Distributed Stochastic Approximation with Constant Communication

Feng Zhu
North Carolina State University
fzhu5@ncsu.edu

Aritra Mitra
North Carolina State University
amitra2@ncsu.edu

Robert W. Heath Jr.
University of California, San Diego
rwheathjr@ucsd.edu

***Abstract*—We study a general distributed stochastic approximation problem involving M agents, each with a distinct, potentially non-linear local operator. The goal is for the agents to collaboratively find the root of the average of their local operators via intermittent communication with a central server. This formulation captures a broad class of problems due to the nonlinearity and heterogeneity of local operators. Existing methods either fail to achieve linear speedup, suffer from heterogeneity-induced biases, or require high communication overhead. We propose `DisSACC`, an algorithm that overcomes all three limitations: it converges to the desired solution with a M-fold linear speedup in sample-complexity, eliminates heterogeneity bias, and requires only logarithmic (near-constant) communication.**

I. Introduction

We consider a distributed stochastic approximation (SA) framework involving M agents that communicate via a central server, as is standard in federated learning (FL) settings [1]. Each agent $i \in [M]$ is associated with a potentially *nonlinear and distinct* operator $\bar{G}_i : \mathbb{R}^d \to \mathbb{R}^d$, which we refer to as the *true operator*. However, agent i does not have direct access to $\bar{G}_i$; instead, it observes a *noisy version* $G_i : \mathbb{R}^d \times \mathcal{X}_i \to \mathbb{R}^d$ of $\bar{G}_i$, where $\mathcal{X}_i$ denotes the sample space of agent i. The noisy operator satisfies the following unbiasedness condition:

$$\bar{G}_i(\theta) = \mathbb{E}_{o\sim\mu_i}\left[G_i(\theta, o)\right], \forall \theta \in \mathbb{R}^d, \tag{1}$$

where $o \in \mathcal{X}_i$ is a random variable drawn from an *unknown* distribution μ_i specific to agent i, and $\theta \in \mathbb{R}^d$ is a parameter. Equivalently, $G_i(\cdot, o)$ is an unbiased estimate of $\bar{G}_i(\cdot)$.

Locally, each agent i observes a sequence of noisy samples $\{G_i(\cdot, o_{i,t})\}_{t\geq 0}$. For simplicity of presentation, we assume that the sequence $\{o_{i,t}\}_{t\geq 0}$ is generated I.I.D. (identically and independently distributed) from the distribution μ_i. This assumption can be relaxed to account for temporal correlations between samples (referred to as *Markovian sampling*), as will be discussed in Section IV. The overall goal, as summarized below, is to find the root $\theta^\star \in \mathbb{R}^d$ of the *average operator* $\bar{G}$:

$$\boxed{\text{Find } \theta^\star \text{ s.t. } \bar{G}(\theta^\star) = 0, \quad \text{where } \bar{G} := \frac{1}{M}\sum_{i\in[M]} \bar{G}_i.} \tag{2}$$

When $M = 1$, Problem (2) specializes to the classical single-agent SA formulation [2]. The standard update rule proposed in [2] takes the following form:

$$\theta^{(t+1)} = \theta^{(t)} + \alpha_t G(\theta^{(t)}, o_t), \quad t = 0, \cdots, T-1, \tag{3}$$

where T denotes the total number of iterations, $\theta^{(t)} \in \mathbb{R}^d$ denotes the parameter estimate at time-step t, and $\{\alpha_t\}_{t\geq 0}$ is the sequence of step-sizes. Finite-time rates for the update rule (3) have been established in [3]–[6]. Specifically, it has been shown that under suitable regularity assumptions, running T iterations of (3) with a constant step-size $\alpha_t = \alpha$ yields the following mean-squared error (MSE) bound:

$$\mathbb{E}\left[\left\|\theta^{(T)} - \theta^\star\right\|_2^2\right] \leq \underbrace{C_1 \exp\left(-\alpha C_2 T\right)}_{\text{bias}} + \underbrace{\alpha C_3 \sigma^2}_{\text{variance}}, \tag{4}$$

where C_1, C_2, C_3 are problem-specific constants, and σ^2 captures the variance of the noise model. The MSE bound in (4) consists of (i) an exponentially decaying *bias* term and (ii) a *variance* term capturing the effect of noise. In simple words, (4) ensures convergence at a linear rate to a ball of radius $\mathcal{O}\left(\alpha\sigma^2\right)$ centered around $\theta^\star$. By selecting α to be on the order of $\mathcal{O}\left(\log T/T\right)$, exact convergence to $\theta^\star$ can be obtained at a sub-linear rate of $\tilde{\mathcal{O}}\left(1/T\right)$.

