Proceedings of Machine Learning Research vol 120:1–11, 2020

Finite-Time Performance of Distributed Two-Time-Scale Stochastic Approximation

Thinh T. Doan THINHDOAN@GATECH.EDU

Justin Romberg JROM@ECE.GATECH.EDU

School of Electrical and Computer Engineering
Georgia Institute of Technology, GA, 30332, USA

Abstract

Two-time-scale stochastic approximation is a popular iterative method for finding the solution of a system of two equations. Such methods have found broad applications in many areas, especially in machine learning and reinforcement learning. In this paper, we propose a distributed variant of this method over a network of agents, where the agents use two graphs representing their communication at different speeds due to the nature of their two-time-scale updates. Our main contribution is to provide a finite-time analysis for the performance of the proposed method. In particular, we establish an upper bound for the convergence rates of the mean square errors at the agents to zero as a function of the step sizes and the network topology.

1. Introduction

Two-time-scale stochastic approximation (SA) is a recursive algorithm for finding the solution of a system of two equations Borkar (2008). In this algorithm, one iterate is updated using step sizes that are very small compared to the ones used to update the other iterate. One can view that the update associated with the small step sizes is implemented at a "slow" time-scale, while the other is executed at a "fast" time-scale. In this paper, our focus is to consider a distributed variant of this two-time-scale SA in the context of multi-agent systems, where a group of agents can communicate at different speeds through two possibly different connected graphs. Our main goal is to study a finite-time analysis for the performance of the proposed method, where we provide an upper bound for its convergence rate as a function of the two step sizes and the two network topology.

Two-time-scale SA and its distributed counterpart have received a surge of interests due to their broad applications in many areas, some examples include optimization Wang et al. (2017); Polyak (1987), distributed optimization on multi-agent systems Doan et al. (2018a, 2017), power control for wireless networks Long et al. (2007), and especially in reinforcement learning Sutton and Barto (1998); Konda and Tsitsiklis (2003); Sutton et al. (2009b); Lee and He (2019). In these applications, it has been observed that using two-time-scale iterations one can achieve a better performance than the one-time-scale counterpart; for example, the iterates may converge faster Polyak (1987), the algorithm performs better under communication constraints Doan et al. (2017, 2018a), and the algorithm is more stable under the so-called off-policy in reinforcement learning Sutton et al. (2009b).

The existing literature has only focused on the convergence of the centralized two-time-scale SA. The asymptotic convergence of this two-time-scale SA can be achieved by using the ODE methods Borkar and Meyn (2000), while its rates of convergence has been studied in Konda and

Tsitsiklis (2004); Dalal et al. (2018); Karmakar and Bhatnagar (2018); Gupta et al. (2019); Doan and Romberg (2019). In particular, the work in Dalal et al. (2018); Karmakar and Bhatnagar (2018) provides a concentration bound for the finite-time analysis of this method, while an asymptotic rate has been studied in Konda and Tsitsiklis (2004). Recently, a finite-time analysis for the performance of the centralized two-time-scale SA has been provided in Gupta et al. (2019) under constant step sizes and in Doan and Romberg (2019) under time-varying step sizes.

We also note some relevant works on time-scale separations on network consensus problems Awad et al. (2015); Jardón-Kojakhmetov and Kuehn (2019), where the authors consider continuous-time dynamics and utilize tools from singular perturbation theory to study the asymptotic convergence of their algorithms. However, this singular perturbation theory does not immediately give the rate of the algorithms, which is the main focus of this paper.

Main Contribution. In this paper, we propose a distributed variant of the linear two-time-scale SA over a multi-agent system. Due to the two-time-scale updates, the agents use two different graphs representing their communication at two different speeds. Our focus is to provide a finite-time analysis for the proposed method. In particular, we provide an upper bound for the rates of the average of the mean square errors at the nodes to zero, as a function of the two step sizes and the two network topology. We show that this method converges at a rate $\mathcal{O}(1/(1-\sigma)^2k^{2/3})$ under some proper choice of the two step sizes, where σ represents the slower mixing time of the two communication graphs and k is the number of iterations. Our theoretical results explicitly show the impacts of the two step sizes and network topology on the performance of the proposed algorithm.

2. Distributed linear two-time-scale stochastic approximation

We consider the problem of finding the solution $(x^*, y^*) \in \mathbb{R}^d \times \mathbb{R}^d$ of a linear system of equations defined over a network of N nodes. Associated with each node i is a matrix $\mathbf{A}$ and a vector $\mathbf{b}^i$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \in \mathbb{R}^{2d\times 2d}, \qquad \mathbf{b}^i = \begin{bmatrix} b_1^i \\ b_2^i \end{bmatrix} \in \mathbb{R}^{2d}.$$

The goal of the nodes is to cooperatively find the solution (x^*, y^*) of the system of linear equations

$$\mathbf{A}_{11}x^* + \mathbf{A}_{21}y^* = \sum_{i=1}^N b_1^i \qquad \text{and} \qquad \mathbf{A}_{21}x^* + \mathbf{A}_{22}y^* = \sum_{i=1}^N b_2^i. \tag{1}$$

