2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Distributed Online System Identification for LTI Systems Using Reverse Experience Replay

Ting-Jui Chang and Shahin Shahrampour, *Senior Member*, *IEEE*

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992456

***Abstract*— Identification of linear time-invariant (LTI) systems plays an important role in control and reinforcement learning. Both asymptotic and finite-time offline system identification are well-studied in the literature. For online system identification, the idea of stochastic-gradient descent with reverse experience replay (SGD-RER) was recently proposed, where the data sequence is stored in several buffers and the stochastic-gradient descent (SGD) update performs backward in each buffer to break the time dependency between data points. Inspired by this work, we study distributed online system identification of LTI systems over a multi-agent network. We consider agents as identical LTI systems, and the network goal is to jointly estimate the system parameters by leveraging the communication between agents. We propose DSGD-RER, a distributed variant of the SGD-RER algorithm, and theoretically characterize the improvement of the estimation error with respect to the network size. Our numerical experiments certify the reduction of estimation error as the network size grows.**

I. Introduction

System identification, the process of estimating the *unknown* parameters of a dynamical system from the observed input-output sequence, is a classical problem in control, reinforcement learning and time-series analysis. Among this class of problems, learning the transition matrix of a LTI system is a prominent well-studied case, and classical results characterize the asymptotic properties of these estimators [1]–[4].

Recently, there has been a renewed interest in the problem of identification of LTI systems, and modern statistical techniques are applied to achieve *finite-time* sample complexity guarantees [5]–[9]. However, the aforementioned works focus on the offline setup, where the estimator has access to the entire data sequence from the outset. The offline estimator cannot be directly extended to the streaming/online setup, where the system parameters need to be estimated on-the-fly. To this end, the idea of stochastic-gradient descent with reverse experience replay (SGD-RER) is proposed [10] to build an online estimator, where the data sequence is stored in several buffers, and the SGD update performs backward in each buffer to break the time dependency between data points. This online method achieves the optimal sample complexity of the offline setup up to log factors.

In this paper, we study the distributed online identification problem for a network of identical LTI systems with unknown dynamics. Each system is modeled as an agent in a multi-agent network, which receives its own data sequence from the underlying system. The goal of this network is to jointly estimate the system parameters by leveraging the communication between agents. We propose DSGD-RER, a distributed version of the online system identification algorithm in [10], where every agent applies SGD in the reverse order and communicates its estimate with its neighbors. We show that this decentralized scheme can improve the estimation error bound as the network size increases, and the simulation results demonstrate this theoretical property.

Related Work: Recently, several works have studied the finite-time properties of LTI system identification. In [5], it is shown that the dynamics of a fully observable system can be recovered from multiple trajectories by a least-squares estimator when the number of trajectories scales linearly with the system dimension. Furthermore, system identification using a single trajectory for stable [6] and unstable [7]–[9] LTI systems has been studied. The works of [8], [11] establish the theoretical lower bound of the sample complexity for fully observable LTI systems. The case of partially observable LTI systems is also addressed for stable [12]–[15] and unstable [16] systems. By applying an ℓ_1-regularized estimator, the work of [17] improves the dependency of the sample complexity on the system dimension to poly-logarithmic scaling. Contrary to the aforementioned works, which focus on the finite-time analysis of *offline* estimators, [10] proposes an online estimation algorithm, where the data sequence is split into several buffers, and in each buffer the SGD update is applied in the reverse order to remove the bias due to the time-dependency of data points. As mentioned before, our work builds on [10] for *distributed* online system identification. Finite-time analysis of system identification has also been studied for non-linear dynamical systems more recently. In the works of [18], [19], the sample complexity bounds are derived for dynamical systems described via generalized linear models, and [20] further improves the dependence of the complexity bound on the mixing time constant.

II. Problem Formulation

A. Notation

$[m]$	The set of $\{1, 2, \ldots, m\}$ for any integer m
$\sigma_{\max}(\mathbf{A})$	The largest singular value of $\mathbf{A}$
$\sigma_{\min}(\mathbf{A})$	The smallest singular value of $\mathbf{A}$
$\|\cdot\|$	Euclidean (spectral) norm of a vector (matrix)
$\mathrm{E}[\cdot]$	The expectation operator
$[\mathbf{A}]_{ij}$	The entry in i-th row j-th column of $\mathbf{A}$
$\mathbf{A} \succeq \mathbf{B}$	$(\mathbf{A} - \mathbf{B})$ is positive semi-definite
$\mathbf{I}_d$	Identity matrix with dimension $d \times d$

T.J. Chang and S. Shahrampour are with the Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115, USA. email:{chang.tin,s.shahrampour}@northeastern.edu.

