An Arrow–Hurwicz–Uzawa type flow as least squares solver for network linear equations

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

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 and random graph. 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.

Keywords: 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, Souravlas, & Roumeliotis, 2007), the Kaczmarz method (Kaczmarz, 1937) and the parallel algorithms for LU-decomposition in sparse linear systems (Saad & Sosonkina, 1999; Stappening, Bisseling, & Vorst, 1993). 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 Liu, Mou, and Morse (2013), Lu and Tang (2009), Mou, Liu, and Morse (2015) and Mou and Morse (2000) and continuous-time algorithms are presented in Anderson, Mou, Morse, and Helmke (2016) and Shi, Anderson, and Helmke (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 equations are closely related to, and sometimes even special cases of, the study of distributed optimization methods on nonlinear models (Gharesifard & Cortés, 2014; Jakovetić, Xavier, & Moura, 2014; Nedic & Ozdaglar, 2009; Nedic, Ozdaglar, & Parrilo, 2010; Tsitsiklis, Bertsekas, & Athans, 1986), 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 (Anderson et al., 2016; Kaczmarz, 1937; Liu et al., 2013; Lu & Tang, 2009; Margaris et al., 2007; Mou et al., 2015; Mou & Morse, 2000), or can only produce least square solutions in the approximate sense or for limited graph structures (Shi et al., 2017; Wang & Elia, 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 (Arrow, Hurwicz, & Uzawa, 1958; Kose, 1956), 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/Charesifard–Cortés (Charesifard & Cortés, 2014; Wang & Elia, 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, Lageman, Anderson, & Shi, 2017). The paper begins by 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 $y \in \mathbb{R}^m$:

$$z = Hy,$$

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

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

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

2.2. Network

Denote

$$H = \begin{bmatrix} h_1^T \\ \vdots \\ h_N^T \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}.$$ (3)

We can rewrite (1) as

$$h_i^T y = z_i, \quad i = 1, \ldots, N.$$ (4)

Let $G = (V, E)$ be a constant, undirected and connected graph with the set of nodes $V = \{1, 2, \ldots, N\}$ and the set of edges $E \subseteq \{(i, j) : i \neq j \in V\}$. Each node $i$ holds the equation $h_i^T y = z_i$ and also holds a vector $x_i(t) \in \mathbb{R}^m$ that varies as a function of time $t$. Note that $x_i(t)$ will turn out to be part of the state of node $i$ at time $t$. Let $N_i$ be the set of neighbor nodes that are connected to node $i$, i.e., $N_i = \{j : (i, j) \in E\}$. Define a diagonal matrix $D = \text{diag}(|N_1|, |N_2|, \ldots, |N_N|)$ and an incidence matrix $A$ of the graph $G$ by $[A]_{ij} = 1$ if $(i, j) \in E$ and $[A]_{ij} = 0$ otherwise. Then $L = D - A$ is the Laplacian of graph $G$.

2.3. Distributed flows

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

$$U(x_1, \ldots, x_N) = \frac{1}{2} \sum_{i=1}^N ||h_i^T x_i - z_i||^2.$$ (5)

Let $x(t) = [x_1^T(t) \ldots x_N^T(t)]^T$ and introduce $v(t) = [v_1^T(t) \ldots v_N^T(t)]^T$ with $v_i(t) \in \mathbb{R}^m$ for $i = 1, \ldots, N$. The vector $v(t)$ is also held by node $i$, and the $2m$-dimensional vector $[x_i^T(t) v_i^T(t)]^T$ represents the state of node $i$. Evidently, the least squares optimization problem (2) is equivalent to

$$\min_x \quad U(x)$$

s.t. $$(L \otimes I_m)x = 0.$$ (6)

