linear algebraic equation in the form of $Ax = b$ by various distributed methods [18]–[22]. With the help of a network structure, every node only gets access to the local information, for example, a row [18]–[20] or a column [22] of the matrix A . In [21], there is a double-layered framework for all nodes and it enhances the flexibility of the access to information. There is not much work about solving matrix equations in a distributed way yet. Though we could transform a matrix equation into the form of a linear algebraic equation sometimes, its partition structure has less flexibility to a certain extent. Due to matrix multiplication rules, there are some differences between matrix equations and linear algebraic equations. A class of matrix equations formed as $AXB = F$ was first discussed in [23], with different distributed algorithms according to different partition structures. However, there are no results on the distributed computation of the Sylvester equation, which is more complicated than $AXB = F$, to our knowledge.

The objective of this paper is to solve the Sylvester equation in a specific distributed formulation. Notice that these methods in $Ax = b$ [18]–[20] and $AXB = F$ [23] cannot be applied directly to the Sylvester equation $AX + XB = C$ considered in this paper, because properties of matrix multiplication and the given partition structure make it different from $Ax = b$ and $AXB = F$. In the distributed structure we consider, each agent only needs to know partial rows or columns of information matrices A, B, C and communicate with its neighbors to exchange

information. Therefore, appropriate conversion methods are introduced to deal with the given partition structure of $AX + XB = C$ in our problem formulation. Main contributions of this paper are summarized as follows:

- We consider the distributed computation problem for the Sylvester equation with a special distributed structure in two main cases: the least squares solution case and the regularization case. To design the distributed algorithm, in which each agent exactly knows the corresponding rows of A and columns of B , we introduce two new variables to deal with the inconsistency caused by different partitions of rows or columns, and then reformulate the problem as some distributed optimization problems by constructing an appropriate equivalent transformation.
- We propose distributed algorithms to solve the Sylvester equation in different cases based on saddle-point dynamics, and moreover, we propose a simplified algorithm for the case when there exist exact solutions. It is worthwhile to mention that we employ the derivative feedback technique in the (nonsmooth) regularization case for the optimization problem with the penalty term for variable X , to solve the problem.
- We provide exponential convergence analysis of proposed algorithms for the least squares (exact) solution case and the regularization case, respectively, under mild conditions. Since solutions of the Sylvester equation may be not unique and the optimal solution obtained by the algorithm trajectory may depend on initial conditions, we employ the semi-stability theory of dynamical systems, which tackles convergence properties of dynamical systems having a continuum of equilibria.

The rest paper is organized as follows. Section II introduces relevant mathematical preliminaries, while Section III formulates distributed computation problems of the Sylvester equation and reformulated distributed optimization problems. Then Sections IV and V present distributed algorithms for the least squares solution case, the exact solution case, and the regularization case with corresponding convergence analysis. Following that, Section VI provides a short discussion and three numerical examples, and Section VII concludes this paper briefly.

II. PRELIMINARIES

In this section, we give related notations, some basic concepts, and some lemmas for the analysis.

Let $\mathbb{R}^{m \times n}$ denote the set of $m \times n$ real matrices, $\mathbf{1}_n$ (or $\mathbf{0}_n$) and I_n denote a vector in \mathbb{R}^n with all elements of 1 (or 0) and the $n \times n$ identity matrix, respectively. Let M^T , $\text{Im}(M)$, $\ker(M)$ and $\text{spec}(M)$ denote the transpose, the image, the kernel and the set of all eigenvalues of the matrix M , respectively. Let $\text{vec}(X) \in \mathbb{R}^{mn}$ denote the vector that is a stack of all columns in matrix $X \in \mathbb{R}^{m \times n}$ and $\text{col} \{M_1, \dots, M_n\} = [M_1^T, \dots, M_n^T]^T$ for every $M_i \in \mathbb{R}^{m_i \times n}$. Let \otimes denote the Kronecker product. Let $\|\cdot\|_F$ denote the Frobenius norm of a matrix, $\|M\|_F := \sqrt{\sum_{i,j} M_{i,j}^2}$ and $\langle \cdot, \cdot \rangle_F$ denote the Frobenius inner product of real matrices: $\langle M_1, M_2 \rangle_F = \sum_{i,j} (M_1)_{i,j} (M_2)_{i,j}$ with $M_1, M_2 \in \mathbb{R}^{m \times n}$. Define augmented matrices $[Y_i]_{\mathcal{R}}, [Z_i]_{\mathcal{C}} \in \mathbb{R}^{m \times r}$ as

$$\begin{aligned} [Y_i]_{\mathcal{R}} &= \left[0_{m_1 \times r}^T, \dots, 0_{m_{i-1} \times r}^T, Y_i^T, 0_{m_{i+1} \times r}^T, \dots, 0_{m_n \times r}^T \right]^T, \\ [Z_i]_{\mathcal{C}} &= \left[0_{m \times r_1}, \dots, 0_{m \times r_{i-1}}, Z_i, 0_{m \times r_{i+1}}, \dots, 0_{m \times r_n} \right], \end{aligned}$$

with $Y_i \in \mathbb{R}^{m_i \times r}$, $Z_i \in \mathbb{R}^{m \times r_i}$.

Definition 1 (see [24]): Let f be a convex function. A vector h is called a subgradient of function f at point $X_0 \in \text{dom} f$ if, for any $X \in \text{dom} f$,

$$f(X) \geq f(X_0) + \langle h, X - X_0 \rangle. \quad (1)$$

The set of all subgradients $\partial f(X_0)$ is said to be the subdifferential of function f at point X_0 .

Consider a time-invariant dynamical system

$$\dot{x}(t) = \phi(x(t)), \quad x(0) = x_0, \quad t \geq 0, \quad (2)$$

where $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous. Given $x(t) : [0, \infty) \rightarrow \mathbb{R}^d$ as a trajectory of system (2), the point $z \in \mathbb{R}^d$ is a limit point of a solution $x(t)$ if there exists a positive increasing divergent sequence $\{t_i\}_{i=1}^{\infty} \subset \mathbb{R}$ such that $z = \lim_{i \rightarrow \infty} x(t_i)$. A set \mathcal{M} is positively invariant with respect to (2) if, for every $x_0 \in \mathcal{M}$, \mathcal{M} contains the solution $x(t)$ of (2) for all $t > 0$ with $x(0) = x_0$. Then we introduce the following definition from [25] and Lemma 1, which are useful for convergence analysis of algorithms in the least squares case.

Definition 2: A point z is semi-stable if z is Lyapunov stable and there exists an open subset \mathcal{D}_0 of \mathcal{D} containing z such that, for all initial conditions in \mathcal{D}_0 , the solution of (2) converge to a Lyapunov stable equilibrium point. Then system (2) is semi-stable with respect to \mathcal{D} if every solution with initial condition in \mathcal{D} converges to a Lyapunov stable equilibrium. Moreover, (2) is said to be globally semi-stable if it is semi-stable with respect to \mathbb{R}^d .

Lemma 1 (Theorem 3.1 in [25]): Let \mathcal{D} be an open positively invariant set with respect to (2), $V : \mathcal{D} \rightarrow \mathbb{R}$ be a continuously differentiable function, and $x(t)$ be a solution of (2) with

$x(0) \in \mathcal{D}$, contained in a compact subset of \mathcal{D} . Assume $\frac{d}{dt}V(x(t)) \leq 0$, for all $x \in \mathcal{D}$ and define $\mathcal{Z} = \{x \in \mathcal{D} : \frac{d}{dt}V(x) = 0\}$. If every point in the largest invariant subset \mathcal{M} of $\bar{\mathcal{Z}} \cap \mathcal{D}$ is Lyapunov stable equilibrium, where $\bar{\mathcal{Z}}$ is the closure of \mathcal{Z} , then the system (2) is semi-stable with respect to \mathcal{D} .

Consider a time-invariant differential inclusion

$$\dot{x}(t) \in \mathcal{F}(x(t)), \quad x(0) = x_0, \quad t \geq 0, \quad (3)$$

where $\mathcal{F} : \mathbb{R}^d \rightarrow \mathcal{B}(\mathbb{R}^d)$ is upper-semicontinuous, and the set $\mathcal{F}(x)$ is nonempty, compact and convex for any x , which can guarantee the existence of solutions of (3) [26]. A set \mathcal{M} is weakly (strongly) positively invariant with respect to (3) if, for every $x_0 \in \mathcal{M}$, \mathcal{M} contains a solution $x(t)$ (all solutions) of (3) for all $t > 0$ with $x(0) = x_0$. An equilibrium x^* of (3) satisfies $0_d \in \mathcal{F}(x^*)$. The definition of semi-stable about (3) [25] is similar to Definition 2. Then we introduce the following Definition 3 and Lemmas 2 and 3 from [26], [27], which are useful for convergence analysis of the algorithm in the regularization case.

Definition 3: The set-valued Lie derivative $\tilde{\mathcal{L}}_{\mathcal{F}}f : \mathbb{R}^d \rightarrow \mathcal{B}(\mathbb{R})$ of a locally Lipschitz function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with respect to $\mathcal{F} : \mathbb{R}^d \rightarrow \mathcal{B}(\mathbb{R}^d)$ at x is defined as

$$\begin{aligned} \tilde{\mathcal{L}}_{\mathcal{F}}f(x) = \{a \in \mathbb{R} : \text{there exists } y \in \mathcal{F}(x) \text{ such that} \\ h^T y = a, \quad \forall h \in \partial f(x)\}. \end{aligned} \quad (4)$$

Remark 1: If f is differentiable at x , then $\tilde{\mathcal{L}}_{\mathcal{F}}f(x) = \{\nabla f(x)^T y : y \in \mathcal{F}(x)\}$.

Lemma 2: Let $x(t)$ be a solution of different inclusion (3), and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a locally Lipschitz continuous and regular function [26]. Then $\frac{d}{dt}f(x(t))$ exists and $\frac{d}{dt}f(x(t)) \in \tilde{\mathcal{L}}_{\mathcal{F}}f(x)$ almost everywhere.

Lemma 3: Suppose that (3) has at least one solution and x^* is an equilibrium of the (3). Let f be locally Lipschitz continuous and regular, and satisfy (i) $f(x) \geq 0$, $f(x) = 0$ if and only if $x = x^*$; (ii) $\max \tilde{\mathcal{L}}_{\mathcal{F}}f(x) \leq 0$. Then x^* is a Lyapunov stable equilibrium of (3).