We are interested in the situation where a central coordinator is absent, therefore, the nodes have to cooperatively solve this problem. In addition, we assume that the matrices $\mathbf{A}_{ij}$ and b_1^i, for all i, j, are unknown to node i and each node can only have an access to a noisy observation of these matrices and vectors. Therefore, we consider distributed iterative methods for solving this problem. In particular, we are interested in studying the distributed variant of the linear two-time-scale SA Konda and Tsitsiklis (2004); Doan and Romberg (2019); Gupta et al. (2019); Dalal et al. (2018), where each node i maintains an estimate (x^i, y^i) of (x^*, y^*) and iteratively updates its estimates as

$$x_{k+1}^i = \sum_{j=1}^N w_{ij}x_k^j - \alpha_k(\mathbf{A}_{11}x_k^i + \mathbf{A}_{12}y_k^i - b_1^i + \xi_k^i) \tag{2}$$

$$y_{k+1}^i = \sum_{j=1}^N v_{ij}y_k^j - \beta_k(\mathbf{A}_{21}x_k^i + \mathbf{A}_{22}y_k^i - b_2^i + \psi_k^i), \tag{3}$$

where $\beta_k \ll \alpha_k$ are two different nonnegative step sizes. In addition, (w_{ij}, v_{ij}) are the weights that node i assigns for the iterate (x^j, y^j) received from node j, a neighbor of node i. We denote by $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{N\times N}$ and $\mathbf{V} = [v_{ij}] \in \mathbb{R}^{N\times N}$ the two adjacency matrices imposed the communication structures between the nodes, that is, nodes i and j can interact with each other if and only if $w_{ij} > 0$ or $v_{ij} > 0$. Note that $\mathbf{W}$ and $\mathbf{V}$ can represent two different graphs, i.e., the nodes can exchange information in different speeds. In addition, $\{\xi_k^i, \psi_k^i\}$ are the noise sequences corresponding to observations at each node i. Here, the goal of the nodes is to obtain (x^*, y^*), i.e.,

$$\lim_{k\to\infty} x_k^i = x^* \qquad \text{and} \qquad \lim_{k\to\infty} y_k^i = y^* \qquad \text{a.s.,} \quad \forall i \in [1, N].$$

Here β_k is much smaller than α_k, implying that x_k^i is updated at a faster time scale than y_k^i. Finally, the adjacency matrices $\mathbf{W}, \mathbf{V}$ is used to present different communication speeds between the nodes.

2.1. Motivating applications

We are motivated by the wide applications of (2) and (3) in many applications, especially the recent interests in multi-agent reinforcement learning Mathkar and Borkar (2017); Doan et al. (2019b,a); Zhang et al. (2019); Yang et al. (2018); Kar et al. (2013); Wai et al. (2018); Ding et al. (2019). One fundamental and important problem in this area is to estimate the total accumulative return rewards of a stationary policy using linear function approximations, which is referred to as the policy evaluation problems. In this context, two-time-scale algorithms (e.g., gradient temporal difference learning (GTD)) have been observed to be more stable and perform better compared to the single-time-scale counterpart (e.g., temporal difference learning (TD)) in the so-called off-policy settings; see for example Sutton et al. (2009b,a). Motivated by the distributed variant of TD studied in Doan et al. (2019b), we consider a distributed version of GTD formulated under the forms of (2) and (3). In particular, a team of agents act in a common environment, get rewarded, update their local estimates of the value function, and then communicate with their neighbors. Let X_k be the state of environment, γ be the discount factor, $\phi(X_k)$ be the feature vector of state X_k, and $R^i(\cdot)$ be the local reward return at agent i. Given a sequence of samples $\{X_k\}$, the updates at the agents can be viewed as a distributed variant of the GTD studied in Sutton et al. (2009a) given as

$$\begin{aligned}
x_{k+1}^i &= \sum_{j=1}^N w_{ij} x_k^j - \alpha_k \Big(\phi(X_k)^T x_k^i + \big[\phi(X_k) - \gamma\phi(X_{k+1})\big]^T y_k^i - R^i(X_k)\Big)\phi(X_k)\\
y_{k+1}^i &= \sum_{j=1}^N v_{ij} y_k^j - \beta_k \big[\gamma\phi(X_{k+1}) - \phi(X_k)\big]\phi^T(X_k) x_k^i.
\end{aligned}$$

At each agent i, y_k^i is the main variable used to estimate the optimal solution, while x_k^i is an additional auxiliary variable. To put these updates in the form of (2) and (3), we introduce the notation

$$\begin{aligned}
&\mathbf{A}_{11}(X_k) = \phi(X_k)\phi^T(X_k),\ \mathbf{A}_{12}(X_k) = \phi(X_k)\big[\phi(X_k) - \gamma\phi(X_{k+1})\big]^T,\ b_1^i(X_k) = R^i(X_k)\phi(X_k)\\
&\mathbf{A}_{21}(X_k) = [\gamma\phi(X_{k+1}) - \phi(X_k)]\phi^T(X_k), \quad \mathbf{A}_{22}(X_k) = 0, \quad b_2^i(X_k) = 0.
\end{aligned}$$

