

An Arrow-Hurwicz-Uzawa Type Flow as Least Squares Solver for Network Linear Equations [★]

Yang Liu ^a, Christian Lageman ^b, Brian D. O. Anderson ^{a,c}, Guodong Shi ^a

^a*Research School of Engineering, The Australian National University, Canberra 0200, Australia*

^b*Institute of Mathematics, University of Würzburg, 97070 Würzburg, Germany*

^c*Hangzhou Dianzi University, Hangzhou 310018, China*

Abstract

We study the approach to obtaining least squares solutions to systems of linear algebraic equations over networks by using distributed algorithms. Each node has access to one of the linear equations and holds a dynamic state. The aim for the node states is to reach a consensus as a least squares solution of the linear equations by exchanging their states with neighbors over an underlying interaction graph. A continuous-time distributed least squares solver over networks is developed in the form of the famous Arrow-Hurwicz-Uzawa flow. A necessary and sufficient condition is established on the graph Laplacian for the continuous-time distributed algorithm to give the least squares solution in the limit, with an exponentially fast convergence rate. The feasibility of different fundamental graphs is discussed including path graph, random graph, etc. Moreover, a discrete-time distributed algorithm is developed by Euler's method, converging exponentially to the least squares solution at the node states with suitable step size and graph conditions.

Key words: Distributed Algorithms; Dynamical Systems; Linear Equations; Least Squares.

1 Introduction

Linear algebraic equations are foundational for various computation tasks arising from practical engineering problems. In recent years much interest has developed in finding out how to solve linear equations using multiple processing units or over a network. Major efforts have been made in the development of parallel or distributed algorithms as linear-equation solvers. On the one hand one aims at faster algorithms in view of the intuition that multiple processing units working in parallel under smart arrangements might provide significant improvements in computation efficiency. On the other hand various distributed systems induce structure constraints, e.g., one node holds one equation and it cannot or does not want to share the exact

equation with other nodes, a constraint which excludes the feasibility of classical centralized algorithms.

Parallel algorithms for linear equations have been developed in the spirit of high-performance computing, e.g., the Jacobi method (Margaris et al. (2014)), the Kaczmarz method (Kaczmarz. (1937)) and the parallel algorithms for LU-decomposition in sparse linear systems (Stappen et al. (1993), Saad et al. (1999)). In these algorithms, the state of each node can give an entry of the solution to a linear equation after a suitably long running time.

Meanwhile, discrete and continuous-time algorithms for linear equations known to have a unique solution are also established from the point of view of distributed control and optimization. A variety of distributed algorithms are presented, among which discrete-time algorithms are given by (Mou et al. (2013, 2015), Liu et al. (2013), Lu et al. (2009)) and continuous-time algorithms are presented in (Anderson et al. (2016), Shi et al. (2017)). In these network distributed algorithms, compared with the development in parallel computing, each node state asymptotically converges to the solution to the linear equation. Such developments on network linear

[★] This work was supported by the DAAD with funds of the German Federal Ministry of Education and Research (BMBF), by the Australian Research Council (ARC) under grants DP-130103610 and DP-160104500.

Email addresses: yang.liu@anu.edu.au (Yang Liu), christian.lageman@mathematik.uni-wuerzburg.de (Christian Lageman), brian.anderson@anu.edu.au (Brian D. O. Anderson), guodong.shi@anu.edu.au (Guodong Shi).

equations are closely related to, and sometimes even special cases of, the study of distributed optimization methods on nonlinear models (Tsitsiklis et al. (1986), Nedic et al. (2009, 2010), Ghahserifard et al. (2014), Jakoveti et al. (2014)), due to the natural connection between solving equations and optimizing objective functions.

Most of the existing work for parallel and distributed algorithms assumes that the linear equations have exact solutions (Margaris et al. (2014), Kaczmarz. (1937), Mou et al. (2013, 2015), Liu et al. (2013), Lu et al. (2009), Anderson et al. (2016)), or can only produce least square solutions in the approximate sense or for limited graph structures (Shi et al. (2017), Wang et al. (2012)). Although one can obtain the least squares solution by expanding an unsolvable linear equation and applying the algorithms mentioned above, strong requirements apply on network structure, communication performance and local storage (Mou et al. (2015)). Few results have been obtained on direct exact distributed least squares solvers for network linear equations. In this paper, inspired by the method of applying differential equations to obtain the saddle point of an optimization problem (Kose. (1956), Arrow et al. (1958)), we present distributed continuous and discrete-time algorithms that can compute the least squares solution to a linear equation over a network. Though the proposed algorithm is noted to be similar to the algorithm of Wang-Elia/Ghahserifard-Cortés (Ghahserifard et al. (2014), Wang et al. (2010)), we illustrate that it has an advantage in communication cost and the analysis in this paper is a complementary content of their work. The contributions of our work are summarized as follows.

- By recognizing the least squares problem as a constrained optimization problem over a network, a continuous-time flow is presented in the form of the classical Arrow-Hurwicz-Uzawa flow (Arrow et al. (1958)), for which we establish necessary and sufficient conditions for the flow to yield convergence.
- Generic feasibility of linear equations and networks are studied, for the purpose of investigating the applicability of the proposed algorithm.
- By Euler's method, a discrete-time algorithm is presented and the properties of its convergence are also specified and proved.

A preliminary version of the current work was presented at the IFAC Congress in 2017 (Liu et al. (2017)). The paper begins by the formulating the network linear equations in Section 2, in addition to explaining the relation between the Arrow-Hurwicz-Uzawa flow and the proposed flow. In Section 3, a necessary and sufficient condition for the continuous-time flow to converge to the least squares solution is established. In Section 4, a discrete-time algorithm is obtained by Euler's method and the necessary and sufficient conditions for

its convergence conditions are established. Finally the conclusion is given in Section 5.

2 Problem Definition

2.1 Linear Equation

Consider the following linear algebraic equation with unknown $\mathbf{y} \in \mathbb{R}^m$:

$$\mathbf{z} = \mathbf{H}\mathbf{y} \quad (1)$$

and where $\mathbf{z} \in \mathbb{R}^N$ and $\mathbf{H} \in \mathbb{R}^{N \times m}$ are known. Denote the column space of a matrix \mathbf{M} by $\text{colsp}(\mathbf{M})$. If $\mathbf{z} \in \text{colsp}(\mathbf{H})$, then the equation (1) always has (one or many) exact solutions. If $\mathbf{z} \notin \text{colsp}(\mathbf{H})$, the least squares solution is defined by the solution of the following optimization problem:

$$\min_{\mathbf{y} \in \mathbb{R}^m} \|\mathbf{z} - \mathbf{H}\mathbf{y}\|^2. \quad (2)$$

It is well known that if $\text{rank}(\mathbf{H}) = m$, then (2) yields a unique solution $\mathbf{y}^* = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{z}$.