Lemma 4 (Theorem 3.1 in [25]): Let \mathcal{D} be an open strongly positively invariant set with respect to (3), $V : \mathcal{D} \rightarrow \mathbb{R}$ be a locally Lipschitz continuous and regular function, and $x(t)$ be a solution of (3) with $x(0) \in \mathcal{D}$, contained in a compact subset of \mathcal{D} . Assume $\tilde{\mathcal{L}}_{\mathcal{F}}V(x) \leq 0$ for all $x \in \mathcal{D}$ and define $\mathcal{Z} = \{x \in \mathcal{D} : 0 \in \tilde{\mathcal{L}}_{\mathcal{F}}V(x)\}$. If every point in the largest weakly positively invariant subset \mathcal{M} of $\bar{\mathcal{Z}} \cap \mathcal{D}$ is a Lyapunov stable equilibrium, where $\bar{\mathcal{Z}}$ is the closure of \mathcal{Z} , then the system (3) is semi-stable with respect to \mathcal{D} .

III. PROBLEM FORMULATION

In this section, we first formulate our problem, and then introduce some transformations which will be helpful in the distributed reformulation.

Consider a continuous-time Sylvester equation of the following form

$$AX + XB = C, \quad (5)$$

with $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{r \times r}$, $C \in \mathbb{R}^{m \times r}$ and the unknown variable $X \in \mathbb{R}^{m \times r}$. Here we consider a specific partition for n parts, Left-Row-Right-Column (LRRC) case, to solve the Sylvester equation: dividing matrix A by row and matrix B by column. Specifically, the partition is

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad B = [B_1, \dots, B_n] \in \mathbb{R}^{r \times r},$$

with $A_i \in \mathbb{R}^{m_i \times m}$, $B_i \in \mathbb{R}^{r \times r_i}$ and $\sum_{i=1}^n m_i = m$, $\sum_{i=1}^n r_i = r$. If there is a Lyapunov equation with $B = A^T$, it only needs to divide A by row, naturally. As for matrix C , we can divide it like A by row or like B by column as required. The treatments for different partitions of C are similar, then we focus on the partition of C by column

$$C = [C_1, \dots, C_n] \in \mathbb{R}^{m \times r}, \quad C_i \in \mathbb{R}^{m \times r_i}. \quad (6)$$

This paper aims to solve two main problems for the Sylvester equation:

(I) Least squares solution and exact solution case: Aim to get a least squares solution or an exact solution of the Sylvester equation (5) by solving a convex unconstrained optimization problem:

$$\min_X \|AX + XB - C\|_F^2, \quad (7)$$

which minimizes the squared Frobenius norm of residual matrix.

(II) Regularization case: Aim to solve the equation with penalty requirement to the unknown variable X by solving a regularized convex unconstrained optimization problem:

$$\min_X \frac{1}{2} \|AX + XB - C\|_F^2 + \alpha g(X), \quad (8)$$

where $\alpha > 0$ is a tradeoff between minimizing the squared residual and the penalty term, $g(\cdot)$ is a general convex function representing the regularization penalty of variables.

About the computation of matrix equations, an immediate idea is to use the Kronecker product to rewrite (5) as a standard linear algebraic equation $(I_m \otimes A + B^T \otimes I_r)\text{vec}(X) = \text{vec}(C)$. It

can be obtained that the solution of (5) for each $\text{vec}(C)$ is unique if and only if the matrix $I_m \otimes A + B^T \otimes I_r$ is nonsingular, which is equivalent to requiring $\text{spec}(A) \cap \text{spec}(-B) = \emptyset$ [28]. Different from centralized situations, the distributed situation may spoil the information structures because the sub-blocks of A and B are mixed up for the Kronecker product. Therefore, inspired by the work about distributed algorithms of the linear algebraic equation [18]–[22], we have to study a new method, distinguished from the centralized and parallel algorithms, to solve the Sylvester equation (5) for given partition in a distributed way.

Consider a multi-agent network consisting of n agents, which is described by an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, A_{\mathcal{G}})$, where $\mathcal{V} = \{1, \dots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges, and $A_{\mathcal{G}} = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is the adjacency matrix with $a_{i,j} = a_{j,i} > 0$ if $(i, j) \in \mathcal{E}$ and $a_{i,j} = 0$ otherwise. Denote the neighbor set of agent i as $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}, j \in \mathcal{V}\}$. The Laplacian matrix $L_n = D - A_{\mathcal{G}}$, where $D \in \mathbb{R}^{n \times n}$ is diagonal with $D_{i,i} = \sum_{j=1}^n a_{i,j}$.

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

Suppose that each agent knows a row sub-block of A , a column sub-block of B , and a column sub-block of C , that is, agent i holds A_i, B_i, C_i . As a result, agent i in the network needs to estimate one common variable X^* , which is a solution for the Sylvester equation and their final estimates are supposed to achieve consensus, that is, $X_i = X^*$ for all $i \in \mathcal{V}$.

Different from the cases given in [23], to make AX and XB simultaneously distributed in design for the LRRC case, here we introduce two new variables Y, Z for coordinating the inconsistency due to the different block partitions of A and B . The following statement shows the idea.

For the LRRC case of matrices A and B , and C divided as (6), the Sylvester equation (5) is equivalent to

$$\begin{cases} AX = Y \\ XB = C - Z \\ Y = Z \end{cases} \quad (9)$$

where supplementary variables Y and Z have corresponding block structures, $Y = \text{col}\{Y_1, \dots, Y_n\}$, $Z = [Z_1, \dots, Z_n]$ with $Y_i \in \mathbb{R}^{m_i \times r}$, $Z_i \in \mathbb{R}^{m \times r_i}$.

In the considered multi-agent network, agent i in the network \mathcal{G} has knowledge of information A_i, B_i, C_i and state set (X_i, Y_i, Z_i) , and cooperates with its neighbors to compute X_i through the exchange of local information.

Remark 2: If C is divided by row like A , then the equivalent form of the Sylvester equation (5) is $AX = C - Z$, $XB = Y$, $Z = Y$ where $Z = \text{col}\{Z_1, \dots, Z_n\}$, $Y = [Y_1, \dots, Y_n]$ with $Z_i \in \mathbb{R}^{m_i \times r}$, $Y_i \in \mathbb{R}^{m \times r_i}$. Algorithms and analysis ideas are analogous to the case in (9), so we will not discuss the case for space limitations.

In order to reformulate the equivalent form (9) as an optimization problem, the idea is to take one or more residuals of the equations as an objective function and take others as equality constrains. Unlike the unconstrained optimization problem (7), the reformulated problems are standard optimization problems with equality constraints. Further, we need to consider the given distributed situation. Since the coupling equality constraint $Y = Z$ is not separable for each agent, and consensus constraints $X_i = X_j$ for all $i, j \in \mathcal{V}$ are hard to deal with directly, following transformations are necessary to assign each constraint to each agent clearly.

Lemma 5: These two equivalent transformations hold for all $i \in \mathcal{V}$.

$$Y = Z \iff \exists \{W_i\}_{i=1}^n, \text{ s.t.}$$

$$[Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j}(W_i - W_j) = 0_{m \times r}, \quad (10a)$$

$$X_i = X_j, \forall j \in \mathcal{V} \iff \sum_{j=1}^n a_{ij}(X_i - X_j) = 0_{m \times r}, \quad (10b)$$

where $[a_{i,j}]$ is the corresponding adjacency matrix of an undirected and connected graph \mathcal{G} , $W_i \in \mathbb{R}^{m \times r}$ are introduced to make up for the inconsistencies between $[Y_i]_{\mathcal{R}}$ and $[Z_i]_{\mathcal{C}}$.

Proof: As for (10a), it is trivial that $Y = Z$ holds by summing the right side of (10a) from $i = 1$ to $i = n$. On the other hand, according to the structures of Y and Z , we have $Y = \sum_{i=1}^n [Y_i]_{\mathcal{R}}$, $Z = \sum_{i=1}^n [Z_i]_{\mathcal{C}}$. Define $L = L_n \otimes I_m$, where L_n is the Laplacian matrix of graph \mathcal{G} . Therefore, $\ker(L) = \{k\mathbf{1}_n \otimes I_m : k \in \mathbb{R}\}$ and $\text{Im}(L) = \ker(L)^\perp = \{w \in \mathbb{R}^{nm} : \sum_{i=1}^n w_i = 0_m, w = \text{col}\{w_1, \dots, w_n\}, w_i \in \mathbb{R}^m, i \in \mathcal{V}\}$. Because $\sum_{i=1}^n [Y_i]_{\mathcal{R}} = \sum_{i=1}^n [Z_i]_{\mathcal{C}}$, the equation $\sum_{i=1}^n ([Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}}) = 0_{m \times r}$ infers that every column of $\text{col}\{[Y_1]_{\mathcal{R}} - [Z_1]_{\mathcal{C}}, \dots, [Y_n]_{\mathcal{R}} - [Z_n]_{\mathcal{C}}\}$ belongs to $\text{Im}(L)$. Thus, there exists $W = \text{col}\{W_1, \dots, W_n\} \in \mathbb{R}^{nm \times r}$, such that $LW = \text{col}\{[Y_1]_{\mathcal{R}} - [Z_1]_{\mathcal{C}}, \dots, [Y_n]_{\mathcal{R}} - [Z_n]_{\mathcal{C}}\}$. Then we have (10a).

As for (10b), the equivalence is straightforward, and its proof is omitted here. \square

Remark 3: The equality constraint (10a) is a typical coupling constraint, and (10b) is a typical consensus constraint. The transformations in Lemma 5 are always effective to deal with these two kinds of constraints.

In brief, to reformulate the problem clearly, we have already introduced three types of variables, where their purposes are presented in Table I.

TABLE I
THE PURPOSES OF INTRODUCED VARIABLES

| Variable | Purpose |
|-------------------------------------|--|
| $Y = \text{col}\{Y_1, \dots, Y_n\}$ | $Y_i = A_i X$, result of $A_i X$ stored by sub-row of Y |
| $Z = [Z_1, \dots, Z_n]$ | $Z_i = C_i - X B_i$, result of $C_i - X B_i$ stored by sub-column of Z |
| $W = \text{col}\{W_1, \dots, W_n\}$ | $(L_n \otimes I_m)W = \text{col}\{[Y_1]_{\mathcal{R}} - [Z_1]_{\mathcal{C}}, \dots, [Y_n]_{\mathcal{R}} - [Z_n]_{\mathcal{C}}\}$ |

IV. THE CASES OF LEAST SQUARES SOLUTION AND EXACT SOLUTION

A. Least Squares Solution Case

We first consider solving a least squares solution of the Sylvester equation (5) because it may not have exact solutions. Additionally, since local information is stored by different agents, it is not clear for each agent whether there are exact solutions. Of course, the proposed method for a least squares solution still works when there exist exact solutions.