We consider the following differential equations describing a network flow:

$$\dot{x} = -(L \otimes I_m)v - \nabla U(x)$$

$$\dot{v} = (L \otimes I_m)x,$$

with initial values $x(0) = x_0 \in \mathbb{R}^{Nm}, v(0) = v_0 \in \mathbb{R}^{Nm}$. Note that $\nabla U = Hx - z_0$ is Lipschitz continuous where $H = \text{diag}(h_1^T, \ldots, h_N^T)$. In the flow (5), the state variable $[x_i^T(t) v_i^T(t)]^T$ of node $i$ obeys the evolution

$$\dot{x}_i(t) = -\sum_{j \in N_i} (v_j(t) - v_i(t)) - (h_i^T x_i(t) - z_i h_i)$$

$$\dot{v}_i(t) = \sum_{j \in N_i} (x_j(t) - x_i(t)).$$ (7)

Therefore, besides the equation $h_i^T y = z_i$ 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

$$\min_x \quad f(x)$$

s.t. $Fx = b$. (8)

$$\min_x \quad f(x)$$

s.t. $Fx = b$. (8)
where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function, $F \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The well-known Arrow-Hurwitz-Uzawa (A–H–U) flow introduced in Arrow et al. (1958) provides under appropriate conditions a continuous-time solver defined by

$$\dot{x} = -\nabla f(x) - F^T v$$

$$\dot{v} = Fx - b.$$  

Convergence properties of (8) are studied in many works, e.g., Arrow et al. (1958), Wang and Elia (2011), Holding and Lestas (2014) and Cherukuri, Gharesifard, and Cortes (2017). In particular, if $f$ is strictly convex and $F$ has full row rank, then (see Arrow et al., 1958 Wang & Elia, 2011) along the flow (8), $x(t)$ will converge to the unique minimizer of (7) and $v(t)$ 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 $Fx = b$ given by (L ⊗ I_m)x = 0. However, the Laplacian $L$ 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 (5).

2.4.2. Relation to Wang–Elia/Gharesifard–Cortés Flows

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

$$\dot{x_i}(t) = -\alpha \sum_{j \in N_i}(x_i(t) - x_j(t))$$

$$- \sum_{j \in N_i}(v_i(t) - v_j(t)) - (h_i^\top h_i)x_i(t) - z_i$$

$$\dot{v}_i(t) = \sum_{j \in N_i}(x_i(t) - x_j(t)).$$

where $\alpha$ is a positive number. With fixed interaction graph and suitable choice of $\alpha$, the flow (9) is a least squares solver for (1) even for balanced directed networks (Theorem 5.4, Gharesifard & Cortés, 2014). In Fig. 1, we show the way that two neighbor nodes $i$ and $j$ share their states under the assumption that the dynamics of $x_i(t)$ and $v_i(t)$ are computed by two individual processors. It can be seen that under the flow (5), external information transmission occurs only between $x_i(t)$ and $v_j(t)$, $x_j(t)$ and $v_i(t)$. However, information transmission occurs additionally between $x_i(t)$ and $x_j(t)$ 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 rank$(H) = m$. Let $y^* = (H^\top H)^{-1}H^\top z$ be the unique least squares solution of (1). Define $S_\lambda$ as the set of all complex eigenvectors of $L$ and for $\alpha \in S_\lambda$ with $[\alpha]$ denoting the $i$th entry,

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

(i) If $\operatorname{span}(h_i : i \in \mathcal{I}_\alpha) = \mathbb{R}^m$ for all $\alpha \in S_\lambda$ then along (5)

$$\lim_{t \to \infty} x_i(t) = y^*, \quad t = 1, \ldots, N$$

$$\lim_{t \to \infty} v_i(t) = v^*_i, \quad t = 1, \ldots, N$$

where $[v_1 \ldots v_N]^\top$ is a Lagrange multiplier associated with the optimization problem (4). Moreover, the convergence rates of both $x_i(t)$ and $v_i(t)$ are exponential.

(ii) If there exists $\alpha \in S_\lambda$ such that $\dim(\operatorname{span}(h_i : i \in \mathcal{I}_\alpha)) < m$, then there exist trajectories of $x(t)$ along (5) which are oscillatory.