Given this premise, we return to our distributed SA setting with M heterogeneous local operators and ask: *Is it possible to simultaneously achieve (i) exact convergence to $\theta^\star$, (ii) an M-fold linear speedup—manifested through a variance term of order $\mathcal{O}\left(\alpha\sigma^2/M\right)$, and (iii) both of these benefits with minimal communication overhead?*

We are now in a position to summarize the **contributions** of this work. We propose a novel algorithm `DisSACC` which provably achieves exact convergence to $\theta^\star$, along with an M-fold reduction in the variance term. Most notably, these guarantees are obtained with only a *near-constant* communication overhead of order $\tilde{\mathcal{O}}\left(1\right)$, which is a significant improvement over prior work, as will be discussed subsequently.

Before delving into the prior art, we briefly motivate the study of the distributed SA problem in (2). The general formulation in (2) captures a wide spectrum of applications, including optimization problems in federated learning, fixed-point formulations that arise in nonlinear inverse problems and variational inequalities, and federated counterparts of classical reinforcement learning (FRL) algorithms such as temporal difference (TD) learning and Q-learning with function approximation. The above problem classes collectively span practical applications across diverse domains, including robotics, multi-agent systems, gaming, and autonomous driving. We now discuss the literature of distributed SA.

Related Works. Distributed SA studies are typically classified into *homogeneous* (identical $\bar{G}_i$'s) and *heterogeneous* (distinct $\bar{G}_i$'s) settings, which we discuss separately as follows.

Homogeneous SA. While [7]–[9] investigate FRL algorithms, only [8] and [9] establish a linear speedup under I.I.D. sampling. The first result showing a linear speedup under Markovian sampling for contractive SA is [10], followed by [11] in the federated tabular Q-learning setting. Subsequent works examine the communication cost required for achieving linear speedups [12], [13], and the impact of imperfect channels [14].

Heterogeneous SA. Since agents typically interact with *different* environments, the heterogeneous SA setting—where local operators vary across agents—is more realistic, but remains underexplored relative to the homogeneous case. Convergence guarantees for federated Q-learning were given in [15], though without linear speedups. Recent works [16], [17] establish linear speedups for federated SARSA and TD, but their bounds feature a heterogeneity-induced bias term that scales with discrepancies across agents' environments, thus negating collaborative benefits. This bias also appears in [15].

Our prior work [18] eliminates this bias using a correction technique and proves linear speedup for general (nonlinear) heterogeneous contractive SA under Markov sampling. The work [19] also achieves bias-free speedups, but only in a restrictive linear SA setting, where their analysis relies heavily on the linearity of the operator. However, [18] requires $\tilde{\mathcal{O}}\left(\sqrt{MR}\right)$ communication, where R is the number of samples per agent. In contrast, the present work shows that the same guarantees can be achieved with only $\mathcal{O}\left(\log MR\right)$ communication.

II. Problem Formulation

In this section, we formally set up the problem of interest. We consider the root-finding problem in (2), where the objective is to find the root of the average operator $\bar{G}$ in a federated setting. In this framework, agents communicate with a central server under *stringent communication constraints*.

Specifically, agents are only allowed to communicate with the server *intermittently*. Suppose each agent $i \in [M]$ has access to R samples in total; these samples are partitioned into T communication rounds, with H samples per round, so that $R = TH$. Furthermore, due to privacy considerations, the raw observation sequence $\{o_{i,t}\}_{t\geq 0}$ must remain local and cannot be shared. Both the communication and privacy constraints described above are standard practices in the FL literature [1].

Next, we make some standard assumptions on the agents' operators and the observation processes.

Assumption 1 (Lipschitzness)**.** *The local true operator $\bar{G}_i$ for each agent $i \in [M]$ is L-Lipschitz, i.e., there exists a constant $L \geq 1$ such that for all $\theta_1, \theta_2 \in \mathbb{R}^d$, we have*

$$\left\|\bar{G}_i(\theta_1) - \bar{G}_i(\theta_2)\right\|_2 \leq L \left\|\theta_1 - \theta_2\right\|_2 . \tag{5}$$

Furthermore, for each $i \in [M]$, there exists $\sigma_i \geq 1$ such that for any given $\theta \in \mathbb{R}^d, o \in \mathcal{X}_i$, the following holds:

$$\max\{\|\bar{G}_i(\theta)\|_2, \|G_i(\theta, o)\|_2\} \leq L(\|\theta\|_2 + \sigma_i). \tag{6}$$

