SCIENCE CHINA
Information Sciences

• **Supplementary File** •

An effective distributed algorithm for solving linear matrix equations

Songsong CHENG[1], Jinlong LEI[2*], Xianlin ZENG[3], Yuan FAN[1] & Yiguang HONG[2]

[1]*Anhui Engineering Laboratory of Human-Robot Integration System and Intelligent Equipment, School of Electrical Engineering and Automation, Anhui University, Hefei* 230601*, China;*
[2]*Department of Control Science and Engineering, Tongji University, Shanghai 200092, China;*
[3]*Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, 100081, Beijing, China.*

Appendix A Notations and Preliminaries

$\mathbf{0}_n \in \mathbb{R}^n$ ($\mathbf{1}_n \in \mathbb{R}^n$, $\mathbf{0}_{n\times m} \in \mathbb{R}^{n\times m}$) is a vector (vector, matrix) with all elements of 0 (1, 0). $I_n \in \mathbb{R}^{n\times n}$ denotes an identity matrix. For these vectors and matrices, we omit the subscript if it is clear from the context. For any vector $\mathbf{x} = [x_1, \cdots, x_n]^\top \in \mathbb{R}^n$, $\|\mathbf{x}\|_\infty = \max\{|x_1|, \cdots, |x_n|\}$. For any $A = [a_{ij}] \in \mathbb{R}^{n\times m}$, a_{ij} denotes the i-th row and j-th column entry, $A^\top$ means the transpose of matrix A, $\|A\|_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^m |a_{ij}|$, $\mathrm{vec}(A)$ is an augmented column vector stacked by all of the column vectors of A. For $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m}$, we define $\|A\|_{\mathrm{F}} = (\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2)^{\frac{1}{2}}$ and $\langle A, B\rangle_{\mathrm{F}} = \sum_{i=1}^n \sum_{j=1}^m a_{ij} b_{ij}$. diag$\{\cdot\}$ (blkdiag$\{\cdot\}$) takes vector arguments (matrix arguments) and constructs a matrix with these entries (matrix inputs) on the diagonal (diagonals). Define $\mathbb{I}_{m_i} = \mathrm{diag}\{\mathbf{0}^\top_{m_{[i-1]}}, \mathbf{1}^\top_{m_i}, \mathbf{0}^\top_{m-m_{[i]}}\} \in \mathbb{R}^{m\times m}$, $\mathbb{I}_{r_i} = \mathrm{diag}\{\mathbf{0}^\top_{r_{[i-1]}}, \mathbf{1}^\top_{r_i}, \mathbf{0}^\top_{r-r_{[i]}}\} \in \mathbb{R}^{r\times r}$, $\mathcal{I}_{m_i} = [\mathbf{0}^\top_{m_{[i-1]}\times m_i}, I_{m_i}, \mathbf{0}^\top_{(m-m_{[i]})\times m_i}]^\top \in \mathbb{R}^{m\times m_i}$, where $\sum_{i=1}^n m_i = m$, $\sum_{i=1}^n r_i = r$, $m_{[i]} = \sum_{j=1}^i m_j$, and $r_{[i]} = \sum_{j=1}^i r_j$. x^t is the t-th power of x and x_t means the value of x at the t-th iteration.

The following properties are necessary for analyzing the convergence of the proposed algorithm.

Lemma 1. [1] For connected and undirected graph, there exists an orthogonal matrix $V = [V_1, V_2]$ with $V_1 \in \mathbb{R}^{n\times(n-1)}$ and $V_2 = \frac{\mathbf{1}}{\sqrt{n}} \in \mathbb{R}^n$, such that

$$V^\top L V = \begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}, LV = [V_1 S, \mathbf{0}], V^\top L = \begin{bmatrix} S V_1^\top \\ \mathbf{0} \end{bmatrix}, \tag{A1}$$

where $S = \mathrm{diag}\{s_1, \cdots, s_{n-1}\}$ with $0 < s_{n-1} \leqslant \cdots \leqslant s_1$.

Define

$$M_s = \begin{bmatrix} H + \mathbb{L} & \mathbb{V}_1 \mathbb{S} \\ -\mathbb{S}\mathbb{V}_1^\top & \mathbf{0} \end{bmatrix} \tag{A2}$$

where $H = \mathrm{blkdiag}\{H_1^\top H_1, \cdots, H_n^\top H_n\}$.

