# A Communication Efficient Hierarchical Distributed Optimization Algorithm for Multi-Agent Reinforcement Learning

**Jineng Ren** [1] **Jarvis Haupt** [1]

## Abstract

Policy evaluation problems in multi-agent reinforcement learning (MARL) have attracted growing interest recently. In these settings, agents collaborate to learn the value of a given policy with private local rewards and jointly observed state-action pairs. However, most fully decentralized algorithms treat each agent equally, without considering the communication structure of the agents over a given network, and the corresponding effects on communication efficiency. In this paper, we propose a hierarchical distributed algorithm that differentiates the roles of each of the agents during the evaluation process. This allows us to freely choose various mixing schemes (and corresponding mixing matrices that are not necessarily doubly stochastic), in order to reduce the communication cost, while still maintaining convergence at rates as fast as or even faster than the previous approaches. Theoretically, we show the proposed method, which contains existing distributed methods as a special case, achieves the same order of convergence rate as state-of-the-art methods. Numerical experiments on real datasets demonstrate the superior performances of our approach over other advanced algorithms in terms of convergence and total communication efficiency.

## 1. Introduction

Reinforcement learning has recently witnessed enormous progress in applications like robotics (Kober et al., 2013) and video games (Mnih et al., 2015). Compared with single-agent reinforcement leaning, multi-agent reinforcement learning (MARL) requires that each agent interacts with not only the environment, but also other agents, which makes the learning tasks more challenging. In this paper, we study the policy evaluation problem in MARL with local rewards. In such scenarios, all the agents share a joint state whose transition depends on the local rewards and actions of individual agents. However, because of practical constraints, each agent only observes its own local rewards but doesn't know those of the others. To achieve the optimal global rewards, which is the sum of local rewards, the agents need to exchange their information with others. This type of setting is motivated by broad applications like traffic signal control (Prabuchandran et al., 2014), sensor networks (Rabbat & Nowak, 2004), and swarm robotics (Corke et al., 2005).

To handle the robustness and scalability problems of centralized policy evaluation methods in MARL, (Lee et al., 2018; Wai et al., 2018) propose fully decentralized methods, where each agent only communicates with its neighbors over a network. However, like most of their predecessors, their approach requires undirected connections or doubly stochastic mixing matrices generated directly from the adjacency matrices of undirected graphs. For directed graphs (digraphs), however, only a special type of them admits a doubly stochastic mixing matrix, and even for them, a general method to construct such matrices is still lacking (Gharesifard & Cortés, 2010; 2012). Moreover, in previous work that requires doubly stochastic mixing matrices, all agents' roles are essentially the same in the sense that they all need to send variables and gradients information to neighbors and receive such information from them. This may cause the agents to communicate too frequently when they are already well connected. In this paper, a hierarchical algorithm is proposed to handle these problems. By properly designing the mixing matrices in the proposed algorithm, it is possible for an agent to receive information from another agent to update its variable without sending such information to it (analogous to the Master-Workers relationship (Li et al., 2014) in a star network with a central node). Another advantage of the proposed algorithm is that it not only includes the mixing scheme in the previous work as a special case, but also includes other mixing schemes that haven't been considered before. The experiments show that the hierarchical structure can save communication in each iteration and still have better convergence performance than the previous approaches. Theoretically, the proposed algorithm is

[1]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, 55455, USA. Correspondence to: Jineng Ren <renxx282@umn.edu>, Jarvis Haupt <jdhaupt@umn.edu>.

*Real-world Sequential Decision Making workshop at $36^{th}$ International Conference on Machine Learning (ICML 2019)*, Long Beach, California, USA, 2019. 

proved to converges linearly with the same order of rate as the state-of-the-art approach though it is much more general than the latter. The analysis is of independent interest for solving general saddle-point problems with convex-concave cost on decentralized agents.

**Related Work** The study of MARL dates back to (Littman, 1994; 2001; Lauer & Riedmiller, 2000). For more recent works see also (Hu & Wellman, 2003; Arslan & Yüksel, 2017). However, most of them suffer from the curse of dimensionality since they exploit the tabular setting. To resolve this issue in the policy evaluation under the MARL framework, linear function approximation and actor-critic algorithms are studied by (Lee et al., 2018) and (Zhang et al., 2018) respectively. For the policy evaluation in signal-agent RL, the primal-dual formulation is studied in papers like (Lian et al., 2016; Dai et al., 2017a; Chen & Wang, 2016; Wang, 2017; Dai et al., 2017b; Du et al., 2017). In the multi-agent setting it is studied by (Macua et al., 2015; 2017; Wai et al., 2018). Our work is more related to (Wai et al., 2018) with the difference that our algorithm allows hierarchical structures of communication and more general mixing schemes. In the works that generally minimize a sum of convex local cost functions (Tsitsiklis et al., 1986; Nedic & Ozdaglar, 2009; Shi et al., 2015; Qu & Li, 2018; Ren et al., 2017; Ren & Haupt, 2018; 2019; Ma et al., 2018b;a), our algorithm is closely related to stochastic average or incremental gradient (Pu et al., 2018; Schmidt et al., 2017; Defazio et al., 2014), with the difference that our objective function is a double sum of convex-concave functions and we consider hierarchical structures and new mixing schemes as well as their effects on efficiency. As far as we know, the proposed algorithm is the first to work on directed graphs with hierarchical structures, and allows rich options of mixing schemes to solve decentralized convex-concave saddle-point problems in MARL.

**Notation** For a vector $\mathbf{a} = (a_1, \cdots, a_s)^\top \in \mathbb{R}^s$ we denote $\|\mathbf{a}\| := (\sum_{i=1}^s |a_i|^2)^{1/2}$. For a matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{s\times k}$, we define $\|\mathbf{A}\| := (\sum_{i=1}^s \sum_{j=1}^k |a_{ij}|^2)^{1/2}$ and $\|\mathbf{A}\|_{1,\infty} := \max_{i=1,\cdots,s}\{\sum_{j=1}^k |a_{ij}|\}$. Given a matrix $\mathbf{H} \in \mathbb{R}^{s\times s}$, we define $\|\mathbf{v}\|_{\mathbf{H}} := \sqrt{\mathbf{v}^\top \mathbf{H}\mathbf{v}}$ for any vector $\mathbf{v} \in \mathbb{R}^s$.

## 2. Problem Formulation

In this section, we introduce the multi-agent Markov decision process (MDP). Then as shown in (Du et al., 2017), we can reformulate the policy evaluation problem as a primal-dual convex-concave optimization problem.

Now consider a network of N agents. We are interested in the multi-agent finite MDP: $\left(\mathcal{P}^{\vec{\mathbf{a}}}, \mathcal{S}, \{\mathcal{A}_i\}_{i=1}^N, \{\mathcal{R}_i\}_{i=1}^N, \gamma\right)$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}_i$ is the action space for agent $i$, $\mathcal{R}_i$ is the reward space for agent $i$, and $\gamma \in (0,1)$ is the discount factor. $\mathcal{P}^{\vec{\mathbf{a}}} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is the state transition matrix under a joint action $\vec{\mathbf{a}} \in \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_N$, where $\mathcal{P}^{\vec{\mathbf{a}}}(\mathbf{s}, \mathbf{s}')$ denotes the transition probability from $\mathbf{s}$ to $\mathbf{s}'$. The local reward received by agent $i$ after taking joint action $\vec{\mathbf{a}}$ at state $\mathbf{s}$ is denoted by $r_i(\mathbf{s}, \vec{\mathbf{a}})$. It is private for agent $i$, while both $\mathbf{s}$ and $\vec{\mathbf{a}}$ are observable by all agents.

We know in this setting the agents are coupled together by the state transition matrix $\mathcal{P}^{\vec{\mathbf{a}}}$. As a motivation, this scenario arises from large-scale applications such as sensor networks (Rabbat & Nowak, 2004; Cortes et al., 2004), robotics (Kober et al., 2013), and power grids (Dall'Anese et al., 2013). Moreover, a policy $\boldsymbol{\pi}(\vec{\mathbf{a}}|\mathbf{s})$ is the conditional probability of taking joint action $\vec{\mathbf{a}}$ given the current state $\mathbf{s}$. We define the reward function of $\boldsymbol{\pi}$ as an average of the local rewards: $R_c^{\boldsymbol{\pi}}(\mathbf{s}) := \frac{1}{N}\sum_{i=1}^N R_i^{\boldsymbol{\pi}}(\mathbf{s})$, where $R_i^{\boldsymbol{\pi}}(\mathbf{s}) := \mathbb{E}_{\vec{\mathbf{a}}\sim\boldsymbol{\pi}(\cdot|\mathbf{s})}\left[r_i(\mathbf{s}, \vec{\mathbf{a}})\right]$. The goal for the agents is to collaboratively find a joint action-selection $\boldsymbol{\pi}$ that maximizes the global return. Note that a policy $\boldsymbol{\pi}$ induces a transition matrix $\mathbf{P}^{\boldsymbol{\pi}}$ over $\mathcal{S}$, whose $(\mathbf{s}, \mathbf{s}')$-th element is given by $[\mathbf{P}^{\boldsymbol{\pi}}]_{\mathbf{s},\mathbf{s}'} := \mathbb{E}_{\boldsymbol{\pi}(\cdot|\mathbf{s})}[\mathcal{P}^{\vec{\mathbf{a}}}]_{\mathbf{s},\mathbf{s}'}$.

**Policy Evaluation** A key step in reinforcement learning is policy evaluation. Efficient estimation of the value function of a given policy is important for MARL. For any given joint policy $\boldsymbol{\pi}$, the value function $V^{\boldsymbol{\pi}}\colon \mathcal{S} \to \mathbb{R}$, is defined as

$$V^{\boldsymbol{\pi}}(\mathbf{s}) := \mathbb{E}\left[\sum_{p=1}^\infty \gamma^{p-1} R_c^{\boldsymbol{\pi}}(\mathbf{s}_p)|\mathbf{s}_1 = \mathbf{s}, \boldsymbol{\pi}\right]. \quad (1)$$

Let us construct the vector $\mathbf{V}^{\boldsymbol{\pi}} \in \mathbb{R}^{|\mathcal{S}|}$ by stacking up $V^{\boldsymbol{\pi}}(\mathbf{s})$ in (1) for all $\mathbf{s}$. Then, $\mathbf{V}^{\boldsymbol{\pi}}$ satisfies the Bellman equation

$$\mathbf{V}^{\boldsymbol{\pi}} = \mathbf{R}_c^{\boldsymbol{\pi}} + \gamma\mathbf{P}^{\boldsymbol{\pi}}\mathbf{V}^{\boldsymbol{\pi}}, \quad (2)$$

where $\mathbf{R}_c^{\boldsymbol{\pi}}$ is formed by stacking up $R_c^{\boldsymbol{\pi}}(\mathbf{s})$ and $\mathbf{P}^{\boldsymbol{\pi}}$ is defined above. Furthermore, $\mathbf{V}^{\boldsymbol{\pi}}$ can be shown as the unique solution of (2).