Assumption 2 (1-point strong monotonicity)**.** *The true average operator $\bar{G}$ is 1-point strongly monotone w.r.t. the fixed point $\theta^\star$, i.e., there exists some constant $\mu \in (0, 1]$ such that for any $\theta \in \mathbb{R}^d$, we have*

$$\left\langle \theta - \theta^\star,\, \bar{G}(\theta) \right\rangle \leq -\mu \left\|\theta - \theta^\star\right\|_2^2 . \tag{7}$$

In the optimization literature, Assumption 1 corresponds to a standard smoothness condition [20], and Assumption 2 corresponds to strong convexity of the objective/loss function. In the context of reinforcement learning (RL), both Assumptions 1 and 2 are **known to hold** for TD-learning with linear function approximation (LFA) [3], [4], [21], and certain variants of Q-learning with LFA [5], [22], where $\bar{G}_i$ and G_i correspond to the non-noisy and noisy versions of the TD/Q-learning update rules, respectively. Our final assumption is as follows.

Assumption 3. *For every pair of agents $i \neq j$, the observation processes $\{o_{i,t}\}$ and $\{o_{j,t}\}$ are statistically independent.*

Assumption 3 is needed to establish linear speedups w.r.t. the number of agents [11], [16], [17]. With the above assumptions in place, we are now in a position to introduce and analyze our proposed `DisSACC` algorithm.

III. Algorithm

In this section, we present the details of our proposed algorithm, `Distributed Stochastic Approximation with Constant Communication (DisSACC)`, which is specifically designed for multi-agent systems operating under *intermittent communication* constraints. We now elaborate on the algorithmic details of `DisSACC`, abstracted out in Algorithm 1. As most algorithms in FL and FRL, `DisSACC` also respects the standard intermittent communication protocol. To this end, suppose that each agent $i \in [M]$ has R samples in total. Agents then equally divide their R samples into T communication rounds with H samples each, i.e., $R = TH$. Here, T and H are design parameters that will be specified later. At the beginning of each communication round $t = 0, 1, \ldots, T-1$, the central server broadcasts the current global iterate $\bar{\theta}^{(t)}$ to all agents. Each agent then performs local computation for H time-steps initialized from this global model. At local step ℓ of round t, agent i observes a sample denoted $o_{i,\ell}^{(t)}$, which is equivalently indexed as $o_{i,tH+\ell}$.

Motivation. Before introducing the core idea of `DisSACC`, let us reiterate the motivation for developing a new algorithm. A common thread in recent FRL works [10], [11], [15]–[17] is the adoption of the following local update rule:

$$\theta_{i,\ell+1}^{(t)} = \theta_{i,\ell}^{(t)} + \alpha G_i(\theta_{i,\ell}^{(t)}, o_{i,\ell}^{(t)}), \tag{8}$$

where $\theta_{i,\ell}^{(t)}$ denotes the local model of agent i at local iteration ℓ of communication round t, initialized from the global model as $\theta_{i,0}^{(t)} = \bar{\theta}^{(t)}$, and α is the step-size. As established in our previous work [18], when each agent i follows (8) for H local time-steps **without synchronization**, the local model $\theta_{i,\ell}^{(t)}$ tends to drift towards the *agent-specific* root $\theta_i^\star$ of its own local operator $\bar{G}_i$. Consequently, naïvely aggregating the local

models at the server leads to a *bias term* in the final bound—which prevents achieving the desired MSE bound in (4) with an M-fold variance reduction. While [18] addresses this issue by introducing a correction technique applied at each local update, enabling convergence to the global root $\theta^\star$ with linear speedup and no heterogeneity-induced bias, the required communication cost is $\tilde{\mathcal{O}}\left(\sqrt{MR}\right)$—which may be prohibitive in practice.

This naturally raises a key question: *Can we achieve exact convergence to $\theta^\star$ with linear speedup, no heterogeneity bias, and minimal communication overhead?*

Core Idea. To address the question posed above, we now develop a variance-reduction strategy at the heart of the DisSACC algorithm. Concretely, the failure of prior algorithms can be attributed to their adoption of H consecutive local updates using highly noisy operators. Zooming in again at (8), we can observe that each local update is driven by the noisy operator G_i, constructed with **a single observation sample**. Due to the high variance of the noise in that operator, the cumulative drift over H steps becomes challenging to control, necessitating frequent communication.

