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

A Discrete-time Distributed Algorithm for Minimum l_1-Norm Solution of an Under-determined Linear Equation Set ⋆

Xuan Wang * **Shaoshuai Mou** * **Brian D. O. Anderson** **

* *School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47906 USA (e-mail: wang3156@purdue.edu, mous@purdue.edu)*
** *Hangzhou Dianzi University, Hangzhou, China, and The Australian National University (e-mail: Brian.Anderson@anu.edu.au)*

Abstract: This paper proposes a discrete-time, distributed algorithm for multi-agent networks to achieve the minimum l_1-norm solution to a group of linear equations known to possess a family of solutions. We assume each agent in the network knows only one equation and can communicate with only its neighbors. The algorithm is developed based on a combination of the projection-consensus idea and the sub-gradient descent method. Given the underlying network graph to be directed and strongly connected, we prove that the algorithm enables all agents to achieve a common minimum l_1-norm solution. The major difficulty to be dealt with is the non-smooth nature of the norm and the lack of strict convexity of the associated relevant performance index.

Keywords: Distributed Algorithms; Minimum l_1-Norm Solutions.

1. INTRODUCTION

Distributed algorithms proposed by Mou et al. (2015); Shi et al. (2016); Wang et al. (2019b) solve linear equations via multi-agent networks, in which each agent knows one private equation and can only communicate with its nearby neighbors. The key idea of these distributed algorithms is a so-called "agreement principle" Mou and Morse (2013) or "projection-consensus flow" Shi et al. (2016) , in which each agent limits the update of its state to satisfy its own equation while trying to reach a consensus with its nearby neighbors' states. Various extensions have been made: these include elimination of the initialization step Wang et al. (2019a), reduction of state vector size Mou et al. (2016), computing solutions with the minimum Euclidean norm Wang et al. (2017, 2018) and achieving least-squares solutions of an over-determined equation set Wang et al. (2019c). These algorithms are however not applicable to achieve a sparse (minimum l_0-norm) solution of an under-determined equation set, which is of particular interest in many engineering applications including earthquake location detection Shearer (1997), analysis of statistical data Dodge (2012), solving biomagnetic inverse problems Beucker and Schlitt (1996), compressive sensing Baron et al. (2009), and so on. Challenged by the fact that the l_0-norm minimization problem is NP-hard Ge et al. (2011), researchers usually turn to achieve solutions with minimum l_1-norm instead, for which the function to be minimized is convex; the obtained solution, is almost surely unique and equals the sparse (minimum l_0-norm) solution Candes and Tao (2005). Existing results for achieving minimum l_1-norm solutions are usually based on the idea of LASSO Tibshirani (1996), and they usually require a centralized coordinator and are not easily generalized to the distributed case. Existing results in distributed optimization are also not directly applicable since they either assume all agents hold the same constraints Nedic et al. (2010) or different but compact constraints Lin et al. (2016), and they typically require the weighting matrix associated with the network graph to be doubly stochastic Nedic et al. (2010); Nedić et al. (2018); Lin et al. (2016) or at least weighted balanced Gharesifard and Cortés (2013). However, when solving under-determined equation sets via multi-agent networks, the local equations known by different agents can not be the same; the solution set to the local equation constraint is an affine subspace which is not compact; and as illustrated in Dominguez-Garcia and Hadjicostis (2013), for a directed graph, additional cooperations among agents are usually required to guarantee its weighting matrix is doubly stochastic. Further, the non-smooth nature of the l_1-norm can be problematic; to handle these issues in continuous-time, we have developed a distributed algorithm for minimum l_1-norm solution Zhou et al. (2018) based on Fillipov Set Value maps, which is a considerable technical complication. In this paper, we achieve further progress by devising a discrete-time algorithm based on a combination of the projection-consensus flow and the sub-gradient method. Moreover, compared with the results in Nedic et al. (2010); Nedić et al. (2018); Lin et al. (2016), we remove the requirement for the weighting matrix associated with the network graph to

⋆ This work was supported by fundings from Northrop Grumman Corporation (NGC-REALM and NGCRC), Data61-CSIRO, and the Australian Research Council's Discovery Projects DP-160104500 and DP190100887. Corresponding Author: Shaoshuai Mou.

be doubly stochastic, so that the algorithm is applicable to any directed network that is strongly connected.

The organization of the paper is as follows. We formulate the problem of interest in Section II. In section III, we present a discrete-time, distributed algorithm for achieving the minimum l_1-norm solution. The effectiveness of this algorithm is theoretically proved in Section IV, which is further validated by a simulation in Section V. Concluding remarks are made in Section VI. Proofs of all Lemmas are given in the Appendix.

Notation: Let $\mathbf{1}_r$ denote the vector in $\mathbb{R}^r$ with all entries equal to 1. Let I_r denote the $r \times r$ identity matrix. We let $\text{col}\,\{A_1, A_2, \cdots, A_r\}$ be a stack of matrices A_i possessing the same number of columns with the index in a top-down ascending order, $i = 1, 2, \cdots, r$. Let $\text{diag}\,\{A_1, A_2, \cdots, A_r\}$ denote a block diagonal matrix with A_i the ith diagonal block entry, $i = 1, 2, \cdots, r$. By $M > 0$ and $M \geq 0$ are meant that the square matrix M is positive definite and positive semi-definite, respectively. By $M^\top$ is meant the transpose of a matrix M. Let $\ker M$ and image M denote the kernel and image of a matrix M, respectively. Let $\otimes$ denote the Kronecker product. Let $\|\cdot\|_1$ denote the l_1-norm of a vector, and $[\,\cdot\,]_j$ denote the jth entry of a vector.

2. PROBLEM FORMULATION

Consider a network of m agents, $i = 1, 2, \cdots, m$, where each agent in the network is able to receive information from certain other agents called its *neighbors*. Let $\mathcal{N}_i$ denote the set of agent i's neighbors. The neighbor relation can be characterized by a *directed* graph $\mathbb{G}$, where we assume $\mathbb{G}$ is *strongly connected*. Let $S \in \mathbb{R}^{m\times m}$ denote a row stochastic weighted adjacency matrix associated with $\mathbb{G}$, namely, for $i, j = 1, 2, ..., m$, the entries of S satisfy $\sum_{j=1}^{m} s_{ij} = 1$, where $s_{ij} > 0$ if $j \in \mathcal{N}_i$, and $s_{ij} = 0$ otherwise. Note that for any strongly connected $\mathbb{G}$, one such S can be simply constructed in a distributed manner by letting $s_{ij} = \frac{1}{d_i}$, for $\forall\ j \in \mathcal{N}_i$, where $d_i = |\mathcal{N}_i|$ is the number of agent i's neighbors. Suppose each agent i knows $A_i \in \mathbb{R}^{n_i\times n}$ and $b_i \in \mathbb{R}^{n_i}$, and controls a state vector $x_i(t) \in \mathbb{R}^n$. For an underlying under-determined linear equation $Ax = b$, let x^* denote a minimum l_1-norm solution such that

$$x^* = \underset{Ax=b}{\arg\min}\, \|x\|_1, \tag{1}$$

where

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix},\ b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \tag{2}$$

Assumption 1. The equation

$$Ax = b \tag{3}$$

is under-determined and it has a unique minimum l_1-norm solution.

According to the applications Shearer (1997); Beucker and Schlitt (1996), most under-determined linear equation sets found in practice have a unique minimum l_1-norm solution. Our assumption does not lead to a degradation on generality.

The problem of interest is to develop an iterative update to each agent's state vector by only using its neighbors' states such that all $x_i(t)$ converges to a common value, viz the minimum l_1 solution x^* of equation (3).

3. THE UPDATE

A distributed update for achieving a solution to $Ax = b$ is developed in Mou et al. (2015) based on the projection-consensus flow as follows:

$$x_i(t+1) = x_i(t) - P_i\left(x_i(t) - s_{ij}\sum_{j\in\mathcal{N}_i} x_j(t)\right),\ i = 1, 2, ..., m \tag{4}$$

where $A_i x_i(0) = b_i$ and $P_i \in \mathbb{R}^{n\times n}$ is the projection matrix to $\ker A_i$. The above update enables all $x_i(t)$ to reach a consensus value exponentially fast which is a solution to $Ax = b$. This procedure imposes an a priori requirement on existence of a solution, or possibly a family of solutions. To further guide this consensus value to be the minimum l_1-norm solution, we add the subgradient of $\|x\|_1$ subject to $A_i x = b_i$, namely, $P_i\text{sgn}\,(x_i(t))$, to (4) and have the following

$$x_i(t+1) = x_i(t) - P_i\left(x_i(t) - s_{ij}\sum_{j\in\mathcal{N}_i} x_j(t)\right) - \frac{P_i}{t+1}\text{sgn}\,(x_i(t)) \tag{5}$$

with $A_i x_i(0) = b_i$, $i = 1, 2, ..., m$.

