Improvement of a Distributed Algorithm for Solving Linear Equations

Xuan Wang, Shaoshuai Mou, Member, IEEE, and Dengfeng Sun

Abstract—This paper performs further improvement to a distributed algorithm for solving linear algebraic equations via multi-agent networks recently developed in [1], in which all agents’ states converge exponentially fast to the same solution to a group of linear equations by assuming each agent only knows part of the linear equations and its nearby neighbors’ states. We first prove that the algorithm in [1] with special initialization is able to achieve the solution that is closest to a given point in the Euclidean distance. Second, we eliminate the initialization step required in [1] by a modification to the update equation. Both analytical and numerical results are provided for validation.

Index Terms—Distributed Algorithms, Linear Equations, Multi-agent Network.

I. MOTIVATIONS

MANY research problems in science and engineering can be reformulated as controlling a group of interconnected agents constrained by a network topology, in which each link could represent the leader-follower relation [2]–[7], the information flow in distributed computations [8]–[12], and so on. These network constraints among agents preclude centralized processing and give rise to distributed control with the goal of achieving global objectives through local coordination among nearby neighboring agents [13]. Developing distributed control strategies for solving linear algebraic equations via multi-agent networks are attractive since it allows decomposing a large systems of linear equations into smaller ones that can be cooperatively solved by multi-agent networks [14]–[17]. To be more specific, we are interested in the problem of solving a linear equation $Ax = b$ by employing a network of $m$ agents, where each agent $i$ knows a private equation $A_i x_i = b_i$ and receives states from its nearby neighbors. Here, $[A \ b]$ is partitioned as follows

$$
[A \ b] = \begin{bmatrix}
A_1 & b_1 \\
A_2 & b_2 \\
\vdots & \vdots \\
A_m & b_m
\end{bmatrix}, \quad A \in \mathbb{R}^{n \times n}
$$

Suppose each agent $i$ controls a state vector $x_i(t) \in \mathbb{R}^n$, which could be looked as the estimate of agent $i$ to the solution of the overall equation $Ax = b$. A natural idea to enable all agents to achieve a solution of $Ax = b$ is by employing a so-called “agreement principle”, in which each agent $i$’s state $x_i$ satisfy is own private equation $A_i x_i(t) = b_i$ while all $x_i(t), i = 1, 2, ..., m$ reach a consensus [18]–[24]. Under the agreement principle, the authors of [1] have devised the following Distributed Algorithm for solving Linear Equations (DALE):

1) Initialization: At $t = 0$, each agent $i$ initializes $x_i(0)$ such that $A_i x_i(0) = b_i$.
2) Update: At $t + 1$, $t = 0, 1, ...$, each agent $i$ receives $x_j(t)$ from certain other agents denoted by $\mathcal{N}_i(t)$, and then update its own state to be

$$
x_i(t + 1) = x_i(t) - P_i \left( x_i(t) - \frac{1}{d_i(t)} \sum_{j \in \mathcal{N}_i(t)} x_j(t) \right)
$$

Here, $P_i$ is the orthogonal projection matrix to the kernel of $A_i$; $\mathcal{N}_i(t)$ denotes the set of agent $i$’s neighbors at time $t$, from which agent $i$ can receive information. We always assume that each agent is a neighbor of itself, that is, $i \in \mathcal{N}_i(t)$; $d_i(t)$ denotes the cardinality of $\mathcal{N}_i(t)$.

A major result of the DALE is the following theorem: Theorem 1: [1] Suppose $Ax = b$ has at least one solution and assume that the sequence of neighbor graphs $\mathcal{N}(t)$, $t \geq 1$, is repeatedly jointly strongly connected. Then the DALE drives all $x_i(t)$ converges exponentially fast to the same solution to $Ax = b$ as $t \to \infty$.

We refer to [1] for introduction of “repeatedly jointly strongly connected”, which has been shown to be not only the sufficient but also necessary requirement of the network connectivity for the above theorem to hold. Throughout this paper we also assume that the underlying neighbor graphs $\mathcal{N}(t)$ is repeatedly jointly strongly connected.