From this observation, we propose the DisSACC algorithm, the core idea of which is to **refine the operator with more samples and update the estimate fewer times**, instead of frequently updating with a noisy operator constructed from a single sample. To see how, during each communication round t, agent i **collects H observations**, and constructs a *refined operator* $\hat{G}_{i,t}$ as follows:

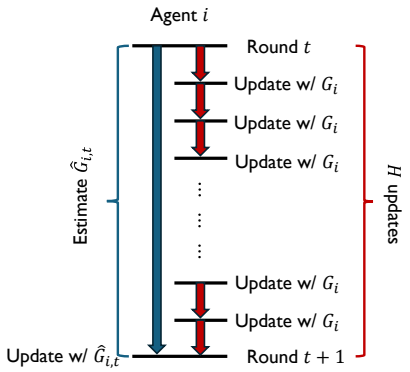

Fig. 1: Illustration of DisSACC in blue (with one local step), and standard FL/FRL schemes in red (with multiple local SA steps).

$$\hat{G}_{i,t}(\cdot) = \frac{1}{H}\sum_{\ell=0}^{H-1} G_i(\cdot, o_{i,\ell}^{(t)}). \tag{9}$$

Since $\{o_{i,\ell}^{(t)}\}_{\ell=0}^{H-1}$ are I.I.D., the variance of the refined operator $\hat{G}_{i,t}$ is reduced by a factor of H compared to G_i (proved in Fact 2). After obtaining $\hat{G}_{i,t}$, agent i updates the global model $\bar{\theta}^{(t)}$ **only once** at the end of round t:

$$\theta_i^{(t)} = \bar{\theta}^{(t)} + \alpha \hat{G}_{i,t}(\bar{\theta}^{(t)}), \tag{10}$$

where $\theta_i^{(t)}$ represents the local model of agent i at the end of round t, and $\bar{\theta}^{(t)}$ is the global model broadcasted by the server at the beginning of round t. Finally, the server simply aggregates the local models across agents, and updates the global model as:

$$\bar{\theta}^{(t+1)} = \frac{1}{M}\sum_{i\in[M]} \theta_i^{(t)}. \tag{11}$$

Algorithm 1 DisSACC

Input: Step-size α, initial parameter $\bar{\theta}^{(0)} = 0$.
for $t = 0, \ldots, T-1$ **do**
  for $i = 1, \ldots, M$ **do**
    for $\ell = 0, \ldots, H-1$ **do**
      Agent i observes and collects sample $o_{i,\ell}^{(t)}$.
    end for
  Agent i constructs refined operator $\hat{G}_{i,t}$ as per (9).
  Agent i obtains local model $\theta_i^{(t)}$ via (10).
  end for
  Server broadcasts $\bar{\theta}^{(t+1)}$ computed as in (11).
end for

Remark. *An illustration comparing DisSACC with the standard FL/FRL update rule in (8) is provided in Fig. 1. Crucially, instead of performing multiple local model updates as in (8), DisSACC aggregates all H samples within each round to construct a refined operator, which is then used for only one model update. This ensures that updates across all agents remain **synchronized at every communication round**, thereby eliminating any heterogeneity-induced bias. As will be shown in the next section, this design allows the algorithm to operate with only $\tilde{\mathcal{O}}(1)$ rounds of communication.*

IV. Main Result and Discussion

The main convergence results and some key takeaways are discussed in this section. Defining $s_t := \bar{\theta}^{(t)} - \theta^\star$ and $\sigma := \max\{\{\sigma_i\}_{i\in[M]}, \|\theta^\star\|_2, 1\}$, we have the following results.

Theorem 1 (Main Result). *Suppose Assumptions 1 to 3 hold. Then, there exists a universal constant $C \geq 1$, such that with $\alpha \leq \mu/(CL^2)$, DisSACC guarantees $\forall T \geq 0$:*

$$\mathbb{E}\left[\|s_T\|_2^2\right] \leq (1-\alpha\mu)^T \|s_0\|_2^2 + \mathcal{O}\left(\frac{\alpha L^2\sigma^2}{\mu HM}\right). \tag{12}$$

The next result is an immediate corollary of Theorem 1.

Corollary 1 (Linear Speedup). *Suppose all conditions in Theorem 1 hold. Then, by choosing $\alpha = \mu/(CL^2)$ and $T = \log MR/(\mu\alpha)$, DisSACC guarantees for any $T \geq 0$:*

$$\mathbb{E}\left[\|s_T\|_2^2\right] \leq \mathcal{O}\left(\frac{\sigma^2 L^2 \log MR}{\mu^2 MR}\right) \leq \tilde{\mathcal{O}}\left(\frac{\sigma^2 L^2}{\mu^2}\cdot\frac{1}{MR}\right). \tag{13}$$

The convergence proof of Theorem 1 is provided in the next section. Before that, a few key takeaways are in order.