Remark 1. Because $A_i x_i(0) = b_i$, and image $P_i = \ker A_i$, under the distributed update (5), one has $A_i x_i(t) = b_i$ for $\forall t > 0$. Note that $\frac{1}{t+1}$ is introduced to adjust impact of the term $P_i\text{sgn}\,(x_i(t))$ to the original update (4), a device which is commonly used in many distributed optimization algorithms Nedic et al. (2010). This takes care of the fact that $P_i\text{sgn}\,(x_i(t))$ cannot be expected to tend to zero. Without such adjusted term $\frac{1}{t+1}$, we could never secure a consensus steady state solution $x_i(t) = x^*$ with $Ax^* = b$.

Remark 2. The update (5) is different from the algorithms proposed in Nedic et al. (2010); Lin et al. (2016). In update (5), the gradient term $\frac{1}{t+1}P_i\text{sgn}\,(x_i(t))$ is computed with respect to the current state $x_i(t)$ of agent i, thus, it is independent of the current round of communication. In Nedic et al. (2010); Lin et al. (2016), the gradient follows the form of $\frac{1}{t+1}P_i\text{sgn}\,(s_{ij}\sum_{j\in\mathcal{N}_i} x_j(t))$, which is computed using a weighted average of agent i's neighbors states.

4. MAIN RESULT

In this section, we will present our main result under the distributed update (5), for the problem identified in Section 2.

Theorem 1. Suppose the equation $Ax = b$ is under-determined with a unique minimum l_1-norm solution. Suppose the graph $\mathbb{G}$ of an associated m-agent network is directed and strongly connected, and its associated weighted adjacency matrix is row stochastic. Let each agent knows A_i and b_i, defined in the partition (2). Initialize $x_i(0)$ such that $A_i x_i(0) = b_i$. Then, under the distributed update (5), all $x_i(t)$ converge asymptotically to a constant given by x^*, which is the unique minimum l_1-norm solution[1] to equation $Ax = b$.

For the convenience of establishing Theorem 1, let $\boldsymbol{x}(t) = \text{col}\{x_1(t), \cdots, x_m(t)\}$ denote a stack of all $x_i(t)$; let $\bar{P} = \text{diag}\{P_1, \cdots, P_m\}$ denote a block-diagonal matrix with the ith diagonal block equal to P_i. Further let $\bar{S} = S \otimes I_n$, where $S \in \mathbb{R}^{m\times m}$ is a weighted adjacency matrix of the graph $\mathbb{G}$. Then based on equation (5), the evolution of all the states in the network can be rewritten in a compact form as

$$\boldsymbol{x}(t+1) = \bar{Q}\boldsymbol{x}(t) - \frac{1}{t+1}\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \tag{6}$$

where

$$\bar{Q} = I - \bar{P} + \bar{P}\bar{S}.$$

To prove Theorem 1, it is sufficient to show that $\boldsymbol{x}(t) \to \boldsymbol{x}^*$, where $\boldsymbol{x}^* \triangleq \mathbf{1} \otimes x^*$. Towards this end, our proof is divided into three steps, which progressively lead to the fact that $\boldsymbol{x}(t) \to \boldsymbol{x}^*$. Firstly, for the purposes of proving the correctness of the algorithm, but not something computed in the course of executing the algorithm, we introduce a trajectory $\boldsymbol{z}(t)$ linked in a certain way to $\boldsymbol{x}(t)$. Then based on this $\boldsymbol{z}(t)$, we show that $\boldsymbol{x}(t) \to \boldsymbol{z}(t)$. Finally, we show that $\boldsymbol{z}(t) \to \boldsymbol{x}^*$. The details of these steps are provided in the following subsections.

4.1 Introducing a trajectory $\boldsymbol{z}(t)$.

For each time step t, define

$$\boldsymbol{z}(t,k) \triangleq \bar{Q}^k\boldsymbol{x}(t). \tag{7}$$

Then the following Proposition holds.

Proposition 1. For each fixed t, as $k \to \infty$, the following limit of $\boldsymbol{z}(t,k)$ exists,

$$\boldsymbol{z}(t) \triangleq \lim_{k\to\infty} \boldsymbol{z}(t,k) = \lim_{k\to\infty} \bar{Q}^k\boldsymbol{x}(t). \tag{8}$$

Moreover, $\bar{Q}\boldsymbol{z}(t) = \boldsymbol{z}(t)$ and for $\forall t$, there holds $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$ where $z(t) \in \mathbb{R}^n$ is a solution to $Ax = b$.

To prove Proposition 1, we propose the following lemma, for which we provide the proof in the Appendix.

Lemma 1. The following statements hold:

(a) All eigenvalues of $\bar{P}\bar{S}$ have magnitude less than or equal to 1.
(b) $\lambda^\star = 1$ is the only eigenvalue of $\bar{P}\bar{S}$ with magnitude 1. It is non-defective and any corresponding eigenvector satisfies $\bar{P}\bar{S}\boldsymbol{u} = \bar{S}\boldsymbol{u} = \boldsymbol{u}$ where $\boldsymbol{u} = \mathbf{1}_m \otimes u$ and $u \in \ker A$.

Proof of Proposition 1: Let t be arbitrary but fixed. By definition (7), one has

$$\boldsymbol{z}(t,k+1) = \bar{Q}\boldsymbol{z}(t,k) = (I - \bar{P} + \bar{P}\bar{S})\boldsymbol{z}(t,k) \tag{9}$$

with $\boldsymbol{z}(t,0) = \boldsymbol{x}(t)$. Since $\bar{P} = \text{diag}\{P_1, \cdots, P_m\}$, and setting $\boldsymbol{z} = \text{col}\{z_1, \cdots, z_m\}$, then update (9) can be rewritten as

$$z_i(t,k+1) = z_i(t,k) - P_i\left(z_i(t,k) - s_{ij}\sum_{j\in\mathcal{N}_i} z_j(t,k)\right). \tag{10}$$

[1] Note that if the minimum l_1-norm solution is non-unique, the algorithm will converge to one of the minimum l_1-norm solutions.

From update (10) and the fact that P_i is a projection matrix to $\ker A_i$, one has $A_iz_i(t,k+1) = A_iz_i(t,k)$. Since also $z_i(t,0) = x_i(t)$ is a solution to $A_ix_i = b_i$, one has for $\forall t,k$, $z_i(t,k)$ is a solution to $A_iz_i = b_i$.

To continue, define $\boldsymbol{z}^* \triangleq \mathbf{1}_m \otimes z^*$, where $z^* \in \mathbb{R}^n$ is an arbitrary solution to $Ax = b$. Since S is row stochastic, for any $z^* \in \mathbb{R}^n$, one has $\bar{S}\boldsymbol{z}^* = (S \otimes I_n)(\mathbf{1}_m \otimes z^*) = \boldsymbol{z}^*$. Further, define $\boldsymbol{\eta}(t,k) \triangleq \boldsymbol{z}(t,k) - \boldsymbol{z}^*$, where $\boldsymbol{\eta}(t,k) = \text{col}\{\eta_1(t,k), \cdots, \eta_m(t,k)\}$ and $\eta_i(t,k) = z_i(t,k) - z^*$ for all $i = 1, \cdots, m$. Recall that both $z_i(t,k)$ and z^* are solutions to $A_iz_i = b_i$; then $\eta_i(t,k) \in \ker A_i$. Because P_i is a projection matrix to $\ker A_i$, one has $P_i\eta_i(t,k) = \eta_i(t,k)$, that is $\bar{P}\boldsymbol{\eta}(t,k) = \boldsymbol{\eta}(t,k)$. Then, by subtracting $\boldsymbol{z}^*$ on both sides of (9), one has

$$\begin{aligned}\boldsymbol{\eta}(t,k+1) &= \boldsymbol{\eta}(t,k) - (\bar{P} - \bar{P}\bar{S})\boldsymbol{\eta}(t,k) - (\bar{P} - \bar{P}\bar{S})\boldsymbol{z}^* \\ &= \boldsymbol{\eta}(t,k) - (\bar{P} - \bar{P}\bar{S})\boldsymbol{\eta}(t,k) \\ &= \bar{P}\bar{S}\boldsymbol{\eta}(t,k)\end{aligned} \tag{11}$$

By Lemma 1, there exists a non-singular matrix T such that

$$\bar{P}\bar{S} = T\begin{bmatrix} I & 0 \\ 0 & R\end{bmatrix}T^{-1} \tag{12}$$

where all the eigenvalues of R are the eigenvalues of $\bar{P}\bar{S}$ with magnitude less than 1. Let

$$M = \lim_{k\to\infty}\left(\bar{P}\bar{S}\right)^k = T\begin{bmatrix} I & 0 \\ 0 & 0\end{bmatrix}T^{-1} \tag{13}$$