In addition, we denote by $\mathbf{A}_{\ell u} = \mathbb{E}[\mathbf{A}_{\ell u}(X_k)]$ and $b_\ell^i = \mathbb{E}[b_\ell^i(X_k)]$, for all $\ell, u = 1, 2$ and $i \in [1, N]$. One can reformulate the distributed GTD above by introducing ξ_k^i and ψ_k^i as

$$\xi_k^i = [\mathbf{A}_{11}(X_k) - \mathbf{A}_{11}]x_k^i + [\mathbf{A}_{12}(X_k) - \mathbf{A}_{12}]y_k^i + b_1^i(X_k) - b_1^i, \quad \psi_k^i = [\mathbf{A}_{21}(X_k) - \mathbf{A}_{21}]x_k^i$$

Let $b_1 = 1/N\sum_i b_1^i$. The goal of the distributed GTD is tried to have all (x_k^i, y_k^i) converge to (x^*, y^*), where $x^* = \mathbf{A}_{11}^{-1}\left(\mathbf{A}_{21}^T y^* + b_1\right)$ and $y^* = \mathbf{A}_{12}^{-1} b_1$.
Another motivating example of using such distributed two-time-scale algorithms 2 and 3 is to solve distributed optimization problems under communication constraints, where another step size in addition to the one associated with the gradients of the functions is introduced to stabilize the algorithm due to the imperfect communication between agents Doan et al. (2017, 2018a). Finally, the distributed two-time-scale method studied in this paper can be used to solve a convex relaxation of the popular pose graph estimation in robotic networks Choudhary et al. (2016).

2.2. Assumptions and notation

We introduce in this section various assumptions, which are necessary to our analysis given later. We first state an assumption on the matrices $\mathbf{A}_{ij}$ to guarantee the existence and uniqueness of (x^*, y^*).

Assumption 1 *The matrices $\mathbf{A}_{11}$ and $\Delta = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ are positive definite but not necessarily symmetric, i.e., their eigenvalues are strictly positive.*

One can relax Assumption 1 to cover the case of complex eigenvalues, i.e., $\mathbf{A}_{11}$ and Δ have eigenvalues with strictly positive real parts. To simplify the notation of our analysis we, however, assume that these matrices are positive definite. An extension of this work to the case of complex eigenvalues is straightforward; see for example Konda and Tsitsiklis (2004); Doan and Romberg (2019).

Assumption 2 *All the matrices $\mathbf{A}_{ij}$ and vectors b_1^i are uniformly bounded, i.e., $\|\mathbf{A}_{ij}\| \leq 1$ and there exists a positive constant R such that $\max\{\|b_1^i\|, \|b_2^i\|\} \leq R$ for all $i \in [1, N]$.*

Assumption 3 *The matrix $\mathbf{W}$, whose (i,j)-th entries are w_{ij}, is doubly stochastic with positive diagonal, i.e., $\sum_{j=1}^n w_{ij} = \sum_{i=1}^n w_{ij} = 1$. Moreover, the graph $\mathcal{G}_{\mathbf{W}}$ associated with $\mathbf{W}$ is connected, and $w_{ij} > 0$ if and only if (i,j) is an edge of $\mathcal{G}_{\mathbf{W}}$. Similar conditions are assumed for $\mathbf{V}$.*

Assumption 4 *The sequence of random variables (ξ_k^i, ψ_k^i), for all $i \in [1, N]$ and $k \geq 0$, is independent of each other, with zero mean and uniformly bounded, i.e., there exists a positive constant C s.t. $\max\{\|\xi_k^i\|, \|\psi_k^i\|\} \leq C$ for all $i \in [1, N]$. Moreover, they have common variances given as*

$$\mathbb{E}[(\xi_k^i)^T \xi_k^i] = \Gamma_{11}, \quad \mathbb{E}[(\xi_k^i)^T \psi_k^i] = \Gamma_{12} = \Gamma_{21}^T, \quad \mathbb{E}[(\psi_k^i)^T \psi_k^i] = \Gamma_{22}. \tag{4}$$

Assumption 2 can be guaranteed through a proper scaling step, while Assumption 3 is a standard assumption in distributed consensus algorithms Doan et al. (2018b). Finally, we consider the noise model similar to the one in Konda and Tsitsiklis (2004); Doan and Romberg (2019).

We denote by $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N\times d}$ the matrices whose $i-$th rows are $(x^i)^T$ and $(y^i)^T$ in $\mathbb{R}^{1\times d}$, respectively. Then, the matrix forms of Eqs. (2) and (3) are given as

$$\mathbf{X}_{k+1} = \mathbf{W}\mathbf{X}_k - \alpha_k\left(\mathbf{X}_k\mathbf{A}_{11}^T + \mathbf{Y}_k\mathbf{A}_{12}^T - \mathbf{B}_1 + \Xi_k\right) \tag{5}$$

$$\mathbf{Y}_{k+1} = \mathbf{V}\mathbf{Y}_k - \alpha_k\left(\mathbf{X}_k\mathbf{A}_{21}^T + \mathbf{Y}_k\mathbf{A}_{22}^T - \mathbf{B}_2 + \Psi_k\right), \tag{6}$$