**Proof.** Recall $\nabla U(x) = \dot{H}x - z_H$. Suppose there exists an equilibrium $(x^*, v^*)$ of (5), i.e.

$$0 = -(L \otimes I_m)v^* - \dot{H}x^* + z_H$$

$$0 = (L \otimes I_m)x^*.$$  

(10)

It is worth noting that (10) specifies exactly the Karush–Kuhn–Tucker conditions on $(x^*, v^*)$ 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 & Vandenberghe, 2004). Therefore $x^*$ is an optimal solution to (4) and any optimal solution of (4) must have the form $1 \otimes y^*$. Since $y^*$ is unique, $x^*$ is also unique. Note however that $v^*$ is not necessarily unique. Define the variables $\tilde{x} = x - x^*$, $\tilde{v} = v - v^*$. Then

$$\dot{\tilde{x}} = -(L \otimes I_m)\tilde{v} - \dot{H}\tilde{x}$$

$$\dot{\tilde{v}} = (L \otimes I_m)\tilde{x}.$$  

(11)

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

$$M = \begin{bmatrix}
-L \otimes I_m & -L \otimes I_m \\
-L \otimes I_m & 0
\end{bmatrix}.$$  

(12)

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

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

Since

$$\dot{V} = -\tilde{x}^\top (L \otimes I_m)\tilde{v} - \tilde{x}^\top \dot{H}\tilde{x} + \tilde{v}^\top (L \otimes I_m)\tilde{x}$$

$$= -\tilde{x}^\top \dot{H}\tilde{x} \leq 0,$$  

(13)

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

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

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

(i) Suppose $\operatorname{span}(h_i : i \in \mathcal{I}_\alpha) = \mathbb{R}^m$ for all $\alpha \in S_\lambda$. We proceed to prove the convergence of $\tilde{x}(t)$ and $\tilde{v}(t)$. The proof contains two steps.
Step 1. We prove $M$ does not have a purely imaginary eigenvalue using a contradiction argument. Suppose $\lambda = \alpha + i \beta$ where $\alpha \neq 0$ and $\beta \neq 0$ is an eigenvalue of $M$ with a corresponding eigenvector $v = [\beta_x \beta_y] \in \mathbb{C}^{2nm}$, where $\beta_x \in \mathbb{C}^{nm}$, $\beta_y \in \mathbb{C}^{nm}$. Let $u(0) = \beta$.

Then $\dot{u}(t) = e^{tM}u(0) = e^{t\beta}$.

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

On the other hand, according to (13),

$$\frac{1}{2} \frac{d}{dt} \|u(t)\|^2 = -e^{2t\beta} \beta_y \dot{H} \beta_y.$$

Consequently, there must hold $H \beta_y = 0$. Next, based on $M \beta = \alpha \beta$, we know

$$- (L \otimes I_m) \beta_y = \alpha \beta_y,$$

$$L \otimes I_m \beta_y = \alpha \beta_y.$$

Since $\beta \neq 0$, neither of $\beta_x$ nor $\beta_y$ can be zero. By simple calculation, we have

$$(L \otimes I_m) \beta_y = r^2 \beta_y,$$

$$(L \otimes I_m) \beta_y = r^2 \beta_y,$$

i.e. $\beta_y$ and $\beta_x$ are both eigenvectors of $(L \otimes I_m)$ corresponding to $r^2$. From (15), we know

$$(L^T \otimes I_m) \beta_x = r^2 \beta_x.$$.

Based on the properties for eigenvectors of the Kronecker product of two matrices (Theorem 13.12 Laub, 2005), we know there exist $(r^2, \alpha_x)$ and $\eta_x$ such that $L^2 \alpha_x = r^2 \alpha_x$ and $\alpha_x = \alpha_x \otimes \eta_x$ with $\alpha_x \in \mathbb{C}^{nm}$ and $\eta_x \in \mathbb{C}^{nm}$. It is trivial that if $L^2 \alpha_x = r^2 \alpha_x$, $L \alpha_x = |r| \alpha_x$, i.e. $\alpha_x$ is an eigenvector of $L$ corresponding to eigenvalue $|r|$. Denote $\beta_x = \begin{bmatrix} \beta_x[1] & \ldots & \beta_x[N] \end{bmatrix}$, $\beta_x[i] \in \mathbb{C}^{nm}$ and $\eta_x = \begin{bmatrix} \eta_x[1] & \ldots & \eta_x[N] \end{bmatrix}$, $\eta_x[i] \in \mathbb{C}$ with $i = 1, \ldots, N$. It is evident that $\beta_x[i] = \alpha_x[i] \eta_x[i]$. If $h \in \mathcal{A}_x$, and $\beta_x[i] = 0$ otherwise. Then noting that