1) *Optimization Model:* Reformulate (9) as an optimization problem

$$\begin{aligned} \min_{X, Y, Z} \quad & \frac{1}{2} \|XB - C + Z\|_F^2 \\ \text{s.t.} \quad & AX = Y, \quad Y = Z. \end{aligned} \tag{11}$$

Next, we distribute the information to every agent i and add the consensus constraints $X_i = X_j$ for all $i, j \in \mathcal{V}$. According to Lemma 5, (11) can be rewritten as a distributed optimization problem

$$\begin{aligned} \min_{\bar{X}, Y, Z, W} \quad & \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 \\ \text{s.t.} \quad & \sum_{j=1}^n a_{i,j} (X_i - X_j) = 0_{m \times r}, \quad A_i X_i = Y_i, \\ & [Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i - W_j) = 0_{m \times r}, \quad \forall i \in \mathcal{V}, \end{aligned} \tag{12}$$

with $\bar{X} = \text{col}\{X_1, \dots, X_n\}$, $W = \text{col}\{W_1, \dots, W_n\}$ and Y, Z as in (9). In the optimization problem (12), $\Lambda_i \in \mathbb{R}^{m \times r}$, $\Upsilon_i \in \mathbb{R}^{m_i \times r}$ and $\Theta_i \in \mathbb{R}^{m \times r}$, $i \in \mathcal{V}$ are introduced as Lagrange multipliers associated with these equality constraints and $\Lambda = \text{col}\{\Lambda_1, \dots, \Lambda_n\}$, $\Upsilon = \text{col}\{\Upsilon_1, \dots, \Upsilon_n\}$, $\Theta =$

$\text{col}\{\Theta_1, \dots, \Theta_n\}$. Let $(\bar{X}^*, Y^*, Z^*, W^*)$ and $(\Lambda^*, \Upsilon^*, \Theta^*)$ be a pair of primal and dual optimal points. Then they satisfy the Karush-Kuhn-Tucker (KKT) condition, which leads to a necessary and sufficient condition for optimality, for all $i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} 0_{m \times r} = \sum_{j=1}^n a_{i,j} (X_i^* - X_j^*), \\ 0_{m_i \times r} = A_i X_i^* - Y_i^*, \\ 0_{m \times r} = [Y_i^*]_{\mathcal{R}} - [Z_i^*]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i^* - W_j^*), \\ 0_{m \times r} = -(X_i^* B_i - C_i + Z_i^*) B_i^T - \sum_{j=1}^n a_{i,j} (\Lambda_i^* - \Lambda_j^*) \\ \quad - A_i^T \Upsilon_i^*, \\ 0_{m_i \times r} = \Upsilon_i^* - [I_{m_i}]_{\mathcal{C}} \Theta_i^*, \\ 0_{m \times r_i} = -(X_i^* B_i - C_i + Z_i^*) + \Theta_i^* [I_{r_i}]_{\mathcal{R}}, \\ 0_{m \times r} = \sum_{j=1}^n a_{i,j} (\Theta_i^* - \Theta_j^*). \end{array} \right. \quad (13)$$

2) *Distributed Algorithm*: Each agent i in the network knows A_i, B_i, C_i and has its own state $(X_i, Y_i, Z_i, W_i, \Lambda_i, \Upsilon_i, \Theta_i)$. Through the exchange of information, agent i can receive partial state information $(X_j, W_j, \Theta_j), j \in \mathcal{N}_i$ from its neighbor agents. Then we propose Algorithm 1 as the distributed algorithm for each agent i in the least squares solution case, where

Algorithm 1 Algorithm for the least squares solution case

$X_i(0) \in \mathbb{R}^{m \times r}, Y_i(0) \in \mathbb{R}^{m_i \times r}, Z_i(0) \in \mathbb{R}^{m \times r_i}, W_i(0) \in \mathbb{R}^{m \times r}, \Lambda_i(0) \in \mathbb{R}^{m \times r}, \Upsilon_i(0) \in \mathbb{R}^{m_i \times r}, \Theta_i(0) \in \mathbb{R}^{m \times r}, i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} \dot{X}_i(t) = -(X_i(t) B_i - C_i + Z_i(t)) B_i^T - A_i^T (A_i X_i(t) - Y_i(t)) \\ \quad - A_i^T \Upsilon_i(t) - \sum_{j=1}^n a_{i,j} (\Lambda_i(t) - \Lambda_j(t)) \\ \quad - \sum_{j=1}^n a_{i,j} (X_i(t) - X_j(t)), \\ \dot{Y}_i(t) = \Upsilon_i(t) - [I_{m_i}]_{\mathcal{C}} \Theta_i(t) + A_i X_i(t) - Y_i(t), \\ \dot{Z}_i(t) = -(X_i(t) B_i - C_i + Z_i(t)) + \Theta_i(t) [I_{r_i}]_{\mathcal{R}}, \\ \dot{W}_i(t) = \sum_{j=1}^n a_{i,j} (\Theta_i(t) - \Theta_j(t)), \\ \dot{\Lambda}_i(t) = \sum_{j=1}^n a_{i,j} (X_i(t) - X_j(t)), \\ \dot{\Upsilon}_i(t) = A_i X_i(t) - Y_i(t), \\ \dot{\Theta}_i(t) = [Y_i(t)]_{\mathcal{R}} - [Z_i(t)]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i(t) - W_j(t)) \\ \quad - \sum_{j=1}^n a_{i,j} (\Theta_i(t) - \Theta_j(t)). \end{array} \right. \quad (14)$$

$(X_i(t), Y_i(t), Z_i(t), W_i(t))$ and $(\Lambda_i(t), \Upsilon_i(t), \Theta_i(t))$ are the estimates of solutions to problem (12) and the estimates of Lagrange multipliers by agent i at time t , respectively. Algorithm 1 can be

viewed as the saddle-point dynamics of an augmented Lagrangian function \mathcal{L}_l , that is, for all $i \in \mathcal{V}$,

$$\dot{\Phi}_i = -\nabla_{\Phi_i} \mathcal{L}_l, \text{ when } \Phi_i = X_i, Y_i, Z_i \text{ or } W_i,$$

$$\dot{\Psi}_i = \nabla_{\Psi_i} \mathcal{L}_l, \text{ when } \Psi_i = \Lambda_i, \Upsilon_i, \text{ or } \Theta_i,$$

where \mathcal{L}_l is defined in (15) and the last three terms are augmented terms.

$$\begin{aligned} \mathcal{L}_l = & \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \sum_{i=1}^n \langle \Lambda_i, \sum_{j=1}^n a_{i,j} (X_i - X_j) \rangle_F + \sum_{i=1}^n \langle \Upsilon_i, A_i X_i - Y_i \rangle_F \\ & + \sum_{i=1}^n \langle \Theta_i, [Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i - W_j) \rangle_F - \frac{1}{2} \sum_{i=1}^n \langle \Theta_i, \sum_{j=1}^n a_{i,j} (\Theta_i - \Theta_j) \rangle_F \\ & + \frac{1}{2} \sum_{i=1}^n \langle X_i, \sum_{j=1}^n a_{i,j} (X_i - X_j) \rangle_F + \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2 \end{aligned}$$

Lemma 6: Under Assumption 1, let $P_l^* = \text{col}\{\bar{X}^*, Y^*, Z^*, W^*, \Lambda^*, \Upsilon^*, \Theta^*\}$. Then $(\bar{X}^*, Y^*, Z^*, W^*)$ is an optimal solution of problem (12) if and only if there exist $\Lambda^* \in \mathbb{R}^{mn \times r}$, $\Upsilon^* \in \mathbb{R}^{m \times r}$, and $\Theta^* \in \mathbb{R}^{mn \times r}$, such that P_l^* is an equilibrium of system (14).

This lemma holds obviously because of the KKT condition (13).

Denote $P_l = \text{col}\{\bar{X}, Y, Z, W, \Lambda, \Upsilon, \Theta\}$. The following result shows main convergence properties of Algorithm 1.

Theorem 1: Under Assumption 1, the following hold with Algorithm 1.

- (i) Every equilibrium of the system (14) is Lyapunov stable and every solution is bounded;
- (ii) The system (14) is globally semi-stable. Moreover, the solution $P_l(t)$ with initial condition $P_l(0)$ converges to an equilibrium of (14) exponentially;
- (iii) For every agent i , the limit of its estimation $\lim_{t \rightarrow \infty} X_i(t) = X^*$ is a least squares solution of the Sylvester equation (5).

Its proof is in Subsection IV-C.

B. Exact Solution Case

In the case when there exist exact solutions of the Sylvester equation (5), we can provide a simplified algorithm with some conversions.

Assumption 2: The Sylvester equation (5) has at least one exact solution.

1) *Optimization Model:* Let Assumptions 1 and 2 hold. The set $\{(X, Y, Z) : AX = Y, XB = C - Z, Y = Z\}$ has at least one element. In this case, we can formulate the optimization problem (11) as follows:

$$\begin{aligned} \min_{X,Y,Z} \quad & \frac{1}{2} \|XB - C + Z\|_F^2 + \frac{1}{2} \|AX - Y\|_F^2 \\ \text{s.t.} \quad & Y = Z. \end{aligned} \quad (15)$$

Equivalently, the problems (11) and (15) have the same solution, because (11) can reach the optimal value 0. Rewrite the objective function as the sum of each agent's objective function; then $\frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2$ can reach the minimum 0. In fact,

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2 \\ & \leq \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2 \\ & \quad + \frac{1}{2} \sum_{i=1}^n \left\langle \sum_{j=1}^n a_{i,j} (X_i - X_j), X_i \right\rangle_F. \end{aligned} \quad (16)$$

When the right-hand side term reaches 0, $X_i = X_j$ must hold for all $i, j \in \mathcal{V}$. If the left-hand side term reaches 0 with $X_i = X_j$ as an exact solution of the original equation, the right-hand side term equals 0 as well. Clearly, we can use the right-hand side of (16) as the objective function to replace consensus constraints $X_i = X_j$. Then the following distributed optimization problem is formulated:

$$\begin{aligned} \min_{\bar{X}, Y, Z, W} \quad & \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2 \\ & + \frac{1}{2} \sum_{i=1}^n \left\langle \sum_{j=1}^n a_{i,j} (X_i - X_j), X_i \right\rangle_F \\ \text{s.t.} \quad & [Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i - W_j) = 0_{m \times r}, \forall i \in \mathcal{V}. \end{aligned} \quad (17)$$

Remark 4: Denote an optimal point for problem (17) by $(\bar{X}^*, Y^*, Z^*, W^*)$. According to Assumption 2, the objective function can reach 0, that is, $X_i^* B_i - C_i + Z_i^* = 0, A_i X_i^* - Y_i^* = 0$ and $X_i^* = X_j^*$, where $X_i^* = X_j^* = X^*$ is an exact solution of (5) and (X^*, Y^*, Z^*) is an exact solution of (9).