2.2 Network

Denote

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^\top \\ \vdots \\ \mathbf{h}_N^\top \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}.$$

We can rewrite (1) as

$$\mathbf{h}_i^\top \mathbf{y} = z_i, \quad i = 1, \dots, N.$$

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a constant, undirected and connected graph with the set of nodes $\mathcal{V} = \{1, 2, \dots, N\}$ and the set of edges $\mathcal{E} \subset \{\{i, j\} : i \neq j \in \mathcal{V}\}$. Each node i holds the equation $\mathbf{h}_i^\top \mathbf{y} = z_i$ and also holds a vector $\mathbf{x}_i(t) \in \mathbb{R}^m$ that varies as a function of time t . Note that $\mathbf{x}_i(t)$ will turn out to be part of the state of node i at time t . Let \mathcal{N}_i be the set of neighbor nodes that are connected to node i , i.e., $\mathcal{N}_i = \{j : \{i, j\} \in \mathcal{E}\}$. Define a diagonal matrix $\mathbf{D} = \text{diag}(|\mathcal{N}_1|, |\mathcal{N}_2|, \dots, |\mathcal{N}_N|)$ and an incidence matrix \mathbf{A} of the graph \mathcal{G} by $[\mathbf{A}]_{ij} = 1$ if $\{i, j\} \in \mathcal{E}$ and $[\mathbf{A}]_{ij} = 0$ otherwise. Then $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian of graph \mathcal{G} .

2.3 Distributed Flows

Consider a cost function $U : \mathbb{R}^m \times \dots \times \mathbb{R}^m \rightarrow \mathbb{R}$

$$U(\mathbf{x}_1, \dots, \mathbf{x}_N) = \frac{1}{2} \sum_{i=1}^N |\mathbf{h}_i^\top \mathbf{x}_i - z_i|^2. \quad (3)$$

Let $\mathbf{x}(t) = [\mathbf{x}_1^\top(t) \dots \mathbf{x}_N^\top(t)]^\top$ and introduce $\mathbf{v}(t) = [\mathbf{v}_1^\top(t) \dots \mathbf{v}_N^\top(t)]^\top$ with $\mathbf{v}_i(t) \in \mathbb{R}^m$ for $i = 1, \dots, N$.

The vector $\mathbf{v}_i(t)$ is also held by node i , and the $2m$ -dimensional vector $[\mathbf{x}_i(t)^\top \mathbf{v}_i(t)^\top]^\top$ represents the state of node i . Evidently, the least squares optimization problem (2) is equivalent to

$$\begin{aligned} \min_{\mathbf{x}} \quad & U(\mathbf{x}) \\ \text{s.t.} \quad & (\mathbf{L} \otimes \mathbf{I}_m)\mathbf{x} = 0 \end{aligned} \quad (4)$$

We consider the following differential equations describing a network flow:

$$\begin{aligned} \dot{\mathbf{x}} &= -(\mathbf{L} \otimes \mathbf{I}_m)\mathbf{v} - \nabla U(\mathbf{x}) \\ \dot{\mathbf{v}} &= (\mathbf{L} \otimes \mathbf{I}_m)\mathbf{x}, \end{aligned} \quad (5)$$

with initial values $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^{Nm}$, $\mathbf{v}(0) = \mathbf{v}_0 \in \mathbb{R}^{Nm}$. Note that $\nabla U = \mathbf{H}\mathbf{x} - \mathbf{z}_\mathbf{H}$ is Lipschitz continuous where $\mathbf{H} = \text{diag}(\mathbf{h}_1\mathbf{h}_1^\top, \dots, \mathbf{h}_N\mathbf{h}_N^\top)$, $\mathbf{z}_\mathbf{H} = [z_1\mathbf{h}_1^\top \dots z_N\mathbf{h}_N^\top]^\top$. In the flow (5), the state variable $[\mathbf{x}_i^\top(t) \mathbf{v}_i^\top(t)]^\top$ of node i obeys the evolution

$$\begin{aligned} \dot{\mathbf{x}}_i(t) &= - \sum_{j \in \mathcal{N}_i} (\mathbf{v}_i(t) - \mathbf{v}_j(t)) \\ &\quad - (\mathbf{h}_i\mathbf{h}_i^\top \mathbf{x}_i(t) - z_i\mathbf{h}_i) \\ \dot{\mathbf{v}}_i(t) &= \sum_{j \in \mathcal{N}_i} (\mathbf{x}_i(t) - \mathbf{x}_j(t)). \end{aligned} \quad (6)$$

Therefore, besides the equation $\mathbf{h}_i^\top \mathbf{y} = z_i$ that node i possesses, it only needs to communicate with its neighbors to obtain their states in order to implement (5). The flow (5) is *distributed* in this sense.

2.4 Discussions

2.4.1 Relation to A-H-U Flow

Consider a constrained optimization problem as

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{F}\mathbf{x} = \mathbf{b} \end{aligned} \quad (7)$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable function, $\mathbf{F} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. The well-known Arrow-Hurwicz-Uzawa (A-H-U) flow introduced in Arrow et al. (1958) provides under appropriate conditions a continuous-time solver defined by

$$\begin{aligned} \dot{\mathbf{x}} &= -\nabla f(\mathbf{x}) - \mathbf{F}^\top \mathbf{v} \\ \dot{\mathbf{v}} &= \mathbf{F}\mathbf{x} - \mathbf{b}. \end{aligned} \quad (8)$$

Convergence properties of (8) are studied in many works, e.g., Arrow et al. (1958), Wang et al. (2011), Holding et al. (2014), Cherukuri et al. (2017). In particular, if f is strictly convex and \mathbf{F} has full row rank, then (see Arrow et al. (1958) Wang et al. (2011)) along the flow (8), $\mathbf{x}(t)$ will converge to the unique minimizer of (7) and $\mathbf{v}(t)$ will converge to a Lagrangian multiplier of (7).

As one can see, the flow (5) is a form of the A-H-U flow (8) with the cost function f being the given U and the constraint $\mathbf{F}\mathbf{x} = \mathbf{b}$ given by $(\mathbf{L} \otimes \mathbf{I}_m)\mathbf{x} = 0$. However, the Laplacian \mathbf{L} is not a full-rank matrix. Therefore, the sufficiency results and analysis for the A-H-U flow established in Wang et al. (2011) cannot be applied directly to the flow (5).