The DALE developed in [1] is applicable to all linear equations that have at least one solution; converges exponentially fast; works for a large class of time-varying directed graphs; operates asynchronously; and does not involve any time-varying step-size. The aim of this paper is to achieve further improvement to the DALE by addressing the following two problems:

- When $Ax = b$ has more than one solutions, the DALE achieves one of them, however, which solution to be achieved by DALE is unclear. In this paper, we will utilize a special initialization step to enable the DALE to achieve a specific solution that is closest to a prescribed point in $\mathbb{R}^n$ in the Euclidean distance. The significance of such solution is that if $x$ represents the state vector needed to
be controlled in industrial process [25]–[27], the obtained solution ensures the system to achieve its goal with the lowest cost.

- The DALE heavily depends on that each state vector \( x_i(t) \) is initialized to be a solution to \( A_i x = b_i \). When there are round-off errors in such initialization step [28], [29], the DALE might fail to achieve the exact solution to \( Ax = b \). Moreover, such initialization step requires each agent to be capable of finding an exact solution to \( A_i x = b_i \), which may be out of the computation capability of low-cost agents [30]. In this paper, we will eliminate such initialization step and allows \( x_i(0) \) can be chosen arbitrarily from \( \mathbb{R}^n \).

II. ACHIEVE THE SOLUTION CLOSEST TO A SPECIFIED POINT

When \( Ax = b \) has multiple solutions, the DALE developed in [1] enables each agent to achieve one of its solutions. But which one is to be achieved is not clear. In this section, we will modify the initialization step of the DALE to achieve a specific solution \( x^q_{\text{min}} \) which minimizes \( \frac{1}{2} \| x - q \|^2 \) subject to \( Ax = b \). That is,

\[
x^q_{\text{min}} = \arg\min_{Ax=b} \frac{1}{2} \| x - q \|^2
\]

(3)

where \( \| \cdot \| \) denotes the Euclidean norm. Specially, when \( q = 0 \), \( x^q_{\text{min}} \) becomes the solution to \( Ax = b \) with the minimum Euclidean norm. Finding such a \( x^q_{\text{min}} \) can be formulated as solving a convex optimization problem by the Dykstras cyclic projection method [31], [32]. It is noted that such cyclic projection method usually requires a centralized scheduling among all agents in the network. In this section we will show that the update (2) achieves \( x^q_{\text{min}} \) exponentially fast under the same network connectivity requirement as in [1] by utilizing the following special initialization step:

Initialization (♣): Each agent \( i \) initializes its \( x_i(0) \) to minimize

\[
\frac{1}{2} \| x - q \|^2 \text{ subject to } A_i x = b_i
\]

The above initialization (♣) could be easily completed by solving a linear equation \( A_i x = b_i \) and \( P_x = P_q \) according to the following lemma:

Lemma 1: \( x = x^q_{\text{min}} \) if and only if it satisfies \( Ax = b \) and \( P_x = P_q \), where \( P_x \) is the orthogonal projection to \( \text{ker} A \).

Proof of Lemma 1: By the standard Lagrange multiplier method for convex optimization subject to linear constraints [33], there must exist a \( \lambda \) and \( x^q_{\text{min}} \) such that

\[
Ax = b
\]

(4)

\[
x - q + A^T \lambda = 0
\]

(5)

Multiplying \( P_x \) to (5), one has \( P_x (x - q) + (AP_x)^T \lambda = 0 \) which and \( AP_x = 0 \) imply

\[
P_x x = P_x q
\]

(6)

Thus \( x^q_{\text{min}} \) must satisfies linear equations (4) and (6), which has the unique solution since \( \text{ker} A \cap \text{ker} A = 0 \) because of the fact image \( A_x = \text{ker} A \).

To summarize, the point in \( \mathbb{R}^n \) which is a solution to \( Ax = b \) and minimizes \( \frac{1}{2} \| x - q \|^2 \) always exists, and must be the unique solution of (4) and (6). Thus Lemma 1 is true.

The main result of this section is the following theorem

Theorem 2: With the initialization (♣) and the update (2), all \( x_i(t) \) converge exponentially fast to be \( x^q_{\text{min}} \).

