Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 2543–2548

Exponential Least Squares Solvers for Linear Equations over Networks

Yang Liu * Christian Lageman **
Brian D. O. Anderson *,***,**** Guodong Shi *

* Research School of Engineering, The Australian National University, Canberra, Australia (E-mail: yang.liu@anu.edu.au, brian.anderson@anu.edu.au, guodong.shi@anu.edu.au).
** Institute of Mathematics, University of Würzburg, Würzburg, Germany (E-mail: christian.lageman@mathematik.uni-wuerzburg.de)
*** DATA61 – CSIRO, Canberra, Australia
**** School of Automation, 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 for the graph Laplacian, regarding whether the continuous-time distributed algorithm can give the least squares solution. The feasibility of different fundamental graphs is discussed including path graph, star 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. The convergence rate is exponential for both the continuous-time and discrete-time algorithms under the established conditions.

© 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Distributed Algorithms, Dynamical Systems, Linear Equations.

1. INTRODUCTION

Systems of linear algebraic equations arise from various practical engineering problems Garland et al. (2008); Keckler et al. (2011); Partl et al. (2011); Preparata and Vuillemin (1981); Elbirt and Paar (2005); Ayari et al. (2016); De Rose et al. (2007). 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 algorithms and distributed algorithms as linear-equation solvers.

Parallel algorithms have been developed, starting many years ago, in the spirit of high-performance computing, including Jacobi method Margaris et al. (2014); Yang and Mittal (2014), successive over relaxations method Young (1954), Kaczmarz method Kaczmarz (1937) and randomized iterative method Gower and Richtárik (2015). In these algorithms, the state of each node can give an entry of the solution to a linear equation after a suitably long time, via successive information exchange with other nodes and parallel independent computing. There are two restrictions in these parallel algorithms. The first restriction is that often each node is implicitly required to have access to all the other ones, i.e., the network graph is naturally complete Margaris et al. (2014); Yang and Mittal (2014); Kaczmarz (1937); Gower and Richtárik (2015). Second, linear equations are restricted in many parallel algorithms. It is somewhat unsatisfactory that, for example in the Jacobi method, a sufficient condition for convergence is that the linear equations must be strictly or irreducibly diagonally dominant.

On the other hand, discrete and continuous-time algorithms for linear equations 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 and Morse (2013); Mou et al. (2015); Liu et al. (2013); Lu and Tang (2009a,b) and continuous-time algorithms are presented in Anderson et al. (2015); Shi et al. (2015). In these network distributed algorithms, compared with the development in parallel computing, each node state asymptotically converges to the solution to the linear equation. However, most of the existing work for parallel and distributed algorithms assumes that the linear equations have exact solutions Margaris et al. (2014); Yang and Mittal (2014); Young (1954); Kaczmarz (1937); Gower and Richtárik (2015); Mou and Morse (2013); Mou et al. (2015); Liu et al. (2013); Lu and Tang (2009a,b); Anderson et al. (2015); Shi et al. (2015). In Wang and Elia (2012), a distributed least squares solver is proposed for networks with stochastically broken communication links. In Shi et al. (2015), a continuous-time flow is shown to be able to calculate approximations to least squares solutions with

2405-8963 © 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2017.08.073

a high gain approach. However, some of them can only produce the least squares solution in approximate sense, or have strong requirements of the network topology, and only a few results have been obtained on exact distributed least squares solvers for network linear equations.

In this paper, distributed continuous and discrete-time algorithms that can compute the least squares solution to a linear equation over a network are presented. By recognizing a least-squares problem for a linear equation as a constrained optimization problem over a network, we propose a continuous-time flow in the form of the classical Arrow-Hurwicz-Uzawa flow Arrow et al. (1958). Necessary and sufficient conditions for the underlying interaction graphs regarding whether the continuous-time network flow converges to the least squares solution are established. Further, by Euler's method, a discrete-time algorithm is presented and the properties of its convergence are also specified and proved.

The paper begins by the formulating the network linear equations in Section 2, in addition to explaining how the Arrow-Hurwicz-Uzawa flow can be used to derive a continuous-time network 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 proposed.