To scale up when the state space size $|\mathcal{S}|$ is large, here we approximate $V^{\boldsymbol{\pi}}(\mathbf{s})$ using the family of linear functions $\left\{V_{\boldsymbol{\theta}}(\mathbf{s}) := \boldsymbol{\phi}^\top(\mathbf{s})\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^d\right\}$, where $\boldsymbol{\theta} \in \mathbb{R}^d$ is the parameter, $\boldsymbol{\phi}(\mathbf{s}) : \mathcal{S} \to \mathbb{R}^d$ is a feature map consisting of $d$ features, e.g., a dictionary induced by tile coding (Sutton & Barto, 2018). We define $\boldsymbol{\Phi} := \left(\cdots; \boldsymbol{\phi}^\top(\mathbf{s}); \cdots\right) \in \mathbb{R}^{|\mathcal{S}|\times d}$ and let $\mathbf{V}_{\boldsymbol{\theta}} \in \mathbb{R}^{|\mathcal{S}|}$ be formed by stacking up $\left\{V_{\boldsymbol{\theta}}(\mathbf{s})\right\}_{\mathbf{s}\in\mathcal{S}}$. Our aim is to find $\boldsymbol{\theta} \in \mathbb{R}^d$ such that $\mathbf{V}_{\boldsymbol{\theta}} \approx \mathbf{V}^{\boldsymbol{\pi}}$, which minimizes the mean squared projected Bellman error (MSPBE)

$$\begin{aligned}&\text{MSPBE}^*(\boldsymbol{\theta})\\&:= \tfrac{1}{2}\left\|\boldsymbol{\Pi}_{\boldsymbol{\Phi}}\left(\mathbf{V}_{\boldsymbol{\theta}} - \gamma\mathbf{P}^{\boldsymbol{\pi}}\mathbf{V}_{\boldsymbol{\theta}} - \mathbf{R}_c^{\boldsymbol{\pi}}\right)\right\|_{\mathbf{H}} + \tfrac{\rho}{2}\|\boldsymbol{\theta}\|^2, \quad (3)\end{aligned}$$

where $\mathbf{H} = \text{diag}[\{\mu^{\boldsymbol{\pi}}(\mathbf{s})\}_{\mathbf{s}\in\mathcal{S}}] \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ is a diagonal matrix whose diagonal elements are the stationary distribution of $\boldsymbol{\pi}$, $\boldsymbol{\Pi}_{\boldsymbol{\Phi}} : \mathbb{R}^{|\mathcal{S}|} \to \mathbb{R}^{|\mathcal{S}|}$ is the projection onto subspace $\{\boldsymbol{\Phi}\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^d\}$, and $\rho \geq 0$ is a regularization parameter. When $\boldsymbol{\Phi}^\top\mathbf{H}\boldsymbol{\Phi}$ is invertible, (3) can be further reformed as

$$\text{MSPBE}^*(\boldsymbol{\theta}) = \tfrac{1}{2}\|\mathbf{A}\boldsymbol{\theta} - \boldsymbol{b}\|_{\mathbf{D}^{-1}} + \tfrac{\rho}{2}\|\boldsymbol{\theta}\|^2, \quad (4)$$

where $\mathbf{A} := \mathbb{E}\left[\boldsymbol{\phi}(\mathbf{s}_p)\left(\boldsymbol{\phi}(\mathbf{s}_p) - \gamma\boldsymbol{\phi}(\mathbf{s}_{p+1})\right)^\top\right]$, $\mathbf{D} := \mathbb{E}\left[\boldsymbol{\phi}(\mathbf{s}_p)\boldsymbol{\phi}^\top(\mathbf{s}_p)\right]$, and $\boldsymbol{b} := \mathbb{E}\left[\mathcal{R}_c^{\boldsymbol{\pi}}(\mathbf{s}_p)\boldsymbol{\phi}(\mathbf{s}_p)\right]$. Here the expectations are taken with respect to the stationary distribution $\mu^{\boldsymbol{\pi}}$. In practice, these expectations are estimated by a finite dataset with $M$ transitions $\{\mathbf{s}_p, \mathbf{a}_p\}_{p=1}^M$ simulated from the multi-agent MDP using joint policy $\boldsymbol{\pi}$. The next state $\mathbf{s}_{M+1}$ of $\mathbf{s}_M$ is also observed. Then the empirical versions of $\mathbf{A}$, $\mathbf{D}$, $\boldsymbol{b}$, denoted respectively by $\widehat{\mathbf{A}}$, $\widehat{\mathbf{D}}$, $\widehat{\boldsymbol{b}}$, are defined as

$$\widehat{\mathbf{A}} := \frac{1}{M}\sum_{p=1}^{M}\mathbf{A}_p,\ \widehat{\mathbf{D}} := \frac{1}{M}\sum_{p=1}^{M}\mathbf{D}_p,\ \widehat{\boldsymbol{b}} := \frac{1}{M}\sum_{p=1}^{M}\boldsymbol{b}_p, \text{with}$$

$$\mathbf{A}_p := \boldsymbol{\phi}(\mathbf{s}_p)\big(\boldsymbol{\phi}(\mathbf{s}_p) - \gamma\boldsymbol{\phi}(\mathbf{s}_{p+1})\big)^\top,\ \mathbf{D}_p := \boldsymbol{\phi}(\mathbf{s}_p)\boldsymbol{\phi}^\top(\mathbf{s}_p),$$
$$\boldsymbol{b}_p := r_c(\mathbf{s}_p, \mathbf{a}_p)\boldsymbol{\phi}(\mathbf{s}_p), \tag{5}$$

where $r_c(\mathbf{s}_p, \mathbf{a}_p) := N^{-1}\sum_{i=1}^N r_i(\mathbf{s}_p, \mathbf{a}_p)$. Here we assume that $M$ is sufficiently large such that $\widehat{\mathbf{D}}$ is invertible and $\widehat{\mathbf{A}}$ is full rank. With the terms defined in (5), the empirical MSPBE is given by

$$\text{EM-MSPBE}(\boldsymbol{\theta}) := \tfrac{1}{2}\|\widehat{\mathbf{A}}\boldsymbol{\theta} - \widehat{\boldsymbol{b}}\|_{\widehat{\mathbf{D}}^{-1}} + \tfrac{\rho}{2}\|\boldsymbol{\theta}\|^2. \tag{6}$$

**Primal-dual Formulation of EM-MSPBE** For any $i \in \{1, \cdots, N\}$ and any $p \in \{1, \cdots, M\}$, we define $\boldsymbol{b}_{p,i} := r_i(\mathbf{s}_p, \mathbf{a}_p)\boldsymbol{\phi}(\mathbf{s}_p)$ and $\widehat{\boldsymbol{b}}_i := M^{-1}\sum_{p=1}^M \boldsymbol{b}_{p,i}$. Recall that $\widehat{\boldsymbol{b}}_i$ is only accessible to agent $i$. To address this, we first notice that minimizing (6) is equivalent to

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \tfrac{1}{N}\textstyle\sum_{i=1}^N \text{EM-MSPBE}_i(\boldsymbol{\theta}) \tag{7}$$

where $\text{EM-MSPBE}_i(\boldsymbol{\theta}) := \frac{1}{2}\|\widehat{\mathbf{A}}\boldsymbol{\theta} - \widehat{\boldsymbol{b}}_i\|^2_{\widehat{\mathbf{D}}^{-1}} + \frac{\rho}{2}\|\boldsymbol{\theta}\|^2$. It is the cost that is private to agent $i$.

As inspired by (Nedic & Bertsekas, 2003; Du et al., 2017), we transform $\text{EM-MSPBE}_i(\boldsymbol{\theta})$ to its conjugate form using Fenchel duality. Then problem (7) is equivalent to

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^d}\max_{\mathbf{w}_i\in\mathbb{R}^d, i=1,\cdots,N} \tfrac{1}{NM}\textstyle\sum_{i=1}^N\sum_{p=1}^M \big(\mathbf{w}_i\mathbf{A}_p\boldsymbol{\theta} - \boldsymbol{b}_{p,i}^\top\mathbf{w}_i - \tfrac{1}{2}\mathbf{w}_i^\top\mathbf{D}_p\mathbf{w}_i + \tfrac{\rho}{2}\|\boldsymbol{\theta}\|^2\big). \tag{8}$$

Define $J_{i,p}(\boldsymbol{\theta}, \mathbf{w}_i) := \big(\mathbf{w}_i\mathbf{A}_p\boldsymbol{\theta} - \boldsymbol{b}_{p,i}^\top\mathbf{w}_i - \frac{1}{2}\mathbf{w}_i^\top\mathbf{D}_p\mathbf{w}_i + \frac{\rho}{2}\|\boldsymbol{\theta}\|^2\big)$, then the global objective function is denoted by $\mathbf{J}(\boldsymbol{\theta}, \{\mathbf{w}_i\}_{i=1}^N) := 1/(NM)\sum_{i=1}^N\sum_{p=1}^M J_{i,p}(\boldsymbol{\theta}, \mathbf{w}_i)$, which is convex $w.r.t.$ the primal variable $\boldsymbol{\theta}$ and is concave $w.r.t.$ the dual variable $\{\mathbf{w}_i\}_{i=1}^N$. In the following, we study this problem by proposing a general hierarchical decentralized first-order algorithm that endows more flexibility of mixing matrices and schemes.

## 3. Algorithm and Analysis

Assume that the $N$ agents communicate over a network specified by a connected digraph $G = (V, E)$, where $V = [N] := \{1, \cdots, N\}$ and $E \subseteq V \times V$ are the vertex set and directed edge set, respectively. Over G, define two row stochastic matrices $R_1$, $R_2$ (and two column stochastic matrices $C_1, C_2$) such that $(R_1)_{ij} = (R_2)_{ij} = 0$ $\big((C_1)_{ij} = (C_2)_{ij} = 0$ respectively$\big)$ if $(j, i) \notin E$. Here a matrix $A$ is said to be row (column) stochastic if $A\mathbf{1} = \mathbf{1}$ ($\mathbf{1}^\top A = \mathbf{1}^\top$). In addition, we assume $R_1$ and $R_2$ share a same left eigenvector. Note the choices of such matrices are abundant, e.g., let $R_1 = R_2$ be row stochastic and $C_1 = C_2 = R_1^\top$. It is also worth noting that the proposed algorithm includes PD-DistIAG as a special case by letting $R_1 = W, R_2 = \mathbf{I}, C_1 = W$, and $C_2 = \mathbf{I}$ (Wai et al., 2018). Moreover, it is much more flexible to construct such matrices over a given graph $G$ compared with the doubly stochastic ones in PD-DistIAG. Thus the proposed approach works on directed graphs to which PD-DistIAG may not be applicable.

**Algorithm 1** Primal Dual Hierarchical Gradient Method (PD-H) for Multi-agent Policy Evaluation

**Input:** Initial variables $\{\boldsymbol{\theta}_i^1, \mathbf{w}_i^1\}_{i\in[N]}$, gradient surrogates $\mathbf{s}_i^0 = \mathbf{d}_i^0 = 0, \forall i \in [N]$, step sizes $\gamma_1, \gamma_2 > 0$, and counter $\tau_p^0 = 0, \forall p \in [M]$. Proper row stochastic matrices $(R_1, R_2)$ and column stochastic matrices $(C_1, C_2)$.

**for** $k \geq 1$ **do**

The agents pick a common sample with index $p_k \in \{1, \cdots, M\}$.

Update the counter variables by

$$\tau_p^k = \begin{cases} k, & p = p_k \\ \tau_p^{k-1}, & p \neq p_k. \end{cases} \tag{9}$$

**for** $i \in \{1, 2, \cdots, N\}$ **do**

Agent $i$ updates the gradient surrogates as:

$$\mathbf{s}_i^k = \textstyle\sum_{j=1}^N (C_1)_{ij}\mathbf{s}_j^{k-1} + \frac{1}{M}\sum_{j=1}^N (C_2)_{ij} \cdot \big[\nabla_{\boldsymbol{\theta}} J_{j,p_k}(\boldsymbol{\theta}_j^k, \mathbf{w}_j^k) - \nabla_{\boldsymbol{\theta}} J_{j,p_k}(\boldsymbol{\theta}_j^{\tau_{p_k}^{k-1}}, \mathbf{w}_j^{\tau_{p_k}^{k-1}})\big] \tag{10}$$

$$\mathbf{d}_i^k = \mathbf{d}_i^{k-1} + \tfrac{1}{M}\big[\nabla_{\mathbf{w}_i} J_{i,p_k}(\boldsymbol{\theta}_i^k, \mathbf{w}_i^k) - \nabla_{\mathbf{w}_i} J_{i,p_k}(\boldsymbol{\theta}_i^{\tau_{p_k}^{k-1}}, \mathbf{w}_i^{\tau_{p_k}^{k-1}})\big], \tag{11}$$

where we define the initial values of gradients $\nabla_{\boldsymbol{\theta}} J_{i,p}(\boldsymbol{\theta}_i^0, \mathbf{w}_i^0) := 0$ and $\nabla_{\mathbf{w}_i} J_{i,p}(\boldsymbol{\theta}_i^0, \mathbf{w}_i^0) := 0$, for $p \in [M]$. Update the primal and dual variables using the gradients surrogates $\mathbf{s}_i^k$ and $\mathbf{d}_i^k$ w.r.t. $\boldsymbol{\theta}_i$ and $\mathbf{w}_i$:

$$\boldsymbol{\theta}_i^{k+1} = \textstyle\sum_{j=1}^N (R_1)_{ij}\boldsymbol{\theta}_j^k - \gamma_1\sum_{j=1}^N (R_2)_{ij}\mathbf{s}_j^k \tag{12}$$

$$\mathbf{w}_i^{k+1} = \mathbf{w}_i^k + \gamma_2\mathbf{d}_i^k. \tag{13}$$

**end for**

**end for**

Generally speaking, our method utilizes new mixing matrices to estimate the parameter variable and track the gradient over space (across $N$ agents) and time (across $M$ samples).