• *Matching Centralized Rates.* Theorem 1 reveals that our DisSACC algorithm achieves convergence to a ball around $\theta^\star$ at an exponential rate. This matches the centralized rate comparing (12) to (4), thus recovering the known finite-time bounds for single-agent SA in prior work [3]–[5].

• *Linear Speedup Effect.* The bound in (12) consists of an exponentially decaying bias term and a variance term as in the centralized case in (4). Notably, the variance term $\mathcal{O}\left(\alpha L^2\sigma^2/(\mu MH)\right)$ gets scaled down by the number of agents M, corroborating the linear speedup effect. Furthermore, as

shown in Corollary 1, with an appropriate choice for α and T, the sample-complexity of DisSACC is $\tilde{\mathcal{O}}\left(\sigma^2/(MR)\right)$, *which is essentially the best one could hope for* since the total number of samples across M agents is exactly MR.

• *No Heterogeneity Bias.* Is is also clear from the expression in (12) that DisSACC effectively removes any heterogeneity-induced bias term, which is a result of the fact that all updates from agents are *synchronized at every time-step*. Moreover, as shown in Section V, the analysis is also significantly simplified, as there is no need to perform drift control—each update directly modifies the global model $\bar{\theta}^{(t)}$ without accumulating agent-specific deviations.

• *Logarithmic Communication.* While our prior work [18] also achieves convergence to $\theta^\star$ with linear speedup and no heterogeneity bias, it requires $\tilde{\mathcal{O}}\left(\sqrt{MR}\right)$ communication rounds, which can be prohibitively large in practice. In contrast, DisSACC attains the same convergence guarantees, but with only $T = \mathcal{O}(\log MR) = \tilde{\mathcal{O}}(1)$ rounds of communication, yielding a substantial reduction in communication cost.

Remark. *Our algorithm can be naturally extended to accommodate the more challenging Markovian sampling model where each agent i's observations are drawn from an ergodic Markov chain of which μ_i is the stationary distribution. In this case, one can modify our algorithm so that for agent i, it operates on every τ_i-th sample and drops the rest, where τ_i is the mixing-time of agent i's Markov chain. Due to geometric mixing, the sub-sampled trajectory essentially behaves as an I.I.D. process, allowing one to port guarantees under I.I.D. sampling via a simple coupling argument [23].*

V. Analysis

The goal of this section is to provide a detailed analysis supporting the main results stated in Section IV. We begin by introducing several key technical facts, which serve as critical components in the proof of the main theorem.

Fact 1 (Unbiasedness). *For every $i \in [M]$ and $t = 0, \cdots, T-1$, it holds for any **fixed** $\theta \in \mathbb{R}^d$ that* $\mathbb{E}\left[\hat{G}_{i,t}(\theta)\right] = \bar{G}_i(\theta)$.

Proof. From the definition of $\hat{G}_{i,t}$, we can write that

$$\mathbb{E}\left[\hat{G}_{i,t}(\theta)\right] = \mathbb{E}\left[\frac{1}{H}\sum_{\ell=0}^{H-1} G_i(\theta, o_{i,\ell}^{(t)})\right] = \bar{G}_i(\theta), \tag{14}$$

where we used the linearity of expectation, the fact that $o_{i,\ell}^{(t)}$ is drawn I.I.D. from μ_i, and the unbiasedness property in (1). □

Fact 1 essentially establishes that the refined operator $\hat{G}_{i,t}$ remains an unbiased estimate of the true local operator $\bar{G}_i$.

Fact 2 (Variance Reduction by H). *For every $i \in [M]$, $t = 0, \cdots, T-1$, and any $\theta \in \mathbb{R}^d$ whose randomness is independent of $\{o_{i,\ell}^{(t)}\}, \ell = 0, \cdots, H-1$, it holds that*

$$\mathbb{E}\left[\left\|\hat{G}_{i,t}(\theta) - \bar{G}_i(\theta)\right\|_2^2\right] \le \mathcal{O}\left(\frac{L^2\mathbb{E}\left[\|\theta - \theta^\star\|_2^2\right] + L^2\sigma^2}{H}\right). \tag{15}$$