Lemma 2. [1] For connected and undirected graph, if $\sum_{i=1}^n H_i^\top H_i$ is positive definite, then $-M_s$ is Hurwitz.

Appendix B Proof of Theorem 1

Proof. Consider the following optimization problem

$$\begin{aligned} \min_{X,Y} \quad & \frac{1}{2}\sum_{i=1}^n \left(\|A_{\mathrm{h}i} X_i - Y_{i\mathrm{h}i}\|_{\mathrm{F}}^2 + \|Y_i B_{\mathrm{v}i} - C_{\mathrm{v}i}\|_{\mathrm{F}}^2\right), \\ \text{s.t.} \quad & X_i = X_j, \\ & Y_i = Y_j. \end{aligned} \tag{B1}$$

Clearly, (B1) can be rewritten as

$$\begin{aligned} \min_{X,Y} \quad & \frac{1}{2}\sum_{i=1}^n \left(\|Y_i B \mathbb{I}_{r_i} - C\mathbb{I}_{r_i}\|_{\mathrm{F}}^2 + \|\mathbb{I}_{m_i} A X_i - \mathbb{I}_{m_i} Y_i\|_{\mathrm{F}}^2\right), \\ \text{s.t.} \quad & X_i = X_j, \\ & Y_i = Y_j, \end{aligned} \tag{B2}$$

* Corresponding author (email: leijinlong@tongji.edu.cn)

where $\mathbb{I}_{m_i} = \mathrm{diag}\{\mathbf{0}_{[m_{i-1}]}^\top, \mathbf{1}_{m_i}^\top, \mathbf{0}_{m-[m_i]}^\top\} \in \mathbb{R}^{m\times m}$ and $\mathbb{I}_{r_i} = \mathrm{diag}\{\mathbf{0}_{[r_{i-1}]}^\top, \mathbf{1}_{r_i}^\top, \mathbf{0}_{r-[r_i]}^\top\} \in \mathbb{R}^{r\times r}$.

Then $(\mathbf{1}_n \otimes X_o^*, \mathbf{1}_n \otimes Y_o^*)$ is the optimal solution of (B2) if and only if the following KKT condition holds

$$\begin{cases} (Y_o^* B - C)\mathbb{I}_{r_i} B^\top + \mathbb{I}_{m_i}(Y_o^* - AX_o^*) + \sum_{j=1}^n l_{ij}\Lambda_{2j}^* = \mathbf{0}, & \text{(B3a)} \\ A^\top \mathbb{I}_{m_i}(AX_o^* - Y_o^*) + \sum_{j=1}^n l_{ij}\Lambda_{1j}^* = \mathbf{0}. & \text{(B3b)} \end{cases}$$

Since $\sum_{i=1}^n \mathbb{I}_{m_i} = I_m$, $\sum_{i=1}^n \mathbb{I}_{r_i} = I_r$, and $\sum_{i=1}^n l_{ij} = 0$, taking the summation of (B3a) and (B3b) from $i = 1$ to n, we have

$$\begin{cases} (Y_o^* B - C)B^\top + (Y_o^* - AX_o^*) = \mathbf{0}, & \text{(B4a)} \\ A^\top(AX_o^* - Y_o^*) = \mathbf{0}. & \text{(B4b)} \end{cases}$$

Left multiplying $A^\top$ on the both sides of (B4a) and according to $A^\top Y_o^* = A^\top AX_o^*$ in (B4b), we have

$$A^\top(Y_o^* B - C)B^\top + A^\top(Y_o^* - AX_o^*) = A^\top(AX_o^* B - C)B^\top = \mathbf{0}, \tag{B5}$$

which means that X_o^* is the least squares solution of $AX_oB = C$. □

Appendix C Proof of Theorem 2

As discussed in Theorem 1, the optimal solution $(\mathbf{1}_n \otimes X_o^*, \mathbf{1}_n \otimes Y_o^*)$ to (B1) must satisfy

$$\begin{cases} \mathbf{0} = A_{\mathrm{h}i}^\top(A_{\mathrm{h}i}X_o^* - Y_{o\mathrm{h}i}^*) + \sum_{j=1}^n l_{ij}\Lambda_{1j}^*, \\ \mathbf{0} = (Y_o^* B_{\mathrm{v}i} - C_{\mathrm{v}i})B_{\mathrm{v}i}^\top + \mathbb{I}_{m_i}(Y_o^* - AX_o^*) + \sum_{j=1}^n l_{ij}\Lambda_{2j}^*. \end{cases} \tag{C1}$$