$$H \beta_y = \begin{bmatrix} \alpha_x[1] h_i \eta_x \eta_x \quad \cdots \\ \alpha_x[N] h_i \eta_x \eta_x \end{bmatrix} = 0,$$

we get $\alpha_x[i] h_i \eta_x = 0$ for $i = 1, 2, \ldots, N$, which implies that

$$(16)$$

$$h_i \eta_x = 0, \quad i \in \mathcal{A}_x.$$

Because span$(h_i : i \in \mathcal{A}_x) = \mathbb{R}^m$, there must hold $\eta_x = 0$. In turn, $\beta_y$ must be zero, leading to $\beta_y = 0$ with (14). Therefore $M$ does not have purely imaginary eigenvalues. Based on C1, C2 and the fact that $M$ has no purely imaginary eigenvalues, it follows that $\mathbf{X}(t)$ and $\mathbf{V}(t)$ converge. Furthermore, by the basic property of linear time-invariant systems, $\mathbf{X}(t)$ and $\mathbf{V}(t)$ exponentially converge, implying that $\mathbf{X}(t)$ and $\mathbf{V}(t)$ exponentially converge.

Step 2. In this step, we establish the limits of $\mathbf{X}(t)$ and $\mathbf{V}(t)$ by studying the zero eigenspace of $M$, thereby obtaining the convergence property for $\mathbf{X}(t)$ and $\mathbf{V}(t)$. Suppose $\delta = [\delta_x \delta_y]^T$ is one of the eigenvectors of $M$ corresponding to zero eigenvalue with $\delta \in \mathbb{R}^{2nm}$ and $\delta_x, \delta_y \in \mathbb{R}^{nm}$, i.e. $M \delta = 0$. Consider a solution $\mathbf{u}(t)$ of (11) with $\mathbf{u}(0) = \delta$. We see from the derivative of the Lyapunov function and $\delta H\delta = (L \otimes I_m) \delta_y = (L \otimes I_m) \delta_y = 0$ that

$$\dot{\mathbf{H}} \delta = (L \otimes I_m) \delta_y = (L \otimes I_m) \delta_y = 0.$$

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

$$(17)$$

$$(\mathbf{J}) \mathbf{T}^{-1} = [\delta_1 \delta_2 \ldots \delta_m \ldots]^T,$$

where $\delta_i$ and $\delta_j^*$ with $i = 1, 2, \ldots, m$ are mutually orthogonal right and left eigenvectors respectively of $M$ all corresponding to zero eigenvalues and all with the form of $\delta_i = [0 \delta_{i1}^*]^T$ and $\delta_{j1}^* = [0 \delta_{i1}^*]$. Then

$$\lim_{t \to \infty} \mathbf{V}(t) = \sum_{i=1}^{m} \delta_i \delta_i^T \mathbf{u}(0),$$

which implies that

$$\lim_{t \to \infty} \mathbf{X}(t) = 0,$$

$$\lim_{t \to \infty} \mathbf{V}(t) = \sum_{i=1}^{m} \delta_i \delta_i^T \mathbf{u}(0).$$