Define $\boldsymbol{\eta}(t)^* = \lim_{k\to\infty}\boldsymbol{\eta}(t,k)$; then by update (11),

$$\boldsymbol{\eta}(t)^* = \lim_{k\to\infty}(\bar{P}\bar{S})^k\boldsymbol{\eta}(t,0) = T\begin{bmatrix} I & 0 \\ 0 & 0\end{bmatrix}T^{-1}\boldsymbol{\eta}(t,0) \tag{14}$$

Further, by the definition of $\boldsymbol{\eta}$, one has

$$\lim_{k\to\infty}\boldsymbol{z}(t,k) = \lim_{k\to\infty}\boldsymbol{\eta}(t,k) + \boldsymbol{z}^* = \boldsymbol{\eta}(t)^* + \boldsymbol{z}^* \tag{15}$$

Equation (15) verifies the existence of $\lim_{k\to\infty}\boldsymbol{z}(t,k)$, namely $\boldsymbol{z}(t)$ as defined in (8). As a consequence, $\bar{Q}\boldsymbol{z}(t) = \lim_{k\to\infty}\bar{Q}^{k+1}\boldsymbol{x}(t) = \boldsymbol{z}(t)$. To further show that $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$ and for $\forall t$, $z(t)$ is a solution to $Ax = b$, recall from equations (12) and (14) that

$$\bar{P}\bar{S}\boldsymbol{\eta}(t)^* = T\begin{bmatrix} I & 0 \\ 0 & R\end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & 0\end{bmatrix}T^{-1}\boldsymbol{\eta}(t,0) = \boldsymbol{\eta}(t)^* \tag{16}$$

Thus, by Lemma 1 (b), one has $\boldsymbol{\eta}(t)^* = \mathbf{1}_m \otimes \eta(t)$ and for $\forall t$, $\eta(t) \in \ker A$. This, along with the definition of $\boldsymbol{z}^*$ yields

$$\boldsymbol{z}(t) = \boldsymbol{\eta}(t)^* + \boldsymbol{z}^* = \mathbf{1}_m \otimes (\eta(t) + z^*) = \mathbf{1}_m \otimes z(t) \tag{17}$$

Because $z^* \in \mathbb{R}^n$ is a solution to $Ax = b$ and $\eta(t) \in \ker A$, it follows that $z(t)$ is also a solution to $Ax = b$. This completes the proof. ■

4.2 Proof of $\boldsymbol{x}(t) \to \boldsymbol{z}(t)$.

Proposition 2. For update (6), given any initial $\boldsymbol{x}(0)$ and the $\boldsymbol{z}(t)$ defined in (8), there always exists a positive constant β independent ot t such that

$$\|\boldsymbol{x}(t) - \boldsymbol{z}(t)\|_2 \le \frac{\beta}{t+1} \tag{18}$$

Proof of Proposition 2: Pre-multiply equation (6) on the left by $\bar{Q}^k$, leading to

$$\bar{Q}^k\boldsymbol{x}(t+1) = \bar{Q}^{k+1}\boldsymbol{x}(t) - \frac{1}{t+1}\bar{Q}^k\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \tag{19}$$

By taking $k \to \infty$ and recalling from definition (8) that $\boldsymbol{z}(t) = \lim_{k\to\infty} \bar{Q}^k \boldsymbol{x}(t) = \lim_{k\to\infty} \bar{Q}^{k+1}\boldsymbol{x}(t)$, one has

$$\boldsymbol{z}(t+1) = \boldsymbol{z}(t) - \frac{1}{t+1} \lim_{k\to\infty} \bar{Q}^k \bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) \tag{20}$$

Recall that $\bar{Q} = I - \bar{P} + \bar{P}\bar{S}$ and $\bar{P}^2 = \bar{P}$ (a property of projection matrices); then

$$\begin{aligned}
\lim_{k\to\infty} \bar{Q}^k \bar{P} &= \lim_{k\to\infty} (I - \bar{P} + \bar{P}\bar{S})^t \bar{P} \\
&= \lim_{k\to\infty} (I - \bar{P} + \bar{P}\bar{S})^{(k-1)} (I - \bar{P} + \bar{P}\bar{S})\bar{P} \\
&= \lim_{k\to\infty} (I - \bar{P} + \bar{P}\bar{S})^{(k-1)} \bar{P}\bar{S}\bar{P} \\
&= \lim_{k\to\infty} (I - \bar{P} + \bar{P}\bar{S})^{(k-2)} (\bar{P}\bar{S})^2\bar{P} \\
&= \lim_{k\to\infty} (\bar{P}\bar{S})^k\bar{P} = M\bar{P}
\end{aligned} \tag{21}$$

where M is defined in (13). Using this, equation (20) can be further written as

$$\boldsymbol{z}(t+1) = \bar{Q}\boldsymbol{z}(t) - \frac{1}{t+1} M\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) \tag{22}$$

Define $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \boldsymbol{z}(t)$. Subtracting (22) from (6) yields

$$\boldsymbol{e}(t+1) = \bar{Q}\boldsymbol{e}(t) - \frac{1}{t+1}(I - M)\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) \tag{23}$$

Now, recall from Remark 1 and Proposition 1 that $x_i(t)$ and $z(t)$ are solutions to $A_i x_i = b_i$, which implies that $P_i(x_i(t) - z(t)) = x_i(t) - z(t)$, that is, $\bar{P}\boldsymbol{e}(t) = \boldsymbol{e}(t)$. Thus, by using (21) in reverse,

$$\begin{aligned}
M\boldsymbol{e}(t) = M\bar{P}\boldsymbol{e}(t) &= \lim_{k\to\infty} \bar{Q}^k \bar{P}\boldsymbol{e}(t) \\
&= \lim_{k\to\infty} \bar{Q}^k \boldsymbol{e}(t) = \lim_{k\to\infty} \bar{Q}^k (\boldsymbol{x}(t) - \boldsymbol{z}(t)) \\
&= \boldsymbol{z}(t) - \boldsymbol{z}(t) = 0
\end{aligned}$$

Bringing this, equations (12) and (13) into update (23), yields

$$\begin{aligned}
\boldsymbol{e}(t+1) &= (I - \bar{P} + \bar{P}\bar{S} - M)\boldsymbol{e}(t) - \frac{I-M}{t+1}\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) \\
&= \bar{P}\bar{S}\boldsymbol{e}(t) - \frac{1}{t+1}(I - M)\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) \\
&= T\begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix} T^{-1}\boldsymbol{e}(t) - \frac{I-M}{t+1}\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right)
\end{aligned} \tag{24}$$

Since all eigenvalues of R have magnitude strictly less than one, there must exist a scalar $0 < \rho < 1$ such that

$$\|\boldsymbol{e}(t+1)\|_2 \le \rho\|\boldsymbol{e}(t)\|_2 + \frac{c_e}{t+1} \tag{25}$$

where $c_e > 0$ is the upper bound of $\|(I - M)\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right)\|_2$.

Now, given equation (25), for $t = 2\tau$, $\tau = 0, 1, 2, \cdots$, one has

$$\begin{aligned}
\|\boldsymbol{e}(2\tau)\|_2 &\le \rho^{2\tau}\|\boldsymbol{e}(0)\|_2 + c_e \sum_{j=1}^{2\tau} \frac{\rho^{2\tau-j}}{j+1} \\
&= \rho^{2\tau}\|\boldsymbol{e}(0)\|_2 + c_e\rho^{\tau} \sum_{j=1}^{\tau} \frac{\rho^{\tau-j}}{j+1} + c_e \sum_{j=\tau+1}^{2\tau} \frac{\rho^{2\tau-j}}{j+1} \\
&\le \rho^{2\tau}\|\boldsymbol{e}(0)\|_2 + c_e\frac{\rho^{\tau}}{1-\rho} + c_e\frac{1}{\tau+2}\frac{1}{1-\rho}
\end{aligned} \tag{26}$$

For $t = 2\tau + 1$, $\tau = 0, 1, 2, \cdots$, one has

$$\begin{aligned}
\|\boldsymbol{e}(2\tau+1)\|_2 &\le \rho\|\boldsymbol{e}(2\tau)\|_2 + \frac{c_e}{2\tau+2} \\
&\le \rho^{2k+1}\|\boldsymbol{e}(0)\|_2 + c_e\frac{\rho^{\tau+1}}{1-\rho} + \rho c_e\frac{1}{\tau+2}\frac{1}{1-\rho} + c_e\frac{1}{2\tau+2}
\end{aligned} \tag{27}$$

Since $\boldsymbol{e}(t) = \boldsymbol{x}(t) - \boldsymbol{z}(t)$, by combining equations (26) and (27), for $\forall t > 0$, it is evident there must exist a positive constant β such that (18) is true. This completes the proof. ■