2. PROBLEM DEFINITION

2.1 Linear Equation

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

$$\mathbf{z} = \mathbf{H}\mathbf{y} \tag{1}$$

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 colsp$\{\mathbf{M}\}$. If $\mathbf{z} \in$ colsp$\{\mathbf{H}\}$, then the equation (1) always has (one or many) exact solutions. If $\mathbf{z} \notin$ colsp$\{\mathbf{H}\}$, the least squares solution is defined by the solution of the following optimization problem:

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

It is well known that if 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

$$H = \begin{bmatrix} \mathbf{h}_1^\top \\ \mathbf{h}_2^\top \\ \vdots \\ \mathbf{h}_N^\top \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}.$$

We can rewrite (1) as

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

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a constant, undirected and connected graph with the set of nodes $\mathcal{V} = \{1, 2, \ldots, N\}$ and the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. 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 $\mathbf{D} = \text{diag}(|\mathcal{N}_1|, |\mathcal{N}_2|, \ldots, |\mathcal{N}_N|)$ 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 $U(\cdot) : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}$

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

Let $\mathbf{x}(t) = [\mathbf{x}_1^\top(t) \ \ldots \ \mathbf{x}_N^\top(t)]^\top$ and introduce $\mathbf{v}(t) = [\mathbf{v}_1^\top(t) \ \ldots \ \mathbf{v}_N^\top(t)]^\top$ with $\mathbf{v}_i(t) \in \mathbb{R}^m$ for $i = 1, \ldots, N$. The vector $\mathbf{v}_i(t)$ is also held by node i, and $[\mathbf{x}_i(t)\ \mathbf{v}_i(t)]^\top$ represents the state of node i.

We consider the following flow:

$$\begin{aligned} \dot{\mathbf{x}} &= -(\mathbf{L} \otimes \mathbf{I}_m)\mathbf{v} - \nabla U(\mathbf{x}) \\ \dot{\mathbf{v}} &= (\mathbf{L} \otimes \mathbf{I}_m)\mathbf{x}. \end{aligned} \tag{4}$$

We term (4) as "Oscillation + Gradient Flow" because the equation

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

yields oscillating trajectories for $\mathbf{x}(t)$, while $\dot{\mathbf{x}} = -\nabla U(\mathbf{x})$ is a gradient flow.

Note that in the flow (4), the state variable $[\mathbf{x}_i^\top(t)\ \mathbf{v}_i^\top(t)]^\top$ of node i obeys the evolution

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

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 (4). The flow (4) is *distributed* in that spirit.

2.4 Discussion

Consider a constrained optimization problem as

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{F}\mathbf{x} = \mathbf{b} \end{aligned} \tag{5}$$

where $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ 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

$$\begin{aligned} \dot{\mathbf{x}} &= -\nabla_{\mathbf{x}} f(\mathbf{x}) - \mathbf{F}^\top \mathbf{v} \\ \dot{\mathbf{v}} &= \mathbf{F}\mathbf{x} - \mathbf{b}. \end{aligned} \tag{6}$$

In particular, if f is strictly convex and $\mathbf{F}$ has full rank, then see Arrow et al. (1958)Wang and Elia (2011) along the flow (6), $\mathbf{x}(t)$ will converge to an optimal point of (5) and $\mathbf{v}(t)$ will converge to the unique Lagrangian multiplier of (5).

As one can see, the flow (4) is a form of the A-H-U flow (6) with the cost function $f(\mathbf{x})$ being the given $U(\mathbf{x})$ 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 and Elia (2011) cannot be applied directly to the flow (4).

3. CONTINUOUS FLOW

In this section, we study the behavior of the flow (4) in terms of convergence to a least squares solution for the $\mathbf{x}_i(t)$, and present necessary and sufficient conditions for convergence.

3.1 Convergence Result

We present the following result.

Theorem 1. Assume that $N > m$ and $\mathrm{rank}(\mathbf{H}) = m$. Let $\mathbf{y}^* = (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}\mathbf{z}$ 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,

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

Then :

(i) If $\mathrm{span}\{\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}\} = \mathbb{R}^m$ for all $\alpha \in \mathcal{S}_\mathbf{L}$, there holds
$$\lim_{t\to\infty} \mathbf{x}_i(t) = \mathbf{y}^*, i = 1, \ldots, N$$
along the flow (4). Further $\mathbf{v}(t)$ along (4) converges to a Lagrange multiplier associated with a solution of the optimization problem
$$\begin{aligned} \min_{\mathbf{x}} \quad & U(\mathbf{x}) \\ \text{s.t.} \quad & (\mathbf{L}\otimes\mathbf{I}_m)\mathbf{x} = 0 \end{aligned} \tag{7}$$

(ii) If there exists $\boldsymbol{\alpha} \in \mathcal{S}_\mathbf{L}$ such that $\dim(\mathrm{span}\{\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}\}) < m$, then there exist trajectories of $\mathbf{x}(t)$ along (4) which do not converge.