2.4.2 Relation to Wang-Elia/Gharesifard-Cortés Flows

In Wang et al. (2010) and Gharesifard et al. (2014), a distributed algorithm that can solve (4) was proposed as following:

$$\begin{aligned} \dot{\mathbf{x}}_i(t) &= -\alpha \sum_{j \in \mathcal{N}_i} (\mathbf{x}_i(t) - \mathbf{x}_j(t)) \\ &\quad - \sum_{j \in \mathcal{N}_i} (\mathbf{v}_i(t) - \mathbf{v}_j(t)) - (\mathbf{h}_i\mathbf{h}_i^\top \mathbf{x}_i(t) - z_i\mathbf{h}_i) \\ \dot{\mathbf{v}}_i(t) &= \sum_{j \in \mathcal{N}_i} (\mathbf{x}_i(t) - \mathbf{x}_j(t)). \end{aligned} \quad (9)$$

where α is a positive number. With fixed interaction graph and suitable choice of α , the flow (9) is a least squares solver for (1) even for balanced directed networks (Theorem 5.4, Gharesifard et al. (2014)). In Figure 1,

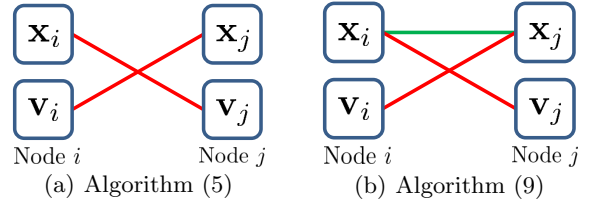


Fig. 1. The way that two neighbor nodes communicate for (a) the proposed flow (5), (b) Wang-Elia/Gharesifard-Cortés flow (9).

we show the way that two neighbor nodes i and j share their states under the assumption that the dynamics of $\mathbf{x}_i(t)$ and $\mathbf{v}_i(t)$ are computed by two individual processors. It can be seen that under the flow (5), external information transmission occurs only between $\mathbf{x}_i(t)$ and $\mathbf{v}_j(t)$, $\mathbf{x}_j(t)$ and $\mathbf{v}_i(t)$. However, information transmission occurs additionally between $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t)$ under flow (9). Therefore, communication cost is reduced with (5). Moreover, it is also useful to establish a clear understanding of the convergence conditions of the system (5) under general conditions, which are currently missing in the literature.

3 Continuous Flow

3.1 Convergence Result

In the following, we provide necessary and sufficient condition for (5) to yield convergence.

Theorem 1 Assume that $N > m$ and $\text{rank}(\mathbf{H}) = m$. Let $\mathbf{y}^* = (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{z}$ be the unique least squares solution of (1). Define $\mathcal{S}_{\mathbf{L}}$ as the set of all complex eigenvectors of \mathbf{L} and for $\boldsymbol{\alpha} \in \mathcal{S}_{\mathbf{L}}$ with $\alpha[i]$ denoting the i -th entry,

$$\mathcal{I}_{\boldsymbol{\alpha}} := \{i : \alpha[i] \neq 0, \boldsymbol{\alpha} = [\alpha[1] \ \alpha[2] \ \dots \ \alpha[N]]^\top\}.$$

(i) If $\text{span}(\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}) = \mathbb{R}^m$ for all $\boldsymbol{\alpha} \in \mathcal{S}_{\mathbf{L}}$, then along (5)

$$\lim_{t \rightarrow \infty} \mathbf{x}_i(t) = \mathbf{y}^*, i = 1, \dots, N$$

$$\lim_{t \rightarrow \infty} \mathbf{v}_i(t) = \mathbf{v}_i^*, i = 1, \dots, N$$

where $[\mathbf{v}_1^* \ \dots \ \mathbf{v}_N^*]^\top$ is a Lagrange multiplier associated with the optimization problem (4). Moreover, the convergence rates of both $\mathbf{x}_i(t)$ and $\mathbf{v}_i(t)$ are exponential.

(ii) If there exists $\boldsymbol{\alpha} \in \mathcal{S}_{\mathbf{L}}$ such that $\dim(\text{span}(\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}})) < m$, then there exist trajectories of $\mathbf{x}(t)$ along (5) which are oscillatory.

PROOF. Recall $\nabla U(\mathbf{x}) = \tilde{\mathbf{H}}\mathbf{x} - \mathbf{z}_{\mathbf{H}}$. Suppose there exists an equilibrium $(\mathbf{x}^*, \mathbf{v}^*)$ of (5), i.e.

$$\begin{aligned} 0 &= -(\mathbf{L} \otimes \mathbf{I}_m) \mathbf{v}^* - \tilde{\mathbf{H}} \mathbf{x}^* + \mathbf{z}_{\mathbf{H}} \\ 0 &= (\mathbf{L} \otimes \mathbf{I}_m) \mathbf{x}^*. \end{aligned} \quad (10)$$

It is worth noting that (10) specifies exactly the Karush-Kuhn-Tucker conditions on $(\mathbf{x}^*, \mathbf{v}^*)$ for the optimization problem (4) (Bertsekas. (1999)). Since U is a convex function and the constraints in (4) are equalities, Slater's condition holds (Boyd et al. (2004)). Therefore \mathbf{x}^* is an optimal solution to (4) and any optimal solution of (4) must have the form $\mathbf{1} \otimes \mathbf{y}^*$. Since \mathbf{y}^* is unique, \mathbf{x}^* is also unique. Note however that \mathbf{v}^* is not necessarily unique. Define the variables $\hat{\mathbf{x}} = \mathbf{x} - \mathbf{x}^*$, $\hat{\mathbf{v}} = \mathbf{v} - \mathbf{v}^*$. Then

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= -(\mathbf{L} \otimes \mathbf{I}_m) \hat{\mathbf{v}} - \tilde{\mathbf{H}} \hat{\mathbf{x}} \\ \dot{\hat{\mathbf{v}}} &= (\mathbf{L} \otimes \mathbf{I}_m) \hat{\mathbf{x}}. \end{aligned} \quad (11)$$

Denote $\hat{\mathbf{u}}(t) = [\hat{\mathbf{x}}(t)^\top \ \hat{\mathbf{v}}(t)^\top]^\top$ and

$$\mathbf{M} = \begin{bmatrix} -\tilde{\mathbf{H}} & -\mathbf{L} \otimes \mathbf{I}_m \\ \mathbf{L} \otimes \mathbf{I}_m & 0 \end{bmatrix}. \quad (12)$$