**Proposed Method** The following primal-dual gradient method is a prototype solving (8) in the single-agent setting:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \gamma_1 \nabla_{\boldsymbol{\theta}} \mathbf{J}(\boldsymbol{\theta}^k, \{\mathbf{w}_i^k\}_{i=1}^N),$$
$$\mathbf{w}_i^{k+1} = \mathbf{w}_i^k + \gamma_2 \nabla_{\mathbf{w}_i} \mathbf{J}(\boldsymbol{\theta}^k, \{\mathbf{w}_i^k\}_{i=1}^N), i \in [N], \quad (14)$$

where $\gamma_1, \gamma_2 > 0$ are step sizes. In the MARL model, however, it is challenging to implement this algorithm. Recall in this setting agent $i$ only has access to the functions of its own and the neighbor agents $\{J_{j,p}(\cdot) : (j,i) \in E\}$. Moreover, computing the batch gradient requires summing up over M samples, which is expensive when $M \gg 1$ as the computation complexity would be $\mathcal{O}(Md)$. We tackle these issues by combining the gradient tracking idea of (Qu & Li, 2018) with an incremental update scheme from (Schmidt et al., 2017) in the following primal-dual hierarchical distributed incremental aggregated gradient (PD-H) method. Here, sequences $\{\mathbf{s}_i^k\}_{k\geq 1}$ and $\{\mathbf{d}_i^k\}_{k\geq 1}$ are introduced to track the gradients *w.r.t.* $\boldsymbol{\theta}_i$ and $\mathbf{w}_i$, respectively. Each agent $i \in [N]$ maintains a local copy of the primal parameter, i.e., $\{\boldsymbol{\theta}_i^k\}_{i\in[N]}$. At the $k$-th iteration, we update the dual variable via gradient update using $\mathbf{d}_i^k$. However, different form previous works, here each primal variable $\boldsymbol{\theta}_i^{k+1}$ is obtained by first averaging both variables $\{\boldsymbol{\theta}_j^k\}$ and $\{\mathbf{s}_j^k\}$ over its neighbors specified by matrix $R_1$ and $R_2$, and then updating the averaged variable with the averaged gradient. The details of our method are presented in Algorithm 1.

Note that here $\mathbf{s}_i^k$ and $\mathbf{d}_i^k$ represent the surrogate functions for the primal and dual gradients; $\underline{\mathbf{s}}^k := [\mathbf{s}_1^k, \cdots, \mathbf{s}_N^k]$ and $\underline{\mathbf{d}}^k := [\mathbf{d}_1^k, \cdots, \mathbf{d}_N^k]$ are the matrices with $\mathbf{s}_i^k$ and $\mathbf{d}_i^k$ being their $i$-th column respectively. We also define $\underline{\boldsymbol{\theta}}^k := [\boldsymbol{\theta}_1^k, \cdots, \boldsymbol{\theta}_N^k]$; $\underline{\mathbf{w}}^k := [\mathbf{w}_1^k, \cdots, \mathbf{w}_N^k]$. Moreover, the counter variable is defined as $\tau_p^k = \max\{l \geq 0 : l \leq k, p_l = p\}$, which represents the last iteration when the $p$-th sample is visited before iteration $k$, and we set $\tau_p^k = 0$ if the $p$-th sample has never been visited.

Now we analyze the convergence of the proposed approach. We begin by defining the following concept that is needed in the analysis.

**Definition 3.1.** *A spanning tree of a directed graph (digraph) is a directed tree that connects the root to all other vertices in the graph (see (Godsil & Royle, 2001)).*

For any nonnegative matrix $Q \in \mathbb{R}^{n\times n}$, we let $G_Q := (V_Q, E_Q)$ denote the directed graph induced by the matrix $Q$, where $V_Q = \{1, 2, ..., n\}$ and $(j,i) \in E_Q$ iff $Q_{ij} > 0$. We define $R_Q$ as the set of roots of directed spanning trees in the graph $G_Q$.

Moreover, The following conditions are used commonly in previous work. The first is to ensure that each sample is picked up frequently in the algorithm.

**Assumption 1.** It holds that $|k - \tau_p^k| \leq M$, $\forall\ p \in [M]$, $k \geq 1$, i.e., each sample is selected at least once per $M$ iterations.

*Remark* 1. The requirement $|k - \tau_p^k| \leq M$ can be relaxed to $|k - \tau_p^k| \leq C \cdot M$ for some constant $C \geq 1$. This assumption is easier to be satisfied but will not change the linear convergence proved latter.

The following assumptions are important to establish the linear convergence rate.

**Assumption 2.** The empirical correlation matrix $\widehat{\mathbf{A}}$ defined in (5) is full rank, and $\widehat{\mathbf{D}}$ defined in (5) is non-singular.

Beside, the proof of the convergence also relies on the assumptions about the mixing matrices.

**Assumption 3.** $R_1, R_2 \in \mathbb{R}^{d\times d}$ are nonnegative and row stochastic with a shared left eigenvector $\mathbf{u}$ of eigenvalue 1 and $C_1, C_2 \in \mathbb{R}^{d\times d}$ are nonnegative and column stochastic. In addition, $(R_1)_{ii} > 0$ and $(C_1)_{ii} > 0$ for all $i \in V$.

*Remark* 2. Given $R_1$, $R_2$ can be set as $\lambda\mathbf{I} + (1-\lambda)R_1$, for $\lambda \in [0,1]$ such that Assumption 3 holds. Moreover, here $(R_1)_{ii} > 0$ and $(C_1)_{ii} > 0$ are necessary to ensure that $\sigma_{R_1} < 1$ and $\sigma_{C_1} < 1$, where $\sigma_{R_1}$ and $\sigma_{C_1}$ are the spectral radii of $(R_1 - \mathbf{1}\mathbf{u}^\top/N)$ and $(C_1 - \mathbf{v}\mathbf{1}^\top/N)$, respectively. Note though not mentioned explicitly, (Wai et al., 2018) also requires the diagonal entries of the doubly stochastic mixing matrix $W$ are positive to ensure $\|W - N^{-1}\mathbf{1}\mathbf{1}^\top\|_{1,\infty} < 1$.

**Assumption 4.** Each of the graphs $G_{R_1}$ and $G_{C_1^\top}$ contains at least one spanning tree. Further, $R_{R_1} \cap R_{C_1^\top} \neq \emptyset$.

Let $(\boldsymbol{\theta}^*, \{\mathbf{w}_i^*\}_{i=1}^N)$ be the optimal solution of (8). Based on these assumptions, now we can present the main theorem.

**Theorem 1.** *Suppose that Assumptions 1-4 hold. Set the step sizes as $\gamma_2 = \beta\gamma, \gamma_1 = \gamma$ with $\beta := 8r'(\rho + \lambda_{\max}(\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}))/\lambda_{\min}(\widehat{\mathbf{D}})$, where $r' := \mathbf{u}^\top\mathbf{v}/N$ and $\gamma > 0$. Define $\bar{\boldsymbol{\theta}}^k = \frac{1}{N}\underline{\boldsymbol{\theta}}^k\mathbf{u}$ as the weighted average of parameters. When the primal step size $\gamma$ is sufficiently small , there exists a constant $\sigma \in (0,1)$ such that*

$$\|\bar{\boldsymbol{\theta}}^k - \boldsymbol{\theta}^*\|^2 + (r')^2/(\beta N)\textstyle\sum_{i=1}^N \|\mathbf{w}_i^k - \mathbf{w}_i^*\|^2 = \mathcal{O}(\sigma^k),$$
$$\tfrac{1}{N}\textstyle\sum_{i=1}^N \|\boldsymbol{\theta}_i^k - \bar{\boldsymbol{\theta}}^k\| = \mathcal{O}(\sigma^k).$$

*If $M, N \gg 1$ and we have $\max\{\sigma_{R_1}, \sigma_{C_1}\} = 1 - e/N$ for some $e > 0$, then setting $\gamma = \mathcal{O}(1/\max\{M^2, N^2\})$ yields a convergence rate $\sigma = 1 - \mathcal{O}(1/\max\{MN^2, M^3\})$.*

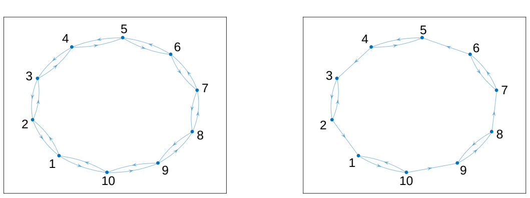
The above theorem shows that the iterates $(\bar{\boldsymbol{\theta}}^k, \{\mathbf{w}_i^k\}_{i=1}^N)$ generated by the proposed algorithm converge to the optimum in a linear rate which is the same as that in (Wai et al., 2018). Therefore, as will be shown in the experiment, the proposed method can achieve better communication efficiency by reducing the information transmission in each epoch. Moreover, the consensus error of local primal variables $\frac{1}{N}\sum_{i=1}^N \|\boldsymbol{\theta}_i^k - \bar{\boldsymbol{\theta}}^k\|$ also converges to 0 linearly. The detailed proof of Theorem 1 is in the Appendix.

## 4. Experiments

In this section we test our algorithm on the policy evaluation problems for the mountaincar task (Du et al., 2017). We obtain $M = 5000$ samples with $d = 300$ features generated by a known policy. Given a sample $p$, each agent is assigned randomly with local reward such that the average of the local rewards equals $r_c(\mathbf{s}_p, \mathbf{a}_p)$. We consider different connection graphs: the ring and the Erdos-Renyi (ER) graph. In these settings, we compare various advanced algorithms: (1) PDBG: the centralized method in (14); (2) SAGA: the centralized method proposed in (Du et al., 2017); (3) PD-DistIAG: the decentralized method in (Wai et al., 2018); (4) PD-H: the proposed method.

For the decentralized methods, we simulate networks of $N = 10$ agents with the ring graph and $N = 100$ agents with the ER graph. We first compare these algorithms' communication cost in each iteration, that is, the total number of transmissions between the nodes in each round of update. Then their optimality gaps vs. epoch number are examined. Here we choose the step sizes $\gamma_1 = 0.005/\lambda_{\max}(\widehat{\mathbf{A}}), \gamma_2 = 0.005$. Figures 1 and 3 show the communication graphs corresponding to the mixing matrices $W$ in the PD-DistIAG and $R_1$ in the proposed method for different topologies. We set $R_2 = C_2 = I$, $C_1 = R_1^\top$ in the first network and $R_2 = R_1, C_1 = C_2 = R_1^\top$ in the second one. Note the left undirected communication graphs of PD-DistIAG in Figure 1 and 3 can be regarded as special digraphs with each undirected edge being equivalent to two in-and-out directed edges. Given the graph in Figure 1a, the right graph in Figure 1b is formed by cutting its directed edges alternatively in one direction. In contrast, if the left graph is considered as an undirected graph as in PD-DistIAG, then cutting any two of its *undirected* edges would result in disconnectedness that prevents PD-DistIAG from converging. In the ER graph case, we generate the graph in Figure 3a with connection probability $1.1\log(N)/N$. Basing on this graph, the right digraph in Figure 3b is formed by randomly cutting 21% of its directed edges. Note that cutting the same ratio of its *undirected* edges reduces the connection probability to $0.98\log(N)/N$, which will also lead to disconnectedness with high probability (Erdos & Rényi, 1960) that prevents PD-DistIAG from converging.

Figures 2 and 4 compare the optimality gaps of the objective function versus the epoch number, which is defined as $t/M$.

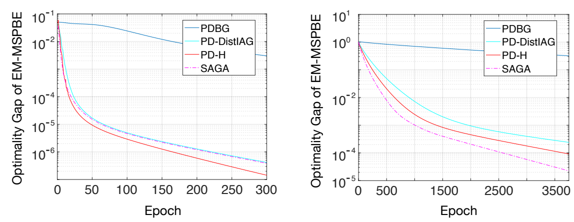



(a) PD-DistIAG (b) PD-H (reduce by **25%**)

*Figure 1.* Communication Graph in Algorithms (Ring graph).

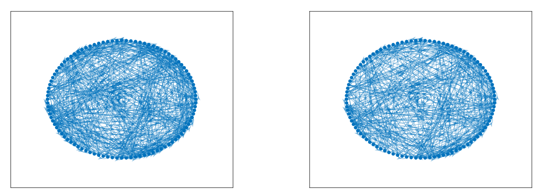
Their communication graphs are in Figure 1 and Figure 3 respectively. Recall the number of links in the graphs is in proportion to the communication cost and the number of summations over agents in each iteration. Therefore it can be observed from Figure 1 and Figure 3 that in both settings the proposed approach saves more than 20% communication per iteration. Moreover, Figure 2 and Figure 4 show that when $\rho > 0$ the convergence of PD-H is comparable to that of the centralized method SAGA. When $\rho = 0$, though slower than the centralized method SAGA, the advantages of PD-H over other methods are more obvious as more epochs are needed to reach the minimum due to the adversarial conditional number. It consistently converges faster than the PD-DistIAG, while requiring less communication in each iteration. Such results can be observed generally in other settings of parameters as shown in Table 1.