Proof. By direct calculation, we know $\nabla U(\mathbf{x}) = \tilde{\mathbf{H}}\mathbf{x} - \mathbf{z}_H$ where $\tilde{\mathbf{H}} = \mathrm{diag}(\mathbf{h}_1\mathbf{h}_1^\top, \ldots, \mathbf{h}_N\mathbf{h}_N^\top)$ and $\mathbf{z}_H = [z_1\mathbf{h}_1^\top \ \ldots \ z_H\mathbf{h}_N^\top]^\top$. Then we rewrite (4) as

$$\begin{aligned} \dot{\mathbf{x}} &= -(\mathbf{L}\otimes\mathbf{I}_m)\mathbf{v} - \tilde{\mathbf{H}}\mathbf{x} + \mathbf{z}_H \\ \dot{\mathbf{v}} &= (\mathbf{L}\otimes\mathbf{I}_m)\mathbf{x}. \end{aligned} \tag{8}$$

Suppose there exists an equilibrium $(\mathbf{x}^*, \mathbf{v}^*)$ of (8), i.e.

$$\begin{aligned} 0 &= -(\mathbf{L}\otimes\mathbf{I}_m)\mathbf{v}^* - \tilde{\mathbf{H}}\mathbf{x}^* + \mathbf{z}_H \\ 0 &= (\mathbf{L}\otimes\mathbf{I}_m)\mathbf{x}^*. \end{aligned} \tag{9}$$

It is worth noting that (9) specifies exactly the Karush-Kuhn-Tucker conditions on $(\mathbf{x}^*, \mathbf{v}^*)$ for the optimization problem (7) Bertsekas (1999). Since $U(x)$ is a convex function and the constraints in (7) are equality constraints, Slater's condition holds Boyd and Vandenberghe (2004). Therefore the optimal points of the primal problem and dual problem are the same, i.e., $\mathbf{x}^*$ is an optimal solution to (7) and any optimal solution of (7) must have the form

$$\mathbf{1}\otimes\mathbf{y}^*,$$

where $\mathbf{y}^*$ is a least squares solution to (1). We know $\mathbf{y}^*$ is unique because $\mathrm{rank}(\mathbf{H}) = m$. Since $\mathbf{x}^* = 1\otimes\mathbf{y}^*$, then $\mathbf{x}^*$ 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

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \dot{\mathbf{x}} - \dot{\mathbf{x}}^* = -(\mathbf{L}\otimes\mathbf{I}_m)\hat{\mathbf{v}} - \tilde{\mathbf{H}}\hat{\mathbf{x}} \\ \dot{\hat{\mathbf{v}}} &= (\mathbf{L}\otimes\mathbf{I}_m)\hat{\mathbf{x}}. \end{aligned} \tag{10}$$

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

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

Then (10) 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

$$\begin{aligned} \dot{V} &= -\hat{\mathbf{x}}^\top(\mathbf{L}\otimes\mathbf{I}_m)\hat{\mathbf{v}} - \hat{\mathbf{x}}^\top\tilde{\mathbf{H}}\mathbf{x} + \hat{\mathbf{v}}^\top(\mathbf{L}\otimes\mathbf{I}_m)\hat{\mathbf{x}} \\ &= -\hat{\mathbf{x}}^\top\tilde{\mathbf{H}}\hat{\mathbf{x}} \leq 0, \end{aligned} \tag{11}$$

$\hat{\mathbf{u}}(t)$ 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 $\mathrm{span}\{\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}\} = \mathbb{R}^m$ 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 if $\mathrm{span}\{\mathbf{h}_i : i \in \mathcal{I}_{\boldsymbol{\alpha}}\} = \mathbb{R}^m$ for all $\boldsymbol{\alpha} \in \mathcal{S}_\mathbf{L}$, using a contradiction argument. Suppose $\lambda = \imath r \neq 0$ where $r \in \mathbb{R}$ is an eigenvalue of $\mathbf{M}$ with a corresponding eigenvector $\boldsymbol{\beta} = [\boldsymbol{\beta}_a^\top \ \boldsymbol{\beta}_b^\top] \in \mathbb{C}^{2Nm}$, where $\boldsymbol{\beta}_a \in \mathbb{C}^{Nm}$, $\boldsymbol{\beta}_b \in \mathbb{C}^{Nm}$. Let $\hat{\mathbf{u}}(0) = \boldsymbol{\beta}$. Then