Then the dynamics of Algorithm 1 can be rewritten with the help of the Kronecker product in a vectorial form

$$\begin{cases} \mathbf{x}_{i,t+1} = \mathbf{x}_{i,t} - \alpha\Big[(I_q \otimes A_{\mathrm{h}i}^\top A_{\mathrm{h}i})\tilde{\mathbf{x}}_{i,t} - (I_q \otimes A^\top \mathbb{I}_{m_i})\tilde{\mathbf{y}}_{i,t} + \sum_{j=1}^n l_{ij}(\tilde{\boldsymbol{\lambda}}_{1j,t} + \tilde{\mathbf{x}}_{j,t})\Big], \\ \mathbf{y}_{i,t+1} = \mathbf{y}_{i,t} - \alpha\Big[(B_{\mathrm{v}i}B_{\mathrm{v}i}^\top \otimes I_m + I_q \otimes \mathbb{I}_{m_i})\tilde{\mathbf{y}}_{i,t} - I_q \otimes \mathbb{I}_{m_i}A\tilde{\mathbf{x}}_{i,t} + \sum_{j=1}^n l_{ij}(\tilde{\boldsymbol{\lambda}}_{2j,t} + \tilde{\mathbf{y}}_{j,t})\Big], \\ \boldsymbol{\lambda}_{1i,t+1} = \boldsymbol{\lambda}_{1i,t} + \alpha\sum_{j=1}^n l_{ij}\tilde{\mathbf{x}}_{j,t}, \\ \boldsymbol{\lambda}_{2i,t+1} = \boldsymbol{\lambda}_{2i,t} + \alpha\sum_{j=1}^n l_{ij}\tilde{\mathbf{y}}_{j,t}, \end{cases} \tag{C2}$$

where $\tilde{X}_{i,t} = X_{i,t} - X_o^*$, $\tilde{Y}_{i,t} = Y_{i,t} - Y_o^*$, $\tilde{\Lambda}_{ji,t} = \Lambda_{ji,t} - \Lambda_{ji}^*$ with $j = 1, 2$ and $\mathbf{x}_{i,t}$, $\mathbf{y}_{i,t}$, $\boldsymbol{\lambda}_{1i,t}$, $\boldsymbol{\lambda}_{2i,t}$, $\tilde{\mathbf{x}}_{i,t}$, $\tilde{\mathbf{y}}_{i,t}$, $\tilde{\boldsymbol{\lambda}}_{1i,t}$, and $\tilde{\boldsymbol{\lambda}}_{2i,t}$ are vectors stacked by the column vectors of matrices $X_{i,t}$, $Y_{i,t}$, $\Lambda_{1i,t}$, $\Lambda_{2i,t}$, $\tilde{X}_{i,t}$, $\tilde{Y}_{i,t}$, $\tilde{\Lambda}_{1i,t}$, and $\tilde{\Lambda}_{2i,t}$, respectively. Taking $\mathbf{w}_{i,t} = [\mathbf{x}_{i,t}^\top, \mathbf{y}_{i,t}^\top]^\top$ and $\boldsymbol{\lambda}_{i,t} = [\boldsymbol{\lambda}_{1i,t}^\top, \boldsymbol{\lambda}_{2i,t}^\top]^\top$ yields

$$\begin{cases} \tilde{\mathbf{w}}_{t+1} = \tilde{\mathbf{w}}_t - \alpha\big[H\tilde{\mathbf{w}}_t + \mathbb{L}(\tilde{\boldsymbol{\lambda}}_t + \tilde{\mathbf{w}}_t)\big], & \text{(C3a)} \\ \tilde{\boldsymbol{\lambda}}_{t+1} = \tilde{\boldsymbol{\lambda}}_t + \alpha\mathbb{L}\tilde{\mathbf{w}}_t, & \text{(C3b)} \end{cases}$$

where $\mathbb{L} = L \otimes I_{(p+m)q}$, $H = \mathrm{blkdiag}\{H_1^\top H_1, \cdots, H_n^\top H_n\}$, and

$$H_i = \begin{bmatrix} I_q \otimes (\mathbb{I}_{m_i}A) & -I_q \otimes \mathbb{I}_{m_i} \\ \mathbf{0} & (\mathbb{I}_{r_i}B^\top) \otimes I_m \end{bmatrix}.$$