**Acknowledgements.** The authors thank the support of the DARPA Young Faculty Award, Grant N66001-14-1-4047.



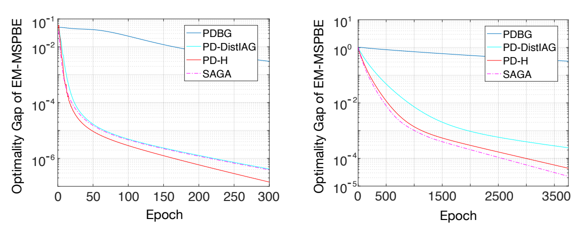

(a) $\rho = 0.01$ (b) $\rho = 0$

*Figure 2.* Convergence Comparison (Ring graph).



(a) PD-DistIAG (b) PD-H (reduce by **21%**)

*Figure 3.* Communication Graph in Algorithms (ER graph).




(a) $\rho = 0.01$ (b) $\rho = 0$

*Figure 4.* Convergence Comparison (ER Graph).

*Table 1.* Comparison in various settings of parameters.

| (N,Epoch,$\rho$) | graph | communication | optimality gap |
|---|---|---|---|
| (10,300,0.01) | Star | reduce by **17%** | improve by 7.3e-09 |
| (10,5000,0) | ER | reduce by **21%** | improve by 6.6e-05 |
| (500,300,0.01) | Ring | reduce by **25%** | improve by 3.9e-09 |
| (500,5000,0) | ER | reduce by **21%** | improve by 5.6e-05 |

## References


Arslan, G. and Yüksel, S. Decentralized q-learning for stochastic teams and games. *IEEE Transactions on Automatic Control*, 62(4):1545–1558, 2017.

Chen, Y. and Wang, M. Stochastic primal-dual methods and sample complexity of reinforcement learning. *arXiv preprint arXiv:1612.02516*, 2016.

Corke, P., Peterson, R., and Rus, D. Networked robots: Flying robot navigation using a sensor net. In *Robotics research. The eleventh international symposium*, pp. 234–243. Springer, 2005.

Cortes, J., Martinez, S., Karatas, T., and Bullo, F. Coverage control for mobile sensing networks. *IEEE Transactions on robotics and Automation*, 20(2):243–255, 2004.

Dai, B., He, N., Pan, Y., Boots, B., and Song, L. Learning from conditional distributions via dual embeddings. In *Artificial Intelligence and Statistics*, pp. 1458–1467, 2017a.

Dai, B., Shaw, A., Li, L., Xiao, L., He, N., Chen, J., and Song, L. Smoothed dual embedding control. *arXiv preprint arXiv:1712.10285*, 2017b.

Dall'Anese, E., Zhu, H., and Giannakis, G. B. Distributed optimal power flow for smart microgrids. *IEEE Trans. Smart Grid*, 4(3):1464–1475, 2013.

Defazio, A., Bach, F., and Lacoste-Julien, S. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in neural information processing systems*, pp. 1646–1654, 2014.

Du, S. S., Chen, J., Li, L., Xiao, L., and Zhou, D. Stochastic variance reduction methods for policy evaluation. *arXiv preprint arXiv:1702.07944*, 2017.

Erdos, P. and Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

Gharesifard, B. and Cortés, J. When does a digraph admit a doubly stochastic adjacency matrix? In *American Control Conference (ACC), 2010*, pp. 2440–2445. IEEE, 2010.

Gharesifard, B. and Cortés, J. Distributed strategies for generating weight-balanced and doubly stochastic digraphs. *European Journal of Control*, 18(6):539–557, 2012.

Godsil, C. D. and Royle, G. F. Algebraic graph theory. In *Graduate texts in mathematics*, 2001.

Horn, R. A. and Johnson, C. R. *Matrix analysis*. Cambridge University Press, 1990.

Hu, J. and Wellman, M. P. Nash q-learning for general-sum stochastic games. *Journal of machine learning research*, 4(Nov):1039–1069, 2003.

Kober, J., Bagnell, J. A., and Peters, J. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

Lauer, M. and Riedmiller, M. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *In Proceedings of the Seventeenth International Conference on Machine Learning*. Citeseer, 2000.

Lee, D., Yoon, H., and Hovakimyan, N. Primal-dual algorithm for distributed reinforcement learning: Distributed gtd2. *arXiv preprint arXiv:1803.08031*, 2018.

Li, M., Andersen, D. G., Smola, A. J., and Yu, K. Communication efficient distributed machine learning with the parameter server. In *Advances in Neural Information Processing Systems*, pp. 19–27, 2014.

Lian, X., Wang, M., and Liu, J. Finite-sum composition optimization via variance reduced gradient descent. *arXiv preprint arXiv:1610.04674*, 2016.

Littman, M. L. Markov games as a framework for multi-agent reinforcement learning. In *Machine Learning Proceedings 1994*, pp. 157–163. Elsevier, 1994.

Littman, M. L. Value-function reinforcement learning in markov games. *Cognitive Systems Research*, 2(1):55–66, 2001.

Ma, M., Nikolakopoulos, A. N., and Giannakis, G. B. Fast decentralized learning via hybrid consensus admm. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018a.

Ma, M., Ren, J., Giannakis, G. B., and Haupt, J. Fast asynchronous decentralized optimization: allowing multiple masters. In *IEEE Global Conference on Signal and Information Processing*, 2018b.

Macua, S. V., Chen, J., Zazo, S., and Sayed, A. H. Distributed policy evaluation under multiple behavior strategies. *IEEE Transactions on Automatic Control*, 60(5):1260–1274, 2015.

Macua, S. V., Tukiainen, A., Hernández, D. G.-O., Baldazo, D., de Cote, E. M., and Zazo, S. Diff-dac: Distributed actor-critic for multitask deep reinforcement learning. *arXiv preprint arXiv:1710.10363*, 2017.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529, 2015.

Nedic, A. and Bertsekas, D. P. Least squares policy evaluation algorithms with linear function approximation. *Discrete Event Dynamic Systems*, 13(1-2):79–110, 2003.

Nedic, A. and Ozdaglar, A. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

Prabuchandran, K., AN, H. K., and Bhatnagar, S. Multi-agent reinforcement learning for traffic signal control. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 2529–2534. IEEE, 2014.

Pu, S., Shi, W., Xu, J., and Nedich, A. A push-pull gradient method for distributed optimization in networks. *arXiv preprint arXiv:1803.07588*, 2018.

Qu, G. and Li, N. Harnessing smoothness to accelerate distributed optimization. *IEEE Transactions on Control of Network Systems*, 5(3):1245–1260, 2018.

Rabbat, M. and Nowak, R. Distributed optimization in sensor networks. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pp. 20–27. ACM, 2004.

Ren, J. and Haupt, J. Provably communication-efficient asynchronous distributed inference for convex and nonconvex problems. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 638–642. IEEE, 2018.

Ren, J. and Haupt, J. A provably communication-efficient asynchronous distributed inference method for convex and nonconvex problems. *arXiv preprint arXiv:1903.06871*, 2019.

Ren, J., Li, X., and Haupt, J. Communication-efficient distributed optimization for sparse learning via two-way truncation. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017.

Schmidt, M., Le Roux, N., and Bach, F. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162(1-2):83–112, 2017.

Shi, W., Ling, Q., Wu, G., and Yin, W. Extra: An exact first-order algorithm for decentralized consensus optimization. *SIAM Journal on Optimization*, 25(2):944–966, 2015.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction, Second Edition*. MIT press, 2018.

Tsitsiklis, J., Bertsekas, D., and Athans, M. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE transactions on automatic control*, 31(9):803–812, 1986.

Wai, H.-T., Yang, Z., Wang, Z., and Hong, M. Multi-agent reinforcement learning via double averaging primal-dual optimization. In *Advances in Neural Information Processing Systems*, pp. 9672–9683. 2018.

Wang, M. Primal-dual $\pi$ learning: Sample complexity and sublinear run time for ergodic Markov decision problems. *arXiv preprint arXiv:1710.06100*, 2017.

Zhang, K., Yang, Z., Liu, H., Zhang, T., and Başar, T. Fully decentralized multi-agent reinforcement learning with networked agents. *arXiv preprint arXiv:1802.08757*, 2018.

## A. Lemmata to Prove Theorem 1

By Assumption 3, we directly have the following result.

**Lemma 1.** *Under Assumption 3, the matrix $C_1$ has a non-negative right eigenvector $\mathbf{v}$ with eigenvalue 1 satisfying $\boldsymbol{1}^\top \mathbf{v} = N$ (see (Horn & Johnson, 1990)).*

The proof of Theorem 1 also needs the following lemmata.

**Lemma 2** ((Pu et al., 2018))**.** *Suppose $R_{R_1} \neq \emptyset$ and $R_{C_1^\top} \neq \emptyset$. Then under Assumption 3, it holds that $R_{R_1} \cap R_{C_1^\top} \neq \emptyset$ if and only if $\mathbf{u}^\top \mathbf{v} \neq 0$.*

We know Assumption 4 essentially ensures sufficient connections given by $R_1$ and $C_1$. By the definition of $r'$ and Lemma 2, this guarantees $r' \neq 0$.

**Lemma 3.** *Under Assumption 1-4, we have the following linear inequalities:*

$$\begin{bmatrix} \|\underline{\boldsymbol{\theta}}^{k+1} - \bar{\boldsymbol{\theta}}^{k+1}\boldsymbol{1}^\top\| \\ \|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \end{bmatrix} \leq Q_0 \begin{bmatrix} \max_{(k-2M)_+ \leq s \leq k} \|\underline{\boldsymbol{\theta}}^s - \bar{\boldsymbol{\theta}}^s \boldsymbol{1}^\top\| \\ \max_{(k-M)_+ \leq s \leq k} \|\underline{\mathbf{s}}^s - \bar{\mathbf{s}}^s \mathbf{v}^\top\| \\ \max_{(k-2M)_+ \leq s \leq k} \|\underline{\boldsymbol{v}}^s\| \end{bmatrix}, \tag{15}$$

*where matrix $Q_0 = [q_{ij}]$ is defined as*

$$\begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix} = \begin{bmatrix} \sigma_{R_1} + \gamma\sigma_{R_2}\|\ \mathbf{v} - \boldsymbol{1}\| \frac{\rho}{\sqrt{N}} \\ \sigma_{C_2}\rho\|R_1 - \mathbf{I}\|_S + \gamma(\sigma_{C_2}\rho^2\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}} + \sigma_{C_2}\bar{A}^2\beta) \end{bmatrix}$$

$$\begin{bmatrix} q_{12} \\ q_{22} \end{bmatrix} = \begin{bmatrix} \gamma\sigma_{R_2} \\ \sigma_{C_1} + \sigma_{C_2}\rho\gamma\|R_2\|_S \end{bmatrix},$$

$$\begin{bmatrix} q_{13} \\ q_{23} \end{bmatrix} = \begin{bmatrix} \sqrt{2}\gamma\sigma_{R_2}\|\ \mathbf{v} - \boldsymbol{1}\|L\max(1, \frac{\sqrt{\beta}}{r'}) \\ \sqrt{2}\big(\sigma_{C_2}L^2\beta\gamma\sqrt{N} + \sigma_{C_2}L^2\gamma\|R_2\mathbf{v}\|\big)\max\{1, \frac{\sqrt{\beta}}{r'}\} \end{bmatrix}, \tag{16}$$

*Here for $s \geq 0$, we define*

$$\underline{\boldsymbol{v}}^s := \begin{bmatrix} \bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^* \\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_1^s - \mathbf{w}_1^*) \\ \vdots \\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_N^s - \mathbf{w}_N^*) \end{bmatrix}, \quad L := \max\{\rho, \bar{A}, \bar{D}\}$$

*with $\bar{A}$ and $\bar{D}$ being defined as*

$$\bar{A} := \max_{1,\cdots,M} \|\mathbf{A}_p\|_S, \qquad \bar{D} := \max_{1,\cdots,M} \|\mathbf{D}_p\|_S. \tag{17}$$

Lemma 3 gives the progress of the consensus errors of $\underline{\boldsymbol{\theta}}$ and $\underline{\mathbf{s}}$. This will help establish the relation between the optimality gap progress and the consensus error progress in one iteration in the proof of Theorem 1. By analyzing this inequality system, we can find the sufficient condition that guarantees the linear convergence of the proposed algorithm.

The proofs of the lemma and the theorem are in the following Appendix B and Appendix C.