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

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

On the other hand, according to (11),

$$\begin{aligned} \frac{d}{dt}(\frac{1}{2}\|\hat{\mathbf{u}}(t)\|^2) &= -\hat{\mathbf{x}}^\top(t)\tilde{\mathbf{H}}\hat{\mathbf{x}}(t) \\ &= -\hat{\mathbf{x}}^\top(0)e^{\imath rt}\tilde{\mathbf{H}}e^{\imath rt}\hat{\mathbf{x}}(0) \\ &= -e^{\imath 2rt}\boldsymbol{\beta}_a^\top\tilde{\mathbf{H}}\boldsymbol{\beta}_a. \end{aligned}$$

Consequently, there must hold $\tilde{\mathbf{H}}\boldsymbol{\beta}_a = 0$. Next, based on

$$\begin{bmatrix} -\tilde{\mathbf{H}} & -\mathbf{L}\otimes\mathbf{I}_m \\ \mathbf{L}\otimes\mathbf{I}_m & 0 \end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b\end{bmatrix} = \imath r\begin{bmatrix}\boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b\end{bmatrix},$$

we know

$$\begin{aligned} -(\mathbf{L}\otimes\mathbf{I}_m)\boldsymbol{\beta}_b &= \imath r\boldsymbol{\beta}_a \\ (\mathbf{L}\otimes\mathbf{I}_m)\boldsymbol{\beta}_a &= \imath r\boldsymbol{\beta}_b. \end{aligned} \tag{12}$$

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

$$\begin{aligned} (\mathbf{L}\otimes\mathbf{I}_m)^2\boldsymbol{\beta}_a &= r^2\boldsymbol{\beta}_a \\ (\mathbf{L}\otimes\mathbf{I}_m)^2\boldsymbol{\beta}_b &= r^2\boldsymbol{\beta}_b, \end{aligned} \tag{13}$$

i.e., $\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_b$ are both eigenvectors of $(\mathbf{L}\otimes\mathbf{I}_m)^2$ corresponding to r^2. From (13), we know

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

Based on the properties for eigenvectors of the Kronecker product of two matrices (Theorem 13.12 Laub (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 $\boldsymbol{\beta}_a = \boldsymbol{\alpha}_a\otimes\boldsymbol{\eta}_a$ 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

$$\boldsymbol{\beta}_a = \begin{bmatrix}\boldsymbol{\beta}_a^{[1]} \\ \boldsymbol{\beta}_a^{[2]} \\ \vdots \\ \boldsymbol{\beta}_a^{[N]}\end{bmatrix},\ \boldsymbol{\beta}_a^{[i]} \in \mathbb{C}^m,\ i = 1, 2, \ldots, N$$

and

$$\boldsymbol{\eta}_a = \begin{bmatrix} \boldsymbol{\eta}_a[1] \\ \boldsymbol{\eta}_a[2] \\ \vdots \\ \boldsymbol{\eta}_a[N] \end{bmatrix}, \ \boldsymbol{\eta}_a[i] \in \mathbb{C}, \ i = 1, 2, \ldots, N.$$

It is apparent 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

$$\tilde{\mathbf{H}}\boldsymbol{\beta}_a = \begin{bmatrix} \boldsymbol{\alpha}_a[1]\mathbf{h}_1\mathbf{h}_1^\top\boldsymbol{\eta}_a \\ \boldsymbol{\alpha}_a[2]\mathbf{h}_2\mathbf{h}_2^\top\boldsymbol{\eta}_a \\ \vdots \\ \boldsymbol{\alpha}_a[N]\mathbf{h}_N\mathbf{h}_N^\top\boldsymbol{\eta}_a \end{bmatrix} = 0,$$

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

$$\mathbf{h}_i^\top\boldsymbol{\eta}_a = 0, \ i \in \mathcal{I}_{\boldsymbol{\alpha}_a}. \tag{14}$$

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 (12). Therefore $\mathbf{M}$ does not have purely imaginary eigenvalues.