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

(ii) Suppose there exists $\alpha_x \in \mathcal{S}_s$ with $\mathbf{L} \alpha_x = \mathbf{r} \alpha_x$ such that dim$(\text{span}(h_i : i \in \mathcal{I}(\alpha_x))) < m$. Then there must exist $\eta_x \neq 0$ satisfying that

$$h_i \eta_x = 0, \quad i \in \mathcal{I}(\alpha_x).$$

Let $\beta = [\beta_x^T \beta_y^T]^T$ with $\beta_x = \alpha_x \otimes \eta_x$ and $\beta_y = \frac{(L \otimes I_m) \beta_y}{\alpha_x}$. It is easy to check that

$$M \beta = \begin{bmatrix} -H & -L \otimes I_m & 0 \\ L \otimes I_m & 0 & \beta_y \end{bmatrix} \beta_y = \alpha \beta_y.$$

Therefore, $M$ 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 $G = (V, E, w)$ is defined by associating $\delta = (\mathcal{V}, \mathcal{E}, w)$ with a function $w : \mathcal{E} \to \mathbb{R}^+$. Introduce $[A_{x,y}^{ij}] = w(i,j)$ if $[i,j] \in \mathcal{E}$ and $[A_{x,y}^{ij}] = 0$ otherwise. Then its weighted Laplacian is given by

$$L_w = \text{diag}(\sum_{j=1}^{N} A_{x,y}^{ij}), \ldots, N \sum_{j=1}^{N} A_{x,y}^{ij}, \ldots - A_{x,y}).$$

Let $\mathcal{S}_L$ be the set of all complex eigenvectors of $L_w$. For $\alpha_{\alpha} \in \mathcal{S}_L$, we define

$$(18)$$

$$(18)$$

$$\mathcal{S}_L = \{i : \alpha_x[i] \neq 0, \quad \alpha_x = [\alpha_x[1] \ldots \alpha_x[N]]^T\}.$$
3.2.2. Generic feasibility of $H$

**Proposition 2.** Let $L$ be the Laplacian of a graph $G$ with $|I_x| = m$ for all $x \in S_L$. Then the convergence condition of Theorem 1, viz. span$\{h_1, \ldots, h_m\} = \mathbb{R}^m$, is satisfied for generic $H \in \mathbb{R}^{N \times m}$, i.e., there does not exist an open nonempty subset in $\mathbb{R}^{N \times m}$ of $H$, for which the convergence condition in Theorem 1 is not satisfied.

**Proof.** Let

$$H_{i_1, \ldots, i_m} = \begin{bmatrix} h_{i_1}^T \\ \vdots \\ h_{i_m}^T \end{bmatrix}, \quad 1 \leq i_1 < \cdots < i_m \leq N.$$ 

Introduce

$$W = \bigcup_{i_1 < \cdots < i_m \leq N} W_{i_1, \ldots, i_m}$$

where

$$W_{i_1, \ldots, i_m} = \{H \in \mathbb{R}^{N \times m} : \det(H_{i_1, \ldots, i_m}) = 0\}.$$ 

It can be noted that $W$ is the set of $H$ for which the convergence condition in Theorem 1 is not satisfied. Since $W_{i_1, \ldots, i_m} \neq \mathbb{R}^{N \times m}$, by identity theorem for holomorphic functions (Abowitz & Fokas, 2003), $W_{i_1, \ldots, i_m}$ does not contain a nonempty open subset of $\mathbb{R}^{N \times m}$, i.e., $W$ 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 $\min_{x \in S_L} |I_x|$ is, the easier that convergence is met. Now we investigate $\min_{x \in S_L} |I_x|$ on several fundamental graphs and random graphs.

**[Path Graph]** It is known from Fuhrmann and Helmeke (2015) that all the eigenvalues of its Laplacian $L$ are distinct with eigenvectors in the set of $S_L = \{\alpha_k : \alpha_k[v] = \cos(k-1)\pi v/N, v = 1, \ldots, N; k = 1, \ldots, N\}$. 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 $\alpha_k[v] = 0$. Therefore, $|I_x| = N$ for all $x$.

(ii) Let $N = 3l$, $l = 1, 2, 3, \ldots$. Then any $\alpha_k \in S_L$ contains at most $l$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = \lceil \frac{N}{l} \rceil$.