4.3 *Proof of $\boldsymbol{z}(t) \to \boldsymbol{x}^*$.*

Proposition 3. Suppose the minimum l_1-norm solution x^* to problem (1) is unique. Let $\boldsymbol{x}^* \triangleq \mathbf{1} \otimes x^*$. Then for update (6), given any initial $\boldsymbol{x}(0)$, and the $\boldsymbol{z}(t)$ defined in (8), one has

$$\|\boldsymbol{z}(t) - \boldsymbol{x}^*\|_2 \to 0 \quad \text{as} \quad t \to \infty \tag{28}$$

Before proving Proposition 3, we define $\boldsymbol{x}^* = \mathbf{1}_m \otimes x^*$ where x^* is the unique minimum l_1-norm solution defined in (1). Let $\boldsymbol{\epsilon}(t) = \boldsymbol{z}(t) - \boldsymbol{x}^*$ and define a function $V(\boldsymbol{\epsilon}(t)) \triangleq \boldsymbol{\epsilon}(t)^\top\boldsymbol{\epsilon}(t)$. Let $\bar{\Pi} \triangleq \Pi \otimes I_n \in \mathbb{R}^{mn\times mn}$ be a diagonal matrix, where $\Pi = \text{diag}\left\{\pi_1, \cdots, \pi_m\right\}$ and $[\pi_1\ \pi_2\ \cdots\ \pi_m] = \pi^\top = \lim_{k\to\infty}\left(\mathbf{1}_m^\top S^k\right) \in \mathbb{R}^{1\times m}$. Then, we introduce the following two lemmas to summarize some useful results with proofs provided in the Appendix.

Lemma 2. The following statements hold:

(a) The row vector $\pi^\top$ is a left eigenvector of S corresponding to eigenvalue 1, $\bar{\Pi}$ is positive definite and
$$\boldsymbol{\epsilon}(t)^\top M\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right) = \boldsymbol{\epsilon}(t)^\top\bar{\Pi}\ \text{sgn}\left(\boldsymbol{x}(t)\right)$$

(b) $$\left[\text{sgn}\left(\boldsymbol{x}(t)\right)\right]^\top \bar{\Pi}(\boldsymbol{x}(t) - \boldsymbol{x}^*) \ge \|\bar{\Pi}\boldsymbol{x}(t)\|_1 - \|\bar{\Pi}\boldsymbol{x}^*\|_1.$$

(c) $\|\bar{\Pi}\boldsymbol{z}(t)\|_1 = \frac{\mathbf{1}_m^\top\pi}{m}\|\boldsymbol{z}(t)\|_1$ and $\|\bar{\Pi}\boldsymbol{x}^*\|_1 = \frac{\mathbf{1}_m^\top\pi}{m}\|\boldsymbol{x}^*\|_1$.

(d) $$\left|\|\bar{\Pi}\boldsymbol{x}(t)\|_1 - \|\bar{\Pi}\boldsymbol{z}(t)\|_1\right| \le \max\{\pi_i\}\frac{\sqrt{mn}\beta}{t+1},$$
where mn is the dimension of $\boldsymbol{x}(t)$ and $\boldsymbol{z}(t)$, with m the number of agents in the network, n the dimension of each $x_i(t)$, $z_i(t)$.

Lemma 3. Suppose the minimum l_1-norm solution x^* of the linear equation $Ax = b$ is unique. Further, suppose $\|\boldsymbol{\epsilon}(t)\|_2 \le \Delta$ is bounded, where $\boldsymbol{\epsilon}(t) = \boldsymbol{z}(t) - \boldsymbol{x}^*$, then there exists a positive constant α such that $\forall t > 0$,

$$\|\boldsymbol{z}(t)\|_1 - \|\boldsymbol{x}^*\|_1 \ge \alpha\boldsymbol{\epsilon}(t)^\top\boldsymbol{\epsilon}(t) = \alpha V(\boldsymbol{\epsilon}(t)). \tag{29}$$

Proof of Proposition 3: By (21), subtracting $\boldsymbol{x}^*$ from both sides of equation (20) yields

$$\boldsymbol{\epsilon}(t+1) = \boldsymbol{\epsilon}(t) - \frac{1}{t+1}M\bar{P}\text{sgn}\left(\boldsymbol{x}(t)\right). \tag{30}$$

Based on (a) of Lemma 2, one has

$$
\begin{aligned}
&V(\boldsymbol{\epsilon}(t+1)) \\
=&\left[\boldsymbol{\epsilon}(t)-\frac{1}{t+1}M\bar{P}\mathrm{sgn}\,(\boldsymbol{x}(t))\right]^{\top}\left[\boldsymbol{\epsilon}(t)-\frac{1}{t+1}M\bar{P}\mathrm{sgn}\,(\boldsymbol{x}(t))\right] \\
=&V(\boldsymbol{\epsilon}(t))-\frac{2}{t+1}\boldsymbol{\epsilon}(t)^{\top}M\bar{P}\mathrm{sgn}\,(\boldsymbol{x}(t))+\frac{1}{(t+1)^2}\|M\bar{P}\mathrm{sgn}\,(\boldsymbol{x}(t))\|_2^2 \\
\leq&V(\boldsymbol{\epsilon}(t))-\frac{2}{t+1}\boldsymbol{\epsilon}(t)^{\top}\bar{\Pi}\mathrm{sgn}\,(\boldsymbol{x}(t))+\frac{\psi}{(t+1)^2} \\
=&V(\boldsymbol{\epsilon}(t))-\frac{2}{t+1}\mathrm{sgn}\,(\boldsymbol{x}(t))^{\top}\bar{\Pi}(\boldsymbol{z}(t)-\boldsymbol{x}^*)+\frac{\psi}{(t+1)^2} \\
=&V(\boldsymbol{\epsilon}(t))-\frac{2}{t+1}\mathrm{sgn}\,(\boldsymbol{x}(t))^{\top}\bar{\Pi}(\boldsymbol{x}(t)-\boldsymbol{x}^*) \\
&-\frac{2}{t+1}\mathrm{sgn}\,(\boldsymbol{x}(t))^{\top}\bar{\Pi}(\boldsymbol{z}(t)-\boldsymbol{x}(t))+\frac{\psi}{(t+1)^2} \\
\leq&V(\boldsymbol{\epsilon}(t))-\frac{2}{t+1}\mathrm{sgn}\,(\boldsymbol{x}(t))^{\top}\bar{\Pi}(\boldsymbol{x}(t)-\boldsymbol{x}^*)+\frac{2\bar{\gamma}\beta}{(t+1)^2}+\frac{\psi}{(t+1)^2}
\end{aligned} \tag{31}
$$

where ψ and $\bar{\gamma}$ are positive constants such that for all $\forall\boldsymbol{x}(t)$, $\|M\bar{P}\mathrm{sgn}\,(\boldsymbol{x}(t))\|_2^2\leq\psi$ and $\|\bar{\Pi}\mathrm{sgn}\,(\boldsymbol{x}(t))\|_2\leq\bar{\gamma}$ (the last inequality in (31) results from (18) of Proposition 2). Then, by bringing (b), (c) and (d) of Lemma 2 into equation (31), one has

$$
\begin{aligned}
&V(\boldsymbol{\epsilon}(t+1))-V(\boldsymbol{\epsilon}(t)) \\
\leq&-\frac{2}{t+1}\left(\|\bar{\Pi}\boldsymbol{x}(t)\|_1-\|\bar{\Pi}\boldsymbol{x}^*\|_1\right)+\frac{2\bar{\gamma}\beta}{(t+1)^2}+\frac{\psi}{(t+1)^2} \\
=&-\frac{2}{t+1}\left(\|\bar{\Pi}\boldsymbol{z}(t)\|_1-\|\bar{\Pi}\boldsymbol{x}^*\|_1-\left(\|\bar{\Pi}\boldsymbol{z}(t)\|_1-\|\bar{\Pi}\boldsymbol{x}(t)\|_1\right)\right)+\frac{2\bar{\gamma}\beta+\psi}{(t+1)^2} \\
\leq&-\frac{2}{t+1}\left(\|\bar{\Pi}\boldsymbol{z}(t)\|_1-\|\bar{\Pi}\boldsymbol{x}^*\|_1-\max\{\pi_i\}\frac{\sqrt{mn}\beta}{t}\right)+\frac{2\bar{\gamma}\beta+\psi}{t^2} \\
\leq&-\frac{2}{t+1}\frac{\mathbf{1}_m^{\top}\pi}{m}\left(\|\boldsymbol{z}(t)\|_1-\|\boldsymbol{x}^*\|_1\right)+\max\{\pi_i\}\frac{2\sqrt{mn}\beta}{(t+1)^2}+\frac{2\bar{\gamma}\beta+\psi}{(t+1)^2} \\
=&-\frac{2\gamma}{t+1}\left(\|\boldsymbol{z}(t)\|_1-\|\boldsymbol{x}^*\|_1\right)+\frac{\bar{\psi}}{(t+1)^2}
\end{aligned} \tag{32}
$$

where $\gamma=\frac{\mathbf{1}_m^{\top}\pi}{m}$ and $\bar{\psi}=2\max\{\pi_i\}\sqrt{mn}\beta+2\bar{\gamma}\beta+\psi$.