Then (11) is a linear system with the form $\dot{\hat{\mathbf{u}}} = \mathbf{M}\hat{\mathbf{u}}$. Consider the following Lyapunov function:

$$V(\hat{\mathbf{x}}, \hat{\mathbf{v}}) = \frac{1}{2} \|\hat{\mathbf{u}}\|^2 = \frac{1}{2} (\|\hat{\mathbf{x}}\|^2 + \|\hat{\mathbf{v}}\|^2).$$

Since

$$\begin{aligned} \dot{V} &= -\hat{\mathbf{x}}^\top (\mathbf{L} \otimes \mathbf{I}_m) \hat{\mathbf{v}} - \hat{\mathbf{x}}^\top \tilde{\mathbf{H}} \hat{\mathbf{x}} + \hat{\mathbf{v}}^\top (\mathbf{L} \otimes \mathbf{I}_m) \hat{\mathbf{x}} \\ &= -\hat{\mathbf{x}}^\top \tilde{\mathbf{H}} \hat{\mathbf{x}} \leq 0, \end{aligned} \quad (13)$$

$\hat{\mathbf{u}}(t)$ is bounded for any finite initial values $\hat{\mathbf{x}}(0)$, $\hat{\mathbf{v}}(0)$, namely $\hat{\mathbf{u}}(0)$. Therefore, we conclude:

C1. $\Re(\lambda) \leq 0$ for all $\lambda \in \sigma(\mathbf{M})$.

C2. If $\Re(\lambda) = 0$, then λ has equal algebraic and geometric multiplicity.

(i). Suppose $\text{span}(\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}) = \mathbb{R}^m$ for all $\boldsymbol{\alpha} \in \mathcal{S}_{\mathbf{L}}$. We proceed to prove the convergence of $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{v}}(t)$. The proof contains two steps.

Step 1. We prove \mathbf{M} does not have a purely imaginary eigenvalue using a contradiction argument. Suppose $\lambda = ir \neq 0$ where $r \in \mathbb{R}$ is an eigenvalue of \mathbf{M} with a corresponding eigenvector $\boldsymbol{\beta} = [\beta_a^\top \ \beta_b^\top] \in \mathbb{C}^{2Nm}$, where $\beta_a \in \mathbb{C}^{Nm}$, $\beta_b \in \mathbb{C}^{Nm}$. Let $\hat{\mathbf{u}}(0) = \boldsymbol{\beta}$. Then

$$\hat{\mathbf{u}}(t) = e^{\mathbf{M}t} \hat{\mathbf{u}}(0) = e^{irt} \hat{\mathbf{u}}(0).$$

Therefore, $\|\hat{\mathbf{u}}(t)\|^2 = \|\hat{\mathbf{u}}(0)\|^2$ for all t .

On the other hand, according to (13),

$$\frac{d}{dt} \left(\frac{1}{2} \|\hat{\mathbf{u}}(t)\|^2 \right) = -e^{2rt} \beta_a^\top \tilde{\mathbf{H}} \beta_a.$$

Consequently, there must hold $\tilde{\mathbf{H}} \beta_a = 0$. Next, based on $\mathbf{M} \boldsymbol{\beta} = ir \boldsymbol{\beta}$, we know

$$\begin{aligned} -(\mathbf{L} \otimes \mathbf{I}_m) \beta_b &= ir \beta_a \\ (\mathbf{L} \otimes \mathbf{I}_m) \beta_a &= ir \beta_b. \end{aligned} \quad (14)$$

Since $\boldsymbol{\beta} \neq 0$, neither of β_a nor β_b can be zero. By simple calculation, we have

$$\begin{aligned} (\mathbf{L} \otimes \mathbf{I}_m)^2 \beta_a &= r^2 \beta_a \\ (\mathbf{L} \otimes \mathbf{I}_m)^2 \beta_b &= r^2 \beta_b, \end{aligned} \quad (15)$$

i.e., β_a and β_b are both eigenvectors of $(\mathbf{L} \otimes \mathbf{I}_m)^2$ corresponding to r^2 . From (15), we know

$$(\mathbf{L}^2 \otimes \mathbf{I}_m) \beta_a = r^2 \beta_a.$$

Based on the properties for eigenvectors of the Kronecker product of two matrices (Theorem 13.12 Laud. (2005)), we know there exist $(r^2, \boldsymbol{\alpha}_a)$ and $\boldsymbol{\eta}_a$ such that $\mathbf{L}^2 \boldsymbol{\alpha}_a = r^2 \boldsymbol{\alpha}_a$ and $\beta_a = \boldsymbol{\alpha}_a \otimes \boldsymbol{\eta}_a$ with $\boldsymbol{\alpha}_a \in \mathbb{C}^N$ and $\boldsymbol{\eta}_a \in \mathbb{C}^m$. It is trivial that if $\mathbf{L}^2 \boldsymbol{\alpha}_a = r^2 \boldsymbol{\alpha}_a$, $\mathbf{L} \boldsymbol{\alpha}_a = |r| \boldsymbol{\alpha}_a$, i.e., $\boldsymbol{\alpha}_a$ is an eigenvector of \mathbf{L} corresponding to eigenvalue $|r|$. Denote $\boldsymbol{\beta}_a = [\beta_a[1]^\top \ \dots \ \beta_a[N]^\top]^\top$, $\beta_a[i] \in \mathbb{C}^m$ and

$\boldsymbol{\eta}_a = [\boldsymbol{\eta}_a[1]^\top \dots \boldsymbol{\eta}_a[N]^\top]^\top$, $\boldsymbol{\eta}_a[i] \in \mathbb{C}$ with $i = 1, \dots, N$. It is evident that $\boldsymbol{\beta}_a[i] = \boldsymbol{\alpha}_a[i]\boldsymbol{\eta}_a$ if $i \in \mathcal{I}_{\boldsymbol{\alpha}_a}$ and $\boldsymbol{\beta}_a[i] = 0$ otherwise. Then noting that

$$\tilde{\mathbf{H}}\boldsymbol{\beta}_a = \begin{bmatrix} \boldsymbol{\alpha}_a[1]\mathbf{h}_1\mathbf{h}_1^\top\boldsymbol{\eta}_a \\ \vdots \\ \boldsymbol{\alpha}_a[N]\mathbf{h}_N\mathbf{h}_N^\top\boldsymbol{\eta}_a \end{bmatrix} = 0,$$

we get $\boldsymbol{\alpha}_a[i]\mathbf{h}_i\mathbf{h}_i^\top\boldsymbol{\eta}_a = 0$ for $i = 1, 2, \dots, N$, which implies that

$$\mathbf{h}_i^\top\boldsymbol{\eta}_a = 0, \quad i \in \mathcal{I}_{\boldsymbol{\alpha}_a}. \quad (16)$$

Because $\text{span}(\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}_a}) = \mathbb{R}^m$, there must hold $\boldsymbol{\eta}_a = 0$. In turn, $\boldsymbol{\beta}_a$ must be zero, leading to $\boldsymbol{\beta}_b = 0$ with (14). Therefore \mathbf{M} does not have purely imaginary eigenvalues. Based on **C1**, **C2** and the fact that \mathbf{M} has no purely imaginary eigenvalue, it follows that $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{v}}(t)$ converge. Furthermore, by the basic property of linear time-invariant systems, $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{v}}(t)$ exponentially converge, implying that $\mathbf{x}(t)$ and $\mathbf{v}(t)$ exponentially converge.