In the optimization problem (17), $\Theta_i \in \mathbb{R}^{m \times r}$ is introduced as the Lagrange multiplier associated with the i -th equality constraint and $\Theta = \text{col}\{\Theta_1, \dots, \Theta_n\}$. Let $(\bar{X}^*, Y^*, Z^*, W^*)$

and Θ^* be a pair of primal and dual optimal points. Then they satisfy the corresponding KKT condition (18), which proposes necessary and sufficient conditions for the optimality, for all $i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} 0_{m \times r} = [Y_i^*]_{\mathcal{R}} - [Z_i^*]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j}(W_i^* - W_j^*), \\ 0_{m \times r} = -(X_i^* B_i - C_i + Z_i^*) B_i^T - \sum_{j=1}^n a_{i,j}(\Lambda_i^* - \Lambda_j^*) \\ \quad - \sum_{j=1}^n a_{i,j}(X_i^* - X_j^*), \\ 0_{m_i \times r} = (A_i X_i^* - Y_i^*) - [I_{m_i}]_{\mathcal{C}} \Theta_i^*, \\ 0_{m \times r_i} = -(X_i^* B_i - C_i + Z_i^*) + \Theta_i^* [I_{r_i}]_{\mathcal{R}}, \\ 0_{m \times r} = \sum_{j=1}^n a_{i,j}(\Theta_i^* - \Theta_j^*). \end{array} \right. \quad (18)$$

2) *Distributed Algorithm*: Each agent i in the network knows A_i, B_i, C_i and has its own state $(X_i, Y_i, Z_i, W_i, \Theta_i)$. Through the exchange of information, agent i can receive partial state information $(X_j, W_j, \Theta_j), j \in \mathcal{N}_i$ from its neighbor agents. Then we propose Algorithm 2, using the similar idea of Algorithm 1, as the distributed algorithm for each agent i in the exact solution case, where $(X_i(t), Y_i(t), Z_i(t), W_i(t))$ and $\Theta_i(t)$ are the estimates of solutions to problem (17)

Algorithm 2 Algorithm for the exact solution case

$X_i(0) \in \mathbb{R}^{m \times r}, Y_i(0) \in \mathbb{R}^{m_i \times r}, Z_i(0) \in \mathbb{R}^{m \times r_i}, W_i(0) \in \mathbb{R}^{m \times r}, \Theta_i(0) \in \mathbb{R}^{m \times r}, i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} \dot{X}_i(t) = -(X_i(t) B_i - C_i + Z_i(t)) B_i^T - A_i^T (A_i X_i(t) - Y_i(t)) \\ \quad - \sum_{j=1}^n a_{i,j}(X_i(t) - X_j(t)), \\ \dot{Y}_i(t) = (A_i X_i(t) - Y_i(t)) - [I_{m_i}]_{\mathcal{C}} \Theta_i(t), \\ \dot{Z}_i(t) = -(X_i(t) B_i - C_i + Z_i(t)) + \Theta_i(t) [I_{r_i}]_{\mathcal{R}}, \\ \dot{W}_i(t) = \sum_{j=1}^n a_{i,j}(\Theta_i(t) - \Theta_j(t)), \\ \dot{\Theta}_i(t) = [Y_i(t)]_{\mathcal{R}} - [Z_i(t)]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j}(W_i(t) - W_j(t)) \\ \quad - \sum_{j=1}^n a_{i,j}(\Theta_i(t) - \Theta_j(t)). \end{array} \right. \quad (19)$$

and the estimate of Lagrange multiplier by agent i at time t , respectively.

Remark 5: Compared with Algorithm 1, the state variables that every agent needs to calculate in Algorithm 2 have changed from $(X_i(t), Y_i(t), Z_i(t), W_i(t), \Lambda_i(t), \Upsilon_i(t), \Theta_i(t))$ to $(X_i(t), Y_i(t), Z_i(t), W_i(t), \Theta_i(t))$. On the premise of exact solutions, the algorithm dimension has been reduced to a certain extent.

Lemma 7: Under Assumptions 1 and 2, let $P_e^* = \text{col}\{\bar{X}^*, Y^*, Z^*, W^*, \Theta^*\}$. Then $(\bar{X}^*, Y^*, Z^*, W^*)$ is an optimal solution of problem (17) if and only if there exists $\Theta^* \in \mathbb{R}^{mn \times r}$ such that P_e^* is an equilibrium of system (19).

This lemma can be easily proved by the KKT optimality condition (18).

Denote $P_e = \text{col}\{\bar{X}, Y, Z, W, \Theta\}$. The following result shows main convergence properties of Algorithm 2.

Theorem 2: Under Assumptions 1 and 2, the following hold with Algorithm 2.

- (i) Every equilibrium of the system (19) is Lyapunov stable and every solution is bounded;
- (ii) The system (19) is globally semi-stable. Moreover, the solution $P_e(t)$ with initial condition $P_e(0)$ converges to an equilibrium of (19) exponentially;
- (iii) For every agent i , the limit of its estimation $\lim_{t \rightarrow \infty} X_i(t) = X^*$ is an exact solution of the Sylvester equation (5).

Its explanation is in the next subsection.

C. Proofs of Theorems 1 and 2

The Proof of Theorem 1: (i) First, note that $P_l^* = \text{col}\{\bar{X}^*, Y^*, Z^*, W^*, \Lambda^*, \Upsilon^*, \Theta^*\}$ is an equilibrium of (14), which satisfies the KKT condition (13). Define a positive definite function: $V_l(P_l) := \frac{1}{2} \|P_l - P_l^*\|_F^2$. By using Algorithm 1 and the KKT condition (13), we calculate the derivative of V_l with respect to time t and obtain

$$\begin{aligned} \frac{d}{dt} V_l &= - \sum_{i=1}^n \|(X_i - X_i^*)B_i + (Z_i - Z_i^*)\|_F^2 \\ &\quad - \sum_{i=1}^n \|A_i(X_i - X_i^*) - (Y_i - Y_i^*)\|_F^2 \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{i,j} (\|X_i - X_j\|_F^2 + \|\Theta_i - \Theta_j\|_F^2) \leq 0. \end{aligned}$$

Because V_l is positive definite and $\frac{d}{dt} V_l \leq 0$, the equilibrium P_l^* of (14) is Lyapunov stable. Moreover, because of $V_l \rightarrow \infty$ as $\|P_l - P_l^*\|_F \rightarrow \infty$, each solution with given initial conditions of (14) is bounded for all $t \geq 0$ (Theorem 4 in [29]).

(ii) With the condition of $\frac{d}{dt} V_l = 0$, we define the set $\mathcal{R}_l = \{P_l : \frac{d}{dt} V_l = 0\} = \{P_l : X_i = X_j, \Theta_i = \Theta_j, (X_i - X_i^*)B_i = -(Z_i - Z_i^*), A_i(X_i - X_i^*) = Y_i - Y_i^*\}$. Let \mathcal{M}_l be the largest invariant subset of $\bar{\mathcal{R}}_l$. It follows from the LaSalle's invariance principle [29] that $P_l(t) = \text{col}\{\bar{X}(t), Y(t), Z(t), W(t), \Lambda(t), \Upsilon(t), \Theta(t)\} \rightarrow \mathcal{M}_l$, as $t \rightarrow \infty$ and \mathcal{M}_l is positively invariant.

Assume that $\hat{P}_l(t) \in \mathcal{M}_l$ for all $t \geq 0$ is a trajectory of (14). For all $i, j \in \mathcal{V}$, $\hat{P}_l(0) \in \mathcal{M}_l$, $\hat{X}_i(t) = \hat{X}_j(t)$, $\hat{\Theta}_i(t) = \hat{\Theta}_j(t)$, $(\hat{X}_i(t) - \hat{X}_i^*)B_i = -(\hat{Z}_i(t) - \hat{Z}_i^*)$, and $A_i(\hat{X}_i(t) - \hat{X}_i^*) = \hat{Y}_i(t) - \hat{Y}_i^*$. Then we have

- a) $\dot{\hat{\Lambda}}_i(t) = 0_{m \times r}$ because $\dot{\hat{X}}_i(t) = \dot{\hat{X}}_j(t)$.
- b) $\dot{\hat{W}}_i(t) = 0_{m \times r}$ because $\dot{\hat{\Theta}}_i(t) = \dot{\hat{\Theta}}_j(t)$.
- c) Since $A_i(\hat{X}_i(t) - \hat{X}_i^*) = \hat{Y}_i(t) - \hat{Y}_i^*$ and $A_i\hat{X}_i^* = \hat{Y}_i^*$, there hold $A_i\hat{X}_i(t) = \hat{Y}_i(t)$ and $\dot{\hat{Y}}_i(t) = 0_{m_i \times r}$.
- d) $\dot{\hat{X}}_i(t) = -(\hat{X}_i(t)B_i - C_i + \hat{Z}_i(t))B_i^T - \sum_{j=1}^n a_{i,j}(\hat{\Lambda}_i(t) - \hat{\Lambda}_j(t)) - A_i^T\hat{Y}_i(t) - \sum_{j=1}^n a_{i,j}(\hat{X}_i(t) - \hat{X}_j(t)) - A_i^T(A_i\hat{X}_i(t) - \hat{Y}_i(t)) = -(\hat{X}_i^*B_i - C_i + \hat{Z}_i^*)B_i^T - \sum_{j=1}^n (\hat{\Lambda}_i(0) - \hat{\Lambda}_j(0)) - A_i^T\hat{Y}_i(0)$ is a constant matrix. If $\dot{\hat{X}}_i(t) \neq 0_{m \times r}$, $\hat{X}_i(t)$ will be unbound, which is a contradiction from the boundedness in (i), so $\dot{\hat{X}}_i(t) = 0_{m \times r}$.
- e) Note that $\hat{Y}_i(t) = A_i\hat{X}_i(t) = A_i\hat{X}_i(0)$ is a constant matrix. Then $\dot{\hat{Y}}_i(t) = 0_{m_i \times r}$.
- f) Due to $(\hat{X}_i(t) - \hat{X}_i^*)B_i = -(\hat{Z}_i(t) - \hat{Z}_i^*)$, and $\dot{\hat{X}}_i(t) = 0_{m \times r}$, $\dot{\hat{Z}}_i(t) = \dot{\hat{Z}}_i^* - (\dot{\hat{X}}_i(0) - \dot{\hat{X}}_i^*)B_i$ is a constant matrix. Because of the boundedness of the solution in (i), $\dot{\hat{Z}}_i(t) = 0_{m \times r_i}$.
- g) Recall that $\dot{\hat{\Theta}}_i(t)$ in (14) is a constant matrix because $\dot{\hat{Y}}_i(t) = 0_{m_i \times r}$, $\dot{\hat{Z}}_i(t) = 0_{m \times r_i}$, $\dot{\hat{W}}_i(t) = 0_{m \times r_i}$. Then $\dot{\hat{\Theta}}_i(t) = 0_{m \times r}$.