Based on **C1**, **C2** and the fact that $\mathbf{M}$ has no purely imaginary eigenvalue, $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{v}}(t)$ 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 (10) with $\hat{\mathbf{u}}(0) = \boldsymbol{\delta}$. We see from the derivative of the Lyapunov function and $\mathbf{M}\boldsymbol{\delta} = 0$ that

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

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., $\boldsymbol{\delta}$ 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 $\mathbf{M} = \mathbf{TJT}^{-1}$:

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

where $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_i'^\top$ with $i = 1, 2, \ldots, m$ 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 $\boldsymbol{\delta}_i'^\top = [0 \ \boldsymbol{\delta}_{ib}'^\top]$. Then

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

which implies that

$$\lim_{t\to\infty} \hat{\mathbf{x}}(t) = 0;$$
$$\lim_{t\to\infty} \hat{\mathbf{v}}(t) = \sum_{i=1}^{m} \boldsymbol{\delta}_i\boldsymbol{\delta}_i'^\top\hat{\mathbf{v}}(0).$$

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

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

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

$$\begin{aligned} \mathbf{M}\boldsymbol{\beta} &= \begin{bmatrix} -\tilde{\mathbf{H}} & -\mathbf{L} \otimes \mathbf{I}_m \\ \mathbf{L} \otimes \mathbf{I}_m & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \end{bmatrix} \\ &= \begin{bmatrix} -(\mathbf{L} \otimes \mathbf{I}_m)\boldsymbol{\beta}_b \\ (\mathbf{L} \otimes \mathbf{I}_m)\boldsymbol{\beta}_a \end{bmatrix} = \imath r \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \end{bmatrix}. \end{aligned} \tag{15}$$

Therefore, $\mathbf{M}$ has a purely imaginary eigenvalue. Hence, $\mathbf{x}(t)$ and $\mathbf{v}(t)$ do not converge for generic initial conditions.

We have now completed the proof of Theorem 1.

3.2 Graph Feasibility

In this section, we consider a few fundamental graphs to investigate the feasibility of the convergence condition presented in Theorem 1. Suppose $N > 2$. For a number of graphs we will first study determine the minimum value of $|\mathcal{I}_{\boldsymbol{\alpha}}|$. The collection of values and the implications for solvability of the least squares problem will be interpreted for all the graphs at the end of the calculations.

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

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

(ii) Let $N = 3l, \ l = 1, 2, 3, \ldots$. Then any $\boldsymbol{\alpha}_k \in \mathcal{S}_\mathbf{L}$ contains at most l zero entries. Therefore $\min\limits_{\boldsymbol{\alpha}\in\mathcal{S}_\mathbf{L}} |\mathcal{I}_{\boldsymbol{\alpha}}| = \frac{2}{3}N$.

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

$$\begin{bmatrix} 1 \\ \cos\dfrac{2k\pi}{N} \\ \cos\dfrac{4k\pi}{N} \\ \vdots \\ \cos\dfrac{2(N-1)k\pi}{N} \end{bmatrix}, \begin{bmatrix} 0 \\ \sin\dfrac{2k\pi}{N} \\ \sin\dfrac{4k\pi}{N} \\ \vdots \\ \sin\dfrac{2(N-1)k\pi}{N} \end{bmatrix} \tag{16}$$

with $k = 1, \ldots, N-1$. If N is even, then zero and the largest eigenvalue are the only two eigenvalues of multiplicity one with eigenvectors $[1 \ 1 \ \ldots \ 1]^\top$ and $[1 \ -1 \ 1 \ \ldots \ 1]^\top$ respectively, while all the other eigenvalues have multiplicity two with a basis of two orthogonal eigenvectors with the same form (16) and $k = 1, \ldots, N-1, \ k \neq \frac{N}{2}$. 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 $\boldsymbol{\alpha} \in \mathcal{S}_\mathbf{L}$ contains at most one zero entry. Therefore $\min\limits_{\boldsymbol{\alpha}\in\mathcal{S}_\mathbf{L}} |\mathcal{I}_{\boldsymbol{\alpha}}| = N - 1$.

(ii) If $N = 3l, \ l = 1, 2, 3, \ldots$, then any $\boldsymbol{\alpha} \in \mathcal{S}_\mathbf{L}$ contains at most l zero entries. Therefore $\min\limits_{\boldsymbol{\alpha}\in\mathcal{S}_\mathbf{L}} |\mathcal{I}_{\boldsymbol{\alpha}}| = \frac{2}{3}N$.