To continue, since $\boldsymbol{z}(t)=\mathbf{1}_m\otimes z(t)$ and $\boldsymbol{x}^*=\mathbf{1}_m\otimes x^*$, where $\boldsymbol{z}(t)$ is a solution to $Ax=b$ and x^* is the unique minimum l_1-norm solution to $Ax=b$, one has $\|\boldsymbol{z}(t)\|_1-\|\boldsymbol{x}^*\|_1\geq 0$. This taken with (32) implies

$$
V(\boldsymbol{\epsilon}(t+1))\leq V(\boldsymbol{\epsilon}(t))+\frac{\bar{\psi}}{(t+1)^2} \tag{33}
$$

Since $\bar{\psi}$ is a constant, and $\sum_{t=1}^{\infty}\frac{1}{t^2}<\infty$, then $V(\boldsymbol{\epsilon}(t))$ must be bounded. Therefore, there exists a constant Δ such that $\|\boldsymbol{\epsilon}(t)\|_2\leq\Delta$. Then, based on Lemma 3, by introducing (29) to (32), one has :

$$
V(\boldsymbol{\epsilon}(t+1))\leq(1-\frac{2\gamma\alpha}{t+1})V(\boldsymbol{\epsilon}(t))+\frac{\bar{\psi}}{(t+1)^2} \tag{34}
$$

The inequality (34) can be 'solved' by writing it in summation form:

$$
V(\boldsymbol{\epsilon}(t+1))\leq\frac{\bar{\psi}}{(t+1)^2}+\sum_{\tau=2}^{t}\left[\frac{\bar{\psi}}{(\tau-1)^2}\prod_{k=\tau}^{t}(1-\frac{2\gamma\alpha}{k})\right] \tag{35}
$$

Define $F(\tau_0,t)=\prod_{k=\tau_0}^{t}(1-\frac{2\gamma\alpha}{k})$, where τ_0 is sufficiently large such that $0<1-\frac{2\gamma\alpha}{k}<1$ for $\forall k\geq\tau_0$. Then

$$
\log F(\tau_0,t)=\sum_{k=\tau_0}^{t}\log(1-\frac{2\gamma\alpha}{k})
$$

Since $0<1-\frac{2\gamma\alpha}{k}<1$ for $\forall k\geq\tau_0$, it follows that $\log(1-\frac{2\gamma\alpha}{k})\leq-\frac{2\gamma\alpha}{k}$. Thus,

$$
\begin{aligned}
\log F(\tau_0,t)&\leq-2\gamma\alpha\sum_{k=\tau_0}^{t}\frac{1}{k}<-2\gamma\alpha\int_{k=\tau_0}^{t}\frac{1}{k+1} \\
&=-2\gamma\alpha\log\left(\frac{t+1}{\tau_0+1}\right).
\end{aligned}
$$

In addition, since $\tau\geq 2$ and $\prod_{k=\tau}^{\tau_0-1}(1-\frac{2\gamma\alpha}{k})$ is a product of finite terms, with each term being bounded by $[1-\gamma\alpha,\ 1]$, this product must be also bounded by a certain $\phi>0$; then,

$$
\begin{aligned}
\prod_{k=\tau}^{t}(1-\frac{2\gamma\alpha}{k})&=\prod_{k=\tau}^{\tau_0-1}(1-\frac{2\gamma\alpha}{k})F(\tau_0,t) \\
&<\phi e^{-2\gamma\alpha\log(\frac{t+1}{\tau_0+2})}=\phi\left(\frac{\tau_0+2}{t+1}\right)^{2\gamma\alpha}
\end{aligned} \tag{36}
$$

Using equation (36) in (35), one has

$$
\begin{aligned}
V(\boldsymbol{\epsilon}(t+1))&<\frac{\bar{\psi}}{t^2}+\sum_{\tau=2}^{t}\left[\frac{\bar{\psi}}{(\tau-1)^2}\phi\left(\frac{\tau_0+2}{t+1}\right)^{2\gamma\alpha}\right] \\
&=\frac{\bar{\psi}}{t^2}+\left(\frac{\tau_0+2}{t+1}\right)^{2\gamma\alpha}\cdot\phi\sum_{\tau=2}^{t}\frac{\psi}{(\tau-1)^2}
\end{aligned} \tag{37}
$$

Since α, γ, ϕ and $\bar{\psi}$ are positive constants, as $t\to\infty$, one has $\frac{\bar{\psi}}{t^2}\to 0$, $\left(\frac{\tau_0+2}{t+1}\right)^{2\alpha}\to 0$ and $\phi\sum_{\tau=2}^{t}\frac{\bar{\psi}}{(\tau-1)^2}$ is bounded. Thus,

$$
V(\boldsymbol{\epsilon}(t+1))\to 0\quad\text{as}\quad t\to\infty.
$$

That is, $\boldsymbol{\epsilon}(t)=(\boldsymbol{z}(t)-\boldsymbol{x}^*)\to 0$ as $t\to\infty$. This completes the proof. ■

4.4 Proof of Theorem 1.

Based on the Proportion 2 and 3, one has for any initial $\boldsymbol{x}(0)$, update (6) will drive $\boldsymbol{x}(t)\to\boldsymbol{x}^*$ as $t\to\infty$. Equivalently, by update (5), the states $x_i(t), i=1,\cdots,m$ in all agents will converge to x^*. This completes the proof. ■

5. SIMULATIONS

In this section, we describe the numerical simulations in MATLAB to validate the main result, noting a particular representative example. The simulations were conducted for a number of randomly generated networks with randomly generated linear equations. More precisely, we employ a directed, strongly connected network of $m=16$ agents, where any two agents are connected with a probability of 0.4. Let $s_{ij}=\frac{1}{d_i}$. Let each agent knows a $A_i\in\mathbb{R}^{2\times 33}$ and $b_i\in\mathbb{R}^2$ with entries randomly selected from the interval $[0,\ 1]$. The equation set $A_ix=b_i$, $i=1,\cdots,16$ has a unique minimum l_1 norm solution x^* with probability 1. Define $V(t)\triangleq\|\boldsymbol{x}(t)-\boldsymbol{x}^*\|_2^2$, where $\boldsymbol{x}(t)=\mathrm{col}\,\{x_1(t),\cdots,x_m(t)\}$ and $\boldsymbol{x}^*=\mathbf{1}\otimes x^*$.

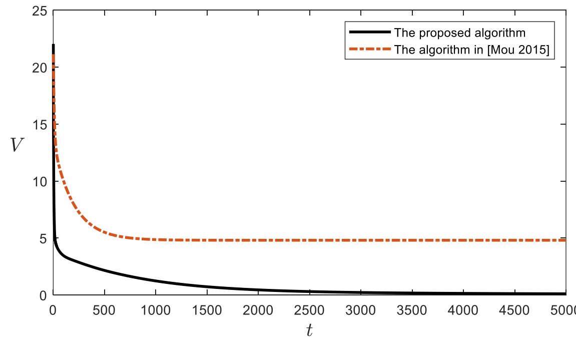

Fig. 1. Convergence of the distributed update (5) under a directed, strongly connected network of 16 agents.

The curves shown in Figure 1 are generally representative of the convergence behavior in the vast majority of simulations we undertook but depict just one example. Comparisons are given for the algorithm proposed in this paper and the one proposed in Mou et al. (2015), as revealed by the results, the distributed update (5) is able to drive the states $x_i(t)$ in all agents to the unique minimum l_1 norm solution x^* of the equation set, which validates Theorem 1. On the contrary, the algorithm introduced in Mou et al. (2015) for solving linear equations is not guaranteed to achieve x^*.

6. CONCLUSION

By combining the projection-consensus and sub-gradient descent aspects, we have developed a discrete-time distributed algorithm for achieving the minimum l_1-norm solutions to under-determined linear equations. Given the network is directed and strongly connected, it has been theoretically proved that the proposed algorithm can drive the states in all agents to the minimum l_1-norm solution of the linear equation. In future, we will focus on the modification of the algorithm to achieve a better convergence rate.