## B. Proof of Lemma 3

To establish the progress of the consensus errors of $\underline{\boldsymbol{\theta}}$ and $\underline{\mathbf{s}}$. Firstly we define the gradient vector

$$\mathbf{J}_p(\underline{\boldsymbol{\theta}}^k, \underline{\mathbf{w}}^k) := \big[J_{1,p}(\boldsymbol{\theta}_1^k, \mathbf{w}_1^k), \cdots, J_{N,p}(\boldsymbol{\theta}_N^k, \mathbf{w}_N^k)\big].$$

Then the update of $\mathbf{s}_i^k$ and $\boldsymbol{\theta}_i^k$ in (10) and (12) for $i \in \{1, \cdots, N\}$ can be written in a compact form:

$$\underline{\mathbf{s}}^k = \underline{\mathbf{s}}^{k-1}C_1^\top + \frac{1}{M}\big[\nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_k}(\underline{\boldsymbol{\theta}}^k, \underline{\mathbf{w}}^k) - \nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_k}(\underline{\boldsymbol{\theta}}^{\tau_{p_k}^{k-1}}, \underline{\mathbf{w}}^{\tau_{p_k}^{k-1}})\big]C_2^\top \tag{18}$$

$$\underline{\boldsymbol{\theta}}^{k+1} = \underline{\boldsymbol{\theta}}^k R_1^\top - \gamma_1 \underline{\mathbf{s}}^k R_2^\top \tag{19}$$

Moreover, we define the weighted average of $\underline{\boldsymbol{\theta}}^{k+1}$ over columns by $\bar{\boldsymbol{\theta}}^{k+1} := \frac{1}{N}\underline{\boldsymbol{\theta}}^{k+1}\mathbf{u}$. Then by the definition of $\mathbf{u}$ we have

$$\begin{aligned} \bar{\boldsymbol{\theta}}^{k+1} &:= \frac{1}{N}(\underline{\boldsymbol{\theta}}^k R_1^\top - \gamma_1\underline{\mathbf{s}}^k R_2^\top)\mathbf{u} \\ &= \bar{\boldsymbol{\theta}}^k - \frac{\gamma_1}{N}\underline{\mathbf{s}}^k R_2^\top\mathbf{u} \\ &= \bar{\boldsymbol{\theta}}^k - \frac{\gamma_1}{N}\underline{\mathbf{s}}^k\mathbf{u}, \end{aligned} \tag{20}$$

where the last two equalities use Assumption 3. It follows from (20) and (19) that

$$\begin{aligned} &\underline{\boldsymbol{\theta}}^{k+1} - \bar{\boldsymbol{\theta}}^{k+1}\mathbf{1}^\top \\ &= (\underline{\boldsymbol{\theta}}^k R_1^\top - \gamma_1\underline{\mathbf{s}}^k R_2^\top) - \big(\bar{\boldsymbol{\theta}}^k + \frac{\gamma_1}{N}\underline{\mathbf{s}}^k\mathbf{u}\big)\mathbf{1}^\top \\ &= (\underline{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^k\mathbf{1}^\top)R_1^\top - \gamma_1\underline{\mathbf{s}}^k\big(R_2 - \frac{\mathbf{1}\mathbf{u}^\top}{N}\big)^\top \\ &= (\underline{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^k\mathbf{1}^\top)\big(R_1 - \frac{\mathbf{1}\mathbf{u}^\top}{N}\big)^\top \\ &- \gamma_1(\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{1}^\top)\big(R_2 - \frac{\mathbf{1}\mathbf{u}^\top}{N}\big)^\top. \end{aligned} \tag{21}$$

Equation (21) gives an expression of the consensus error of the column vectors in $\underline{\boldsymbol{\theta}}^{k+1}$.

Similarly, the average of $\underline{\mathbf{s}}^{k+1}$ is defined as $\bar{\mathbf{s}}^{k+1} := \frac{1}{N}\underline{\mathbf{s}}^{k+1}\mathbf{1}$. We can also consider the consensus error of vectors $\underline{\mathbf{s}}^{k+1}$ with respect to its weighted average

$$
\begin{aligned}
&\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top \\
&= \Big(\frac{1}{M}\big[\nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_{k+1}}(\underline{\boldsymbol{\theta}}^{k+1}, \underline{\mathbf{w}}^{k+1}) - \nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_{k+1}}(\underline{\boldsymbol{\theta}}^{\tau^k_{p_{k+1}}}, \underline{\mathbf{w}}^{\tau^k_{p_{k+1}}})\big]\Big) \\
&\cdot \Big(C_2 - \frac{\mathbf{v}\mathbf{1}^\top}{N}\Big)^\top + \underline{\mathbf{s}}^k C_1^\top - \bar{\mathbf{s}}^k\mathbf{v}^\top \\
&= \Big(\frac{1}{M}\big[\nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_{k+1}}(\underline{\boldsymbol{\theta}}^{k+1}, \underline{\mathbf{w}}^{k+1}) - \nabla_{\underline{\boldsymbol{\theta}}}\mathbf{J}_{p_{k+1}}(\underline{\boldsymbol{\theta}}^{\tau^k_{p_{k+1}}}, \underline{\mathbf{w}}^{\tau^k_{p_{k+1}}})\big]\Big) \\
&\cdot \Big(C_2 - \frac{\mathbf{v}\mathbf{1}^\top}{N}\Big)^\top + (\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top)\Big(C_1 - \frac{\mathbf{v}\mathbf{1}^\top}{N}\Big)^\top. \qquad (22)
\end{aligned}
$$

Recall that we define $\sigma_{R_i}$ as the spectral radii of $(R_i - \mathbf{1}\mathbf{u}^\top/N)$, for $i = 1, 2$. Then By (21), we have

$$
\begin{aligned}
&\|\underline{\boldsymbol{\theta}}^{k+1} - \bar{\boldsymbol{\theta}}^{k+1}\mathbf{1}^\top\| \\
&\le \sigma_{R_1}\|\underline{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^k\mathbf{1}^\top\| + \gamma_1\sigma_{R_2}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{1}^\top\| \\
&\le \sigma_{R_1}\|\underline{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^k\mathbf{1}^\top\| + \gamma_1\sigma_{R_2}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \gamma_1\sigma_{R_2}\|\,\mathbf{v} - \mathbf{1}\|\|\bar{\mathbf{s}}^k\| \\
&\le \sigma_{R_1}\|\underline{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^k\mathbf{1}^\top\| + \gamma_1\sigma_{R_2}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| + \gamma_1\sigma_{R_2}\|\,\mathbf{v} - \mathbf{1}\| \\
&\cdot \Big\|\bar{\mathbf{s}}^k - \frac{1}{NM}\sum_{i=1}^N\sum_{p=1}^M \nabla_{\boldsymbol{\theta}} J_{i,p}(\bar{\boldsymbol{\theta}}^{\tau^k_p}, \mathbf{w}_i^{\tau^k_p})\Big\| + \gamma_1\sigma_{R_2}\|\,\mathbf{v} - \mathbf{1}\| \\
&\cdot \Big\|\frac{1}{NM}\sum_{i=1}^N\sum_{p=1}^M \nabla_{\boldsymbol{\theta}} J_{i,p}(\bar{\boldsymbol{\theta}}^{\tau^k_p}, \mathbf{w}_i^{\tau^k_p}) \\
&- \frac{1}{NM}\sum_{i=1}^N\sum_{p=1}^M \nabla_{\boldsymbol{\theta}} J_{i,p}(\boldsymbol{\theta}^*, \mathbf{w}_i^*)\Big\| \\
&\le \Big(\sigma_{R_1} + \gamma\sigma_{R_2}\|\mathbf{v} - \mathbf{1}\|\frac{\rho}{\sqrt{N}}\Big)\max_{(k-M)_+ \le s \le k}\|\underline{\boldsymbol{\theta}}^s - \bar{\boldsymbol{\theta}}^s\mathbf{1}^\top\| \\
&+ \gamma\sigma_{R_2}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| + \gamma\sigma_{R_2}\|\,\mathbf{v} - \mathbf{1}\|\rho \\
&\cdot \max_{(k-M)_+ \le s \le k}\|\bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\| + \gamma\sigma_{R_2}\|\,\mathbf{v} - \mathbf{1}\|\bar{A} \\
&\cdot \max_{(k-M)_+ \le s \le k}\frac{1}{N}\sum_{i=1}^N\|\mathbf{w}_i^s - \mathbf{w}_i^*\|. \qquad (23)
\end{aligned}
$$

Similarly, we define $\sigma_{C_i}$ as the spectral radii of $(C_i - \mathbf{v}\mathbf{1}^\top/N)$, for $i = 1, 2$. Then it follows from this definition and (22) that

$$
\begin{aligned}
&\|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \\
&\le \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\big(\rho\|\underline{\boldsymbol{\theta}}^{k+1} - \underline{\boldsymbol{\theta}}^{\tau^k_{p_{k+1}}}\| + \bar{A}\|\underline{\mathbf{w}}^{k+1} - \underline{\mathbf{w}}^{\tau^k_{p_{k+1}}}\|\big) \\
&\le \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\sum_{l=\tau^k_{p_{k+1}}}^{k}\big(\rho\|\underline{\boldsymbol{\theta}}^{l+1} - \underline{\boldsymbol{\theta}}^l\| + \bar{A}\|\underline{\mathbf{w}}^{l+1} - \underline{\mathbf{w}}^l\|\big) \\
&\le \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\sum_{l=(k-M)_+}^{k}\big(\rho\|(\underline{\boldsymbol{\theta}}^l(R_1 - \mathbf{I})^\top) - \gamma_1\underline{\mathbf{s}}^l R_2^\top\| \\
&+ \gamma_2\bar{A}\|\underline{\mathbf{d}}^l\|\big),
\end{aligned}
$$

where the last inequality uses the update equations (12)-(13) of the primal dual variables. Thus we have

$$
\begin{aligned}
&\|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \\
&\le \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\big(\rho\sum_{l=(k-M)_+}^{k}\|(\underline{\boldsymbol{\theta}}^l - \bar{\boldsymbol{\theta}}^l\mathbf{1}^\top)(R_1 - \mathbf{I})^\top\| + \gamma\|\underline{\mathbf{s}}^l R_2^\top\| \\
&+ \gamma_2\bar{A}\|\underline{\mathbf{d}}^l\|\big) \\
&\overset{(a)}{\le} \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\Big(\rho\sum_{l=(k-M)_+}^{k}\|(\underline{\boldsymbol{\theta}}^l - \bar{\boldsymbol{\theta}}^l\mathbf{1}^\top)(R_1 - \mathbf{I})^\top\| \\
&+ \gamma\rho\|(\underline{\mathbf{s}}^l - \bar{\mathbf{s}}^l\mathbf{v}^\top)R_2^\top\| + \gamma\rho\|R_2\mathbf{v}\|\|\bar{\mathbf{s}}^l\| + \gamma_2\bar{A}\|\underline{\mathbf{d}}^l - \mathbf{d}^*\|\Big) \\
&\overset{(b)}{\le} \sigma_{C_1}\|\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{1}{M}\Big(\rho\sum_{l=(k-M)_+}^{k}\big(\|(\underline{\boldsymbol{\theta}}^l - \bar{\boldsymbol{\theta}}^l\mathbf{1}^\top)(R_1 - \mathbf{I})^\top\| \\
&+ \gamma\rho\|(\underline{\mathbf{s}}^l - \bar{\mathbf{s}}^l\mathbf{v}^\top)R_2^\top\| \\
&+ \gamma\rho\|R_2\mathbf{v}\|\|\bar{\mathbf{s}}^l - \mathbf{s}^*\| + \gamma_2\bar{A}^2\frac{1}{M}\sum_{p=1}^M\|\underline{\boldsymbol{\theta}}^{\tau^l_p} - \boldsymbol{\theta}^*\mathbf{1}^\top\| \\
&+ \gamma_2\bar{A}\bar{D}\frac{1}{M}\sum_{p=1}^M\|\underline{\mathbf{w}}^{\tau^l_p} - \underline{\mathbf{w}}^*\|\big)\Big)
\end{aligned}
$$

where in the last inequality we use the Lipschitz property of the dual gradient surrogate $\underline{\mathbf{d}}^l$ with respect to ($w.r.t.$) the primal and dual variables. Note that inequalities $(a)$ and $(b)$ follow from the optimality condition of the problem (8) with respect to $\mathbf{w}_i$ for $i \in [N]$:

$$
\mathbf{d}^* := \begin{bmatrix} \frac{1}{M}\sum_{i=1}^M \nabla_{\mathbf{w}_1} J_{1,p}(\boldsymbol{\theta}^*, \mathbf{w}_1^*) \\ \vdots \\ \frac{1}{M}\sum_{i=1}^M \nabla_{\mathbf{w}_N} J_{N,p}(\boldsymbol{\theta}^*, \mathbf{w}_N^*) \end{bmatrix} = \mathbf{0},
$$

and also the optimality condition with respect to $\boldsymbol{\theta}$:

$$\mathbf{s}^* := \frac{1}{NM}\sum_{i=1}^{N}\sum_{p=1}^{M}\nabla_{\boldsymbol{\theta}} J_{i,p}(\boldsymbol{\theta}^*, \mathbf{w}_i^*) = 0.$$