Step 2. In this step, we establish the limits of $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{v}}(t)$ by studying the zero eigenspace of \mathbf{M} , thereby obtaining the convergence property for $\mathbf{x}(t)$ and $\mathbf{v}(t)$. Suppose $\boldsymbol{\delta} = [\boldsymbol{\delta}_a^\top \boldsymbol{\delta}_b^\top]^\top$ is one of the eigenvectors of \mathbf{M} corresponding to zero eigenvalue with $\boldsymbol{\delta} \in \mathbb{R}^{2Nm}$ and $\boldsymbol{\delta}_a, \boldsymbol{\delta}_b \in \mathbb{R}^{Nm}$, i.e., $\mathbf{M}\boldsymbol{\delta} = 0$. Consider a solution $\hat{\mathbf{u}}(t)$ of (11) with $\hat{\mathbf{u}}(0) = \boldsymbol{\delta}$. We see from the derivative of the Lyapunov function and $\mathbf{M}\boldsymbol{\delta} = 0$ that

$$\tilde{\mathbf{H}}\boldsymbol{\delta}_a = (\mathbf{L} \otimes \mathbf{I}_m)\boldsymbol{\delta}_a = (\mathbf{L} \otimes \mathbf{I}_m)\boldsymbol{\delta}_b = 0.$$

Then there exist $\boldsymbol{\eta}_a \in \mathbb{R}^m$ and $\boldsymbol{\eta}_b \in \mathbb{R}^m$ such that $\boldsymbol{\delta}_a = \mathbf{1} \otimes \boldsymbol{\eta}_a$ and $\boldsymbol{\delta}_b = \mathbf{1} \otimes \boldsymbol{\eta}_b$. Since $\tilde{\mathbf{H}}\boldsymbol{\delta}_a = 0$ and $\text{rank}(\tilde{\mathbf{H}}) = m$, $\boldsymbol{\delta}_a = 0$, i.e., $\boldsymbol{\delta}$ must be in the form $\boldsymbol{\delta} = [0 \ \boldsymbol{\delta}_b^\top]^\top$ with $\boldsymbol{\delta}_b = \mathbf{1} \otimes \boldsymbol{\eta}_b$. Note that the algebraic and geometric multiplicity of the zero eigenvalue of \mathbf{M} is m . Now we decompose \mathbf{M} into its Jordan canonical form $\mathbf{M} = \mathbf{T}\mathbf{J}\mathbf{T}^{-1}$:

$$\mathbf{T} = [\boldsymbol{\delta}_1 \ \boldsymbol{\delta}_2 \ \dots \ \boldsymbol{\delta}_m \ \dots],$$

$$\mathbf{T}^{-1} = [\boldsymbol{\delta}'_1 \ \boldsymbol{\delta}'_2 \ \dots \ \boldsymbol{\delta}'_m \ \dots]^\top,$$

where $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}'_i$ with $i = 1, 2, \dots, m$ are mutually orthogonal right and left eigenvectors respectively of \mathbf{M} all corresponding to zero eigenvalues and all with the form of $\boldsymbol{\delta}_i = [0 \ \boldsymbol{\delta}_{ib}^\top]^\top$ and $\boldsymbol{\delta}'_i = [0 \ \boldsymbol{\delta}'_{ib}^\top]$. Then

$$\lim_{t \rightarrow \infty} \hat{\mathbf{u}}(t) = \sum_{i=1}^m \boldsymbol{\delta}_i \boldsymbol{\delta}'_i^\top \hat{\mathbf{u}}(0),$$

which implies that

$$\lim_{t \rightarrow \infty} \hat{\mathbf{x}}(t) = 0,$$

$$\lim_{t \rightarrow \infty} \hat{\mathbf{v}}(t) = \sum_{i=1}^m \boldsymbol{\delta}_{ib} \boldsymbol{\delta}'_{ib}^\top \hat{\mathbf{v}}(0).$$

Thus we can conclude that $\mathbf{x}(t)$ converges to $\mathbf{x}^* = \mathbf{1} \otimes \mathbf{y}^*$ while $\mathbf{v}(t)$ converges to a constant associated with the initial value $\mathbf{v}(0)$, which completes the proof of (i).

(ii). Suppose there exists $\boldsymbol{\alpha}_a \in \mathcal{S}_{\mathbf{L}}$ with $\mathbf{L}\boldsymbol{\alpha}_a = r\boldsymbol{\alpha}_a$ such that $\dim(\text{span}(\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}_a})) < m$. Then there must exist $\boldsymbol{\eta}_a \neq 0$ satisfying that

$$\mathbf{h}_i^\top \boldsymbol{\eta}_a = 0, \quad i \in \mathcal{I}_{\boldsymbol{\alpha}_a}$$

Let $\boldsymbol{\beta} = [\boldsymbol{\beta}_a^\top \ \boldsymbol{\beta}_b^\top]^\top$ with $\boldsymbol{\beta}_a = \boldsymbol{\alpha}_a \otimes \boldsymbol{\eta}_a$ and $\boldsymbol{\beta}_b = \frac{(\mathbf{L} \otimes \mathbf{I}_m)\boldsymbol{\beta}_a}{wr}$. It is easy to check that

$$\mathbf{M}\boldsymbol{\beta} = \begin{bmatrix} -\tilde{\mathbf{H}} & -\mathbf{L} \otimes \mathbf{I}_m \\ \mathbf{L} \otimes \mathbf{I}_m & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \end{bmatrix} = wr \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \end{bmatrix}. \quad (17)$$

Therefore, \mathbf{M} has a purely imaginary eigenvalue. Hence, $\mathbf{x}(t)$ and $\mathbf{v}(t)$ are oscillatory for generic initial conditions.

We have now completed the proof of Theorem 1.