APPENDIX

Proof of Lemma 1

(**a**) Since S is a row stochastic matrix corresponding to the weighted adjacency matrix of a strongly connected graph, by the Perron-Frobenius theorem Perron (1907), S has a simple eigenvalue equal to 1 and all the other eigenvalues of S have magnitude strictly less than 1. Now let $\omega^\top = [\omega_1, \cdots, \omega_m] \neq 0$ be a normalized left Perron-Frobenius eigenvector [2] of S. Let $\Omega = \text{diag}\,\{\omega_1, \cdots, \omega_m\}$ and $H = \Omega - S^\top \Omega S$. Since Ω is diagonal and positive; and all elements of S are non-negative, then all the off-diagonal elements of H are non-positive. Further note that the row sums of H have the property:

$$\begin{aligned} H \cdot \mathbf{1}_m &= (\Omega - S^\top \Omega S)\mathbf{1}_m \\ &= \pi - S^\top \Omega \mathbf{1}_m \\ &= \pi - S^\top \pi = 0 \end{aligned}$$

Thus, H must be a Laplacian matrix of a certain graph so that $H = \Omega - S^\top \Omega S \geq 0$. Recall that the graph $\mathbb{G}$ for our problem is strongly connected, which means $\pi_i > 0$, thus, Ω is positive. By left/right multiplying by $\Omega^{-1/2}$ in the expression for H, one has

$$\Omega^{-1/2} S^\top \Omega S \Omega^{-1/2} \leq I_m. \tag{38}$$

It follows that

$$\sigma_{\max}(\bar{\Omega}^{-1/2} \bar{S}^\top \bar{\Omega} \bar{S} \bar{\Omega}^{-1/2}) \leq 1 \tag{39}$$

where $\bar{\Omega} = \Omega \otimes I_n$, $\bar{S} = S \otimes I_n$ and $\sigma_{\max}(\cdot)$ denotes the largest singular value of a matrix. Note that P_i is the projection to $\ker A_i$ and $\bar{P} = \text{diag}\,\{P_1, \cdots, P_m\}$, then one must have $\sigma_{\max}(P) \leq 1$. This, along with (39), leads to

$$\sigma_{\max}(\bar{P}\bar{\Omega}^{-1/2} \bar{S}^\top \bar{\Omega} \bar{S} \bar{\Omega}^{-1/2} \bar{P}) \leq 1$$

Furthermore, since $\bar{\Omega} = \Omega \otimes I_n$ where Ω is a diagonal matrix and $\bar{P}$ is a block-diagonal matrix with each block $P_i \in \mathbb{R}^{n \times n}$, then one has $\bar{P}\bar{\Omega}^{-1/2} = \bar{\Omega}^{-1/2}\bar{P}$. That is

$$\sigma_{\max}\left((\bar{\Omega}^{-1/2}\bar{P}\bar{S}^\top\bar{\Omega}^{1/2})(\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2})\right) \leq 1 \tag{40}$$

This indicates that all the eigenvalues of $\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}$ has magnitude less than or equal to one. Since for all the eigenvalues, we have $\lambda(\bar{P}\bar{S}) = \lambda(\bar{S}\bar{P}) = \lambda(\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2})$ Horn and Johnson (2012) , this completes the proof of statement (a).

(**b**) Note that the equality $|\lambda(\bar{P}\bar{S})| = 1$ holds if and only if there exist a vector $\boldsymbol{u} \neq 0$ such that

$$\bar{P}\bar{S}\boldsymbol{u} = \lambda^\star \boldsymbol{u} \quad \text{with} \quad |\lambda^\star| = 1 \tag{41}$$

Thus, $\bar{S}\boldsymbol{u} \neq 0$ and

$$\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}\bar{\Omega}^{1/2}\bar{S}\boldsymbol{u} = \lambda^\star \bar{\Omega}^{1/2}\bar{S}\boldsymbol{u}. \tag{42}$$

Let $\boldsymbol{q} = \bar{\Omega}^{1/2}\bar{S}\boldsymbol{u}$, since $\bar{P}\bar{\Omega}^{-1/2} = \bar{\Omega}^{-1/2}\bar{P}$, then (42) can be rewritten as

$$\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}\bar{P}\boldsymbol{q} = \lambda^\star \boldsymbol{q} \tag{43}$$

The equality of (43) holds only if

$$\|\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}\bar{P}\boldsymbol{q}\| = \|\lambda^\star \boldsymbol{q}\| = |\lambda^\star| \cdot \|\boldsymbol{q}\| = \|\boldsymbol{q}\| \tag{44}$$

Recall that $\sigma_{\max}(\bar{P}) \leq 1$ and $\sigma_{\max}(\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}) \leq 1$, thus,

$$\|\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}\bar{P}\boldsymbol{q}\| \leq \|\bar{P}\boldsymbol{q}\| \leq \|\boldsymbol{q}\|. \tag{45}$$

Then, (44) holds if and only if $\|\bar{P}\boldsymbol{q}\| = \|\boldsymbol{q}\|$. Further, recall that $\bar{P}$ is a projection matrix; then additionally

$$\bar{P}\boldsymbol{q} = \boldsymbol{q} \tag{46}$$

Bringing (46) into (43) leads to

$$\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}\boldsymbol{q} = \lambda^\star \boldsymbol{q}, \quad \text{with} \quad |\lambda^\star| = 1. \tag{47}$$

Note that $\bar{\Omega}^{1/2}\bar{S}\bar{\Omega}^{-1/2}$ and $\bar{S}$ are similar matrices with identical eigenvalues. Further, recall the definition of $\bar{S} = S \otimes I_n$ and the fact that $|\lambda(S)| = 1$ if and only if $\lambda(S) = 1$, thus, one has $\lambda^\star = 1$. Bringing this into equation (41) leads to

$$\bar{P}\bar{S}\boldsymbol{u} = \boldsymbol{u} \tag{48}$$

To continue, recall that $\boldsymbol{q} = \bar{\Omega}^{1/2}\bar{S}\boldsymbol{u}$ and $\bar{P}\bar{\Omega}^{1/2} = \bar{\Omega}^{1/2}\bar{P}$. Then equation (46) implies

$$\bar{\Omega}^{1/2}\bar{P}\bar{S}\boldsymbol{u} = \bar{P}\bar{\Omega}^{1/2}\bar{S}\boldsymbol{u} = \bar{\Omega}^{1/2}\bar{S}\boldsymbol{u} \tag{49}$$

Since $\bar{\Omega}^{1/2}$ is positive definite, one has

$$\bar{P}\bar{S}\boldsymbol{u} = \bar{S}\boldsymbol{u} \tag{50}$$

Equations (48) and (50) implies

$$\bar{S}\boldsymbol{u} = \boldsymbol{u}. \tag{51}$$

Thus, $\boldsymbol{u} = \mathbf{1}_m \otimes u$, $u \in \mathbb{R}^n$. Bringing (51) back to (48) yields $\bar{P}\boldsymbol{u} = \boldsymbol{u}$. This, along with $\boldsymbol{u} = \mathbf{1}_m \otimes u$ implies

[2] A Perron-Frobenius eigenvector is an eigenvector of S corresponding to the 1 eigenvalue, furthermore it is positive with entries summing to 1.

that $u \in \bigcap_{i=1}^m \ker A_i = \ker A$. Since $Ax = b$ is under-determined, $\ker A \neq \emptyset$ and such u exists. Till now, we have validated that $|\lambda(\bar{P}\bar{S})| = 1$, which happens if and only if $\bar{P}\bar{S}\boldsymbol{u} = \lambda^\star \boldsymbol{u}$ with $\boldsymbol{u} = \mathbf{1}_m \otimes u,\ u \in \ker A$, and $\lambda^\star = 1$.