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

$$\begin{bmatrix} 1 \\ \cos(2k\pi/N) \\ \vdots \\ \cos((2(N-1)k\pi/N) \\ \sin(2k\pi/N) \\ \vdots \\ \sin((2(N-1)k\pi/N) \end{bmatrix}$$

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 $[1 \ 1 \ldots 1]^T$ and $[-1 \ 1 \ldots 1]^T$, respectively, while all other eigenvalues have multiplicity two with a basis of two orthogonal eigenvectors of the same form (18) and $k = 1, \ldots, N - 1$, $k \neq N$. 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 $\alpha \in S_L$ contains at most one zero entry. Therefore, $\min_{x \in S_L} |I_x| = N - 1$.

(ii) If $N = 3l$, $l = 1, 2, 3, \ldots$, then any $\alpha \in S_L$ contains at most $l$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = \lceil \frac{2}{3} \rceil N$.

(iii) If $N = 2^l$, $l = 3, 4, \ldots$, then any $\alpha \in S_L$ contains at most $2^{l-1}$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = \lceil \frac{2}{3} \rceil N$.

**[Star Graph]** We know that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\alpha_1 = [1 \ 1 \ldots 1]^T$, an eigenvalue $N$ of multiplicity one with eigenvector $\alpha_N = [-1 \ 1 \ldots 1]^T$ and eigenvalue one with multiplicity $N - 2$ and a set of associated eigenvectors $\{\alpha_k : \alpha_k = 0, \alpha_k \neq \alpha_l \forall l \neq k; k = 2, 3, \ldots, N - 1\}$. Thus $\alpha_1$ has at most $N - 2$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = 2$.

**[Complete Graph]** It is known from Cvetković, Doob and Sachs (1980) that its Laplacian has an eigenvalue zero of multiplicity one with eigenvector $\alpha = [1 \ 1 \ldots 1]^T$ and eigenvalue $N$ with multiplicity $N - 1$ and a set of associated eigenvectors $\{\alpha_k : \alpha_k = 0, \alpha_k \neq \alpha_l \forall l \neq k; k = 2, 3, \ldots, N\}$. Then it can be concluded that $\alpha$ has at most $N - 2$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = 2$.

**[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 node, we do $10^4$ experiments and find out the value of $\min_{x \in S_L} |I_x|$ for each experiment. Let $|I_x|_{ave}$ denote the average of $\min_{x \in S_L} |I_x|$ for these $10^4$ experiments. The calculated result is presented in the following table.

| $|I_x|_{ave}$ | $n = 100$ | $n = 200$ | $n = 300$ |
|-------------|------------|------------|------------|
| $p = 0.2$   | 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 $10^5$ samples and the numerical results show the graph Laplacian has a 99.84% chance of satisfying $\min_{x \in S_L} |I_x| = N$. For star and complete graphs, there holds $\min_{x \in S_L} |I_x| = 2$. 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, 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 $k = 0, 1, 2, \ldots$ and propose the following algorithm:

$$x(k+1) = x(k) - \sum_{j \in N_k} \lambda_j (x(k) - \bar{w}(k))$$

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 $[1 \ 1 \ldots 1]^T$ and $[-1 \ 1 \ldots 1]^T$, respectively, while all other eigenvalues have multiplicity two with a basis of two orthogonal eigenvectors of the same form (18) and $k = 1, \ldots, N - 1$, $k \neq N$. 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 $\alpha \in S_L$ contains at most one zero entry. Therefore, $\min_{x \in S_L} |I_x| = N - 1$.

(ii) If $N = 3l$, $l = 1, 2, 3, \ldots$, then any $\alpha \in S_L$ contains at most $l$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = \lceil \frac{2}{3} \rceil N$.

(iii) If $N = 2^l$, $l = 3, 4, \ldots$, then any $\alpha \in S_L$ contains at most $2^{l-1}$ zero entries. Therefore, $\min_{x \in S_L} |I_x| = \lceil \frac{2}{3} \rceil N$.