To sum up, $\mathcal{M}_l \subset \{P_l(t) : \dot{\hat{X}}_i(t) = 0_{m \times r}, \dot{\hat{Y}}_i(t) = 0_{m_i \times r}, \dot{\hat{Z}}_i(t) = 0_{m \times r_i}, \dot{\hat{W}}_i(t) = 0_{m \times r}, \dot{\hat{\Lambda}}_i(t) = 0_{m \times r}, \dot{\hat{Y}}_i(t) = 0_{m_i \times r}, \dot{\hat{\Theta}}_i(t) = 0_{m \times r}\}$ and any point in \mathcal{M}_l is an equilibrium of (14), which is Lyapunov stable. In view of Lemma 1 and the fact that every solution is bounded, (14) is globally semi-stable. In other words, for any initial conditions $P_l(0)$, with Algorithm 1 the solution $P_l(t)$ converges to a Lyapunov stable equilibrium of (14).

If, for any given initial conditions, the solution of a linear time-invariant system converges, then it converges exponentially [23]. In fact, (14) is a linear time-invariant system, hence, every solution for given initial conditions of system (14) converges to an equilibrium exponentially.

(iii) Evidently, due to Lemma 6, for given initial value $P_l(0)$, every trajectory $(\bar{X}(t), Y(t), Z(t), W(t))$ converges to an equilibrium which is an optimal solution of problem (12). Thus, $X_i^* = X_j^*$, for all $i, j \in \mathcal{V}$ from the equality constraints. In other word, for every agent i , the limit of its estimation $\lim_{t \rightarrow \infty} X_i(t) = X^*$ is a least squares solution of the Sylvester equation (5). \square

As for Theorem 2, we define a similar Lyapunov function $V_e(P_e) := \frac{1}{2} \|P_e - P_e^*\|_F^2$, whose derivative with respect to time t satisfying $\frac{d}{dt}V_e \leq 0$. Then Theorem 2 can be proved by using the analogical analytical method about the proof of Theorem 1. More details are omitted here.

V. REGULARIZATION CASE

The regularization technique, such as Tikhonov regularization and LASSO [30], has been widely applied to statistics and machine learning to deal with practical problems. Similar ideas

can also be applied to the study of matrices, whose regularization also has many applications, such as multi-task learning, matrix completion and multivariate regression. Matrix regularization penalty functions are typically chosen to be convex. Specifically, the l_1 -norm can be selected to enforce sparsity; the $l_{2,1}$ -norm can be selected to enforce structured sparsity in feature selection; and the nuclear norm can be selected to enforce low rank approximation properties [31]–[35].

For the problem (8), we present the following regularization distributed optimization problem with a constant $\alpha > 0$,

$$\begin{aligned} \min_{\bar{X}, Y, Z, W} \quad & \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2 + \alpha \sum_{i=1}^n g(X_i) \\ \text{s.t.} \quad & \sum_{j=1}^n a_{i,j} (X_i - X_j) = 0_{m \times r}, \quad A_i X_i = Y_i, \\ & [Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i - W_j) = 0_{m \times r}, \quad \forall i \in \mathcal{V}. \end{aligned} \quad (20)$$

Then we introduce $\Lambda_i \in \mathbb{R}^{m \times r}$, $\Upsilon_i \in \mathbb{R}^{m_i \times r}$ and $\Theta_i \in \mathbb{R}^{m \times r}$, $i \in \mathcal{V}$ as the Lagrange multipliers associated with these equality constraints and $\Lambda = \text{col}\{\Lambda_1, \dots, \Lambda_n\}$, $\Upsilon = \text{col}\{\Upsilon_1, \dots, \Upsilon_n\}$, $\Theta = \text{col}\{\Theta_1, \dots, \Theta_n\}$. Let $(\bar{X}^*, Y^*, Z^*, W^*)$ and $(\Lambda^*, \Upsilon^*, \Theta^*)$ be a pair of primal and dual optimal points. They satisfy the KKT condition, which leads to a necessary and sufficient condition for optimality, for all $i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} 0_{m \times r} = \sum_{j=1}^n a_{i,j} (X_i^* - X_j^*), \\ 0_{m_i \times r} = A_i X_i^* - Y_i^*, \\ 0_{m \times r} = [Y_i^*]_{\mathcal{R}} - [Z_i^*]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i^* - W_j^*), \\ 0_{m \times r} \in -(X_i^* B_i - C_i + Z_i^*) B_i^T - \alpha \partial g(X_i) \\ \quad - \sum_{j=1}^n a_{i,j} (\Lambda_i^* - \Lambda_j^*) - A_i^T \Upsilon_i^*, \\ 0_{m_i \times r} = \Upsilon_i^* - [I_{m_i}]_{\mathcal{C}} \Theta_i^* + A_i X_i^* - Y_i^*, \\ 0_{m \times r_i} = -(X_i^* B_i - C_i + Z_i^*) + \Theta_i^* [I_{r_i}]_{\mathcal{R}}, \\ 0_{m \times r} = \sum_{j=1}^n a_{i,j} (\Theta_i^* - \Theta_j^*). \end{array} \right. \quad (21)$$

The fourth formula in (21) means that there exists $h_i^* \in \partial g(X_i^*)$ such that

$$\begin{aligned} 0_{m \times r} = & -(X_i^* B_i - C_i + Z_i^*) B_i^T - \alpha h_i^* \\ & - \sum_{j=1}^n a_{i,j} (\Lambda_i^* - \Lambda_j^*) - A_i^T \Upsilon_i^*. \end{aligned} \quad (22)$$

Note that each agent i in the network can get access to the information about A_i, B_i, C_i and has its own state $(\dot{X}_i, X_i, Y_i, Z_i, W_i, \Lambda_i, \Upsilon_i, \Theta_i)$. Through the exchange of information, agent i

can receive partial state information $(\dot{X}_j, X_j, W_j, \Theta_j)$, $j \in \mathcal{N}_i$ from its neighbor agents. Then we propose Algorithm 3 as the distributed algorithm for each agent i in the regularization case, where $(X_i(t), Y_i(t), Z_i(t), W_i(t))$ and $(\Lambda_i(t), \Upsilon_i(t), \Theta_i(t))$ are estimates of solutions to problem

Algorithm 3 Algorithm for the regularization case

$X_i(0) \in \mathbb{R}^{m \times r}, Y_i(0) \in \mathbb{R}^{m_i \times r}, Z_i(0) \in \mathbb{R}^{m \times r_i}, W_i(0) \in \mathbb{R}^{m \times r}, \Lambda_i(0) \in \mathbb{R}^{m \times r}, \Upsilon_i(0) \in \mathbb{R}^{m_i \times r}, \Theta_i(0) \in \mathbb{R}^{m \times r}, i \in \mathcal{V}$,

$$\left\{ \begin{array}{l} \dot{X}_i(t) \in - (X_i(t)B_i - C_i + Z_i(t)) B_i^T - \alpha \partial g(X_i) - A_i^T \Upsilon_i(t) \\ \quad - \sum_{j=1}^n a_{i,j} (\Lambda_i(t) - \Lambda_j(t)) - A_i^T (A_i X_i(t) - Y_i(t)) \\ \quad - \sum_{j=1}^n a_{i,j} (X_i(t) - X_j(t)), \\ \dot{Y}_i(t) = \Upsilon_i(t) - [L_{m_i}]_C \Theta_i(t) + A_i (X_i(t) - \dot{X}_i(t)) - Y_i(t), \\ \dot{Z}_i(t) = - \left((X_i(t) - \dot{X}_i(t)) B_i - C_i + Z_i(t) \right) + \Theta_i(t) [L_{r_i}]_R, \\ \dot{W}_i(t) = \sum_{j=1}^n a_{i,j} (\Theta_i(t) - \Theta_j(t)), \\ \dot{\Lambda}_i(t) = \sum_{j=1}^n a_{i,j} (X_i(t) - X_j(t) + \dot{X}_i(t) - \dot{X}_j(t)), \\ \dot{\Upsilon}_i(t) = A_i (X_i(t) + \dot{X}_i(t)) - Y_i(t), \\ \dot{\Theta}_i(t) = [Y_i(t)]_R - [Z_i(t)]_C - \sum_{j=1}^n a_{i,j} (W_i(t) - W_j(t)) \\ \quad - \sum_{j=1}^n a_{i,j} (\Theta_i(t) - \Theta_j(t)). \end{array} \right. \quad (23)$$

(20) and estimates of Lagrange multipliers by agent i at time t , respectively. Algorithm 3 is inspired by the saddle-point dynamics and derivative feedback. Derivative feedback is a useful method to make the algorithm convergent [36], [37].

Lemma 8: Under Assumption 1, let $P_r^* = \text{col}\{\bar{X}^*, Y^*, Z^*, W^*, \Lambda^*, \Upsilon^*, \Theta^*\}$. Then $(\bar{X}^*, Y^*, Z^*, W^*)$ is an optimal solution of problem (20) if and only if there exist $\Lambda^* \in \mathbb{R}^{mn \times r}$, $\Upsilon^* \in \mathbb{R}^{m \times r}$, and $\Theta^* \in \mathbb{R}^{mn \times r}$, such that P_r^* is an equilibrium of system (23).

Obviously, Lemma 8 can be easily proved by the KKT condition (21).

Denote $\dot{X}_i(t) \in \mathcal{F}_{X_i}$, $P_r = \text{col}\{\bar{X}, Y, Z, W, \Lambda, \Upsilon, \Theta\}$ and $\dot{P}_r \in \mathcal{F}_{P_r}(P_r(t))$. Let P_r^* be an equilibrium of system (23). The next result shows main convergence properties of Algorithm 3.

Theorem 3: Under Assumption 1, the following hold with Algorithm 3.