This work is supported by NSF ECCS-2136206 Award.

B. Distributed Online System Identification

We consider a multi-agent network of m identical LTI systems, where the dynamics of agent k is defined as,

$$\mathbf{x}_{t+1}^k = \mathbf{A}\mathbf{x}_t^k + \mathbf{w}_t^k, \quad k \in [m].$$

$\mathbf{A} \in \mathrm{R}^{d\times d}$ is the *unknown* transition matrix, and $\mathbf{w}_t^k$ is the noise sequence generated independently from a distribution with zero mean and finite covariance Σ. The goal of the agents is to recover the matrix $\mathbf{A}$ collaboratively. Though this task can be accomplished by an individual agent (e.g., as in [10]), we will show that the collective system identification improves the theoretical error bound of estimation.

Assumption 1: The considered system is stable in the sense that $\|\mathbf{A}\| < 1$.

With Assumption 1, it can be shown that for any initial state $\mathbf{x}_0^k$, the distribution of $\mathbf{x}_t^k$ converges to a stationary distribution π with the covariance matrix $\mathbf{G} := \mathrm{E}_{\mathbf{x}\sim\pi}[\mathbf{x}\mathbf{x}^\top] = \sum_{t=0}^{\infty} \mathbf{A}^t\Sigma\mathbf{A}^t$. Note that Assumption 1 is more stringent compared to the standard stability assumption (i.e., $\rho(\mathbf{A}) < 1$); however, the analysis is extendable to this case under some modifications of the time horizon (see [10] for more details).

Assumption 2: For any $\mathbf{x} \in \mathrm{R}^d$, $\langle \mathbf{x}, \mathbf{w}_t^k\rangle$ is $C_\mu\langle \mathbf{x}, \Sigma\mathbf{x}\rangle$ sub-Gaussian, i.e.

$$\mathrm{E}_{\mathbf{w}}[\exp(y\langle \mathbf{x}, \mathbf{w}\rangle)] \le \exp(\frac{y^2}{2}C_\mu\langle \mathbf{x}, \Sigma\mathbf{x}\rangle),$$

for any $y \in \mathrm{R}$.

Sub-Gaussian noise is a standard choice for finite-time analysis of system identification (see e.g., [8], [9], [21]). Also, in some existing works, i.i.d. Gaussian noise is applied to ensure the persistence of excitation ([11], [12]). In this paper, we assume that Σ is positive-definite.

Problem Statement: Departing from the classical offline ordinary least squares (OLS) estimator, the goal is to develop a fully decentralized algorithm, where each agent produces estimates of the unknown system $\mathbf{A}$ in an online fashion. Specifically, at each iteration, agent k first updates its estimate based on the current information, and then it communicates its estimate with its neighbors. The communication is modeled via a graph that captures the underlying network topology. As mentioned earlier, the goal of an online distributed system identification algorithm, as opposed to the centralized one, is to leverage the network element to provide a finer estimation guarantee for each agent's estimation error. In this work, we quantify the estimation error of each agent, such that the error bound encapsulates the dependency on the *network size* and *topology*.

Network Structure: During the estimation process, the agents communicate locally based on a symmetric doubly stochastic matrix $\mathbf{P}$, i.e., all elements of $\mathbf{P}$ are non-negative and $\sum_{i=1}^m [\mathbf{P}]_{ji} = \sum_{j=1}^m [\mathbf{P}]_{ji} = 1$. Agents i and j communicate with each other if $[\mathbf{P}]_{ji} > 0$; otherwise $[\mathbf{P}]_{ji} = 0$. Thus, $\mathcal{N}_i := \{j \in [m] : [\mathbf{P}]_{ji} > 0\}$ is the neighborhood of agent i. The network is assumed to be connected, i.e., for any two agents $i, j \in [m]$, there is a (potentially multi-hop) path from i to j. We also assume $\mathbf{P}$ has a positive diagonal. Then, there exists a geometric mixing bound for $\mathbf{P}$ [22], such that $\sum_{j=1}^m \left|[\mathbf{P}^k]_{ji} - 1/m\right| \le \sqrt{m}\beta^k$, $i \in [m]$, where β is the second largest singular value of $\mathbf{P}$. Agents exchange only local system estimates, and they do not share any other information in the process. The communication is consistent with the structure of $\mathbf{P}$. We elaborate on this in the algorithm description.

III. Algorithm and Theoretical Results

We now lay out the distributed online linear system identification algorithm and provide the theoretical bound for the estimation error.