3.2 Feasibilities

The verification of the convergence condition in Theorem 1 is noted to be difficult, especially in a distributed manner. Now we discuss the feasibility of graphs and linear equations to show that the convergence condition verification is unlikely to be a major concern.

3.2.1 Weighted Laplacian

An N -node weighted undirected graph $\mathcal{G}_w = (\mathcal{V}, \mathcal{E}, w)$ is defined by associating $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a function $w : \mathcal{E} \rightarrow \mathbb{R}^+$. Introduce $[\mathbf{A}_w]_{ij} = w(\{i, j\})$ if $\{i, j\} \in \mathcal{E}$ and $[\mathbf{A}_w]_{ij} = 0$ otherwise. Then its weighted Laplacian is given by

$$\mathbf{L}_w = \text{diag}\left(\sum_{j=1}^N [\mathbf{A}_w]_{1j}, \dots, \sum_{j=1}^N [\mathbf{A}_w]_{Nj}\right) - \mathbf{A}_w.$$

Let $\mathcal{S}_{\mathbf{L}_w}$ be the set of all complex eigenvectors of \mathbf{L}_w . For $\boldsymbol{\alpha}_w \in \mathcal{S}_{\mathbf{L}_w}$, we define

$$\mathcal{I}_{\boldsymbol{\alpha}_w} = \{i : \boldsymbol{\alpha}_w[i] \neq 0, \boldsymbol{\alpha}_w = [\boldsymbol{\alpha}_w[1] \ \dots \ \boldsymbol{\alpha}_w[N]]^\top\}.$$

Evidently, if we replace \mathbf{L} with \mathbf{L}_w in (5), Theorem 1 continues to hold for $\mathcal{S}_{\mathbf{L}_w}$ and $\mathcal{I}_{\boldsymbol{\alpha}_w}$. For a graph that

does not satisfy the convergence condition in Theorem 1, there may exist weighted graphs with the same topology on which (5) yields convergent flows. This adds flexibility to guarantee convergence with a given network structure.

3.2.2 Generic Feasibility of \mathbf{H}

Proposition 2 Let \mathbf{L} be the Laplacian of a graph \mathcal{G} with $|\mathcal{I}_\alpha| \geq m$ for all $\alpha \in \mathcal{S}_\mathbf{L}$. Then the convergence condition of Theorem 1, viz. $\text{span}(\mathbf{h}_i : i \in \mathcal{I}_\alpha) = \mathbb{R}^m$, is satisfied for generic $\mathbf{H} \in \mathbb{R}^{N \times m}$, i.e., there does not exist an open nonempty subset in $\mathbb{R}^{N \times m}$ of \mathbf{H} , for which the convergence condition in Theorem 1 is not satisfied.

PROOF. Let

$$\mathbf{H}_{i_1, \dots, i_m} = \begin{bmatrix} \mathbf{h}_{i_1}^\top \\ \vdots \\ \mathbf{h}_{i_m}^\top \end{bmatrix}, \quad 1 \leq i_1 < \dots < i_m \leq N.$$

Introduce

$$\mathcal{W} = \bigcup_{\substack{i_1, \dots, i_m: \\ 1 \leq i_1 < \dots < i_m \leq N}} \mathcal{W}_{i_1, \dots, i_m}$$

where

$$\mathcal{W}_{i_1, \dots, i_m} = \{\mathbf{H} \in \mathbb{R}^{N \times m} : \det(\mathbf{H}_{i_1, \dots, i_m}) = 0\}.$$

It can be noted that \mathcal{W} is the set of \mathbf{H} for which the convergence condition in Theorem 1 is not satisfied. Since $\mathcal{W}_{i_1, \dots, i_m} \neq \mathbb{R}^{N \times m}$, by identity theorem for holomorphic functions (Ablowitz et al. (2003)), $\mathcal{W}_{i_1, \dots, i_m}$ does not contain a nonempty open subset of $\mathbb{R}^{N \times m}$, i.e., \mathcal{W} does not contain a nonempty open subset of $\mathbb{R}^{N \times m}$. This completes the proof.

3.2.3 Graph Feasibility

Theorem 1 shows that the larger $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha|$ is, the easier that convergence is met. Now we investigate $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha|$ on several fundamental graphs and random graphs.

[Path Graph] It is known from Fuhrmann et al. (2015) that all the eigenvalues of its Laplacian \mathbf{L} are distinct with eigenvectors in the set of $\mathcal{S}_\mathbf{L} = \{\alpha_k : \alpha_k[v] = \cos(\frac{(k-1)(2v-1)\pi}{2N}), v = 1, \dots, N; k = 1, \dots, N\}$. We discuss two cases:

- (i) Let $N = 2^l$, $l = 2, 3, 4, \dots$. Then it is obvious that there do not exist v and k such that $\alpha_k[v] = 0$. Therefore, $|\mathcal{I}_\alpha| = N$ for all α .

- (ii) Let $N = 3l$, $l = 1, 2, 3, \dots$. Then any $\alpha_k \in \mathcal{S}_\mathbf{L}$ contains at most l zero entries. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = \frac{2}{3}N$.

[Ring Graph] We know from Fuhrmann et al. (2015) that if N is odd, then zero is the only eigenvalue of multiplicity one with eigenvector $\alpha_1 = [1 \ 1 \ \dots \ 1]^\top$, while all the other eigenvalues have multiplicity two with a basis of two orthogonal eigenvectors

$$\begin{bmatrix} 1 \\ \cos(\frac{2k\pi}{N}) \\ \vdots \\ \cos(\frac{2(N-1)k\pi}{N}) \end{bmatrix}, \quad \begin{bmatrix} 0 \\ \sin(\frac{2k\pi}{N}) \\ \vdots \\ \sin(\frac{2(N-1)k\pi}{N}) \end{bmatrix} \quad (18)$$

with $k = 1, \dots, N-1$. If N is even, then zero and the largest eigenvalue are the only two eigenvalues of multiplicity one with eigenvectors $[1 \ 1 \ \dots \ 1]^\top$ and $[1 \ -1 \ 1 \ \dots \ 1]^\top$ respectively, while all the other eigenvalues have multiplicity two with a basis of two orthogonal eigenvectors with the same form (18) and $k = 1, \dots, N-1$, $k \neq \frac{N}{2}$. Note that the eigenspaces of $k = p$ and $k = q$ are the same if and only if $p + q = N$ and $1 \leq p, q \leq N$.