Proof of Theorem 2: Note that by the initialization (♣) one still has \( A_i x_i(0) = b_i \), which is a special case of the DALE. By Theorem 1, all \( x_i(t) \) converges exponentially fast to be the same \( x^* \) such that \( Ax^* = b \). To prove Theorem 2, we only need to prove that \( x^* \) additionally minimizes \( \frac{1}{2} \| x - q \|^2 \), or equivalently by Lemma 1, \( P_A x^* = P_A q \) holds. Since \( x^* \) is the final value that all \( x_i(t) \) converges to, it suffices to show

\[
P_A x_i(t) = P_A q, \quad i = 1, 2, ..., n
\]

(7)

for all \( t = 0, 1, 2, ... \), for which the induction method will be employed in the followings.

We first show

\[
P_A x_i(0) = P_A q.
\]

(8)

From image \( P_A = \text{ker} A_i \), image \( P_t = \text{ker} A_i \) and \( \text{ker} A \subset \text{ker} A_i \), one has image \( P_A \subset \text{image} P_t \). Then

\[
\text{ker} P_t \subset \text{ker} P_A
\]

which and image \( (I - P_t) = \text{ker} P_t \) imply \( P_A (I - P_t) = 0 \), that is,

\[
P_A P_t = P_A
\]

(9)

By the initialization (♣), one has \( P_t x_i(0) = P_t q \), which and (9) lead to (8).

Now we suppose (7) is true for all \( i \) at \( t \), and show it is true at \( t + 1 \). From (2), (9) and the induction assumption, one has

\[
P_A x_i(t + 1) = P_A x_i(t) - P_A P_t \left( x_i(t) - \frac{1}{d_i(t)} \sum_{j \in N_i(t)} x_j(t) \right)
\]

\[
= P_A x_i(t) - \left( P_A x_i(t) - \frac{1}{d_i(t)} \sum_{j \in N_i(t)} P_A x_j(t) \right)
\]

\[
= \frac{1}{d_i(t)} \sum_{j \in N_i(t)} P_A q
\]

Thus (7) is true.
III. A MODIFIED DALE WITHOUT INITIALIZATION

The DALE in [1] requires an initialization step in which each \( x_i(0) \) is initialized to be a solution to \( A_i x = b_i \). In order to eliminate the initialization step, we modify the DALE by adding an additional term to (2), which drives all \( x_i(t) \) to the manifold \( A x = b \) even when they are not initialized to be so. Similar ideas have also been discussed in [34] and [16].

Let \( \bar{A}, \bar{b} \) denotes a submatrix of \( [A, b] \) such that \( \ker A_i = \ker \bar{A}_i \) and \( \bar{A}_i \bar{A}_i^t \) is non-singular, which also implies that \( A_i x = b_i \) if and only if \( \bar{A}_i x = \bar{b}_i \). Thus the orthogonal projection matrix \( P_i \) to the kernel of \( A_i \) can be expressed as \( P_i = I - \bar{A}_i(\bar{A}_i \bar{A}_i^t)^{-1} \bar{A}_i \). The modified DALE will be: At \( t = 0 \), \( x_i(0) \) is any vector in \( \mathbb{R}^n \); The update for each agent \( i \) at \( t + 1 \) is

\[
x_i(t + 1) = x_i(t) - P_i \left( x_i(t) - \frac{1}{d_i(t)} \sum_{j \in N_i(t)} x_j(t) \right)
- \bar{A}_i' (\bar{A}_i \bar{A}_i^t)^{-1} (\bar{A}_i x_i(t) - \bar{b}_i)
\tag{10}
\]

Remark 2: It is worth mentioning that by multiplying \( \bar{A}_i \) to both sides of algorithm (10), one has \( \bar{A}_i x_i(t+1) = \bar{A}_i x_i(t) - (\bar{A}_i x_i(t) - \bar{b}_i) \) because of \( \bar{A}_i P_i = 0 \). Then \( \bar{A}_i x_i(t+1) = \bar{b}_i \) for \( t = 0, 1, 2, \ldots \). This observation implies that \( x_i(t) \) for \( t \geq 1 \) is always a solution to \( \bar{A}_i x = \bar{b}_i \) even if \( x_i(0) \) is not.