Proof. From the definition of $\hat{G}_{i,t}$, we have

$$\begin{aligned}
\mathbb{E}\left[\left\|\hat{G}_{i,t}(\theta) - \bar{G}_i(\theta)\right\|_2^2\right] &= \mathbb{E}\left[\left\|\frac{1}{H}\sum_{\ell=0}^{H-1}\left(G_i(\theta, o_{i,\ell}^{(t)}) - \bar{G}_i(\theta)\right)\right\|_2^2\right] \\
&\overset{(a)}{=} \frac{1}{H^2}\sum_{\ell=0}^{H-1}\mathbb{E}\left[\left\|G_i(\theta, o_{i,\ell}^{(t)}) - \bar{G}_i(\theta)\right\|_2^2\right] \\
&\overset{(b)}{\le} \frac{L^2}{H}\mathcal{O}\left(\mathbb{E}\left[\|\theta\|_2^2\right] + \sigma_i^2\right) \\
&\overset{(c)}{\le} \mathcal{O}\left(\frac{L^2\mathbb{E}\left[\|\theta - \theta^\star\|_2^2\right] + L^2\sigma^2}{H}\right),
\end{aligned} \tag{16}$$

where (b) follows from Assumption 1, and (c) holds due to the definition of σ. To see why (a) holds, define $d_{i,\ell}^{(t)}(\theta) := G_i(\theta, o_{i,\ell}^{(t)}) - \bar{G}_i(\theta)$. We can then write

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{\ell=0}^{H-1} d_{i,\ell}^{(t)}(\theta)\right\|_2^2\right] &= \mathbb{E}\left[\sum_{\ell=0}^{H-1}\left\|d_{i,\ell}^{(t)}(\theta)\right\|_2^2\right] + \mathbb{E}\left[\sum_{p\neq q}\left\langle d_{i,p}^{(t)}(\theta),\, d_{i,q}^{(t)}(\theta)\right\rangle\right] \\
&= \mathbb{E}\left[\sum_{\ell=0}^{H-1}\left\|d_{i,\ell}^{(t)}(\theta)\right\|_2^2\right] + \mathbb{E}\left[\mathbb{E}\left[\sum_{p\neq q}\left\langle d_{i,p}^{(t)}(\theta),\, d_{i,q}^{(t)}(\theta)\right\rangle \;\middle|\; \theta\right]\right] \\
&= \mathbb{E}\left[\sum_{\ell=0}^{H-1}\left\|d_{i,\ell}^{(t)}(\theta)\right\|_2^2\right],
\end{aligned} \tag{17}$$

where the last equality holds because $\mathbb{E}\left[d_{i,\ell}^{(t)}(\theta) \mid \theta\right] = 0$ since G_i is an unbiased estimate of $\bar{G}_i$ when $o \sim \mu_i$, and $d_{i,p}^{(t)}(\theta), d_{i,q}^{(t)}(\theta)$ are independent (conditioned on θ) since observation samples are drawn I.I.D from μ_i. □

Fact 2 validates that the variance of the refined operator is reduced by a factor of H due to local estimation.

Fact 3 (Variance Reduction by MH). *For each $t = 0, \cdots, T-1$ and any $\theta \in \mathbb{R}^d$ whose randomness is independent of $\{o_{i,\ell}^{(t)}\}, \ell = 0, \cdots, H-1$, $i \in [M]$, the following holds*

$$\mathbb{E}\left[\left\|\tfrac{1}{M}\textstyle\sum_{i\in[M]}\hat{G}_{i,t}(\theta) - \bar{G}(\theta)\right\|_2^2\right] \le \mathcal{O}\left(\tfrac{L^2\mathbb{E}\left[\|\theta-\theta^\star\|_2^2\right]+L^2\sigma^2}{HM}\right). \tag{18}$$

Proof. Define $e_i^{(t)}(\theta) := \hat{G}_{i,t}(\theta) - \bar{G}_i(\theta)$. Then due to the definition of $\bar{G}$ and $e_i^{(t)}(\theta)$, we can write the left-hand-side of (18) multiplied by M^2 as

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{i\in[M]} e_i^{(t)}(\theta)\right\|_2^2\right] &= \mathbb{E}\left[\sum_{i\in[M]}\left\|e_i^{(t)}(\theta)\right\|_2^2\right] + \mathbb{E}\left[\sum_{i\neq j}\left\langle e_i^{(t)}(\theta),\, e_j^{(t)}(\theta)\right\rangle\right] \\
&= \mathbb{E}\left[\sum_{i\in[M]}\left\|e_i^{(t)}(\theta)\right\|_2^2\right] + \mathbb{E}\left[\mathbb{E}\left[\sum_{i\neq j}\left\langle e_i^{(t)}(\theta),\, e_j^{(t)}(\theta)\right\rangle \;\middle|\; \theta\right]\right] \\
&= \sum_{i\in[M]}\mathbb{E}\left[\left\|e_i^{(t)}(\theta)\right\|_2^2\right] \le \mathcal{O}\left(\frac{ML^2\mathbb{E}\left[\|\theta-\theta^\star\|_2^2\right]+ML^2\sigma^2}{H}\right),
\end{aligned} \tag{19}$$

where the last equality holds because $\mathbb{E}\left[e_i^{(t)}(\theta) \mid \theta\right] = 0$ from Fact 1, and $e_i^{(t)}(\theta)$ and $e_j^{(t)}(\theta)$ are independent (conditioned on θ) due to Assumption 3. The inequality holds due to Fact 2. Multiplying both sides with $1/M^2$ yields the desired result. □

Fact 3 is critical for establishing a linear speedup by revealing that the variance of the local models are further brought down by M after aggregation at the server.