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

[Star Graph] We know that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\boldsymbol{\alpha}_1 = [1\ \dots\ 1]^\top$, an eigenvalue N of multiplicity one with eigenvector $\boldsymbol{\alpha}_N = [1-N\ 1\ \dots\ 1]^\top$ and eigenvalue one with multiplicity $N-2$ and a set of associated eigenvectors $\{\boldsymbol{\alpha}_k | \mathbf{1}^\top \boldsymbol{\alpha}_k = 0,\ \boldsymbol{\alpha}_k \neq p[1-N\ 1\ \dots\ 1]^\top,\ p \in \mathbb{R};\ k = 2, 3, \dots, N-1\}$. Thus $\boldsymbol{\alpha}_k$ has at most $N-2$ zero entries. Therefore $\min_{\boldsymbol{\alpha}\in\mathcal{S}_{\mathbf{L}}} |\mathcal{I}_{\boldsymbol{\alpha}}| = 2$.

[Complete Graph] It is known from Kelner (2009 (accessed 11/10/2016) that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\boldsymbol{\alpha}_1 = [1\ \dots\ 1]^\top$ and eigenvalue N with multiplicity $N-1$ and a set of associated eigenvectors $\{\boldsymbol{\alpha}_k | \mathbf{1}^\top \boldsymbol{\alpha}_k = 0;\ k = 2, 3, \dots, N\}$. Then it can be concluded that $\boldsymbol{\alpha}_k$ has at most $N-2$ zero entries. Therefore $\min_{\boldsymbol{\alpha}\in\mathcal{S}_{\mathbf{L}}} |\mathcal{I}_{\boldsymbol{\alpha}}| = 2$.

For star and complete graphs, there holds that $\min_{\boldsymbol{\alpha}\in\mathcal{S}_{\mathbf{L}}} |\mathcal{I}_{\boldsymbol{\alpha}}| = 2$. This means that as long as $m > 2$, the sufficient convergence condition in Theorem 1 will not hold. On the other hand, for path and ring graphs,

$$\min_{\boldsymbol{\alpha}\in\mathcal{S}_{\mathbf{L}}} |\mathcal{I}_{\boldsymbol{\alpha}}| \approx \mathcal{O}(N).$$

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 (4). We index time as $k = 0, 1, 2, \dots$ and propose the following algorithm:

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{x}(k) - \epsilon(\mathbf{L}\otimes\mathbf{I}_m)\mathbf{v}(k) - \epsilon\nabla U(\mathbf{x}(k)) \\ \mathbf{v}(k+1) &= \mathbf{v}(k) + \epsilon(\mathbf{L}\otimes\mathbf{I}_m)\mathbf{x}(k). \end{aligned} \tag{17}$$

For $[\mathbf{x}_i^\top(k)\ \mathbf{v}_i^\top(k)]^\top$ held by node i, (17) gives

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

Therefore, the algorithm (17) inherits the same distributed structure as the flow (4). Note that (17) is an Euler approximation of (4). However, since dynamical system (4) does not have all its modes exponentially stable, we cannot immediately conclude that for a sufficiently small ϵ, the solution to (17) will converge to the same consensus as (4).

4.1 Convergence Result

Recall that $\mathbf{y}^*$ is the unique least squares solution of (1) and denote

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

The following result holds.

Theorem 2. Suppose $\operatorname{span}\{\mathbf{h}_i : \boldsymbol{\alpha}[i] \neq 0\} = \mathbb{R}^m$ for all the eigenvectors $\boldsymbol{\alpha} = \big[\boldsymbol{\alpha}[1]\ \dots\ \boldsymbol{\alpha}[N]\big]^\top \in \mathbb{C}^N$ of $\mathbf{L}$. Then there exists a positive constant ϵ^* such that

(i) If $0 < \epsilon < \epsilon^*$, then along (17) we have

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

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

(ii) If $\epsilon > \epsilon^*$, then along (17) there exist initial values $\mathbf{x}(0)$ and $\mathbf{v}(0)$ under which $[\mathbf{x}(k)\ \mathbf{v}(k)]^\top$ diverges.