To prove \( x_i(t) \) converges to \( x^* \) where \( A x = b \), we prove the error vector \( e_i(t) = x_i(t) - x^* \) converge to 0. From (10), \( \bar{A}_i x^* = \bar{b}_i \), and \( P_i = I - \bar{A}_i (\bar{A}_i \bar{A}_i^t)^{-1} \bar{A}_i \), one has

\[
e_i(t+1) = e_i(t) - P_i (e_i(t) - \frac{1}{d_i(t)} \sum_{j \in N_i(t)} e_j(t)) - \bar{A}_i' (\bar{A}_i \bar{A}_i^t)^{-1} (\bar{A}_i e_i(t) - \bar{b}_i)
\]

which is the same as the error equation in [1], for which one has the following lemma:

**Lemma 2:** All \( e_i(t) \) converge exponentially fast to the same value \( e^* \), where \( e^* = 0 \) when \( A x = b \) has a unique solution.

From Lemma 2 and \( e_i(t) = x_i(t) - x^* \), one has all \( x_i(t) \) converges to be

\[
z^* = e^* + x^*
\]

When \( A x = b \) has a unique solution. One has \( z^* = x^* \) and all \( x_i(t) \) converges to the unique solution \( x^* \). Since \( e^* \) may not be zero when \( A x = b \) has multiple solutions, we will need to show that \( z^* \) is actually also a solution to \( A x = b \). Note that \( z^* \) is the common value to which all \( x_i(t) \) converge to. Then \( z^* \) satisfies the equations (10). By replacing \( x_i(t+1) \) and \( x_i(t) \) by \( z^* \), one has

\[
z^* - P_i (z^* - \frac{1}{d_i(t)} \sum_{j \in N_i(t)} z^*)
- \bar{A}_i' (\bar{A}_i \bar{A}_i^t)^{-1} (\bar{A}_i z^* - \bar{b}_i), \ i = 1, 2, \ldots, m
\tag{11}
\]

which implies

\[
\bar{A}_i' (\bar{A}_i \bar{A}_i^t)^{-1} (\bar{A}_i z^* - \bar{b}_i) = 0, \ i = 1, 2, \ldots, m
\tag{12}
\]

Multiplying \( \bar{A}_i \) to both sides of (12) leads to

\[
\bar{A}_i z^* - \bar{b}_i = 0, \ i = 1, 2, \ldots, m
\tag{13}
\]

Then

\[
\bar{A}_i z^* = \bar{b}_i, \ i = 1, 2, \ldots, m
\]

Thus \( z^* \) is a solution to \( A x = b \). To summarize, we have the following theorem:

**Theorem 3:** The modified DALE (10) with arbitrary \( x_i(0) \in \mathbb{R}^n \) drives all \( x_i(t) \) converge exponentially fast to be a common solution to \( A x = b \).

IV. VALIDATION

Numerical simulations will be provided in this section to demonstrate the effectiveness of the main results in Theorem 2 and Theorem 3.

**Example 1:** Suppose that the following linear equation is solved by the improved DALE with 3 agents:

\[
\begin{bmatrix}
A_1 \\
A_2 \\
A_3
\end{bmatrix} = \begin{bmatrix}
1 & -1 & 1 \\
1 & -1 & -1 \\
0 & 0 & 1
\end{bmatrix}, \begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix} = \begin{bmatrix}
3 \\
3 \\
3
\end{bmatrix}
\tag{14}
\]

\( x^q_{\min} \) and the preferred point \( q \) are:

\[
x^q_{\min} = \begin{bmatrix} 5.5 \\
5.5 \\
3 \end{bmatrix}, \quad q = \begin{bmatrix} 5 \\
6 \\
5 \end{bmatrix}
\tag{15}
\]

Let \( V(t) = \sum_{i=1}^{3} |x_i(t) - x^{q}_{\min}|^2 \). Then \( V(t) = 0 \) if and only if all \( x_i(t) = x^{q}_{\min} \). Fig. 1 shows that the DALE in [1] does not achieve \( x^{q}_{\min} \) while the modified DALE with the special initialization step (3) does.