where $\mathbf{B}_1, \mathbf{B}_2, \Xi_k$, and Ψ_k are the matrices whose $i-$th rows are $(b_1^i)^T, (b_2^i)^T, (\xi_k^i)^T$, and $(\psi_k^i)^T$, respectively. Given a collection of $x^1, \ldots, x^N$, we use $\bar{x}$ to denote its average, i.e., $\bar{x} = \frac{1}{N}\sum_{i=1}^N x^i$. Thus, since $\mathbf{W}$ and $\mathbf{V}$ are doubly stochastic matrices and by Eqs. (2) and (3) we have

$$\bar{x}_{k+1} = \bar{x}_k - \alpha_k\left(\mathbf{A}_{11}\bar{x}_k + \mathbf{A}_{12}\bar{y}_k - \bar{b}_1 + \bar{\xi}_k\right) \tag{7}$$

$$\bar{y}_{k+1} = \bar{y}_k - \alpha_k\left(\mathbf{A}_{21}\bar{x}_k + \mathbf{A}_{22}\bar{y}_k - \bar{b}_2 + \bar{\psi}_k\right). \tag{8}$$

3. Finite-time bounds of distributed linear two-time-scale SA

We present here the convergence rates of the distributed linear two-time-scale SA, where we provide an upper bound for the rates of the average of the mean square errors at the nodes to zero. Our result shows that this quantity decays to zero at a rate $\mathcal{O}(1/(k+1)^{2/3})$. In addition, it also depends on the network topology represented by $1-\sigma$, the algebraic network connectivity of two graphs.

We first introduce a bit more notation. We denote by $\sigma_{\mathbf{W}}$ and $\sigma_{\mathbf{V}}$ the second larges singular values of $\mathbf{W}$ and $\mathbf{V}$, respectively. By Assumption 3 we have $\sigma_{\mathbf{W}},\sigma_{\mathbf{V}}\in(0,1)$; see for example Godsil and Royle (2001). In addition, we denote by σ, the slower mixing speed of these two graphs

$$\sigma\triangleq\max\{\sigma_{\mathbf{W}},\sigma_{\mathbf{V}}\}\in(0,1). \tag{9}$$

Let $\delta\in(\sigma,1)$ and denote by $\mathcal{K}^*$ a positive integer such that

$$\mathcal{K}^*\geq\left\lceil(\alpha_0/(\delta-\sigma))^{3/2}\right\rceil. \tag{10}$$

Finally, since $\lim_{k\to\infty}\sigma^k(k+1)=0$, without loss of generality we assume that $\sigma^k\leq\frac{1}{k+1}$. Our main result, the rate of the distributed two-time-scale SA, is stated in the following theorem.

Theorem 1 *Suppose that Assumptions 1–4 hold. Let $\{x_k^i,y_k^i\}$, for all $i\in[1,N]$, be generated by (2) and (3) with $x_0^i=y_0^i=0$. Let $\{\alpha_k,\beta_k\}$ be the sequence of step sizes chosen as*

$$\alpha_k=\frac{\alpha_0}{(k+1)^{2/3}},\qquad\beta_k=\frac{\beta_0}{k+1}. \tag{11}$$

Then, there exits constants $\mathcal{D},\mathcal{D}_0,\mathcal{D}_1$ given in Lemmas 1 and 2 below such that

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}[\|y_k^i-y^*\|^2]+\frac{\beta_k}{\alpha_k}\mathbb{E}[\|x_k^i-x^*\|^2]\right)\\&\leq\frac{16\mathcal{D}^2\beta_0\alpha_0\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{N(1-\sigma)^2(k+1)^{2/3}}+\frac{16\mathcal{D}^2\beta_0\alpha_0}{N(1-\sigma)^2(k+2)^{5/3}}+\frac{2\mathcal{D}_0}{(k+1)^{2/3}}+\frac{2\mathcal{D}_1\ln(k+1)}{k+1}.\end{aligned} \tag{12}$$

More details about the choice of the two step sizes can be found in Doan and Romberg (2019).

4. Convergence analysis

We now present the analysis for the results presented in Theorem 1. Our analysis is composed of two main steps. We first show that the estimates x_k^i and y_k^i converge to their averages $\bar{x}_k$ and $\bar{y}_k$, respectively. We provide an upper bound for the rates of this convergence. This step is done through considering a residual function, which takes into account the coupling between the two step sizes

$$V_k=\|\mathbf{Y}_k-\mathbf{1}\bar{y}_k^T\|+\frac{\beta_k}{\alpha_k}\|\mathbf{X}_k-\mathbf{1}\bar{x}_k^T\|. \tag{13}$$

Second, we study the convergence of $\bar{x}_k$ and $\bar{y}_k$ to the solutions x^* and y^*, respectively. One can view the updates of (7) and (8) as a centralized approach for solving (21). We, therefore, utilize the results in our previous work to have such convergence Doan and Romberg (2019). Due to the space limit, we skip the analysis of the second step and refer interested readers to Doan and Romberg (2019) for more details. Our focus here is to provide the analysis for the first step as follows.