- (i) If N is a prime number, then any $\alpha \in \mathcal{S}_\mathbf{L}$ contains at most one zero entry. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = N - 1$.
- (ii) If $N = 3l$, $l = 1, 2, 3, \dots$, then any $\alpha \in \mathcal{S}_\mathbf{L}$ contains at most l zero entries. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = \frac{2}{3}N$.
- (iii) If $N = 2^l$, $l = 3, 4, \dots$, then any $\alpha \in \mathcal{S}_\mathbf{L}$ contains at most 2^{l-1} zero entries. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = \frac{1}{2}N$.

[Star Graph] We know that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\alpha_1 = [1 \ \dots \ 1]^\top$, an eigenvalue N of multiplicity one with eigenvector $\alpha_N = [1 - N \ 1 \ \dots \ 1]^\top$ and eigenvalue one with multiplicity $N-2$ and a set of associated eigenvectors $\{\alpha_k : \mathbf{1}^\top \alpha_k = 0, \alpha_k \neq p[1 - N \ 1 \ \dots \ 1]^\top, p \in \mathbb{R}; k = 2, 3, \dots, N-1\}$. Thus α_k has at most $N-2$ zero entries. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = 2$.

[Complete Graph] It is known from Cvetković et al. (1980) that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\alpha_1 = [1 \ \dots \ 1]^\top$ and eigenvalue N with multiplicity $N-1$ and a set of associated eigenvectors $\{\alpha_k : \mathbf{1}^\top \alpha_k = 0; k = 2, 3, \dots, N\}$. Then it can be concluded that α_k has at most $N-2$ zero entries. Therefore, $\min_{\alpha \in \mathcal{S}_\mathbf{L}} |\mathcal{I}_\alpha| = 2$.

[Random Graph] Consider random graphs with $n = 100, 200, 300$ nodes generated by letting every possible edge occur independently with probability

$p = 0.2, 0.5, 0.8$, respectively. For each p and each n , we do 10^4 experiments and find out the value of $\min_{\alpha \in \mathcal{S}_L} |\mathcal{I}_\alpha|$ for each experiment. Let $|\mathcal{I}_{\alpha, \text{ave}}|$ denote the average of $\min_{\alpha \in \mathcal{S}_L} |\mathcal{I}_\alpha|$ for these 10^4 experiments. The calculated result is presented in the following table.

$ \mathcal{I}_{\alpha, \text{ave}} $	$n = 100$	$n = 200$	$n = 300$
$p = 0.2$	95.33	189.84	283.02
$p = 0.5$	95.52	190.58	284.83
$p = 0.8$	95.34	189.86	282.85

[Randomly Weighted Graph] Consider a 4-node complete weighted graph with uniformly distributed edge weights over $[0, 1]$. We draw 10^7 samples and the numerical results show the graph Laplacian has a 99.84% chance of satisfying $\min_{\alpha_w \in \mathcal{S}_L} |\mathcal{I}_{\alpha_w}| = 4$.

For star and complete graphs, there holds that $\min_{\alpha \in \mathcal{S}_L} |\mathcal{I}_\alpha| = 2$. This means that as long as $m > 2$, the sufficient convergence condition in Theorem 1 will not hold. However, the feasibility of non-converging unweighted graphs can be improved by introducing additional edge weights as seen from the random weight graph example.

On the other hand, for path and ring graphs, there holds $\min_{\alpha \in \mathcal{S}_L} |\mathcal{I}_\alpha| \approx \mathcal{O}(N)$, and for random graphs with edge occurrence probability $p = 0.2, 0.5, 0.8$, $|\mathcal{I}_{\alpha, \text{ave}}| \approx \frac{19}{20}N$. Therefore, if $N \gg m$, it is relatively easy for the sufficient condition in Theorem 1 to hold.

4 Discrete-time Algorithm

In this section, we investigate the discrete-time analog of the flow (5). We index time as $k = 0, 1, 2, \dots$ and propose the following algorithm:

$$\begin{aligned} \mathbf{x}_i(k+1) &= \mathbf{x}_i(k) - \epsilon \sum_{j \in \mathcal{N}_i} (\mathbf{v}_i(k) - \mathbf{v}_j(k)) \\ &\quad - \epsilon (\mathbf{h}_i \mathbf{h}_i^\top \mathbf{x}_i(k) - z_i \mathbf{h}_i) \\ \mathbf{v}_i(k+1) &= \mathbf{v}_i(k) + \epsilon \sum_{j \in \mathcal{N}_i} (\mathbf{x}_i(k) - \mathbf{x}_j(k)). \end{aligned} \quad (19)$$

Note that (19) is an Euler approximation of (5). According to the proof of Theorem 1, (5) does not guarantee the convergence of $\mathbf{x}(t)$ and $\mathbf{v}(t)$. Therefore, it cannot be directly concluded that the solution to (19) will converge to the same consensus as (5). Recall that \mathbf{y}^* is the unique least squares solution of (1) and \mathbf{M} is as defined in (12). Then the following result holds.

Theorem 3 Suppose $\text{span}(\mathbf{h}_i : \alpha[i] \neq 0) = \mathbb{R}^m$ for all the eigenvectors $\alpha = [\alpha[1] \dots \alpha[N]]^\top \in \mathbb{C}^N$ of \mathbf{L} . Then there exists a positive constant ϵ^* such that

(i) If $0 < \epsilon < \epsilon^*$, then along (19) we have

$$\lim_{k \rightarrow \infty} \mathbf{x}_i(k) = \mathbf{y}^*, i = 1, \dots, N$$

which converge exponentially for all i . In this case $\mathbf{v}(k)$ converges to a constant.

(ii) If $\epsilon > \epsilon^*$, then along (19) there exist initial values $\mathbf{x}(0)$ and $\mathbf{v}(0)$ under which $[\mathbf{x}(k)^\top \mathbf{v}(k)^\top]^\top$ diverges.

Define $\sigma^*(\mathbf{M}) \subset \sigma(\mathbf{M})$ by $\sigma^*(\mathbf{M}) := \{\lambda \in \sigma(\mathbf{M}) : \Re(\lambda) \neq 0\}$. Then $\epsilon^* = \min_{\lambda \in \sigma^*(\mathbf{M})} \left[-\frac{2\Re(\lambda)}{|\lambda|^2} \right]$.