A. Offline and Online Settings

For the identification of LTI systems, it is well-known that the (centralized) OLS estimator $\mathbf{A}_{OLS} := \mathrm{argmin}_{\mathbf{A}} \sum_{t=0}^{T-1} \|\mathbf{A}\mathbf{x}_t - \mathbf{x}_{t+1}\|^2$ is statistically optimal. OLS is an offline estimator that can also be implemented in an online fashion (i.e., in the form of recursive least squares) by keeping track of the data covariance matrix and the residual based on the current estimate. On the other hand, gradient-based methods provide efficient mechanisms for system identification. In particular, SGD uses the gradient of the current data pair $(\mathbf{x}_{t+1}, \mathbf{x}_t)$ to perform the update

$$\mathbf{A}_{t+1} = \mathbf{A}_t - 2\gamma(\mathbf{A}_t\mathbf{x}_t - \mathbf{x}_{t+1})\mathbf{x}_t^\top,$$

where γ is the step size. Despite the efficient update, SGD suffers from the time-dependency over the data sequence, which leads to biased estimators. To observe this, unrolling the recursive update rule of SGD, we have

$$\begin{aligned}\mathbf{A}_t - \mathbf{A} = (\mathbf{A}_0 - \mathbf{A})\Pi_{s=0}^{t-1}(\mathbf{I} - 2\gamma\mathbf{x}_s\mathbf{x}_s^\top) \\ + 2\gamma\sum_{s=0}^{t-1}\mathbf{w}_s\mathbf{x}_s^\top\Pi_{l=s+1}^{t-1}(\mathbf{I} - 2\gamma\mathbf{x}_l\mathbf{x}_l^\top),\end{aligned}$$

where in the second term, the dependence of later states on previous noise realizations prevents the estimator from being unbiased even if $\mathbf{A}_0$ is initialized with a distribution where $\mathrm{E}[\mathbf{A}_0] = \mathbf{A}$. To deal with this issue, [10] develops the method SGD-RER, which applies SGD in the reverse order of the sequence to break the dependency over time. Suppose that the estimate is updated along the opposite direction of the sequence. In particular, if we have T samples and use the SGD update in the reverse order, the problematic term in the above equation takes the following form

$$2\gamma\sum_{k=0}^{t-1}\mathbf{w}_{T-k}\mathbf{x}_{T-k}^\top\Pi_{l=k+1}^{t-1}(\mathbf{I} - 2\gamma\mathbf{x}_{T-l}\mathbf{x}_{T-l}^\top).$$

whose expectation is equal to zero since the later noise realizations are independent of previous states. Evidently, this approach does not work in an online sense (as the entire sequence of T samples has to be collected first). However, we can mimic this approach by dividing the data into smaller buffers and use reverse SGD for each buffer [10].

B. Distributed SGD with Reverse Experience Replay

We extend SGD-RER to the distributed case and call it the DSGD-RER algorithm. Each agent splits the sequence of data pairs into several buffers of size B, and within each buffer all agents perform SGD in the reverse order

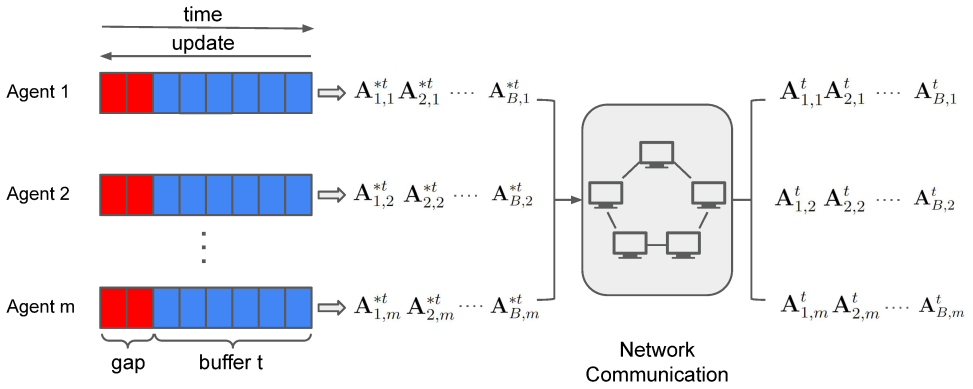

Fig. 1: The illustration of the update scheme within one buffer: each agent splits the received data sequence into buffers and applies SGD reversely within each buffer. The estimate of each agent is then updated locally through the network communication. The communication between agents is based on the network structure captured by $\mathbf{P}$.

collectively based on the network topology. Between two consecutive buffers, u data pairs are discarded in order to decrease the correlation of data in different buffers. The proposed method is outlined in Algorithm 1 and depicted in Fig. 1.