Left multiplying $\mathbb{V}^\top = V^\top \otimes I_{(p+m)q}$ on the both sides of (C3b) and defining $\acute{\boldsymbol{\lambda}}_t = [\acute{\boldsymbol{\lambda}}_{1,t}^\top, \acute{\boldsymbol{\lambda}}_{2,t}^\top]^\top = \mathbb{V}^\top\tilde{\boldsymbol{\lambda}}_t$ with $\acute{\boldsymbol{\lambda}}_{1,t} \in \mathbb{R}^{(n-1)(p+m)q}$ and $\acute{\boldsymbol{\lambda}}_{2,t} \in \mathbb{R}^{(p+m)q}$, we attain

$$\begin{bmatrix} \acute{\boldsymbol{\lambda}}_{1,t+1} \\ \acute{\boldsymbol{\lambda}}_{2,t+1} \end{bmatrix} = \begin{bmatrix} \acute{\boldsymbol{\lambda}}_{1,t} \\ \acute{\boldsymbol{\lambda}}_{2,t} \end{bmatrix} + \alpha \begin{bmatrix} \mathbb{S}\mathbb{V}_1^\top \\ \mathbf{0} \end{bmatrix} \tilde{\mathbf{w}}_t, \tag{C4}$$

where $\mathbb{S} = S \otimes I_{(p+m)q}$ and $\mathbb{V}_1 = V_1 \otimes I_{(p+m)q}$ with V, V_1, and S was defined in Lemma 1. Therefore, we rewrite the dynamics of (C3a) and (C3b)

$$\begin{bmatrix} \tilde{\mathbf{w}}_{t+1} \\ \acute{\boldsymbol{\lambda}}_{1,t+1} \end{bmatrix} = \underbrace{\left(I - \alpha \underbrace{\begin{bmatrix} H + \mathbb{L} & \mathbb{V}_1\mathbb{S} \\ -\mathbb{S}\mathbb{V}_1^\top & \mathbf{0} \end{bmatrix}}_{M_s} \right)}_{M} \begin{bmatrix} \tilde{\mathbf{w}}_t \\ \acute{\boldsymbol{\lambda}}_{1,t} \end{bmatrix}, \tag{C5}$$

where $M \in \mathbb{R}^{\bar{n}\times\bar{n}}$ with $\bar{n} = (2n-1)(p+m)q$. According to Lemma 2, $H + \mathbb{L} \succ 0$. Moreover, noting that $-\mathbb{S}\mathbb{V}_1^\top$ is of full row rank, we have that $-M_s$ is Hurwitz.

Besides, to determine the eigenvalues of M_s, we need to solve the following equation

$$\det(M_s - \sigma_s I) = -\sigma_s \det\big(H + \mathbb{L} - \frac{\mathbb{V}_1 \mathbb{S}\mathbb{S}\mathbb{V}_1^\top}{\sigma_s} - \sigma_s I\big) = 0, \tag{C6}$$

where the first equality follows from the Schur's determinantal formula. Since $-M_s$ is Hurwitz, we have $\mathrm{Re}(\sigma_s) > 0$, namely, $\det(-\sigma_s I) < 0$. Therefore, we just need to consider $\det\big(H + \mathbb{L} - \sigma_s I - \mathbb{V}_1 \mathbb{S}\mathbb{S}\mathbb{V}_1^\top/\sigma_s\big) = 0$. Because $\mathbb{V}_1 \mathbb{S}\mathbb{S}\mathbb{V}_1^\top$ is semi-positive definite, we have $\sigma_s \leqslant (h_{\mathrm{m}} + s_1)$. Therefore, if $\alpha < 1/(h_{\mathrm{m}} + s_1)$, then M is stable. Moreover, let σ_1 being the least eigenvalue of M_s, $X_{i,t}$ generated by Algorithm 1 linearly converges to the least squares solution of X_o^* as follows.

$$\|X_{i,t} - X_o^*\| \leqslant \gamma^t \|X_{i,0} - X_o^*\| \tag{C7}$$

where $\gamma = 1 - \alpha|\sigma_1|$ and σ_1 is the smallest eigenvalue of H.