The proof of Theorem 3 is analogous to the proof of Theorem 1 with the method of analyzing the asymptotic stability of a discrete-time linear system equivalent to (19). It can be shown that there exists $\epsilon^* > 0$ such that (19) is a asymptotic stable linear system if $0 < \epsilon < \epsilon^*$, and is an unstable system if $\epsilon > \epsilon^*$. When $\epsilon = \epsilon^*$, of course $\mathbf{I} + \epsilon^* \mathbf{M}$ might have complex eigenvalues on the unit circle, leading to the possibility of periodic trajectories for $[\mathbf{x}(k)^\top \mathbf{v}(k)^\top]^\top$. From Theorem 3, a small enough ϵ would always guarantee convergence but possibly at a low rate. It can be easily proved that the fastest convergence is given by

$$\bar{\epsilon} = \operatorname{argmax}_{0 < \epsilon < \epsilon^*} \left\{ \max_{\lambda \in \sigma^*(\mathbf{M})} |1 + \epsilon \lambda| \right\}.$$

Inspired by Theorem 1. (ii), we have the following result.

Theorem 4 If there exists $\alpha \in \mathcal{S}_L$ such that

$$\dim(\text{span}(\mathbf{h}_i : \alpha[i] \neq 0)) < m,$$

then for any $\epsilon > 0$, there always exist trajectories $\mathbf{x}(k)$ for the algorithm (19) that diverge as k tends to infinity.

The proof of Theorem 4 can be achieved by showing that matrix $\mathbf{I} + \epsilon \mathbf{M}$ has eigenvalues with modulus larger than one, making (19) an unstable linear system. It can be seen from Theorem 3 and Theorem 4 that apart from the dimensionality condition also given in Theorem 1, the calculation of ϵ^* depends on both the Laplacian \mathbf{L} and matrix \mathbf{H} .

5 Conclusions

We studied the problem of obtaining the least squares solution to a linear algebraic equation using distributed algorithms. Each node has the information of one scalar linear equation and holds a dynamic state. Two

distributed algorithms in continuous time and discrete time respectively were developed as least squares solvers for linear equations. Under certain conditions, all node states can reach a consensus, which gives the least square solution, by exchanging information with neighbors over a network. To show the possibility of employing the proposed algorithms, the feasibility of linear equations and fundamental graphs was discussed. Future directions currently being contemplated include studying the method of verifying the convergence condition in a distributed manner, analyzing the robustness, finding out the approach to distributed computation of the residual vector, which can be of practical interest and modifying the underlying cost function or adding constraints on it to reflect objectives such as outlier suppression.

References

- Ablowitz, M., J., & Fokas, A., S. (2003). *Complex Variables: Introduction and Applications*. Cambridge University Press.
- Anderson, B., & Mou, S., & Morse, A., S., & Helmke, U. (2016). Decentralized gradient algorithm for solution of a linear equation. *Numerical Algebra, Control and Optimization*, 6, 319.
- Arrow, K., J., & Hurwicz, L., & Uzawa, H. (1958). *Studies in Linear and Non-linear Programming*. Stanford University Press.
- Bertsekas, D., P. (1999). *Nonlinear Programming*. Athena Scientific Belmont.
- Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Cherukuri, A., & Gharesifard, B., & Cortes, J. (2017). Saddle-point dynamics: conditions for asymptotic stability of saddle points. *SIAM Journal on Control and Optimization*, 55(1), 486–511.
- Cvetković, D., M., & Doob, M., & Sachs, H. (1980). *Spectra of Graphs: Theory and Application*, volume 87. Academic Pr.
- Fuhrmann, P., A., & Helmke, U. (2015). *The Mathematics of Networks of Linear Systems*. Springer.
- Gharesifard, B., & Cortés, J. (2014). Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3), 781–786.
- Holding, T., & Lestas, I. (2014). On the convergence to saddle points of concave-convex functions, the gradient method and emergence of oscillations. *IEEE 53rd Conference on Decision and Control*, 1143–1148.
- Jakovetić, D., & Xavier, J., & Moura, J., M. (2014). Fast distributed gradient methods. *IEEE Transactions on Automatic Control*, 59(5), 1131–1146.
- Kaczmarz, S. (1937). Angenäherte auflösung von systemen linearer gleichungen. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, 35, 355–357.
- Kose, T. (1956). Solutions of saddle value problems by differential equations. *Econometrica, Journal of the Econometric Society*, 59–70.
- Laub, A., J. (2005) *Matrix Analysis for Scientists and Engineers*. SIAM.
- Liu, J., & Mou, S., & Morse, A., S. (2013). An asynchronous distributed algorithm for solving a linear algebraic equation. *IEEE 52nd Conference on Decision and Control*, 5409–5414.
- Liu, Y., & Lageman, C., & Anderson, B., D., & Shi, G. (2017). Exponential Least Squares Solvers for Linear Equations over Networks. *20th World Congress of the International Federation of Automatic Control*, 50(1), 2543–2548.
- Lu, J., & Tang, C., Y. (2009). Distributed asynchronous algorithms for solving positive definite linear equations over networks—Part I: Agent networks. *IFAC Proceedings Volumes*, 42(20), 252–257.
- Margaris, A., & Souravlas, S., & Roumeliotis, M. (2007). Parallel implementations of the Jacobi linear algebraic systems solve. *In Proc. 3rd Balkan Conference in Informatics*.
- Mou, S., & Morse, A., S. (2013). A fixed-neighbor, distributed algorithm for solving a linear algebraic equation. *European Control Conference*, 2269–2273.
- Mou, S., & Liu, J., & Morse, A., S. (2015). A distributed algorithm for solving a linear algebraic equation. *IEEE Transactions on Automatic Control*, 60(11), 2863–2878.
- Nedic, A., & Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1), 48–61.
- Nedic, A., & Ozdaglar, A., & Parrilo, P., A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4), 922–938.
- Saad, Y., & Sasonkina, M. (1999). Distributed schur complement techniques for general sparse linear systems. *SIAM Journal on Scientific Computing*, 21(4), 1337–1356.
- Shi, G., & Anderson, B., & Helmke, U. (2017). Network flows that solve linear equations. *IEEE Transactions on Automatic Control*, 62(6), 2659–2674.
- Stappen, A., F., & Bisseling, R., H., & Vorst, J., G. (1993). Parallel sparse lu decomposition on a mesh network of transputers. *SIAM Journal on Matrix Analysis and Applications*, 14(3), 853–879.
- Tsitsiklis, J., N., & Bertsekas, D., P., & Athans, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 39(9), 803–812.
- Wang, J., & Elia, N. (2010). Control approach to distributed optimization. *48th Annual Allerton Conference on Communication, Control, and Computing*, 557–561.
- Wang, J., & Elia, N. (2011). A control perspective for centralized and distributed convex optimization. *50th IEEE Conference on Decision and Control and European Control Conference*, 3800–3805.
- Wang, J., & Elia, N. (2012). Distributed least square with intermittent communications. *American Control Conference*, 6479–6484.