Now, we prove that the eigenvalues of $\bar{P}\bar{S}$ equal to 1 must be non-defective by contradiction. Suppose the matrix $\bar{P}\bar{S}$ has defective eigenvalues equal to 1. Then so must the matrix $\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}$ since they are similar. From the definition of defective eigenvalues Horn and Johnson (2012), there always exist vectors $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$, with $\|\boldsymbol{v}_1\| = \|\boldsymbol{v}_2\| = 1$, and $\boldsymbol{v}_1^\top \boldsymbol{v}_2 \geq 0$ such that

$$(\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2} - I)\boldsymbol{v}_1 = 0 \tag{52}$$

$$(\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2} - I)\boldsymbol{v}_2 = \boldsymbol{v}_1 \tag{53}$$

Since $\|\boldsymbol{v}_1\| = \|\boldsymbol{v}_2\| = 1$ and $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$, one can always find a vector $\|\hat{\boldsymbol{v}}\| = 1$ such that $\hat{\boldsymbol{v}} = r_1\boldsymbol{v}_1 + r_2\boldsymbol{v}_2$ with $r_1 + r_2 > 1$. Then one has

$$\begin{aligned}\frac{\|\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}\hat{\boldsymbol{v}}\|}{\|\hat{\boldsymbol{v}}\|} &= \|\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}(r_1\boldsymbol{v}_1 + r_2\boldsymbol{v}_2)\| \\ &= \|(r_1+r_2)\boldsymbol{v}_1 + r_2\boldsymbol{v}_2\| \end{aligned} \tag{54}$$

Recall that $r_1 > 0$, $r_2 > 0$ and $\boldsymbol{v}_1^\top \boldsymbol{v}_2 \geq 0$, then

$$\|(r_1+r_2)\boldsymbol{v}_1 + r_2\boldsymbol{v}_2\| \geq \|(r_1+r_2)\boldsymbol{v}_1\| = r_1 + r_2 > 1.$$

This indicates $\sigma_{\max}\left(\bar{\Omega}^{1/2}\bar{S}\bar{P}\bar{\Omega}^{-1/2}\right) > 1$, which contradicts with equation (40). Thus, the eigenvalues of $\bar{P}\bar{S}$ equal to 1 must be non-defective. This completes the proof. ■

Proof of Lemma 2

(**a**) Since the matrix S is row stochastic, it has an eigenvalue equal to 1. Let $\pi^\top = \lim_{k\to\infty}\left(\mathbf{1}_m^\top S^k\right)$; then one has $\pi^\top S = \lim_{k\to\infty}\left(\mathbf{1}_m^\top S^{k+1}\right) = \pi^\top$. Thus, $\pi^\top$ is a left eigenvector of S corresponding to eigenvalue 1. Further since S is primitive (due to the strong connectedness of the graph), then by the Perron-Frobenius theorem Perron (1907) , all entries of π are positive and the corresponding eigenvalue at 1 is simple.

Now, recall that $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$, $\boldsymbol{x}^* = \mathbf{1}_m \otimes x^*$, thus, $\boldsymbol{\epsilon}(t) = \boldsymbol{z}(t) - \boldsymbol{x}^* = \mathbf{1}_m \otimes (z(t) - x^*) = \mathbf{1}_m \otimes \epsilon(t)$. Furthermore, since both $z(t)$ and x^* are solutions to $Ax = b$, one has $P_i\epsilon(t) = \epsilon(t)$, that is, $\left(v^\top \otimes \epsilon(t)^\top\right)\bar{P} = \left(v^\top \otimes \epsilon(t)^\top\right)$, for any $v \in \mathbb{R}^m$. With this in mind, by the definition of M in (13), one has,

$$\begin{aligned}&\boldsymbol{\epsilon}(t)^\top M\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \\ &= \lim_{k\to\infty}\left(\mathbf{1}_m^\top \otimes \epsilon(t)^\top\right)(\bar{P}\bar{S})^k\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \\ &= \lim_{k\to\infty}\left(\mathbf{1}_m^\top \otimes \epsilon(t)^\top\right)(S\otimes I_n)(\bar{P}\bar{S})^{(k-1)}\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \\ &= \lim_{k\to\infty}\left(\mathbf{1}_m^\top S \otimes \epsilon(t)^\top\right)(\bar{P}\bar{S})^{(k-1)}\bar{P}\text{sgn}\,(\boldsymbol{x}(t)^\top) \\ &= \lim_{k\to\infty}\left(\mathbf{1}_m^\top S^k \otimes \epsilon(t)^\top\right)\bar{P}\text{sgn}\,(\boldsymbol{x}(t)) \\ &= \left(\pi^\top \otimes \epsilon(t)^\top\right)\text{sgn}\,(\boldsymbol{x}(t)) \\ &= \boldsymbol{\epsilon}(t)^\top\bar{\Pi}\ \text{sgn}\,(\boldsymbol{x}(t)) \end{aligned} \tag{55}$$

This completes the proof.

(**b**) Consider a function $G(\boldsymbol{\xi}) = \|\bar{\Pi}\boldsymbol{\xi}\|_1$, where $\boldsymbol{\xi} \in \mathbb{R}^{mn}$. $G(\boldsymbol{\xi})$ is convex because the l_1-norm is convex. It follows that

$$\begin{aligned}&\left[\text{sgn}\,(\bar{\Pi}\boldsymbol{\xi}_1)\right]^\top\bar{\Pi}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) = \left[\left.\frac{\partial G(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_1}\right]^\top(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) \\ &\geq G(\boldsymbol{\xi}_1) - G(\boldsymbol{\xi}_2) = \|\bar{\Pi}\boldsymbol{\xi}_1\|_1 - \|\bar{\Pi}\boldsymbol{\xi}_2\|_1\end{aligned}$$

Recall that $\bar{\Pi}$ is a diagonal matrix with all entries being positive, which means $\text{sgn}\,(\bar{\Pi}\boldsymbol{\xi}_1) = \text{sgn}\,(\boldsymbol{\xi}_1)$; then by letting $\boldsymbol{x}(t) = \boldsymbol{\xi}_1$ and $\boldsymbol{x}^* = \boldsymbol{\xi}_2$ one has

$$[\text{sgn}\,(\boldsymbol{x}(t))]^\top\bar{\Pi}(\boldsymbol{x}(t) - \boldsymbol{x}^*) \geq \|\bar{\Pi}\boldsymbol{x}(t)\|_1 - \|\bar{\Pi}\boldsymbol{x}^*\|_1 \tag{56}$$

This completes the proof.

(**c**) Since $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$, $\boldsymbol{x}^* = \mathbf{1}_m \otimes x^*$, by the definition of $\bar{\Pi}$, one has

$$\begin{aligned}\|\bar{\Pi}\boldsymbol{z}(t)\|_1 &= \|\left(\text{diag}\,\{\pi_1,\ \pi_2,\ \cdots,\ \pi_m\}\otimes I_n\right)\left(\mathbf{1}_m \otimes z(t)\right)\|_1 \\ &= \|\text{col}\,\{\pi_1,\ \pi_2,\ \cdots,\ \pi_m\}\otimes z(t)\|_1 = \|\pi \otimes z(t)\|_1 \\ &= \mathbf{1}_m^\top\pi\|z(t)\|_1 = \frac{\mathbf{1}_m^\top\pi}{m}\|\boldsymbol{z}(t)\|_1\end{aligned} \tag{57}$$

Similarly,

$$\|\bar{\Pi}\boldsymbol{x}^*\|_1 = \frac{\mathbf{1}_m^\top\pi}{m}\|\boldsymbol{x}^*\|_1 \tag{58}$$

This completes the proof.

(**d**) Let $[\ \cdot\ ]_j$ denote the jth entry of a vector. Using the Cauchy-Schwarz inequality, one has

$$\begin{aligned}&|\|\bar{\Pi}\boldsymbol{x}(t)\|_1 - \|\bar{\Pi}\boldsymbol{z}(t)\|_1| = \left|\sum_{j=1}^{mn}|[\bar{\Pi}\boldsymbol{x}(t)]_j| - \sum_{j=1}^{mn}|[\bar{\Pi}\boldsymbol{z}(t)]_j|\right| \\ &= \sum_{j=1}^{mn}\left||[\bar{\Pi}\boldsymbol{x}(t)]_j| - |[\bar{\Pi}\boldsymbol{z}(t)]_j|\right| \leq \sum_{j=1}^{mn}\left|[\bar{\Pi}\boldsymbol{x}(t)]_j - [\bar{\Pi}\boldsymbol{z}(t)]_j\right| \\ &= \sum_{j=1}^{mn}\left|[\bar{\Pi}\boldsymbol{e}(t)]_j\right| \leq \max\{\pi_i\}\sum_{j=1}^{mn}\left|[\boldsymbol{e}(t)]_j\right|\cdot 1 \\ &\leq \max\{\pi_i\}\left(\sum_{j=1}^{mn}\left|[\boldsymbol{e}(t)]_j\right|^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{mn}1^2\right)^{\frac{1}{2}} \\ &= \max\{\pi_i\}\sqrt{mn}\|\boldsymbol{e}(t)\|_2 = \max\{\pi_i\}\frac{\sqrt{mn}\beta}{t+1}\end{aligned}$$