With these facts at hand, we are in a position to prove the main result. From the governing rules of DisSACC in equations (9)–(11), we obtain

$$\begin{aligned}\|s_{t+1}\|_2^2 &= \left\| s_t + \frac{\alpha}{M}\sum_{i\in[M]} \hat{G}_{i,t}(\bar{\theta}^{(t)}) \right\|_2^2 \\ &= \|s_t\|_2^2 + \underbrace{\left\langle s_t, \frac{2\alpha}{M}\sum_{i\in[M]} \hat{G}_{i,t}(\bar{\theta}^{(t)}) \right\rangle}_{T_1} + \underbrace{\left\| \frac{\alpha}{M}\sum_{i\in[M]} \hat{G}_{i,t}(\bar{\theta}^{(t)}) \right\|_2^2}_{T_2}.\end{aligned} \tag{20}$$

For the term T_1, we have

$$\begin{aligned}\mathbb{E}[T_1] &= 2\alpha\mathbb{E}\left[\mathbb{E}\left[\left\langle s_t, \frac{1}{M}\sum_{i\in[M]} \hat{G}_{i,t}(\bar{\theta}^{(t)})\right\rangle \;\middle|\; \bar{\theta}^{(t)}\right]\right] \\ &= 2\alpha\mathbb{E}\left[\left\langle s_t, \frac{1}{M}\sum_{i\in[M]} \bar{G}_{i}(\bar{\theta}^{(t)})\right\rangle\right] \\ &= 2\alpha\mathbb{E}\left[\left\langle s_t, \bar{G}(\bar{\theta}^{(t)})\right\rangle\right] \le -2\alpha\mu\mathbb{E}\left[\|s_t\|_2^2\right],\end{aligned} \tag{21}$$

where the second equality uses Fact 1, and the inequality uses the strong-monotonicity property in Assumption 2.

For the term T_2, it follows that

$$\begin{aligned}\mathbb{E}[T_2] &= \alpha^2\mathbb{E}\left[\left\| \frac{1}{M}\sum_{i\in[M]}\left(\hat{G}_{i,t}(\bar{\theta}^{(t)}) - \bar{G}_i(\bar{\theta}^{(t)})\right) + \frac{1}{M}\sum_{i\in[M]}\bar{G}_i(\bar{\theta}^{(t)})\right\|_2^2\right] \\ &\overset{(a)}{\le} 2\alpha^2\mathbb{E}\left[\left\| \frac{1}{M}\sum_{i\in[M]}\hat{G}_{i,t}(\bar{\theta}^{(t)}) - \bar{G}(\bar{\theta}^{(t)})\right\|_2^2\right] + 2\alpha^2\mathbb{E}\left[\left\|\bar{G}(\bar{\theta}^{(t)})\right\|_2^2\right] \\ &\overset{(b)}{\le} \mathcal{O}\left(\frac{\alpha^2L^2\left(\mathbb{E}\left[\|s_t\|_2^2\right]+\sigma^2\right)}{HM}\right) + \mathcal{O}\left(\alpha^2L^2\right)\mathbb{E}\left[\|s_t\|_2^2\right],\end{aligned} \tag{22}$$

where (a) uses $\|x+y\|_2^2 \le 2\|x\|_2^2 + 2\|y\|_2^2, \forall x, y \in \mathbb{R}^d$, and (b) uses Fact 3 and Assumption 1. Taking expectation on both sides of (20), and using (21) and (22) yields:

$$\begin{aligned}\mathbb{E}\left[\|s_{t+1}\|_2^2\right] &\le \left(1 - 2\alpha\mu + C\alpha^2L^2\right)\mathbb{E}\left[\|s_t\|_2^2\right] + \mathcal{O}\left(\frac{\alpha^2L^2\sigma^2}{HM}\right) \\ &\le (1-\alpha\mu)\mathbb{E}\left[\|s_t\|_2^2\right] + \mathcal{O}\left(\frac{\alpha^2L^2\sigma^2}{HM}\right),\end{aligned} \tag{23}$$

where C is some suitably large universal constant; the second inequality in the above display follows from selecting α such that $\alpha \le \mu/(CL^2)$. Iterating (23) for T steps yields

$$\mathbb{E}\left[\|s_T\|_2^2\right] \le (1-\alpha\mu)^T\|s_0\|_2^2 + \mathcal{O}\left(\frac{\alpha L^2\sigma^2}{\mu HM}\right), \tag{24}$$

which is the desired result.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, pp. 1273–1282, PMLR, 2017.