For any given matrix $Q$, let $\|Q\|_S$ denotes its spectral norm. Therefore it follows that

$$\begin{aligned}
&\|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \\
&\le (\sigma_{C_1} + \sigma_{C_2}\rho\gamma\|R_2\|_S) \max_{(k-M)_+\le l\le k} \|\underline{\mathbf{s}}^l - \bar{\mathbf{s}}^l\mathbf{v}^\top\| \\
&+ \sigma_{C_2}\frac{\rho}{M}\sum_{(k-M)_+\le s\le k}\|R_1 - \mathbf{I}\|_S\|\underline{\boldsymbol{\theta}}^l - \bar{\boldsymbol{\theta}}^l\mathbf{1}^\top\| \\
&+ \sigma_{C_2}\frac{\bar{A}^2}{M}\sum_{(k-M)_+\le l\le k}\gamma_2 \max_{(l-M)_+\le s\le l}\|\underline{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\mathbf{1}^\top\| \\
&+ \sigma_{C_2}\frac{\bar{A}\bar{D}}{M}\sum_{(k-M)_+\le l\le k}\gamma_2 \max_{(l-M)_+\le s\le l}\|\underline{\mathbf{w}}^s - \underline{\mathbf{w}}^*\| \\
&+ \sigma_{C_2}\frac{\rho}{M}\sum_{(k-M)_+\le l\le k}\gamma\|R_2\mathbf{v}\|\Big(\frac{\rho}{\sqrt{N}}\|\underline{\boldsymbol{\theta}}^s - \bar{\boldsymbol{\theta}}^s\mathbf{1}^\top\| \\
&+ \rho \max_{(l-M)_+\le s\le l}\|\bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\| \\
&+ \frac{\bar{A}}{N}\max_{(l-M)_+\le s\le l}\sum_{i=1}^{N}\|\mathbf{w}_i^s - \mathbf{w}_i^*\|\Big),
\end{aligned}$$

Further bound the summations in the right hand side (RHS) of the above inequality we have

$$\begin{aligned}
&\|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \\
&\le (\sigma_{C_1} + \sigma_{C_2}\rho\gamma\|R_2\|_S) \max_{(k-M)_+\le l\le k} \|\underline{\mathbf{s}}^l - \bar{\mathbf{s}}^l\mathbf{v}^\top\| \\
&+ \big(\sigma_{C_2}\rho\|R_1 - \mathbf{I}\|_S + \sigma_{C_2}\rho^2\gamma\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}}\big) \\
&\cdot \max_{(k-2M)_+\le s\le k}\|\underline{\boldsymbol{\theta}}^s - \bar{\boldsymbol{\theta}}^s\mathbf{1}^\top\| \\
&+ \sigma_{C_2}\bar{A}^2\gamma_2 \max_{(k-2M)_+\le s\le k}\|\underline{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\mathbf{1}^\top\| \\
&+ \sigma_{C_2}\rho^2\gamma\|R_2\mathbf{v}\| \max_{(k-2M)_+\le s\le k}\|\bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\| \\
&+ \big(\sigma_{C_2}\bar{A}\bar{D}\gamma_2 + \sigma_C\rho\bar{A}\gamma\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}}\big) \\
&\cdot \max_{(k-2M)_+\le s\le k}\|\underline{\mathbf{w}}^s - \underline{\mathbf{w}}^*\|,
\end{aligned} \tag{24}$$

which further implies

$$\begin{aligned}
&\|\underline{\mathbf{s}}^{k+1} - \bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\| \\
&\overset{(c)}{\le} (\sigma_{C_1} + \sigma_{C_2}\rho\gamma\|R_2\|_S) \max_{(k-M)_+\le l\le k} \|\underline{\mathbf{s}}^l - \bar{\mathbf{s}}^l\mathbf{v}^\top\| \\
&+ \big(\sigma_{C_2}\rho\|R_1 - \mathbf{I}\|_S + \sigma_{C_2}\rho^2\gamma\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}} + \sigma_{C_2}\bar{A}^2\gamma_2\big) \\
&\cdot \max_{(k-2M)_+\le s\le k}\|\underline{\boldsymbol{\theta}}^s - \bar{\boldsymbol{\theta}}^s\mathbf{1}^\top\| \\
&+ \big(\sigma_{C_2}\bar{A}^2\gamma_2\sqrt{N} + \sigma_{C_2}\rho^2\gamma\|R_2\mathbf{v}\|\big) \\
&\cdot \max_{(k-2M)_+\le s\le k}\|\bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^*\| \\
&+ \big(\sigma_{C_2}\bar{A}\bar{D}\gamma_2 + \sigma_{C_2}\rho\bar{A}\gamma\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}}\big) \\
&\cdot \max_{(k-2M)_+\le s\le k}\|\underline{\mathbf{w}}^s - \underline{\mathbf{w}}^*\|,
\end{aligned} \tag{25}$$

where we use the triangular inequality in $(c)$.

## C. Proof of Theorem 1

For any $\beta > 0$, by the optimality condition, the primal-dual optimal solution to the optimal problem (8), $(\boldsymbol{\theta}^*, \{\mathbf{w}_i^*\}_{i=1}^N)$, satisfies

$$\mathbf{G}\begin{bmatrix}\boldsymbol{\theta}^* \\ \frac{r'}{\sqrt{\beta N}}\mathbf{w}_1^* \\ \vdots \\ \frac{r'}{\sqrt{\beta N}}\mathbf{w}_N^*\end{bmatrix} = -\begin{bmatrix}0 \\ \sqrt{\frac{\beta}{N}}\widehat{\boldsymbol{b}}_1 \\ \vdots \\ \sqrt{\frac{\beta}{N}}\widehat{\boldsymbol{b}}_N\end{bmatrix}, \tag{26}$$

where we define the constant $r' := \frac{\mathbf{u}^\top\mathbf{v}}{N}$ and

$$\mathbf{G} = \begin{bmatrix}\rho r'\mathbf{I} & \sqrt{\frac{\beta}{N}}\widehat{\mathbf{A}}^\top & \cdots & \sqrt{\frac{\beta}{N}}\widehat{\mathbf{A}}^\top \\ -\sqrt{\frac{\beta}{N}}\widehat{\mathbf{A}} & \frac{\beta}{r'}\widehat{\mathbf{D}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{\frac{\beta}{N}}\widehat{\mathbf{A}} & 0 & \cdots & \frac{\beta}{r'}\widehat{\mathbf{D}}\end{bmatrix}. \tag{27}$$

For $p \in \{1, \cdots, M\}$, $\mathbf{G}_p$ is defined as

$$\mathbf{G}_p = \begin{bmatrix}\rho r'\mathbf{I} & \sqrt{\frac{\beta}{N}}\mathbf{A}_p^\top & \cdots & \sqrt{\frac{\beta}{N}}\mathbf{A}_p^\top \\ -\sqrt{\frac{\beta}{N}}\mathbf{A}_p & \frac{\beta}{r'}\mathbf{D}_p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\sqrt{\frac{\beta}{N}}\mathbf{A}_p & 0 & \cdots & \frac{\beta}{r'}\mathbf{D}_p\end{bmatrix}. \tag{28}$$

By definition $\mathbf{G} = \frac{1}{M}\sum_{p=1}^{M}\mathbf{G}_p$. Define $\bar{\boldsymbol{\theta}}^k := \frac{1}{N}\mathbf{u}^\top\underline{\boldsymbol{\theta}}^k$ as the weighted average of the parameters at iteration $k$.

Furthermore, define

$$h_{\boldsymbol{\theta}}(k) := \rho\bar{\boldsymbol{\theta}}^k + \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathbf{A}}^\top \mathbf{w}_i^k \tag{29}$$

$$\widehat{g}_{\boldsymbol{\theta}}(k) := \frac{1}{N}\underline{\mathbf{s}}^k\mathbf{u}$$

$$h_{\mathbf{w}_i}(k) := \widehat{\mathbf{A}}\bar{\boldsymbol{\theta}}^k - \widehat{\mathbf{D}}\mathbf{w}_i^k - \widehat{\boldsymbol{b}}_i \tag{30}$$

$$g_{\mathbf{w}_i}(k) := \frac{1}{M}\sum_{p=1}^{M}(\mathbf{A}_p\boldsymbol{\theta}_i^{\tau_p^k} - \mathbf{D}_p\mathbf{w}_i^{\tau_p^k} - \boldsymbol{b}_{p,i}),$$

where $h_{\boldsymbol{\theta}}(k)$ and $h_{\mathbf{w}_i}(k)$ represent the batch gradients $w.r.t$ $\boldsymbol{\theta}$ and $\mathbf{w}_i$ at $(\bar{\boldsymbol{\theta}}^k, \mathbf{w}_i^k)$. It can be checked that $\bar{\boldsymbol{\theta}}^{k+1} = \bar{\boldsymbol{\theta}}^k - \gamma_1\widehat{g}_{\boldsymbol{\theta}}(k)$ and $\mathbf{w}_i^{k+1} = \mathbf{w}_i^k - \gamma_2 g_{\mathbf{w}_i}(k)$ for all $k \geq 1$. That is, the primal-dual variables $\bar{\boldsymbol{\theta}}^{k+1}$ and $\mathbf{w}_i^{k+1}$ are updated with $\widehat{g}_{\boldsymbol{\theta}}(k)$ and $g_{\mathbf{w}_i}(k)$. We also define $\underline{\mathbf{h}}(k)$, $\underline{\mathbf{g}}(k)$, and $\underline{\boldsymbol{v}}^k$ by

$$\underline{\mathbf{h}}(k) = \begin{bmatrix} r'h_{\boldsymbol{\theta}}(k) \\ -\sqrt{\frac{\beta}{N}}h_{\mathbf{w}_1}(k) \\ \vdots \\ -\sqrt{\frac{\beta}{N}}h_{\mathbf{w}_N}(k) \end{bmatrix}, \quad \underline{\mathbf{g}}(k) = \begin{bmatrix} \widehat{g}_{\boldsymbol{\theta}}(k) \\ -\sqrt{\frac{\beta}{N}}g_{\mathbf{w}_1}(k) \\ \vdots \\ -\sqrt{\frac{\beta}{N}}g_{\mathbf{w}_N}(k) \end{bmatrix}, \tag{31}$$

$$\underline{\boldsymbol{v}}^k = \begin{bmatrix} \bar{\boldsymbol{\theta}}^s - \boldsymbol{\theta}^* \\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_1^k - \mathbf{w}_1^*) \\ \vdots \\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_N^k - \mathbf{w}_N^*) \end{bmatrix}. \tag{32}$$

Note (29) and (30) can be written as

$$\mathbf{G}\begin{bmatrix} \bar{\boldsymbol{\theta}}(k) \\ \frac{r'}{\sqrt{\beta N}}\mathbf{w}_1^k \\ \vdots \\ \frac{r'}{\sqrt{\beta N}}\mathbf{w}_N^k \end{bmatrix} = \begin{bmatrix} r'h_{\boldsymbol{\theta}}(k) \\ -\sqrt{\frac{\beta}{N}}h_{\mathbf{w}_1}(k) - \sqrt{\frac{\beta}{N}}\widehat{\boldsymbol{b}}_1 \\ \vdots \\ -\sqrt{\frac{\beta}{N}}h_{\mathbf{w}_N}(k) - \sqrt{\frac{\beta}{N}}\widehat{\boldsymbol{b}}_N \end{bmatrix}. \tag{33}$$

Combining (33) and (26) yields

$$\underline{\mathbf{h}}(k) = \mathbf{G}\underline{\boldsymbol{v}}^k. \tag{34}$$

By the analysis similar to (Du et al., 2017), it can be shown that under Assumption 2 and with $\beta := \frac{8r'(\rho+\lambda_{\max}(\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}))}{\lambda_{\min}(\widehat{\mathbf{D}})}$, $\mathbf{G}$ is full rank whose eigenvalue satisfying

$$\lambda_{\max}(\mathbf{G}) \leq \left|\frac{\lambda_{\max}(\widehat{\mathbf{D}})}{\lambda_{\min}(\widehat{\mathbf{D}})}\right|\lambda_{\max}(\rho r'\mathbf{I} + r'\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}})$$

$$\lambda_{\min}(\mathbf{G}) \geq \frac{8r'}{9}\lambda_{\min}(\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}). \tag{35}$$