This completes the proof. ■

Proof of Lemma 3

If $\boldsymbol{z}(t) - \boldsymbol{x}^* = \boldsymbol{\epsilon}(t) = 0$, then $V(\boldsymbol{\epsilon}(t)) = 0$. Thus, (29) holds. If $\boldsymbol{\epsilon}(t) \neq 0$, recall that $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$, $\boldsymbol{x}^* = \mathbf{1}_m \otimes x^*$, then, $\boldsymbol{\epsilon}(t) = \boldsymbol{z}(t) - \boldsymbol{x}^* = \mathbf{1}_m \otimes (z(t) - x^*) = \mathbf{1}_m \otimes \epsilon(t)$. Since both $z(t)$ and x^* are solutions to $Ax = b$, one has $\epsilon(t) \in \ker A$. Let κ be a positive scalar whose value can be made arbitrarily small. Let

$$h \triangleq \kappa \cdot \frac{\epsilon(t)}{\|\epsilon(t)\|_1}, \tag{59}$$

obviously, $\|h\|_1 = \kappa$ and $h \in \ker A$. Since $\|\cdot\|_1$ is piecewise linear, and given that x^* is fixed, then κ can always be chosen small enough such that for any h in the form of (59), the function value of $\|x\|_1$ varies linearly along the line segment $\{x^* + \mu h \mid \mu \in (0, 1]\}$. That is, for $y = \mu h$ with $\mu \in (0, 1]$, one has $\text{sgn}\,(x^* + y) = \text{sgn}\,(x^* + h)$ and

$$\|x^* + y\|_1 - \|x^*\|_1 = \text{sgn}\,(x^* + h)^\top y. \tag{60}$$

Recall that $\frac{y}{\|y\|_1} = \frac{h}{\|h\|_1}$, then equation (60) can be further written as

$$\|x^* + y\|_1 - \|x^*\|_1 = \eta(x^*, h)\|y\|_1. \tag{61}$$

where

$$\eta(x^*, h) = \frac{\text{sgn}\,(x^* + h)^\top h}{\|h\|_1} \in \mathbb{R}. \tag{62}$$

Since $y \in \ker A$ and x^* is the unique minimum l_1-norm solution to $Ax = b$, from (61), one has $\eta(x^*, h) > 0$. Further since the h in (62) is chosen from a compact set such that $h \in \ker A$ and $\|h\|_1 = \kappa$, the values of $\eta(x^*, h)$ must have a lower bound $\hat{\eta}$, which is strictly positive. Thus, from (61), one has

$$\|x^* + y\|_1 - \|x^*\|_1 \geq \hat{\eta}\|y\|_1. \tag{63}$$

Based on inequality (63), we will consider two possibilities, depending on the magnitude of $\|\epsilon(t)\|_1$. First, if $\|\epsilon(t)\|_1 \leq \kappa$, that is $\frac{\|\epsilon(t)\|_1}{\kappa} \leq 1$, we let $\mu = \frac{\|\epsilon(t)\|_1}{\kappa}$, which leads to $y = \mu h = \frac{\|\epsilon(t)\|_1}{\kappa} \cdot \kappa \frac{\epsilon(t)}{\|\epsilon(t)\|_1} = \epsilon(t)$. Then from (63), one has

$$\|z(t)\|_1 - \|x^*\|_1 = \|x^* + \epsilon(t)\|_1 - \|x^*\|_1 \geq \hat{\eta}\|\epsilon(t)\|_1. \tag{64}$$

Second, if $\|\epsilon(t)\|_1 > \kappa$, it follows (because of (59) and $y = \mu h$) that $\epsilon(t) = \frac{\|\epsilon(t)\|_1}{\mu\kappa} y$, where $\frac{\|\epsilon(t)\|_1}{\mu\kappa} > 1$. Then due to the convexity of $\|\cdot\|_1$, one has

$$\begin{aligned} \|z(t)\|_1 - \|x^*\|_1 &= \left\| x^* + \frac{\|\epsilon(t)\|_1}{\mu\kappa} y \right\|_1 - \|x^*\|_1 \\ &\geq \frac{\|\epsilon(t)\|_1}{\mu\kappa} \left(\|x^* + y\|_1 - \|x^*\|_1\right) \geq \frac{\|\epsilon(t)\|_1}{\mu\kappa} \hat{\eta}\|y\|_1 \\ &= \hat{\eta}\|\epsilon(t)\|_1. \end{aligned} \tag{65}$$

where the last equality follows because $y = \mu h$ and $\|h\|_1 = \kappa$. Then equations (64) and (65), along with the fact that $\boldsymbol{z}(t) = \mathbf{1}_m \otimes z(t)$, $\boldsymbol{x}^* = \mathbf{1}_m \otimes x^*$, lead to

$$\|\boldsymbol{z}(t)\|_1 - \|\boldsymbol{x}^*\|_1 \geq \hat{\eta}\|\boldsymbol{\epsilon}(t)\|_1.$$

Recall that $\|\boldsymbol{\epsilon}(t)\|_1 \geq \|\boldsymbol{\epsilon}(t)\|_2$ Horn and Johnson (2012) and $\|\boldsymbol{\epsilon}(t)\|_2 \leq \Delta$, then

$$\|\boldsymbol{z}(t)\|_1 - \|\boldsymbol{x}^*\|_1 \geq \hat{\eta}\frac{\|\boldsymbol{\epsilon}(t)\|_2^2}{\|\boldsymbol{\epsilon}(t)\|_2} \geq \frac{\hat{\eta}}{\Delta} V(t)$$

Let $\alpha = \dfrac{\hat{\eta}}{\Delta}$. This completes the proof. ∎

REFERENCES

Baron, D., Duarte, M.F., Wakin, M.B., Sarvotham, S., and Baraniuk, R.G. (2009). Distributed compressive sensing. *arXiv preprint arXiv:0901.3403*.

Beucker, R. and Schlitt, H. (1996). On minimal lp-norm solutions of the biomagnetic inverse problem. Technical report, Zentralinstitut für Angewandte Mathematik.

Candes, E.J. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 4203–4215.

Dodge, Y. (2012). *Statistical data analysis based on the L1-norm and related methods*. Birkhäuser.

Dominguez-Garcia, A.D. and Hadjicostis, C.N. (2013). Distributed matrix scaling and application to average consensus in directed graphs. *IEEE Transactions on Automatic Control*, 58(3), 667–681.

Ge, D., Jiang, X., and Ye, Y. (2011). A note on the complexity of l p minimization. *Mathematical programming*, 129(2), 285–299.

Gharesifard, B. and Cortés, J. (2013). Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3), 781–786.

Horn, R.A. and Johnson, C.R. (2012). *Matrix analysis*. Cambridge university press.

Lin, P., Ren, W., and Song, Y. (2016). Distributed multi-agent optimization subject to nonidentical constraints and communication delays. *Automatica*, 65, 120–131.

Mou, S., Lin, Z., Wang, L., Fullmer, D., and Morse, A.S. (2016). A distributed algorithm for efficiently solving linear equations and its applications (special issue jcw). *Systems & Control Letters*, 91, 21–27.

Mou, S. and Morse, A.S. (2013). A fixed-neighbor, distributed algorithm for solving a linear algebraic equation. *European Control Conference*, 2269–2273.

Mou, S., Liu, J., and Morse, A.S. (2015). A distributed algorithm for solving a linear algebraic equation. *IEEE Transactions on Automatic Control*, 60(11), 2863–2878.

Nedić, A., Olshevsky, A., and Rabbat, M.G. (2018). Network topology and communication-computation trade-offs in decentralized optimization. *Proceedings of the IEEE*, 106(5), 953–976.

Nedic, A., Ozdaglar, A., and Parrilo, P.A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4), 922–938.

Perron, O. (1907). Zur theorie der matrices. *Mathematische Annalen*, 64(2), 248–263.

Shearer, P.M. (1997). Improving local earthquake locations using the l1 norm and waveform cross correlation: Application to the whittier narrows, california, aftershock sequence. *Journal of Geophysical Research: Solid Earth*, 102(B4), 8269–8283.

Shi, G., Anderson, B.D., and Helmke, U. (2016). Network flows that solve linear equations. *IEEE Transactions on Automatic Control*, 62(6), 2659–2674.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Wang, P., Ren, W., and Duan, Z. (2018). Distributed algorithm to solve a system of linear equations with unique or multiple solutions from arbitrary initializations. *IEEE Transactions on Control of Network Systems*, 6(1), 82–93.

Wang, P., Ren, W., and Duan, Z. (2019a). Distributed algorithm to solve a system of linear equations with unique or multiple solutions from arbitrary initializations. *IEEE Transactions on Control of Network Systems*, 6(1), 82–93.

Wang, X., Mou, S., and Sun, D. (2017). Improvement of a distributed algorithm for solving linear equations. *IEEE Transactions on Industrial Electronics*, 64(4), 3113–3117.

Wang, X., Mou, S., and Anderson, B.D. (2019b). Scalable, distributed algorithms for solving linear equations via double-layered networks. *IEEE Transactions on Automatic Control*.

Wang, X., Zhou, J., Mou, S., and Corless, M.J. (2019c). A distributed algorithm for least squares solutions. *IEEE Transactions on Automatic Control*, 64(10), 4217–4222.

Zhou, J., Xuan, W., Mou, S., and Anderson, B.D. (2018). Distributed algorithm for achieving minimum l 1 norm solutions of linear equation. In *2018 Annual American Control Conference (ACC)*, 5857–5862. IEEE.