- (i) Every equilibrium of the system (23) is Lyapunov stable and every solution is bounded;
- (ii) The system (23) is globally semi-stable.
- (iii) For every agent i , the limit of its estimation $\lim_{t \rightarrow \infty} X_i(t) = X^*$ is an optimal solution of the problem (8) based on the Sylvester equation (5).

The proof of Theorem 3 follows some statements. According to Lemma 2 and the convexity of g , $\frac{d}{dt}g(X_i(t)) \in \tilde{\mathcal{L}}_{\mathcal{F}_{X_i}}g$, that is $\frac{d}{dt}g(X_i(t)) = \langle h_i, \dot{X}_i \rangle_F$, for all $h_i \in \partial g(X_i)$. Because $g(\cdot)$ is a

convex function, for any $h_i^* \in \partial g(X_i^*)$ and any $h_i \in \partial g(X_i)$, there holds

$$\langle h_i - h_i^*, X_i - X_i^* \rangle_F \geq 0, \quad \forall i \in \mathcal{V}. \quad (24)$$

Define a scalar function

$$\begin{aligned} V_1 := & \sum_{i=1}^r (\alpha g(X_i) - \alpha g(X_i^*) - \alpha \langle h_i^*, X_i - X_i^* \rangle_F) \\ & + \frac{1}{4} \sum_{i=1}^r \sum_{j=1}^r a_{i,j} \|X_i - X_j\|_F^2, \end{aligned}$$

where h_i^* satisfies (22). Then V_1 is nonnegative. Define a scalar function $V_r(P_r) := V_1 + V_2$, where

$$\begin{aligned} V_2 := & \frac{1}{2} \|P_r - P_r^*\|_F^2 + \frac{1}{2} \sum_{i=1}^r \|A_i(X_i - X_i^*)\|_F^2 \\ & + \frac{1}{2} \sum_{i=1}^r \|(X_i - X_i^*)B_i\|_F^2. \end{aligned}$$

Then the function V_r has the following properties:

- (a) $V_r(P_r^*) = 0, V_r(P_r) > 0$ for all $P_r \neq P_r^*$; $V \rightarrow \infty$ as $\|P_r - P_r^*\|_F \rightarrow \infty$.
- (b) $\frac{d}{dt} V_r(P_r(t)) \in \tilde{\mathcal{L}}_{\mathcal{F}_{P_r}} V_r(P_r) = \{\nabla V^T y : y \in \mathcal{F}_{P_r}(P_r)\}$.

Next, it is time to prove Theorem 3.

The Proof of Theorem 3: (i) Note that $P_r^* = \text{col}\{\bar{X}^*, Y^*, Z^*, W^*, \Lambda^*, \Upsilon^*, \Theta^*\}$ is an equilibrium of (23). First, we calculate $\frac{d}{dt} V_1(P_r(t))$, for $h_i \in \partial g(X_i)$, denote $\dot{X}_i = -(X_i B_i - F_i + Z_i) B_i^T - \alpha H_i(X_i) - \sum_{j=1}^n a_{i,j} (\Lambda_i - \Lambda_j) - A_i^T \Upsilon_i - \sum_{j=1}^n a_{i,j} (X_i - X_j) - A_i^T (A_i X_i - Y_i)$, then

$$\begin{aligned} \frac{d}{dt} V_1 &= \sum_{i=1}^r \langle \alpha H_i(X_i) - \alpha h_i^* + \sum_{j=1}^r a_{i,j} (X_i - X_j), \dot{X}_i \rangle_F \\ &= - \sum_{i=1}^r \langle \dot{X}_i, \dot{X}_i \rangle_F - \sum_{i=1}^r \langle (X_i B_i - F_i + Z_i) B_i^T + \sum_{j=1}^r a_{i,j} (\Lambda_i \\ &\quad - \Lambda_j) + A_i^T \Upsilon_i + A_i^T (A_i X_i - Y_i) + \alpha h_i^*, \dot{X}_i \rangle_F. \end{aligned}$$

Denote the second term in $\frac{d}{dt} V_1$ by β . With the help of the definition of h_i^* , β can be rewritten as $\beta = - \sum_{i=1}^n \langle (X_i - X_i^*) B_i B_i^T, \dot{X}_i \rangle_F - \sum_{i=1}^n \langle (Z_i - Z_i^*) B_i^T, \dot{X}_i \rangle_F - \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \langle \Lambda_i - \Lambda_i^*, \dot{X}_i - \dot{X}_j \rangle_F - \sum_{i=1}^n \langle A_i^T (\Upsilon_i - \Upsilon_i^*), \dot{X}_i \rangle_F - \sum_{i=1}^n \langle A_i^T (A_i (X_i - X_i^*)), \dot{X}_i \rangle_F + \sum_{i=1}^n \langle A_i^T (Y_i - Y_i^*), \dot{X}_i \rangle_F$.

Next, we calculate $\frac{d}{dt}V_2$. Then, from $V_r = V_1 + V_2$, we have the derivative of V_r with respect to time t ,

$$\begin{aligned} \frac{d}{dt}V_r &= - \sum_{i=1}^n \|(X_i - X_i^*)B_i + (Z_i - Z_i^*)\|_F^2 \\ &\quad - \sum_{i=1}^n \|A_i(X_i - X_i^*) - (Y_i - Y_i^*)\|_F^2 \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \|X_i - X_j\|_F^2 - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_{i,j} \|\Theta_i - \Theta_j\|_F^2 \\ &\quad - \sum_{i=1}^n \|\dot{X}_i\|_F^2 - \alpha \langle X_i - X_i^*, h_i - h_i^* \rangle_F \leq 0. \end{aligned}$$

Thus, every element in $\tilde{\mathcal{L}}_{\mathcal{F}_{P_r}} V_r$ is nonpositive; that is $\max \tilde{\mathcal{L}}_{\mathcal{F}_{P_r}} V \leq 0$. As a result, from the properties of V_r and Lemma 3, the equilibrium P_r^* is Lyapunov stable. Moreover, because V_r is radially unbounded, each solution with given initial conditions of system (23) is bounded for all $t \geq 0$.

(ii) Define the set $\mathcal{R}_r = \{P_r : \frac{d}{dt}V_r = 0\} = \{P_r : \dot{X}_i = 0, X_i = X_j, \Theta_i = \Theta_j, (X_i - X_i^*)B_i = -(Z_i - Z_i^*), A_i(X_i - X_i^*) = Y_i - Y_i^*, \langle X_i - X_i^*, h_i - h_i^* \rangle_F = 0\} \subset \{P_r : \dot{X}_i = 0, X_i = X_j, \Theta_i = \Theta_j, (X_i - X_i^*)B_i = -(Z_i - Z_i^*), A_i(X_i - X_i^*) = Y_i - Y_i^*\}$. It follows from the invariant principle of the nonsmooth system [27], [38] that $P_r(t) = \text{col}\{\bar{X}(t), Y(t), Z(t), W(t), \Lambda(t), \Upsilon(t), \Theta(t)\} \rightarrow \mathcal{M}_r$, as $t \rightarrow \infty$ and \mathcal{M}_r is the largest weakly positively invariant subset of $\bar{\mathcal{R}}_r$.

Assume that $\hat{P}_r(t) \in \mathcal{M}_r$ for all $t \geq 0$ is a trajectory of (23). Then we have

- a) $\dot{X}_i(t) = 0_{m \times r}$.
- b) $\dot{\Lambda}_i(t) = 0_{m \times r}$, because $\hat{X}_i(t) = \hat{X}_j(t)$ and $\dot{X}_i(t) = 0_{m \times r}$.
- c) $\dot{W}_i(t) = 0_{m \times r}$, because $\hat{\Theta}_i(t) = \hat{\Theta}_j(t)$.
- d) Since $\dot{X}_i(t) = 0$, and $A_i \hat{X}_i(t) = \hat{Y}_i(t)$ due to $A_i(\hat{X}_i(t) - \hat{X}_i^*) = \hat{Y}_i(t) - \hat{Y}_i^*$ and $A_i \hat{X}_i^* = \hat{Y}_i^*$, there holds $\hat{Y}_i(t) = 0_{m_i \times r}$.
- e) Note that $\hat{Y}_i(t) = A_i \hat{X}_i(t) = A_i \hat{X}_i(0)$ is a constant matrix. $\dot{Y}_i(t) = 0_{m_i \times r}$.
- f) Due to $(\hat{X}_i(t) - \hat{X}_i^*)B_i = -(\hat{Z}_i(t) - \hat{Z}_i^*)$, and $\dot{X}_i(t) = 0_{m \times r}$, $\hat{Z}_i(t) = \hat{Z}_i^* - (\hat{X}_i(0) - \hat{X}_i^*)B_i$ is a constant matrix. Because of the boundedness of the solution in (i), $\dot{Z}_i(t) = 0_{m \times r_i}$.
- g) Recall that $\hat{\Theta}_i(t)$ in (14) is a constant matrix because $\dot{Y}_i(t) = 0_{m_i \times r}$, $\dot{Z}_i(t) = 0_{m \times r_i}$, $\dot{W}_i(t) = 0_{m \times r_i}$. Then $\dot{\Theta}_i(t) = 0_{m \times r}$.

To sum up, $\mathcal{M}_r \subset \{P_r(t) : \dot{X}_i(t) = 0_{m \times r}, \dot{Y}_i(t) = 0_{m_i \times r}, \dot{Z}_i(t) = 0_{m \times r_i}, \dot{W}_i(t) = 0_{m \times r_i}, \dot{\Lambda}_i(t) = 0_{m \times r}, \dot{\Upsilon}_i(t) = 0_{m_i \times r}, \dot{\Theta}_i(t) = 0_{m \times r}\}$. That is, any point in \mathcal{M}_r is an equilibrium of (23), which

is Lyapunov stable. In view of Lemma 4 and the fact that every solution is bounded, this system is globally semi-stable. In other words, for any initial value $P_r(0)$, with Algorithm 3 the solution $P_r(t)$ converges to a Lyapunov stable equilibrium of (23).