Lemma 1 *Suppose that all assumptions and step sizes in Theorem 1 hold. Denote by $\mathcal{D}$ a constant*

$$\mathcal{D} \triangleq \frac{2\sqrt{N}(R+C)(6\alpha_0+1)(\mathcal{K}^*)^{1/3}}{1-\delta}, \tag{14}$$

where $\mathcal{K}^*$ is defined in (10). Then we obtain for all $k \geq 0$

$$\sum_{i=1}^{N}\left(\|y_i^k - \bar{y}_k\|^2 + \frac{\beta_k}{\alpha_k}\|x_i^k - \bar{x}_k\|^2\right) \leq \frac{8\mathcal{D}^2\beta_0\alpha_0\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{(1-\sigma)^2(k+1)^{2/3}} + \frac{8\mathcal{D}^2\beta_0\alpha_0}{(1-\sigma)^2(k+2)^{5/3}}. \tag{15}$$

Proof Let $\hat{x}^i = x^i - \bar{x}$ and $\hat{y}^i = y^i - \bar{y}$. Since $\mathbf{W}$ is doubly stochastic Eqs. (2) and (7) gives

$$\hat{x}_{k+1}^i = \sum_{j=1}^{N} w_{ij}\hat{x}_k^j - \alpha_k\mathbf{A}_{11}\hat{x}_i^k - \alpha_k\mathbf{A}_{12}\hat{y}_k^i + \alpha_k(b_1^i - \bar{b}_1) - \alpha_k(\xi_k^i - \bar{\xi}_k),$$

which implies that

$$\hat{\mathbf{X}}_{k+1} = \mathbf{W}\hat{\mathbf{X}}_k - \alpha_k\hat{\mathbf{X}}_k\mathbf{A}_{11}^T - \alpha_k\hat{\mathbf{Y}}_k\mathbf{A}_{12}^T + \alpha_k(\mathbf{B}_1 - \mathbf{1}\bar{b}_1^T) - \alpha_k(\Xi_k - \mathbf{1}\bar{\xi}_k^T). \tag{16}$$

Using Assumption 3 yields $\|\mathbf{W}\hat{\mathbf{X}}_k\| \leq \sigma_{\mathbf{W}}\|\hat{\mathbf{X}}_k\|$. Thus, by (16) and Assumptions 2 and 4 we have

$$\|\hat{\mathbf{X}}_{k+1}\| \leq (\sigma_{\mathbf{W}} + \alpha_k)\|\hat{\mathbf{X}}_k\| + \alpha_k\|\hat{\mathbf{Y}}_k\| + \sqrt{N}(R+C)\alpha_k. \tag{17}$$

Similarly, using Eqs. (3) and (8) we obtain

$$\|\hat{\mathbf{Y}}_{k+1}\| \leq (\sigma_{\mathbf{V}} + \beta_k)\|\hat{\mathbf{Y}}_k\| + \beta_k\|\hat{\mathbf{X}}_k\| + \sqrt{N}(R+C)\beta_k. \tag{18}$$

By (9), $\sigma = \max\{\sigma_{\mathbf{V}}, \sigma_{\mathbf{W}}\} \in (0,1)$, and by (10), $\sigma + 2\alpha_k \leq \delta \in (\sigma, 1), \forall k \geq \mathcal{K}^*$. Then, adding Eq. (17) to Eq. (18) and using $\beta_k \ll \alpha_k$ yield

$$\begin{aligned}
&\|\hat{\mathbf{X}}_{k+1}\| + \|\hat{\mathbf{Y}}_{k+1}\| \\
&\leq (\sigma + 2\alpha_k)(\|\hat{\mathbf{X}}_k\| + \|\hat{\mathbf{Y}}_k\|) + 2\sqrt{N}(R+C)\alpha_k \leq \delta(\|\hat{\mathbf{X}}_k\| + \|\hat{\mathbf{Y}}_k\|) + 2\sqrt{N}(R+C)\alpha_k \\
&\leq \delta^{k+1-\mathcal{K}^*}(\|\hat{\mathbf{X}}_{\mathcal{K}^*}\| + \|\hat{\mathbf{Y}}_{\mathcal{K}^*}\|) + 2\sqrt{N}(R+C)\sum_{t=\mathcal{K}^*}^{k}\alpha_k\delta^{k-t} \leq (\|\hat{\mathbf{X}}_{\mathcal{K}^*}\| + \|\hat{\mathbf{Y}}_{\mathcal{K}^*}\|) + \frac{2\sqrt{N}(R+C)\alpha_0}{1-\delta}.
\end{aligned}$$

Similarly, since $x_0^i = y_0^i = 0$ we can obtain

$$\begin{aligned}
\|\hat{\mathbf{X}}_{\mathcal{K}^*}\| + \|\hat{\mathbf{Y}}_{\mathcal{K}^*}\| &\leq (\sigma + 2\alpha_0)(\|\hat{\mathbf{X}}_{\mathcal{K}^*-1}\| + \|\hat{\mathbf{Y}}_{\mathcal{K}^*-1}\|) + 2\sqrt{N}(R+C)\alpha_{\mathcal{K}^*-1} \\
&\leq 2\sqrt{N}(R+C)\sum_{t=0}^{\mathcal{K}^*-1}\alpha_t \leq 6\sqrt{N}(R+C)\alpha_0(\mathcal{K}^*)^{1/3},
\end{aligned}$$

where the last inequality is due to $\sum_{t=0}^{\mathcal{K}^*-1}\alpha_t \leq 3\alpha_0(\mathcal{K}^*)^{1/3}$. Thus, the two preceding relations give

$$\|\hat{\mathbf{X}}_{\mathcal{K}^*+1}\| + \|\hat{\mathbf{Y}}_{\mathcal{K}^*+1}\| \leq \frac{6\sqrt{N}(R+C)\alpha_0(\mathcal{K}^*)^{1/3}}{1-\delta}. \tag{19}$$