[2] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[3] J. Bhandari, D. Russo, and R. Singal, "A finite time analysis of temporal difference learning with linear function approximation," in *Conference on Learning Theory*, pp. 1691–1692, PMLR, 2018.

[4] R. Srikant and L. Ying, "Finite-time error bounds for linear stochastic approximation and TD learning," in *Conference on Learning Theory*, pp. 2803–2830, PMLR, 2019.

[5] Z. Chen, S. Zhang, T. T. Doan, J.-P. Clarke, and S. T. Maguluri, "Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning," *Automatica*, vol. 146, p. 110623, 2022.

[6] A. Mitra, "A simple finite-time analysis of td learning with linear function approximation," *IEEE Transactions on Automatic Control*, 2024.

[7] T. Doan, S. Maguluri, and J. Romberg, "Finite-time analysis of distributed TD (0) with linear function approximation on multi-agent reinforcement learning," in *International Conference on Machine Learning*, pp. 1626–1635, PMLR, 2019.

[8] R. Liu and A. Olshevsky, "Distributed TD (0) with almost no communication," *IEEE Control Systems Letters*, vol. 7, pp. 2892–2897, 2023.

[9] H. Shen, K. Zhang, M. Hong, and T. Chen, "Towards understanding asynchronous advantage actor-critic: Convergence and linear speedup," *IEEE Transactions on Signal Processing*, 2023.

[10] S. Khodadadian, P. Sharma, G. Joshi, and S. T. Maguluri, "Federated reinforcement learning: Linear speedup under markovian sampling," in *International Conference on Machine Learning*, pp. 10997–11057, PMLR, 2022.

[11] J. Woo, G. Joshi, and Y. Chi, "The blessing of heterogeneity in federated Q-learning: Linear speedup and beyond," in *International Conference on Machine Learning*, pp. 37157–37216, PMLR, 2023.

[12] H. Tian, I. C. Paschalidis, and A. Olshevsky, "One-shot averaging for distributed TD (λ) under markov sampling," *IEEE Control Systems Letters*, 2024.

[13] S. Salgia and Y. Chi, "The sample-communication complexity trade-off in federated q-learning," in *Advances in Neural Information Processing Systems*, 2024.

[14] A. Mitra, G. J. Pappas, and H. Hassani, "Temporal difference learning with compressed updates: Error-feedback meets reinforcement learning," *Transactions on Machine Learning Research*, 2024.

[15] H. Jin, Y. Peng, W. Yang, S. Wang, and Z. Zhang, "Federated reinforcement learning with environment heterogeneity," in *International Conference on Artificial Intelligence and Statistics*, pp. 18–37, PMLR, 2022.

[16] C. Zhang, H. Wang, A. Mitra, and J. Anderson, "Finite-time analysis of on-policy heterogeneous federated reinforcement learning," in *International Conference on Learning Representations*, 2024.

[17] H. Wang, A. Mitra, H. Hassani, G. J. Pappas, and J. Anderson, "Federated temporal difference learning with linear function approximation under environmental heterogeneity," *Transactions on Machine Learning Research*, 2024.

[18] F. Zhu, A. Mitra, and R. W. Heath, "Achieving tighter finite-time rates for heterogeneous federated stochastic approximation under markovian sampling," *arXiv preprint arXiv:2504.11645*, 2025.

[19] P. Mangold, S. Samsonov, S. Labbi, I. Levin, R. Alami, A. Naumov, and E. Moulines, "Scafflsa: Taming heterogeneity in federated linear stochastic approximation and td learning," *Advances in Neural Information Processing Systems*, 2024.

[20] T. T. Doan, "Finite-time analysis of markov gradient descent," *IEEE Transactions on Automatic Control*, 2022.

[21] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," in *IEEE Transactions on Automatic Control*, 1997.

[22] S. Zeng, T. T. Doan, and J. Romberg, "Finite-time convergence rates of decentralized stochastic approximation with applications in multi-agent and multi-task learning," *IEEE Transactions on Automatic Control*, 2022.

[23] R. Dorfman and K. Y. Levy, "Adapting to mixing time in stochastic optimization with markovian data," in *International Conference on Machine Learning*, pp. 5429–5446, PMLR, 2022.