(iii) Evidently, due to Lemma 8, for given initial value $P_r(0)$, every trajectory $(\bar{X}(t), Y(t), Z(t), W(t))$ converges to an equilibrium, which is an optimal solution of problem (20). Thus, $X_i^* = X_j^*$ for all $i, j \in \mathcal{V}$. Furthermore, the conclusion follows. \square

VI. DISCUSSIONS AND NUMERICAL SIMULATIONS

In this section, we first give a brief comparison for three cases in Table II, and then give three examples for illustration of the validity of algorithms and theorems. Denote

$$\begin{aligned} F_1 &= \frac{1}{2} \sum_{i=1}^n \|X_i B_i - C_i + Z_i\|_F^2, \\ F_2 &= \frac{1}{2} \sum_{i=1}^n \|A_i X_i - Y_i\|_F^2, \\ F_3 &= \frac{1}{2} \sum_{i=1}^n \langle \sum_{j=1}^n a_{i,j} (X_i - X_j), X_i \rangle_F, \\ I : \sum_{j=1}^n a_{i,j} (X_i - X_j) &= 0_{m \times r}, \quad II : A_i X_i = Y_i, \\ III : [Y_i]_{\mathcal{R}} - [Z_i]_{\mathcal{C}} - \sum_{j=1}^n a_{i,j} (W_i - W_j) &= 0_{m \times r}. \end{aligned}$$

TABLE II
THE COMPARISON OF THE THREE CASES

| | Least Squares Case | Exact Solution Case | Regularization Case |
|----------------------|--|--|------------------------------------|
| Objective Function | F_1 | $F_1 + F_2 + F_3$ | $F_1 + \alpha \sum_{i=1}^n g(X_i)$ |
| Constraints | I, II, III | III | I, II, III |
| Primal Variables | X_i, Y_i, Z_i, W_i | X_i, Y_i, Z_i, W_i | X_i, Y_i, Z_i, W_i |
| Dual Variables | $\Lambda_i, \Upsilon_i, \Theta_i$ | Θ_i | $\Lambda_i, \Upsilon_i, \Theta_i$ |
| Derivative Feedback | - | - | \dot{X}_i |
| Exchange Status | X_i, W_i, Θ_i | X_i, W_i, Θ_i | $\dot{X}_i, X_i, W_i, \Theta_i$ |
| Convergence of X_i | $X_i \xrightarrow{\text{exponentially}} X^*$ | $X_i \xrightarrow{\text{exponentially}} X^*$ | $X_i \rightarrow X^*$ |

Example 1: Consider a 8-node network in a label set $\mathcal{V} = \{1, \dots, 8\}$, whose interactions form an undirected complete graph (which has $\mathcal{E} = \{(i, j), \forall i, j \in \mathcal{V}\}$). There is a Sylvester equation that has an exact solution: $AX + XB = C, A, B, C \in \mathbb{R}^{8 \times 8}$. Select a random initial value and

appropriate $a_{i,j} > 0$, and plot the estimations for the first row of X_i over time in Fig. 1 using Algorithm 1. As a reference, the first row of the exact solution X^* is

$$X^*(1, :) = [0.5183, 0.9633, 0.2613, 0.1600, \\ 0.6670, 0.2552, 0.1219, 0.4817].$$

Then Fig. 1 shows that the trajectories of the first row of X_i converge to $X^*(1, :)$. In fact, we can also use Algorithm 2 to solve this equation. The results show that both two algorithms derive the solutions that converge to the exact solution. Denote the error of all estimations $E(t) = \frac{1}{n} \sum_{i=1}^n \|X_i - X^*\|_F^2$. Then we plot the evolution of $\log(E(t))$ in Fig. 2, using Algorithms 1 and 2, respectively. Fig. 2 verifies the exponential convergence of proposed distributed algorithms, which is consistent with the aforementioned theorems.

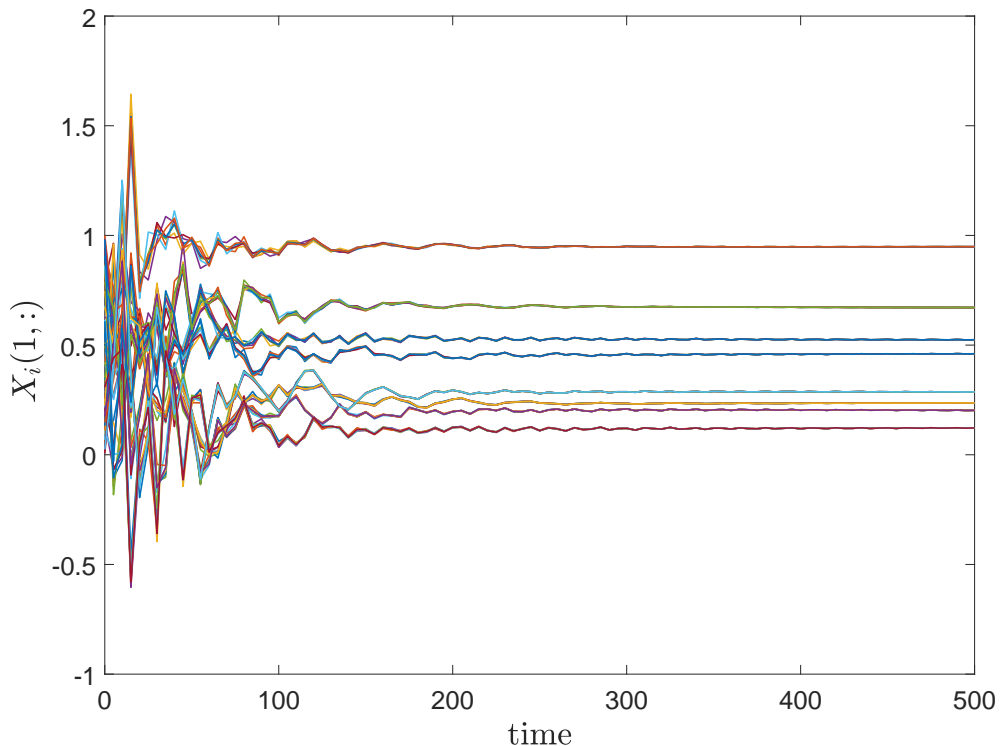


Fig. 1. The estimations of all $X_i(1, :)$ over time using Algorithm 1.

Example 2: Consider a regularization problem for the Sylvester equation $AX + XB = C$, $A, B, C \in \mathbb{R}^{20 \times 20}$ over a 10-node network whose interactions form an undirected complete graph. We assign two rows of A and two columns of B and C to each agent and select the regularization penalty function as $g(X) := |X|_1 = \sum_{i,j} |X_{ij}|$. l_1 -norm of a vector is usually

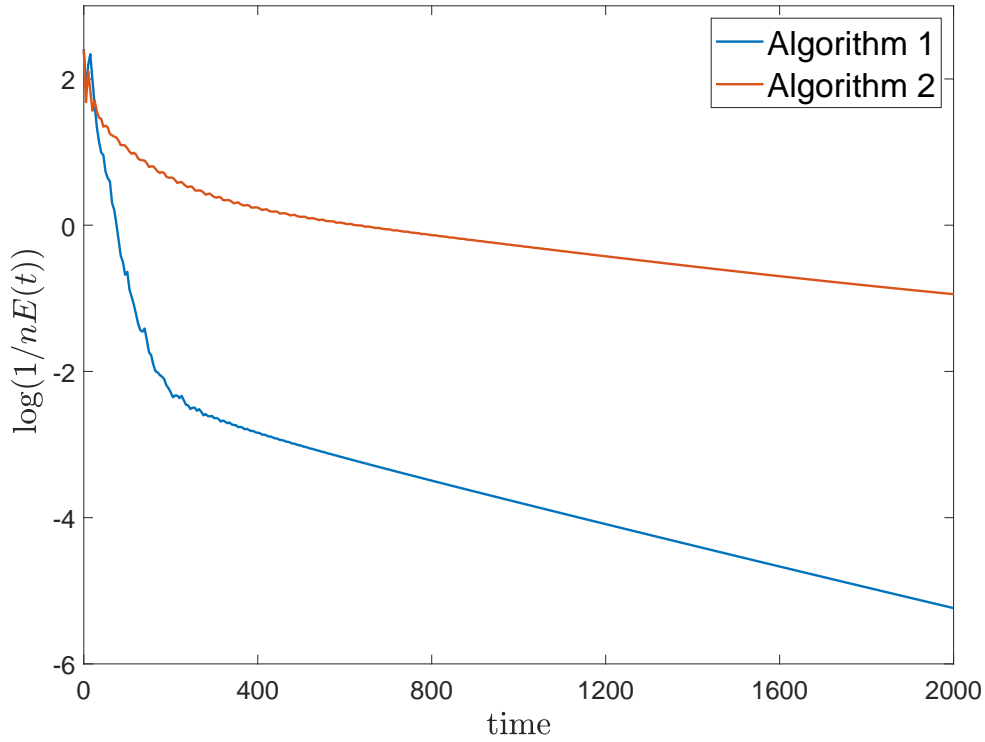


Fig. 2. The evolution of $\log(E(t))$ using Algorithms 1 and 2, respectively.

used to be a regularization term to get a sparse solution; similarly, we denote the sum of the absolute value of each entry of a matrix by the norm $|X|_1$. Choose $h(X)$ as a subgradient of $g(X)$, $h(X) \in \partial g(X)$, $X \in \mathbb{R}^{m \times r}$ and it satisfies

$$h(X) \in \mathbb{R}^{m \times r}, \quad [h(X)]_{ij} = \begin{cases} 1, & \text{if } X_{ij} > 0, \\ 0, & \text{if } X_{ij} = 0, \\ -1 & \text{if } X_{ij} < 0. \end{cases}$$

Our aim is to obtain a solution with smaller $|\cdot|_1$. With $\alpha = 1$, we compare the $|\cdot|_1$ in Fig. 3 between the exact solution and the solution trajectories by Algorithm 3, and plot the error of consensus $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \|X_i - X_j\|_F^2$ in Fig. 4. Figs. 3 and 4 show that the solutions for all agents can reach a consensus and $|\cdot|_1$ of the solution has been reduced relatively.