We denote by $\gamma_k = \beta_k/\alpha_k$, a nonnegative and nonincreasing sequence since $\beta_k \ll \alpha_k$. Moreover, since $\beta_0 \leq \alpha_0$, we have $\gamma_k \leq 1$. We next consider the residual function V in (13). Indeed, using Eqs. (17) and (18) and since $\gamma_{k+1} \leq \gamma_k \leq 1$ we have

$$\begin{aligned} V_{k+1} &= \|\hat{\mathbf{Y}}_{k+1}\| + \gamma_{k+1}\|\hat{\mathbf{X}}_{k+1}\| \leq \|\hat{\mathbf{Y}}_{k+1}\| + \gamma_k\|\hat{\mathbf{X}}_{k+1}\| \\ &\leq (\sigma_{\mathbf{V}} + \beta_k)\|\hat{\mathbf{Y}}_k\| + \beta_k\|\hat{\mathbf{X}}_k\| + 2\sqrt{N}(R+C)\beta_k + \sigma_{\mathbf{W}}\gamma_k\|\hat{\mathbf{X}}_k\| + \beta_k\|\hat{\mathbf{X}}_k\| + \beta_k\|\hat{\mathbf{Y}}_k\| \\ &\leq \sigma V_k + 2\beta_k(\|\hat{\mathbf{Y}}_k\| + \|\hat{\mathbf{X}}_k\|) + 2\sqrt{N}(R+C)\beta_k, \end{aligned}$$

which by using Eq. (19) and $\mathcal{D}$ in (14) we have for all $k \geq \mathcal{K}^*$

$$\begin{aligned} V_{k+1} &\leq \sigma V_k + \mathcal{D}\beta_k \leq \sigma^{k+1-\mathcal{K}^*}V_{\mathcal{K}^*} + \mathcal{D}\sum_{t=\mathcal{K}^*}^{k}\beta_t\sigma^{k-t} \leq \sigma^{k+1-\mathcal{K}^*}V_{\mathcal{K}^*} + \mathcal{D}\sum_{t=\mathcal{K}^*}^{\lfloor k/2\rfloor}\beta_t\sigma^{k-t} + \mathcal{D}\sum_{t=\lceil k/2\rceil}^{k}\beta_t\sigma^{k-t} \\ &\leq \sigma^{k+1-\mathcal{K}^*}V_{\mathcal{K}^*} + \frac{\mathcal{D}\beta_0\sigma^{\lceil k/2\rceil}}{1-\sigma} + \frac{\mathcal{D}\beta_{k/2}}{1-\sigma} \leq \sigma^{k+1-\mathcal{K}^*}V_{\mathcal{K}^*} + \frac{\mathcal{D}\beta_0}{1-\sigma}\sigma^{\lceil k/2\rceil} + \frac{2\mathcal{D}\beta_0}{(1-\sigma)}\frac{1}{k+1}. \end{aligned}$$

Moreover, since $x_0^i = y_0^i = 0$ implying $V_0 = 0$, we have

$$V_{\mathcal{K}^*} \leq \sigma V_{\mathcal{K}^*-1} + \mathcal{D}\beta_{\mathcal{K}^*-1} \leq \mathcal{D}\beta_0\sum_{t=0}^{\mathcal{K}^*-1}\frac{1}{t+1} \leq \mathcal{D}\beta_0\ln(\mathcal{K}^*).$$

Combining these two relations immediately gives

$$\begin{aligned} V_{k+1} &\leq \mathcal{D}\beta_0\ln(\mathcal{K}^*)\sigma^{k+1-\mathcal{K}^*} + \frac{\mathcal{D}\beta_0}{1-\sigma}\sigma^{\lceil k/2\rceil} + \frac{2\mathcal{D}\beta_0}{(1-\sigma)}\frac{1}{k+1} \\ &\leq \frac{2\mathcal{D}\beta_0\ln(\mathcal{K}^*)\sigma^{-\mathcal{K}^*}}{1-\sigma}\sigma^{\lceil k/2\rceil} + \frac{2\mathcal{D}\beta_0}{(1-\sigma)}\frac{1}{k+1}. \end{aligned}$$

Using the preceding relation, the definition of V in (13), and $(x+y)^2 \leq 2x^2 + 2y^2$ we obtain

$$\sum_{i=1}^{N}\|y_i^k - \bar{y}_k\|^2 \leq \frac{4\mathcal{D}^2\beta_0^2\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{(1-\sigma)^2}\sigma^k + \frac{4\mathcal{D}^2\beta_0^2}{(1-\sigma)^2(k+2)^2}.$$