Furthermore, assume $\mathbf{G} := U\Lambda U^{-1}$ to be the eigen-decomposition of $\mathbf{G}$, where $\Lambda$ is the diagonal matrix of $G's$ eigenvalues; the columns of $U$ are the eigenvectors. Then let $U$ satisfy

$$\|U\| \leq 8r'\left|\frac{\lambda_{\max}(\widehat{\mathbf{D}})}{\lambda_{\min}(\widehat{\mathbf{D}})}\right|(\rho + \lambda_{\max}(\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}}))$$

$$\|U^{-1}\| \leq \frac{1}{\rho r' + r'\lambda_{\max}(\widehat{\mathbf{A}}^\top\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{A}})} \tag{36}$$

Furthermore, we define the upper bounds of the spectral norms by

$$G := \|\mathbf{G}\|_S, \qquad \bar{G} = \max_{p=1,\cdots,M}\|\mathbf{G}_p\|_S, \tag{37}$$

$$\bar{A} = \max_{p=1,\cdots,M}\|\mathbf{A}_p\|_S, \quad \bar{D} = \max_{p=1,\cdots,M}\|\mathbf{D}_p\|_S. \tag{38}$$

We also define the Lyapunov function as

$$\boldsymbol{\varepsilon}_c(k) := \frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\theta}_i^k - \bar{\boldsymbol{\theta}}^k\|.$$

Next, recall that $\gamma_1 := \gamma$ and $\gamma_2 := \beta\gamma$. In the following we establish a bound on the optimality gap of the primal-dual variables, $\underline{\boldsymbol{v}}^k$. Note that $\underline{\boldsymbol{v}}^{k+1} = \underline{\boldsymbol{v}}^k - \gamma\underline{\mathbf{g}}(k)$. Thus we have

$$\underline{\boldsymbol{v}}^{k+1} = (\mathbf{I} - \gamma\mathbf{G})\underline{\boldsymbol{v}}^k + \gamma(\underline{\mathbf{h}}(k) - \underline{\mathbf{g}}(k)). \tag{39}$$

Now we consider the difference $\underline{\mathbf{h}}(k) - \underline{\mathbf{g}}(k)$. Its first block can be written as

$$\begin{aligned}
&[\underline{\mathbf{h}}(k) - \underline{\mathbf{g}}(k)]_1 \\
&= r'h_{\boldsymbol{\theta}}(k) - \widehat{g}_{\boldsymbol{\theta}}(k) \\
&= -\frac{1}{N}(\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top)\mathbf{u} - \frac{1}{N}\bar{\mathbf{s}}^k\mathbf{v}^\top\mathbf{u} + r'h_{\boldsymbol{\theta}}(k) \\
&= -\frac{1}{N}(\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top)(\mathbf{u} - \mathbf{1}) + r'(h_{\boldsymbol{\theta}}(k) - \bar{\mathbf{s}}^k).
\end{aligned} \tag{40}$$

For any $i \in \{1, \cdots, N\}$, the $(i+1)$-th block is

$$\begin{aligned}
&[\underline{\mathbf{h}}(k) - \underline{\mathbf{g}}(k)]_{i+1} \\
&= \sqrt{\frac{\beta}{N}}\frac{1}{M}\sum_{p=1}^{M}\mathbf{A}_p\left(\bar{\boldsymbol{\theta}}^k - \boldsymbol{\theta}_i^{\tau_p^k}\right) + \mathbf{D}_p(\mathbf{w}_i^k - \mathbf{w}_i^{\tau_p^k}) \\
&= \sqrt{\frac{\beta}{N}}\frac{1}{M}\sum_{p=1}^{M}\mathbf{A}_p\left(\bar{\boldsymbol{\theta}}^k - \bar{\boldsymbol{\theta}}^{\tau_p^k}\right) + \mathbf{D}_p(\mathbf{w}_i^k - \mathbf{w}_i^{\tau_p^k}) \\
&+ \sqrt{\frac{\beta}{N}}\frac{1}{M}\sum_{p=1}^{M}\mathbf{A}_p\left(\bar{\boldsymbol{\theta}}^{\tau_p^k} - \boldsymbol{\theta}_i^{\tau_p^k}\right).
\end{aligned} \tag{41}$$

We construct the residual vector $\underline{\boldsymbol{\varepsilon}}_c(k)$ as the following: the first block of $\underline{\boldsymbol{\varepsilon}}_c(k)$ is $-\frac{1}{N}(\underline{\mathbf{s}}^k - \bar{\mathbf{s}}^k\mathbf{v}^\top)(\mathbf{u} - \mathbf{1}) +$

$\frac{\rho r'}{NM}\sum_{i=1}^{N}\sum_{p=1}^{M}\left(\boldsymbol{\theta}_i^{\tau_p^k}-\bar{\boldsymbol{\theta}}^{\tau_p^k}\right)$ and the remaining blocks are given by $\sqrt{\frac{\beta}{N}}\frac{1}{M}\sum_{p=1}^{M}\mathbf{A}_p\left(\boldsymbol{\theta}_i^{\tau_p^k}-\bar{\boldsymbol{\theta}}^{\tau_p^k}\right), i\in 1,\cdots,N$. By (40), (41), and the definition of $\mathbf{G}_p$ in (28), we have the following simple form of $\underline{\mathbf{h}}(k)-\underline{\mathbf{g}}(k)$:

$$\underline{\mathbf{h}}(k)-\underline{\mathbf{g}}(k)-\underline{\boldsymbol{\varepsilon}}_c(k)=\frac{1}{M}\sum_{p=1}^{M}\mathbf{G}_p\Big(\sum_{j=\tau_p^k}^{k-1}\Delta\mathbf{v}(j)\Big), \quad (42)$$

where we define

$$\Delta\underline{\boldsymbol{v}}^j=\begin{bmatrix}\bar{\boldsymbol{\theta}}^{j+1}-\bar{\boldsymbol{\theta}}^{j}\\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_1^{j+1}-\mathbf{w}_1^{j})\\ \vdots\\ \frac{r'}{\sqrt{\beta N}}(\mathbf{w}_N^{j+1}-\mathbf{w}_N^{j})\end{bmatrix}. \quad (43)$$

Note it holds that $\Delta\underline{\boldsymbol{v}}^j=\underline{\boldsymbol{v}}^{j+1}-\underline{\boldsymbol{v}}^j$. From (39), $\Delta\underline{\boldsymbol{v}}^j$ in (43) can also be written as

$$\Delta\underline{\boldsymbol{v}}^j=\gamma[\underline{\mathbf{h}}(j)-\underline{\mathbf{g}}(j)]-\gamma\underline{\mathbf{h}}(j). \quad (44)$$

Multiplying $U^{-1}$ on both sides of (39) results in

$$\widehat{\underline{\boldsymbol{v}}}^{k+1}=(\mathbf{I}-\gamma\mathbf{G})\widehat{\underline{\boldsymbol{v}}}^{k}+\gamma(\underline{\mathbf{h}}(k)-\underline{\mathbf{g}}(k)), \quad (45)$$

where $\widehat{\underline{\boldsymbol{v}}}^{k}:=U^{-1}\underline{\boldsymbol{v}}^k$. By Combining (42), (44), and (45), we have

$$\begin{aligned}\|\widehat{\underline{\boldsymbol{v}}}^{k+1}\|&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+\gamma\|U^{-1}\|\|\underline{\mathbf{h}}(k)-\underline{\mathbf{g}}(k)\|\\&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+\gamma\|U^{-1}\|\Big[\|\underline{\boldsymbol{\varepsilon}}_c(k)\|\\&+\frac{1}{M}\sum_{p=1}^{M}\|\mathbf{G}_p\sum_{j=\tau_p^k}^{k-1}\Delta\underline{\boldsymbol{v}}^j\|\Big]\\&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+\gamma\|U^{-1}\|\Big\{\|\underline{\boldsymbol{\varepsilon}}_c(k)\|\\&+\frac{\gamma\bar{G}}{M}\sum_{p=1}^{M}\sum_{j=\tau_p^k}^{k-1}\big[\|\underline{\mathbf{h}}(j)\|+\|\underline{\mathbf{h}}(j)-\underline{\mathbf{g}}(j)\|\big]\Big\},\end{aligned} \quad (46)$$

where $\bar{G}$ is defined in in (37). By further bounding the right-hand side of (46) we have

$$\begin{aligned}\|\widehat{\underline{\boldsymbol{v}}}^{k+1}\|&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+\gamma\|U^{-1}\|\Big\{\|\underline{\boldsymbol{\varepsilon}}_c(k)\|\\&+\gamma\bar{G}\sum_{j=(k-M)_+}^{k-1}\big[\|\underline{\mathbf{h}}(j)\|+\|\underline{\mathbf{h}}(j)-\underline{\mathbf{g}}(j)\|\big]\Big\}\\&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+\gamma\|U^{-1}\|\Big\{\|\underline{\boldsymbol{\varepsilon}}_c(k)\|\\&+\gamma\bar{G}\sum_{j=(k-M)_+}^{k-1}\Big[\|\underline{\boldsymbol{\varepsilon}}_c(j)\|+G\|U\|\|\widehat{\underline{\boldsymbol{v}}}^{j}\|\\&+\bar{G}\|U\|\cdot\sum_{j=(k-M)_+}^{k-1}(\|\widehat{\underline{\boldsymbol{v}}}^{l+1}\|+\|\widehat{\underline{\boldsymbol{v}}}^{l}\|)\Big]\Big\}.\end{aligned} \quad (47)$$

Moreover, $\|\underline{\boldsymbol{\varepsilon}}_c(k)\|$ can be upper bounded by

$$\begin{aligned}&\|\underline{\boldsymbol{\varepsilon}}_c(k)\|\\&\le\frac{1}{M}\sum_{p=1}^{M}\Big[(\rho r'+\bar{A}\sqrt{\beta N})\big(\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\theta}_i^{\tau_p^k}-\bar{\boldsymbol{\theta}}^{\tau_p^k}\|\big)\Big]\\&+\frac{1}{N}\|\mathbf{u}-\mathbf{1}\|\|\underline{\mathbf{s}}^k-\bar{\mathbf{s}}^k\mathbf{v}^\top\|\\&\le(\rho r'+\bar{A}\sqrt{\beta N})\max_{(k-M)_+\le q\le k}\boldsymbol{\varepsilon}_c(q)\\&+\frac{1}{N}\|\mathbf{u}-\mathbf{1}\|\|\underline{\mathbf{s}}^k-\bar{\mathbf{s}}^k\mathbf{v}^\top\|.\end{aligned}$$

Thus, we can bound $\|\widehat{\underline{\boldsymbol{v}}}^{k+1}\|$ by

$$\begin{aligned}&\|\widehat{\underline{\boldsymbol{v}}}^{k+1}\|\\&\le\|\mathbf{I}-\gamma\Lambda\|\|\widehat{\underline{\boldsymbol{v}}}^{k}\|+C_1(\gamma)\max_{(k-2M)_+\le q\le k-1}\|\widehat{\underline{\boldsymbol{v}}}^{q}\|\\&+C_2(\gamma)\max_{(k-2M)_+\le q\le k}\boldsymbol{\varepsilon}_c(q)+C_3(\gamma)\|\underline{\mathbf{s}}^k-\bar{\mathbf{s}}^k\mathbf{v}^\top\|,\end{aligned} \quad (48)$$

where constant $C_1(\gamma)$, $C_2(\gamma)$ , and $C_3(\gamma)$ are defined as

$$\begin{aligned}C_1(\gamma)&=\gamma^2\|U\|\|U^{-1}\|\bar{G}M(G+2\bar{G}M),\\C_2(\gamma)&=\gamma\|U^{-1}\|(1+\gamma\bar{G}M)(\rho r'+\bar{A}\sqrt{\beta N}),\\C_3(\gamma)&=\gamma\|U^{-1}\|\frac{\|\mathbf{u}-\mathbf{1}\|}{N}(1+\gamma\bar{G}M).\end{aligned} \quad (49)$$