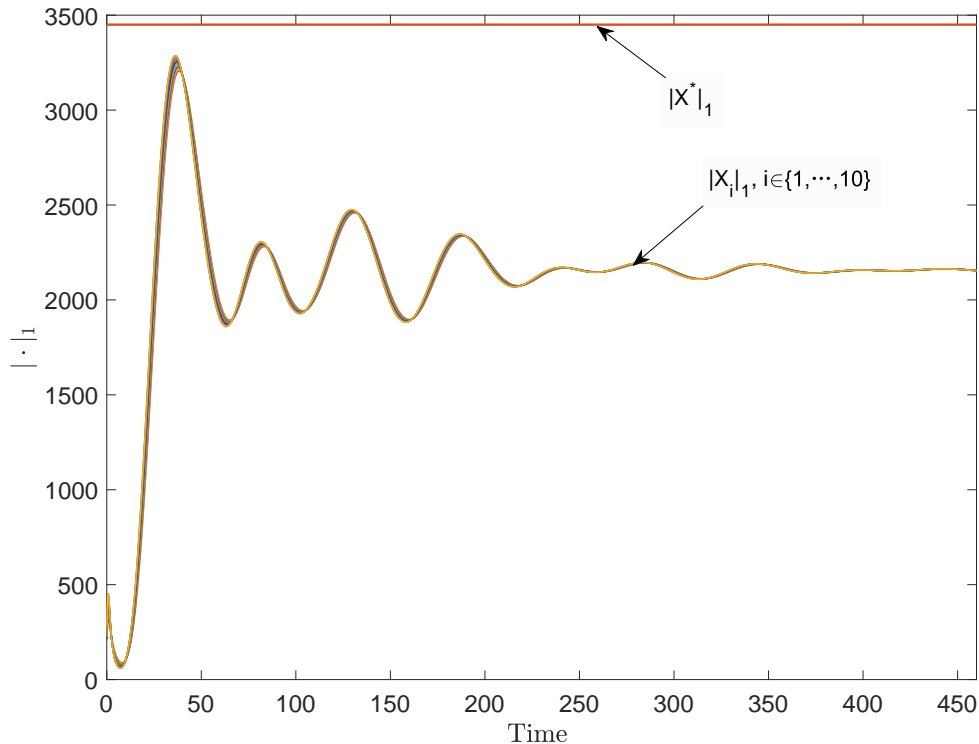


Fig. 3. The evolution of $|X^*|_1$ and all $|X_i(t)|_1, i \in \{1, \dots, 10\}$.

VII. CONCLUSIONS

This paper has proposed distributed algorithms over the multi-agent network with the help of convex optimization for solving the Sylvester equation in different cases, the least squares solution (the exact solution) case and the (nonsmooth) regularization case. In the LRR partition case, each agent only has knowledge of some rows of A and some corresponding columns of B and C , and holds its own state set with exchanging partial state information among its neighbors for achieving the consensus solution. Both theoretical proofs and numerical simulations have been presented to verify the convergence of proposed distributed algorithms.

REFERENCES

- [1] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, p. 48, 2009.
- [2] S. Wang and C. Li, "Distributed robust optimization in networked system," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 2321–2333, 2017.

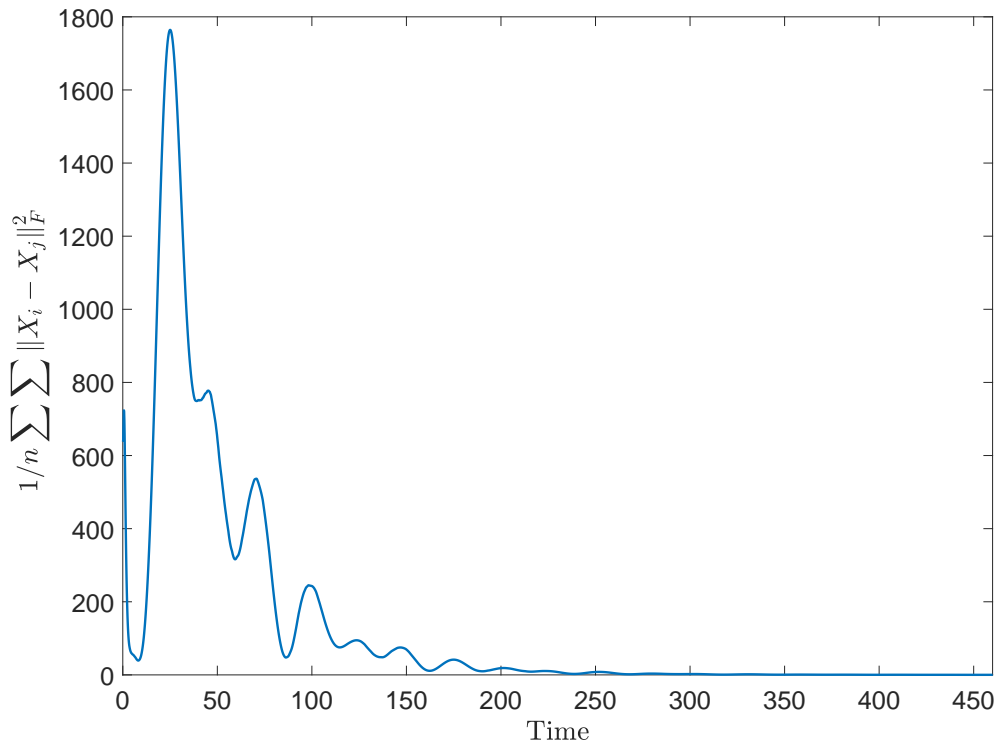


Fig. 4. The error of $1/n \sum_{i=1}^n \sum_{j=1}^n \|X_i - X_j\|_F^2$ over time.

- [3] D. Yuan, D. W. Ho, and S. Xu, “Regularized primal–dual subgradient method for distributed constrained optimization,” *IEEE Trans. Cybern.*, vol. 46, no. 9, pp. 2109–2118, 2016.
- [4] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, 2014.
- [5] S. Jafarizadeh, “Optimizing the diffusion system based on continuous-time consensus algorithm,” *SIAM J. Control Optim.*, vol. 55, no. 3, pp. 2024–2051, 2017.
- [6] Z. Deng, X. Nian, and C. Hu, “Distributed algorithm design for nonsmooth resource allocation problems,” *IEEE Trans. Cybern.*, 2019, DOI: 10.1109/TCYB.2019.2901256.
- [7] B. Datta, *Numerical methods for linear control systems*. Academic Press, 2004, vol. 1.
- [8] R. Bhatia and P. Rosenthal, “How and why to solve the operator equation $AX - XB = Y$,” *Bull. Lond. Math. Soc.*, vol. 29, no. 1, pp. 1–21, 1997.
- [9] V. Simoncini, “Computational methods for linear matrix equations,” *SIAM Rev.*, vol. 58, no. 3, pp. 377–441, 2016.
- [10] M. J. Corless and A. E. Frazho, *Linear Systems and Control: An Operator Perspective*, 1st ed. Boca Raton: CRC Press, 2003.
- [11] J. W. Demmel, “Three methods for refining estimates of invariant subspaces,” *Computing*, vol. 38, pp. 43–57, 1987.
- [12] L. Lv, G. Duan, and B. Zhou, “Parametric pole assignment and robust pole assignment for discrete-time linear periodic systems,” *SIAM J. Control Optim.*, vol. 48, no. 6, pp. 3975–3996, 2010.
- [13] D. C. Sorensen and A. Antoulas, “The Sylvester equation and approximate balanced reduction,” *Linear Algebra Appl.*,

- vol. 351, pp. 671–700, 2002.
- [14] D. Y. Hu and L. Reichel, “Krylov-subspace methods for the Sylvester equation,” *Linear Algebra Appl.*, vol. 172, pp. 283–313, 1992.
- [15] P. Benner, R. C. Li, and N. Truhar, “On the ADI method for Sylvester equations,” *J. Comput. Appl. Math.*, vol. 233, no. 4, pp. 1035–1045, 2009.
- [16] F. Ding and T. Chen, “On iterative solutions of general coupled matrix equations,” *SIAM J. Control Optim.*, vol. 44, no. 6, pp. 2269–2284, 2006.
- [17] R. Granat, B. Kågström, and P. Poromaa, “Parallel scalapack-style algorithms for solving continuous-time sylvester matrix equations,” in *European Conference on Parallel Processing*. Springer, 2003, pp. 800–809.
- [18] G. Shi, B. D. O. Anderson, and U. Helmke, “Network flows that solve linear equations,” *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2659–2674, 2017.
- [19] Y. Liu, C. Lageman, B. D. Anderson, and G. Shi, “An arrow–hurwicz–uzawa type flow as least squares solver for network linear equations,” *Automatica J. IFAC*, vol. 100, pp. 187–193, 2019.
- [20] J. Liu, S. Mou, and A. S. Morse, “Asynchronous distributed algorithms for solving linear algebraic equations,” *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 372–385, 2017.
- [21] X. Wang, S. Mou, and B. D. O. Anderson, “A double-layered framework for distributed coordination in solving linear equations,” 2017, arXiv: 1711.10947.
- [22] K. Cao, X. Zeng, and Y. Hong, “Continuous-time distributed algorithms for solving linear algebraic equation,” in *Control Conference (CCC), 2017 36th Chinese*. IEEE, 2017, pp. 8068–8073.
- [23] X. Zeng, S. Liang, Y. Hong, and J. Chen, “Distributed computation of linear matrix equations: An optimization perspective,” *IEEE Trans. Autom. Control*, 2018, DOI: 10.1109/TAC.2018.2847603.
- [24] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1970.
- [25] Q. Hui, W. Haddad, and S. Bhat, “Semistability, finite-time stability, differential inclusions, and discontinuous dynamical systems having a continuum of equilibria,” *IEEE Trans. Autom. Control*, vol. 54, no. 10, pp. 2465–2470, 2009.
- [26] J. Cortés, “Discontinuous dynamical systems: A tutorial on solutions, nonsmooth analysis, and stability,” *IEEE Control Syst. Mag.*, vol. 28, no. 3, pp. 36–73, 2008.
- [27] A. Bacciotti and F. Ceragioli, “Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions,” *ESAIM Control Optim. Calc. Var.*, vol. 4, pp. 361–376, 1999.
- [28] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [29] J. LaSalle, “Some extensions of Liapunov’s second method,” *IRE Transactions on Circuit Theory*, vol. 7, no. 4, pp. 520–527, Dec. 1960.
- [30] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [31] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for l_1 -minimization with applications to compressed sensing,” *SIAM J. Imaging Sci.*, vol. 1, no. 1, pp. 143–168, 2008.
- [32] F. Nie, H. Huang, X. Cai, and C. H. Ding, “Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization,” in *Advances in neural information processing systems*, 2010, pp. 1813–1821.
- [33] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [34] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

- [35] P. Li, J. Yu, M. Wang, L. Zhang, D. Cai, and X. Li, “Constrained low-rank learning using least squares-based regularization,” *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4250–4262, 2017.
- [36] A. Antipin, “Feedback-controlled saddle gradient processes,” *Autom. Remote Control*, vol. 55, no. 3, pp. 311–320, 1994.
- [37] X. Zeng, P. Yi, Y. Hong, and L. Xie, “Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems,” *SIAM J. Control Optim.*, vol. 56, no. 6, pp. 3973–3993, 2018.
- [38] D. Shevitz and B. Paden, “Lyapunov stability theory of nonsmooth systems,” *IEEE Trans. Autom. Control*, vol. 39, no. 9, pp. 1910–1914, 1994.