We also provide Fig. 2 to illustrate the detailed process of the improved DALE to achieve \( x^{q}_{\min} \). By \( P_i x_i(0) = P_i q \) and
$A_i x_i(0) = b_i$, each $x_i(0)$ is chosen as the closest point to $q$ subject to $A_i x = b_i$. All the $x_i(t)$ then converge to reach a consensus value which is $x_{\text{min}}^q$ as indicated in Fig. 2.

Example 2: Given the following linear equation,

\[
\begin{bmatrix}
A_1 \\
A_2 \\
A_3
\end{bmatrix} = 
\begin{bmatrix}
19 & 0 & 12 & 5 & 10 \\
12 & 16 & 10 & 5 & 5 \\
5 & 12 & 16 & 14 & 16 \\
4 & 16 & 16 & 17 & 4 \\
13 & 4 & 5 & 7 & 4 
\end{bmatrix},
\begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix} = 
\begin{bmatrix}
5 \\
17 \\
99 \\
9 \\
51
\end{bmatrix}
\]

with the unique solution $x^* = \begin{bmatrix} 0.962777 \ 12.4675 \ 10.10 \end{bmatrix}^T$. Let $V(t) = \sum_{i=1}^{3} |x_i(t) - x^*|^2$. Then $V(t) = 0$ if and only if all $x_i(t) = x^*$. Fig. 3 shows that the modified DALE (10) without any initialization step drives all agents exponentially fast to $x^*$.

In addition, Table I about an index $D_i(t) = |A_i x_i(t) - b_i|^2$ demonstrates that for any $t \geq 1$, $x(t)$ is a solution to $A_i x = b_i$.

| Number of Agent | $D_i(t) = |A_i x_i(t) - b_i|^2$ |
|-----------------|---------------------------------|
| Agent 1         | $314 \times 10^{-15}$          |
| Agent 2         | $7.3 \times 10^{-14}$          |
| Agent 3         | $1.5 \times 10^{-14}$          |

V. Conclusion

In this paper we have performed further improvements to the distributed algorithm for solving linear equations (DALE) developed in [1]: For equations with more than one solutions, the modified initialization (green) guarantees the DALE to find the solution that has the smallest Euclidean distance to a given point in $\mathbb{R}^n$; For agents with limited computation capability, the improved DALE (10) enables all agents to achieve exponentially fast a solution to the whole linear equations without any initialization step. Our future work includes further improvement on DALE for achieving finite-time convergence using the ideas appearing in [35], [36], approaching the least square solution even if the overall linear equations do not have solutions as in [37], [38] and being resilient to cyber attacks to agents in the network [39].

REFERENCES

Shaoshuai Mou (M’08) received his B. S. and M. S. degree in Control Science and Engineering from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively. He finished his Ph.D. study under the supervision of Prof. A. Stephen Morse at Yale University in 2014. He then joined MIT as a postdoctoral researcher. Since July 2015 he has started working as a tenure-track assistant professor at Purdue University.

Dengfeng Sun is an Associate Professor in School of Aeronautics and Astronautics at Purdue University, West Lafayette, Indiana. Before joining Purdue, he was an Associate Scientist with University of California Santa Cruz at NASA Ames Research Center. He received a bachelor’s degree in precision instruments and mechatronics from Tsinghua University in China, a master’s degree in industrial and systems engineering from the Ohio State University, and a PhD degree in civil engineering from University of California at Berkeley. Dr. Sun’s research interests include control and optimization of large scale systems and their applications in aerospace engineering.

Xuan Wang received his B. S. and M. S. degree in Control Science and Engineering from Harbin Institute of Technology, Harbin, China, in 2014 and 2016, respectively. Between February 2013 and July 2013, he was a exchange student in College of Electrical Engineering & Computer Science, National Taiwan University, Taiwan, China. Currently, he is an Ph.D. student in School of Aeronautics and Astronautics at Purdue University.