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

Proof. Let $\mathbf{x}^* = 1\otimes\mathbf{y}^*$ and $\mathbf{v}^*$ satisfy $\nabla U(\mathbf{x}^*) + (\mathbf{L}\otimes\mathbf{I}_m)\mathbf{v}^* = 0$. We continue to use the change of variables defined by $\hat{\mathbf{x}}(k) = \mathbf{x}(k) - \mathbf{x}^*$ and $\hat{\mathbf{v}}(k) = \mathbf{v}(k) - \mathbf{v}^*$ so that the equilibrium of (17) is shifted. We have

$$\begin{bmatrix} \hat{\mathbf{x}}(k+1) \\ \hat{\mathbf{v}}(k+1) \end{bmatrix} = (\mathbf{I} + \epsilon\mathbf{M}) \begin{bmatrix} \hat{\mathbf{x}}(k) \\ \hat{\mathbf{v}}(k) \end{bmatrix}.$$

It is straightforward that

$$\sigma(\mathbf{I} + \epsilon\mathbf{M}) = \{1 + \epsilon\lambda : \lambda \in \sigma(\mathbf{M})\}.$$

and then the eigenvalues of M, coupled with the continuity of $\sigma(1+\epsilon\mathbf{M})$ as a function of ϵ, imply that there exists $\epsilon^* > 0$ such that

(i) when $0 < \epsilon < \epsilon^*$, there hold

- $|\lambda| < 1$ for all $\lambda \in \sigma(1+\epsilon\mathbf{M})$ with $\lambda \neq 1$;
- 1 is an eigenvalue of $1+\epsilon\mathbf{M}$ with equal algebraic and geometric multiplicity.

Moreover, the eigenspace of $1+\epsilon\mathbf{M}$ corresponding to eigenvalue one is the same as the eigenspace of $\mathbf{M}$ corresponding to eigenvalue zero.

Consequently, $[\hat{\mathbf{x}}(k)\ \ \hat{\mathbf{v}}(k)]^\top$ converges to a vector in $\mathbb{R}^{2Nm}$, which implies, together with the structure of the eigenspace for the eigenvalue 1, the desired convergence for $[\mathbf{x}(k)\ \mathbf{v}(k)]^\top$.

(ii) when $\epsilon > \epsilon^*$, there exists $\lambda \in \sigma(1+\epsilon\mathbf{M})$ with $|\lambda| > 1$. Therefore, $[\hat{\mathbf{x}}(k)\ \hat{\mathbf{v}}(k)]^\top$ will diverge for certain initial values, so in turn $[\mathbf{x}(k)\ \mathbf{v}(k)]^\top$ will also diverge.

Finally, we compute the value of ϵ^*. Consider the following set of functions of ϵ: $\sigma_\lambda(\epsilon) = 1 + \epsilon(\Re(\lambda) + \imath\Im(\lambda))$ with $\lambda \in \sigma^*(\mathbf{M})$. According to the definition of ϵ^*, ϵ^* must be the smallest ϵ^*_λ for which

$$|\sigma_\lambda(\epsilon^*_\lambda)| = \sqrt{(\epsilon^*_\lambda\Re(\lambda)+1)^2 + (\epsilon^*_\lambda\Im(\lambda))^2} = 1.$$

Therefore, we conclude $\epsilon^* = \min_{\lambda\in\sigma^*(\mathbf{M})} \big[-\frac{2\Re(\lambda)}{|\lambda|^2}\big]$.

We have now completed the proof of Theorem 2.

When $\epsilon = \epsilon^*$, of course $1+\epsilon^*\mathbf{M}$ might have complex eigenvalues on the unit circle, leading to the possibility of periodic trajectories for $[\mathbf{x}(k)\ \mathbf{v}(k)]^\top$. Based on Theorem 1. (ii), one might also expect that periodic trajectories could occur in discrete time when the dimensionality condition is fulfilled. In contrast however, we have the following result, whose proof is simple and omitted due to space limitations.

Theorem 3. If $\dim(\operatorname{span}\{\mathbf{h}_i : \boldsymbol{\alpha}[i] \neq 0\}) < m$, then for any $\epsilon > 0$, there always exist trajectories $\mathbf{x}(k)$ for the algorithm (17) that diverge as k tends to infinity.

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. Besides, the feasibility of several fundamental graphs was discussed. Future directions currently being contemplated include establishing the convergence rate of the distributed algorithms, distributed identification of the residual vector, which can be of practical interest and modifying the underlying cost function or adding constraints on it to reflect objective such as outlier suppression.

ACKNOWLEDGEMENTS

This work was supported by the DAAD with funds of the German Federal Ministry of Education and Research (BMBF).