Similarly, we obtain

$$\begin{aligned} \frac{\beta_k}{\alpha_k}\sum_{i=1}^{N}\|x_i^k - \bar{x}_k\|^2 &\leq \frac{4\mathcal{D}^2\beta_0^2\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{(1-\sigma)^2}\frac{\sigma^k\alpha_k}{\beta_k} + \frac{4\mathcal{D}^2\beta_0^2}{(1-\sigma)^2}\frac{\alpha_k}{(k+2)^2\beta_k} \\ &\leq \frac{4\mathcal{D}^2\beta_0\alpha_0\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{(1-\sigma)^2}\frac{1}{(k+1)^{2/3}} + \frac{4\mathcal{D}^2\beta_0\alpha_0}{(1-\sigma)^2}\frac{1}{(k+2)^{5/3}}, \end{aligned}$$

where recall that we assume $\sigma^k \leq 1/(k+1)$. Adding the preceding two relations give Eq. (15). ■

We next utilize the following result about the convergence of of $(\bar{x}_k, \bar{y}_k)$ to the solutions (x^*, y^*).

Lemma 2 (Theorem 1 in Doan and Romberg (2019)) *Suppose that all assumptions and step sizes in Theorem 1 hold. Then there exists two absolute constants $\mathcal{D}_0$ and $\mathcal{D}_1$ such that*

$$\mathbb{E}[\|\bar{y}_k - y^*\|^2] + \frac{\beta_k}{\alpha_k}\mathbb{E}[\|\bar{x}_k - x^*\|^2] \leq \frac{\mathcal{D}_0}{(k+1)^{2/3}} + \frac{\mathcal{D}_1\ln(k+1)}{k+1}. \tag{20}$$

Using the results in Lemmas 1 and 2, we immediately have the proof of Theorem 1 as follows. By using Eqs. (15) and (20) we obtain Eq. (12), i.e.,

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{E}[\|y_k^i - y^*\|^2] + \frac{\beta_k}{\alpha_k}\mathbb{E}[\|x_k^i - x^*\|^2]\right)\\ &\leq \frac{16\mathcal{D}^2\beta_0\alpha_0\ln^2(\mathcal{K}^*)\sigma^{-2\mathcal{K}^*}}{N(1-\sigma)^2(k+1)^{2/3}} + \frac{16\mathcal{D}^2\beta_0\alpha_0}{N(1-\sigma)^2(k+2)^{5/3}} + \frac{2\mathcal{D}_0}{(k+1)^{2/3}} + \frac{2\mathcal{D}_1\ln(k+1)}{k+1}.\end{aligned}$$

Remark 3 *Note that the analysis studied in this paper can be extended to cover the case when each node i knows a different matrix $\mathbf{A}^i$, i.e., associated with each node i is a matrix $\mathbf{A}^i$ and a vector $\mathbf{b}^i$*

$$\mathbf{A}^i = \begin{bmatrix} \mathbf{A}^i_{11} & \mathbf{A}^i_{12} \\ \mathbf{A}^i_{21} & \mathbf{A}^i_{22} \end{bmatrix} \in \mathbb{R}^{2d\times 2d}, \qquad \mathbf{b}^i = \begin{bmatrix} b^i_1 \\ b^i_2 \end{bmatrix} \in \mathbb{R}^{2d}.$$

The goal of the nodes is to cooperatively find the solution (x^*, y^*) of the linear equations

$$\sum_{i=1}^{N}\mathbf{A}^i_{11}x^* + \mathbf{A}^i_{21}y^* - b^i_1 = 0, \quad \textit{and} \quad \sum_{i=1}^{N}\mathbf{A}^i_{21}x^* + \mathbf{A}^i_{22}y^* - b^i_2 = 0. \tag{21}$$

However, an additional projection step to a compact set $\mathcal{X}$ containing (x^*, y^*) is needed in this case

$$\begin{aligned} x^i_{k+1} &= \left[\sum_{j=1}^{N} w_{ij}x^j_k - \alpha_k(\mathbf{A}^i_{11}x^i_k + \mathbf{A}^i_{12}y^i_k - b^i_1 + \xi^i_k)\right]_{\mathcal{X}} \\ y^i_{k+1} &= \left[\sum_{j=1}^{N} v_{ij}y^j_k - \beta_k(\mathbf{A}_{21}x^i_k + \mathbf{A}^i_{22}y^i_k - b^i_2 + \psi^i_k)\right]_{\mathcal{X}}. \end{aligned}$$

This projection is often used in the context of distributed optimization Doan et al. (2017, 2018a). However, this step may not be practical in reinforcement learning since $\mathcal{X}$ is often difficult to decide.

5. Concluding Remarks

We proposed a distributed variant of the two-time-scale SA for finding the root of a system of two linear equations. Our main contribution is to provide a finite-time analysis of the proposed method, where we show that this method converges at a rate $\mathcal{O}(1/k^{2/3})$. Future interesting problems include the finite-time analysis of nonlinear counterparts, impacts of Markovian noise Gupta et al. (2019), and applications to multi-agent reinforcement learning.