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

**Theorem 3.** Suppose $\text{span}(h_i : \alpha(i) \neq 0) = \mathbb{R}^m$ for all the eigenvalues $\alpha = [\alpha(1), \ldots, \alpha(N)]^{\top} \in \mathbb{C}^N$ of $L$. Then there exists a positive constant $\epsilon^*$ such that

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

$$\lim_{k \to \infty} x_i(k) = y^*, \quad i = 1, \ldots, N,$$

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

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

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

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 an asymptotically stable linear system if $0 < \epsilon < \epsilon^*$ and is an unstable system if $\epsilon > \epsilon^*$. When $\epsilon = \epsilon^*$, of course $I + \epsilon M$ might have complex eigenvalues on the unit circle, leading to the possibility of periodic trajectories for $|x(k)|^{\top} v(k)^{\top} L^{\top}$. From Theorem 3, a small enough $\epsilon$ would always guarantee convergence but possibly at a low rate. It can be easily proved that the fastest convergence is given by

$$\bar{\epsilon} = \arg\max_{0 < \epsilon < \epsilon^*} \left\{ \max_{\lambda \in \sigma^*(M)} |1 - \epsilon \lambda| \right\}.$$

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

**Theorem 4.** If there exists $\alpha \in \mathbb{S}_+$ such that

$$\dim(\text{span}(h_i : \alpha(i) \neq 0)) < m,$$

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

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

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**


Holding, T., & Lestas, I. (2014). On the convergence to saddle points of concave-concave functions, the gradient method and emergence of oscillations. In IEEE 53rd Conference on decision and control (pp. 1143–1148).


Yang Liu received the B.Eng. degree from the Department of Microelectronics, Tsinghua University, Beijing, China in 2015. Since Dec. 2015, he has been a Ph.D student at the Research School of Engineering, College of Engineering and Computer Science, The Australian National University, Canberra, ACT, Australia. His research interests include distributed computation and optimization, and privacy-aware machine learning.

Christian Lageman received the Dipl.-Inf. degree in Computer Science and the Dipl.-Math. and Dr. rer. nat. degrees in Mathematics from the University of Würzburg, Germany, in 2001, 2002 and 2007, respectively. From 2007 to 2008 he worked as a Research Fellow at the Centre for Mathematics and its Applications, The Australian National University. From 2008 to August 2009 he was on a postdoctoral fellowship at the Department of Electrical Engineering and Computer Science, University of Liège, Belgium. From 2009 to 2017 he was a postdoctoral researcher at the Institute of Mathematics, University of Würzburg, Germany. Since 2017 he is working as a consultant in the IT industry. His research interests are in nonlinear control theory, cooperative control, optimization and estimation on manifolds.

Brian D.O Anderson was born in Sydney, Australia, and educated at Sydney University in Mathematics and Electrical Engineering, with Ph.D in Electrical Engineering from Stanford University in 1966. He is an Emeritus Professor at the Australian National University (having retired as Distinguished Professor in 2016), Distinguished Professor at Hangzhou Dianzi University, and Distinguished Researcher in National ICT Australia. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H Mulligan, Jr Education Medal, and the Bode Prize of the IEEE Control System Society in 1992, as well as several IEEE and other best paper prizes. He is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Royal Society, and a foreign member of the US National Academy of Engineering. He holds honorary doctorates from a number of universities, including Université Catholique de Louvain, Belgium, and ETH, Zürich. He is a past president of the International Federation of Automatic Control and the Australian Academy of Science. His current research interests are in distributed control, social networks and econometric modeling.

Guodong Shi received the B.Sc. degree in Mathematics and Applied Mathematics from the School of Mathematics, Shandong University, Jinan, China in 2005, and the Ph.D. degree in Systems Theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China in 2010. From 2010 to 2014, he was a Postdoctoral Researcher at the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. Since May 2014, he has been with the Research School of Engineering, The Australian National University, Canberra, ACT, Australia, where he is now a Senior Lecturer and Future Engineering Research Leadership Fellow. His research interests include distributed control systems, quantum networking and decisions, and social opinion dynamics.