Appendix D Illustrative Examples

Consider the linear matrix equation $AX_oB = C$ with dimensions $m_1 = \cdots = m_5 = r_1 = \cdots = r_5 = 1$ and $p = q = 5$. All entries of A, B, and C are randomly chosen from a uniform distribution on $(-5, 5)$ with $\mathrm{rank}(A) = \mathrm{rank}(B) = 5$. We achieve the solution with a multi-agent system containing 5 agents, connected by an undirected circle graph. We utilized the following methods to solve this equation.

1) Method 1: the proposed algorithm;
2) Method 2: the primal-dual algorithm based on the optimization model in [2];
3) Method 3: the ADMM algorithm based on the optimization model in [2];

The stepsizes of those three methods are designed as 0.01, 0.01, and 0.02, respectively. Both of those three methods are run in Personal Computer (PC) MacBook Pro with 2.7 GHz 4-core Intel Core i7 CPU and 16GB RAM..

The simulation result is shown in Fig. D1, where $e_{i,t} = \|X^i_{1,t} - X_o^*\|_{\mathrm{F}} / \|X^i_{1,0} - X_o^*\|_{\mathrm{F}}$ is the relative computational error of method i. Moreover, we summarize iteration time at the machine for some certain number of iteration of both three methods in Table D1, where "Alg", "Tim", and "Num" mean "the **alg**orithm of methods 1, 2, and 3", "the iteration **tim**e in PC" and "the **num**ber of iteration".

As shown in Fig. D1 and Table D1, Algorithm 1 distributedly achieves the least squares solution of $AX_oB = C$ with a faster convergence rate than those of the primal-dual and ADMM algorithms based on the optimization model in [2].

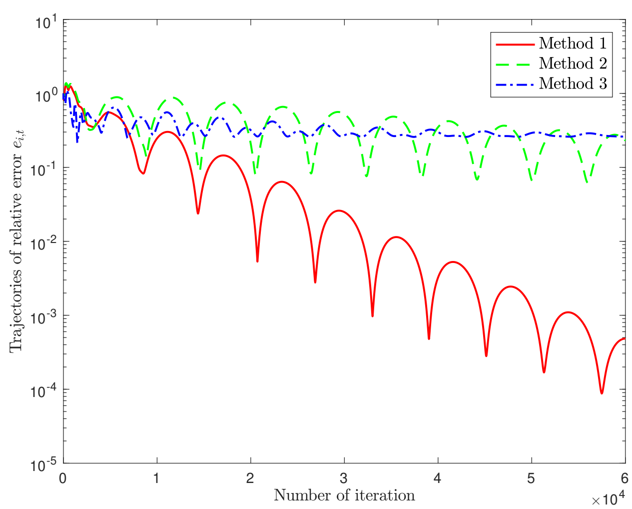

Figure D1 Trajectories of relative errors $e_{i,t}, \forall i = 1, 2, 3$.

Table D1 The actual iteration time (sec) comparison in PC Appendix Example

Tim / Num / Alg	6000	12000	18000	24000	30000	48000	60000
Method 1	0.240	0.439	0.642	0.843	1.044	1.447	2.065
Method 2	0.239	0.468	0.690	0.914	1.147	1.596	2.266
Method 3	1.956	3.830	5.726	7.612	9.484	13.279	18.986

References

1 Lei J, Chen H, Fang H. Asymptotic properties of primal-dual algorithm for distributed stochastic optimization over random networks with imperfect communications. SIAM J Control Opt, 2018, 56(3): 2159-2188.
2 Zeng X, Liang S, Hong Y, et al. Distributed computation of linear matrix equations: An optimization perspective. IEEE Trans Automat Contr, 2018, 64(5): 1858-1873.