REFERENCES

Anderson, B., Mou, S., Morse, A.S., and Helmke, U. (2015). Decentralized gradient algorithm for solution of a linear equation. *arXiv preprint arXiv:1509.04538.*

Arrow, K.J., Hurwicz, L., and Uzawa, H. (1958). *Studies in linear and non-linear programming.* With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford University Press, Stanford, Calif.

Ayari, R., Hafnaoui, I., Aguiar, A., Gilbert, P., Galibois, M., Rousseau, J.P., Beltrame, G., and Nicolescu, G. (2016). Multi-objective mapping of full-mission simulators on heterogeneous distributed multi-processor systems. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 1548512916657907.

Bertsekas, D.P. (1999). *Nonlinear programming.* Athena scientific Belmont.

Boyd, S. and Vandenberghe, L. (2004). In *Convex optimization*, 215–227. Cambridge university press.

De Rose, C.A., Heiss, H.U., and Linnert, B. (2007). Distributed dynamic processor allocation for multicomputers. *Parallel Computing*, 33(3), 145–158.

Elbirt, A.J. and Paar, C. (2005). An instruction-level distributed processor for symmetric-key cryptography. *IEEE Transactions on Parallel and distributed Systems*, 16(5), 468–480.

Fuhrmann, P.A. and Helmke, U. (2015). *The mathematics of networks of linear systems.* Springer.

Garland, M., Le Grand, S., Nickolls, J., Anderson, J., Hardwick, J., Morton, S., Phillips, E., Zhang, Y., and Volkov, V. (2008). Parallel computing experiences with cuda. *Micro, IEEE*, 28(4), 13–27.

Gower, R.M. and Richtárik, P. (2015). Randomized iterative methods for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 36(4), 1660–1690.

Kaczmarz, S. (1937). Angenäherte auflösung von systemen linearer gleichungen. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, 35, 355–357.

Keckler, S.W., Dally, W.J., Khailany, B., Garland, M., and Glasco, D. (2011). Gpus and the future of parallel computing. *IEEE Micro*, 31(5), 7–17.

Kelner, J. (2009 (accessed 11/10/2016)). *Lecture 2.* `https://ocw.mit.edu/courses/mathematics/`.

Laub, A.J. (2005). In *Matrix analysis for scientists and engineers*, 139–148. Siam.

Liu, J., Mou, S., and Morse, A.S. (2013). An asynchronous distributed algorithm for solving a linear algebraic equation. In *52nd IEEE Conference on Decision and Control*, 5409–5414. IEEE.

Lu, J. and Tang, C.Y. (2009a). Distributed asynchronous algorithms for solving positive definite linear equations over networks—part i: Agent networks. *IFAC Proceedings Volumes*, 42(20), 252–257.

Lu, J. and Tang, C.Y. (2009b). Distributed asynchronous algorithms for solving positive definite linear equations over networks—part ii: Wireless networks. *IFAC Proceedings Volumes*, 42(20), 258–263.

Margaris, A., Souravlas, S., and Roumeliotis, M. (2014). Parallel implementations of the jacobi linear algebraic systems solve. *arXiv preprint arXiv:1403.5805.*

Mou, S. and Morse, A. (2013). A fixed-neighbor, distributed algorithm for solving a linear algebraic equation. In *Proc. 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.

Partl, A.M., Maselli, A., Ciardi, B., Ferrara, A., and Müller, V. (2011). Enabling parallel computing in crash. *Monthly Notices of the Royal Astronomical Society*, 414(1), 428–444.

Preparata, F.P. and Vuillemin, J. (1981). The cube-connected cycles: a versatile network for parallel computation. *Communications of the ACM*, 24(5), 300–309.

Shi, G., Anderson, B.D.O., and Helmke, U. (2015). Network Flows that Solve Linear Equations. *ArXiv e-prints.* URL `http://arxiv.org/abs/1510.05176`.

Wang, J. and Elia, N. (2011). A control perspective for centralized and distributed convex optimization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 3800–3805. doi:10.1109/CDC.2011.6161503.

Wang, J. and Elia, N. (2012). Distributed least square with intermittent communications. In *American Control Conference (ACC), 2012*, 6479–6484. IEEE.

Yang, X.I. and Mittal, R. (2014). Acceleration of the jacobi iterative method by factors exceeding 100 using scheduled relaxation. *Journal of Computational Physics*, 274, 695–708.

Young, D. (1954). Iterative methods for solving partial difference equations of elliptic type. *Transactions of the American Mathematical Society*, 76(1), 92–111.