Now combining (23), (25), and (48) yields

$$\begin{bmatrix}\|\boldsymbol{\theta}^{k+1}-\bar{\boldsymbol{\theta}}^{k+1}\mathbf{1}^\top\|\\\|\underline{\mathbf{s}}^{k+1}-\bar{\mathbf{s}}^{k+1}\mathbf{v}^\top\|\\\|\widehat{\underline{\boldsymbol{v}}}^{k+1}\|\end{bmatrix}\le Q(\gamma)\begin{bmatrix}\max_{(k-2M)_+\le s\le k}\|\underline{\boldsymbol{\theta}}^s-\bar{\boldsymbol{\theta}}^s\mathbf{1}^\top\|\\\max_{(k-M)_+\le s\le k}\|\underline{\mathbf{s}}^s-\bar{\mathbf{s}}^s\mathbf{v}^\top\|\\\max_{(k-2M)_+\le s\le k}\|\widehat{\underline{\boldsymbol{v}}}^{s}\|\end{bmatrix}, \quad (50)$$

where matrix $Q(\gamma)$ is defined by

$$Q(\gamma):=\begin{bmatrix}\sigma_{R_1}+\gamma\sigma_{R_2}\|\mathbf{v}-\mathbf{1}\|\frac{\rho}{\sqrt{N}} & \gamma\sigma_{R_2} & C_6(\gamma)\\ C_7(\gamma) & C_4(\gamma) & C_5(\gamma)\\ \frac{1}{\sqrt{N}}C_2(\gamma) & C_3(\gamma) & C_8(\gamma)\end{bmatrix},$$

and

$$\begin{aligned}&C_4(\gamma):=\sigma_{C_1}+\sigma_{C_2}\rho\gamma\|R_2\|_S\\&C_5(\gamma):=\sqrt{2}\big(\sigma_{C_2}L^2\gamma_2\sqrt{N}+\sigma_{C_2}L^2\gamma\|R_2\mathbf{v}\|\big)\max\{1,\frac{\sqrt{\beta}}{r'}\}\\&C_6(\gamma):=\sqrt{2}\gamma\sigma_{R_2}\|\mathbf{v}-\mathbf{1}\|L\max(1,\frac{\sqrt{\beta}}{r'})\\&C_7(\gamma):=\sigma_{C_2}\rho\|R_1-\mathbf{I}\|_S+\gamma(\sigma_{C_2}\rho^2\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}}+\sigma_{C_2}\bar{A}^2\beta)\\&C_8(\gamma):=\|I-\gamma\Lambda\|+C_1(\gamma),\end{aligned} \quad (51)$$

with $L := \max\{\rho, \bar{A}, \bar{D}\}$. Note that $\mathbf{G}$'s eigenvalues are bounded in (35). Thus by setting the step size $\gamma$ to be small enough we can ensure $\|\mathbf{1} - \gamma\Lambda\| < 1$. Hence there exists some $\alpha > 0$ such that $\|\mathbf{1} - \gamma\Lambda\| = 1 - \gamma\alpha$. Moreover, the upper bound of $\|U\|$ and $\|U^{-1}\|$ are given in (36).

Now define

$$
\begin{aligned}
a_1 &:= \sigma_{R_2}\| \mathbf{v} - \mathbf{1}\| \frac{\rho}{\sqrt{N}} \\
a_2 &:= \sqrt{2}\sigma_{R_2}\| \mathbf{v} - \mathbf{1}\| L \max(1, \sqrt{\beta}/r') \\
a_3 &:= \sigma_{C_2}\rho\|R_1 - \mathbf{I}\|_S \\
a_4 &:= \sigma_{C_2}\rho^2\|R_2\mathbf{v}\|\frac{1}{\sqrt{N}} + \sigma_{C_2}\bar{A}^2\beta \\
a_5 &:= \sigma_{C_2}\rho\|R_2\|_S \\
a_6 &:= \sqrt{2}\big(\sigma_{C_2}L^2\beta\sqrt{N} + \sigma_{C_2}L^2\|R_2\mathbf{v}\|\big)\max\{1, \sqrt{\beta}/r'\} \\
a_7 &:= \frac{1}{\sqrt{N}}\|U^{-1}\|\big(\rho r' + \bar{A}\sqrt{\beta N}\big) \\
a_8 &:= \frac{1}{\sqrt{N}}\|U^{-1}\|2\bar{G}M\big(\rho r' + \bar{A}\sqrt{\beta N}\big) \\
a_9 &:= \|U\|\|U^{-1}\|\bar{G}M(G + 2\bar{G}M) \\
a_{10} &:= \frac{\|\mathbf{u} - \mathbf{1}\|}{N}\|U^{-1}\|(1 + \bar{G}).
\end{aligned}
$$

Observe that when $\gamma < \frac{1}{M}$ we have

$$
Q \leq Q_1 := \begin{bmatrix} \sigma_{R_1} + a_1\gamma & \sigma_{R_2}\gamma & a_2\gamma \\ a_3 + a_4\gamma & \sigma_{C_1} + a_5\gamma & a_6\gamma \\ a_7\gamma + a_8\gamma^2 & a_{10}\gamma & (1 - \gamma\alpha + a_9\gamma^2) \end{bmatrix},
$$

when the inequality means $Q$ is entrywise less than $Q_1$.

Now consider

$$
\mathbf{g}(\sigma) := |\sigma\mathbf{I} - Q_1| =
\begin{vmatrix} \sigma - \sigma_{R_1} - a_1\gamma & -\sigma_{R_2}\gamma & -a_2\gamma \\ -a_3 - a_4\gamma & \sigma - \sigma_{C_1} - a_5\gamma & -a_6\gamma \\ -a_7\gamma - a_8\gamma^2 & -a_{10}\gamma & \sigma - (1 - \gamma\alpha + a_9\gamma^2) \end{vmatrix}.
$$

Therefore

$$
\begin{aligned}
\mathbf{g}(\sigma) = {} & \big(\sigma - (1 - \gamma\alpha + \gamma^2 a_9)\big)\mathbf{g}_0(\sigma) \\
& - (a_7\gamma + a_8\gamma^2)\big(\sigma_{R_2}a_6\gamma^2 + a_2\gamma(\sigma - \sigma_{C_1} - a_5\gamma)\big) \\
& - a_{10}\gamma\big(a_2\gamma(a_3 + a_4\gamma) + a_6\gamma(\sigma - \sigma_{R_1} - a_1\gamma)\big).
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathbf{g}_0(\sigma) \\
&:= (\sigma - \sigma_{R_1} - a_1\gamma)(\sigma - \sigma_{C_1} - a_5\gamma) - \sigma_{R_2}\gamma(a_3 + a_4\gamma).
\end{aligned}
$$

Without loss of generality, assume $\sigma_{R_1} \leq \sigma_{C_1}$. We define $\bar{\sigma} := \sigma_{C_1} + (a_1 + a_5)\gamma + \sqrt{\sigma_{R_2}\gamma(a_3 + a_4\gamma)}$. It is easy to checked that for all $\sigma \geq \bar{\sigma}$, we have

$$
\mathbf{g}_0(\sigma) \geq (\sigma - \bar{\sigma})^2.
$$

Now we define

$$
\begin{aligned}
\sigma^* := \max\Bigg\{ & \frac{\gamma\alpha}{4} + (1 - \gamma\alpha + \gamma^2 a_9), \bar{\sigma} + \frac{2a_2\gamma(a_7 + a_8)}{\alpha} \\
& + \frac{2a_6a_{10}\gamma}{\alpha} + \sqrt{\gamma}\Bigg\{ \frac{\frac{a_2^2\gamma^2(a_7 + a_8\gamma)^2}{\alpha} + (a_7\gamma + a_8\gamma^2)\sigma_{R_2}a_6\gamma}{\frac{\gamma\alpha}{4}} \\
& + \frac{a_2a_{10}\gamma(a_3 + a_4\gamma)}{\frac{\gamma\alpha}{4}} \\
& + \frac{a_2\gamma(a_7 + a_8\gamma)\big(a_1\gamma + \sqrt{\sigma_{R_2}\gamma(a_3 + a_4\gamma)}\big)}{\frac{\gamma\alpha}{4}} \\
& + \frac{a_6a_{10}\gamma\big(\sigma_{C_1} - \sigma_{R_1} + a_5\gamma + \sqrt{\sigma_{R_2}\gamma(a_3 + a_4\gamma)}\big)}{\frac{\gamma\alpha}{4}} \Bigg\}^{\frac{1}{2}} \Bigg\}.
\end{aligned}
\tag{52}
$$

For all $\sigma \geq \sigma^*$, it holds that

$$
\begin{aligned}
\mathbf{g}(\sigma) \geq {} & \big(\sigma - (1 - \gamma\alpha + \gamma^2 a_9)\big)(\sigma - \bar{\sigma})^2 \\
& - (a_7\gamma + a_8\gamma^2)\big(\sigma_{R_2}a_6\gamma^2 + a_2\gamma(\sigma - \sigma_{C_1} - a_5\gamma)\big) \\
& - a_{10}\gamma\big(a_2\gamma(a_3 + a_4\gamma) + a_6\gamma(\sigma - \sigma_{R_1} - a_1\gamma)\big) \\
\geq {} & \frac{\gamma\alpha}{4}\big(\sigma - \bar{\sigma} - \frac{2a_2\gamma(a_7 + a_8\gamma)}{\alpha} - \frac{2a_6a_{10}\gamma}{\alpha}\big)^2 \\
& - \frac{a_2^2\gamma^3(a_7 + a_8\gamma)^2}{\alpha} - a_2\gamma^2(a_7 + a_8\gamma)(\bar{\sigma} - \sigma_C - a_5\gamma) \\
& - (a_7\gamma + a_8\gamma^2)\sigma_{R_2}a_6\gamma^2 - a_2a_{10}\gamma^2(a_3 + a_4\gamma) \\
& - a_6a_{10}\gamma^2\big(\bar{\sigma} - \sigma_{R_1} - a_1\gamma\big) \\
\geq {} & 0.
\end{aligned}
$$

Now from the Perron Frobenius theorem one can conclude that $\rho(\mathbf{Q}) < \sigma^*$. Moreover, by Assumptions 3-4, one has $\max\{\sigma_{R_1}, \sigma_{C_1}\} < 1$ (Pu et al., 2018). As $\alpha > 0$, there exists a sufficiently small $\gamma$ such that $\sigma^* < 1$, which implies $\rho(\mathbf{Q}) < 1$.

Next, let us consider the asymptotic rate when $M, N \gg 1$. Note that the proposed algorithm converges if $\sigma^* < 1$. Let us consider the first operand in the $\max\{\cdot\}$ of (52). The first operand will be less than 1 if $0 < \gamma < \frac{\alpha}{2a_9}$ since

$$
\frac{\gamma\alpha}{4} + 1 - \gamma\alpha - \gamma^2 a_9 \leq 1 - \frac{\gamma\alpha}{4} < 1. \tag{53}
$$

By the definition of $a_9$, this requires $\gamma = \mathcal{O}(1/M^2)$ if $M \gg 1$.

Note that we have $\sigma_{C_1} = 1 - \frac{e}{N}$ for some positive $e$. Substituting this into the second operand in (52) yields

$$
\begin{aligned}
&1-\frac{e}{N}+(2a_1+a_5)\gamma+\sqrt{\sigma_{R_2}\gamma(a_3+a_4\gamma)}\\
&+\frac{2a_2\gamma(a_7+a_8\gamma)}{\alpha}+\frac{2a_6a_{10}\gamma}{\alpha}\\
&+\sqrt{\gamma}\Big\{\frac{4}{\alpha}\Big[a_2^2\gamma(a_2+a_8\gamma)^2+(a_7\gamma+a_8\gamma^2)\sigma_{R_2}a_6\\
&+a_2(a_7+a_8\gamma)\big(a_1\gamma+\sqrt{\sigma_{R_2}\gamma(a_3+a_4\gamma)}\big)\\
&+a_2a_{10}(a_3+a_4\gamma)+a_6a_{10}\big(\bar{\sigma}-\sigma_{R_1}-a_1\gamma\big)\Big]\Big\}^{\frac{1}{2}}.
\end{aligned} \tag{54}
$$

Therefore, to guarantee $\sigma^* < 1$, it's sufficient to let $\gamma$ satisfy (53) and (54).

Now let us consider the asymptotic rate when $N, M \gg 1$. Observe that $a_1 = \Theta(1)$, $a_2 = \Theta(\sqrt{N})$, $a_3 = \mathcal{O}(1)$, $a_4 = \Theta(1)$, $a_5 = \Theta(1)$, $a_6 = \Theta(\sqrt{N})$, $a_7 = \Theta(1)$, $a_8 = \Theta(M)$, and $a_{10} = o(1)$. Therefore Eq. (54) can be approximated by

$$
\begin{aligned}
&1-\frac{e}{N}+\Theta(\gamma)+\Theta(\sqrt{\gamma})+\gamma\Theta(\sqrt{N})+\Theta(\gamma\frac{1}{\sqrt{N}})\\
&+\sqrt{\gamma}\Theta\big(\{N^2\gamma+\sqrt{N\gamma}\}^{\frac{1}{2}}\big).
\end{aligned} \tag{55}
$$

To ensure (55) is less than $1-\frac{e}{2N}$, it requires that $\gamma = \mathcal{O}\big(\frac{1}{N^2}\big)$. In summary, by setting $\gamma = \mathcal{O}\big(1/\max\{N^2, M^2\}\big)$ we have $\sigma^* \leq \max\{1-\gamma\frac{\alpha}{4}, 1-e/(2N)\} = 1-\mathcal{O}(1/\max\{N^2, M^2\})$.