Acknowledgment

This work was supported in part by ARL DCIST CRA W911NF-17-2-018, the Semiconductor Research Corporation (SRC), and DARPA.

References

A. Awad, A. Chapman, , E. Schoof, A. Narang-Siddarth, and M. Mesbahi. Time-scale separation on networks: Consensus, tracking, and state-dependent interactions. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015.

V. Borkar and S. Meyn. The o.d.e. method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000.

V.S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

S. Choudhary, L. Carlone, C. Nieto, J. Rogers, H. I. Christensen, and F. Dellaert. Distributed trajectory estimation with privacy and communication constraints: A two-stage distributed gauss-seidel approach. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

G. Dalal, G. Thoppe, B. Szörényi, and S. Mannor. Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In *COLT*, 2018.

D. Ding, X. Wei, Z. Yang, Z. Wang, and M. R. Jovanović. Fast multi-agent temporal-difference learning via homotopy stochastic primal-dual optimization. available at: https://arxiv.org/abs/1908.02805, 2019.

T. T. Doan and J. Romberg. Linear two-time-scale stochastic approximation: A finite-time analysis. In *Proceedings of Allerton Conference on Communication, Control, and Computing, Monticello, IL. Available at: https://proceedings.allerton.csl.illinois.edu/media/files/0108.pdf*, 2019.

T. T. Doan, C. L. Beck, and R. Srikant. On the convergence rate of distributed gradient methods for finite-sum optimization under communication delays. *Proceedings ACM Meas. Anal. Comput. Syst.*, 1(2):37:1–37:27, 2017.

T. T. Doan, S. T. Maguluri, and J. Romberg. Distributed stochastic approximation for solving network optimization problems under random quantization. Available at: https://arxiv.org/abs/1810.11568, 2018a.

T. T. Doan, S. T. Maguluri, and J. Romberg. Fast convergence rates of distributed subgradient methods with adaptive quantization. Available at: https://arxiv.org/abs/1810.13245, 2018b.

T. T. Doan, S. T. Maguluri, and J. Romberg. Finite-time performance of distributed temporal difference learning with linear function approximation. available at: https://arxiv.org/abs/1907.12530, 2019a.

T. T. Doan, S. T. Maguluri, and J. Romberg. Finite-time analysis of distributed TD(0) with linear function approximation on multi-agent reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1626–1635, Long Beach, California, USA, 2019b.

C. Godsil and G. Royle. *Algebraic Graph Theory*, volume 207 of *Graduate Texts in Mathematics*. Springe-Verlag, New York, 2001.

H. Gupta, R. Srikant, and L. Ying. Finite-time performance bounds and adaptive learning rate selection for two time-scale reinforcement learning. In *Advances in Neural Information Processing Systems*, 2019.

H. Jardón-Kojakhmetov and C. Kuehn. On fast-slow consensus networks with a dynamic weight . available at: https://arxiv.org/abs/1904.02690, 2019.

S. Kar, J. M. F. Moura, and H. V. Poor. Qd-learning: A collaborative distributed strategy for multi-agent reinforcement learning through consensus + innovations. *IEEE Trans. Signal Processing*, 61:1848–1862, 2013.

P. Karmakar and S. Bhatnagar. Two time-scale stochastic approximation with controlled markov noise and off-policy temporal-difference learning. *Math. Oper. Res.*, 43(1):130–151, February 2018.

V. R. Konda and J. N. Tsitsiklis. On actor-critic algorithms. *SIAM J. Control Optim.*, 42(4), 2003.

Vijay R. Konda and John N. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *The Annals of Applied Probability*, 14(2):796–819, 2004.

Donghwan Lee and Niao He. Target-based temporal-difference learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3713–3722, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

C. Long, Q. Zhang, B. Li, H. Yang, and X. Guan. Non-cooperative power control for wireless ad hoc networks with repeated games. *IEEE Journal on Selected Areas in Communications*, 25(6): 1101–1112, 2007.

A. Mathkar and V. S. Borkar. Distributed reinforcement learning via gossip. *IEEE Transactions on Automatic Control*, 62(3):1465–1470, 2017.

B. Polyak. *Introduction to Optimization*. Optimization software, Inc., Publication division, New York, 1987, 1987.

R. Sutton, H. R. Maei, D. Precup, S. Bhatnagar, D. Silver, C. Szepesvri, and E. Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. volume 382, 01 2009a.

R. Sutton, H. R. Maei, and C. Szepesvári. A convergent o(n) temporal-difference algorithm for off-policy learning with linear function approximation. In *Advances in Neural Information Processing Systems 21*. 2009b.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1st edition, 1998.

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

Mengdi Wang, Ethan X. Fang, and Han Liu. Stochastic compositional gradient descent: algorithms for minimizing compositions of expected-value functions. *Mathematical Programming*, 161(1): 419–449, Jan 2017.

Z. Yang, K. Zhang, M. Hong, and T. Basar. A finite sample analysis of the actor-critic algorithm. In *2018 IEEE Conference on Decision and Control (CDC)*, 2018.

K. Zhang, Z. Yang, and T. Baar. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. available at: https://arxiv.org/abs/1911.10635, 2019.