Averaged Iterate over Buffers: To further improve the convergence rate, we average the iterates. Each agent computes as the estimation output the tail-averaged estimate, i.e., the average of the last estimates of all buffers (line 11 of Algorithm 1).

Coupled Process: Despite the gap between buffers, there is still dependency across data points in various buffers, which makes the analysis challenging. To simplify the analysis, we consider the idea of coupled process [10], where for the state sequence $\{\mathbf{x}_i^k : i = 0, \dots, T\}$ received by agent k, the corresponding coupled sequence $\tilde{\mathbf{x}}_i^k$ is defined as follows.

1) For each buffer t, the starting state $\mathbf{x}_0^{k,t}$ is independently generated from π, the stationary distribution of the state.
2) The coupled process evolves according to the noise sequence $\mathbf{w}_i^{k,t}$ of the actual process, such that

$$\tilde{\mathbf{x}}_{i+1}^{k,t} = \mathbf{A}\tilde{\mathbf{x}}_i^{k,t} + \mathbf{w}_i^{k,t},\ i = 0, \dots, S-1\ (S = B + u).$$

We will see in the analysis that it is more straightforward to first compute the estimation error based on the coupled process and then quantify the excessive error introduced by replacing the coupled process with the actual one. Note that based on the dynamics of the coupled process, the excessive error is related to the gap size u.

C. Theoretical Guarantee

In this section, we present our main theoretical result. By running Algorithm 1 with specified hyper-parameters, we show that for the estimation error upper bound (of any agent), the term corresponding to the leading term in [10] can be improved by increasing the network size. There exists a (high probability) upper bound R for the norm of state, such that $\left\|\mathbf{x}_i^{k,t}\right\| \leq \sqrt{R}$, which is also one of the parameters to be tuned.

Algorithm 1 DSGD-RER

1: **Require:** number of agents m, doubly stochastic matrix $\mathbf{P} \in \mathrm{R}^{m\times m}$, step size γ, buffer size B, gap size u, time horizon T, the parameter τ.
2: **Initialize:** $\mathbf{A}_{0,k}^0$ is initialized as a zero matrix for all agents $k \in [m]$. The number of buffers $N = T/S$, where $S = B + u$.
3: **for** $t = 0, 1, \dots, N-1$ **do**
4: Each agent k collects its data sequence of buffer t, $\{\mathbf{x}_0^{k,t}, \mathbf{x}_1^{k,t}, \dots, \mathbf{x}_{S-1}^{k,t}\}$, where $\mathbf{x}_i^{k,t} := \mathbf{x}_{S\cdot t+i}^k$, and we also define $\mathbf{x}_{-i}^{k,t} := \mathbf{x}_{(S-1)-i}^{k,t}$.
5: **for** $i = 0, 1, \dots, B-1$ **do**
6: **for** $k = 1, 2 \dots, m$ **do**
7: $\mathbf{A}_{i+1,k}^{*t} = \mathbf{A}_{i,k}^t - 2\gamma(\mathbf{A}_{i,k}^t \mathbf{x}_{-i}^{k,t} - \mathbf{x}_{-(i-1)}^{k,t})(\mathbf{x}_{-i}^{k,t})^\top$
8: **end for**
9: Each agent k communicates with its neighbors based on $\mathbf{P}$ to update its estimate:

$$\mathbf{A}_{i+1,k}^t = \sum_{j\in\mathcal{N}_i} [\mathbf{P}^\tau]_{jk} \mathbf{A}_{i+1,j}^{*t}.$$

10: **end for**
11: For each agent k, compute the tail-averaged estimate until the current buffer as $\hat{\mathbf{A}}_{0,t}^k = \frac{1}{t+1}\sum_{s=0}^t \mathbf{A}_{B,k}^s$ and define $\mathbf{A}_{0,k}^{t+1} = \mathbf{A}_{B,k}^t$.
12: **end for**

Theorem 1: Let Assumptions 1-2 hold and consider the following hyper-parameter setup:

1) $d \leq O(\log(T))$.
2) $\gamma RB < \frac{1}{8}$.
3) $c_1 \gamma B \sigma_{\min}(\mathbf{G}) \leq \frac{1}{2}$, where c_1 is a problem-dependent constant.
4) $R = \Omega(C_\mu \sigma_{\max}(\mathbf{G}) \log(T))$,
$B = u = \Theta(\sqrt{\frac{T}{\log(T)}})$,
$\gamma = \Theta(\frac{1}{\sqrt{T\log T}} \frac{\log\sqrt{T\log(T)}}{\log(T)})$, and
$\tau = \Theta(\log(T))$.

Then, by applying Algorithm 1, with probability at least $1 -$

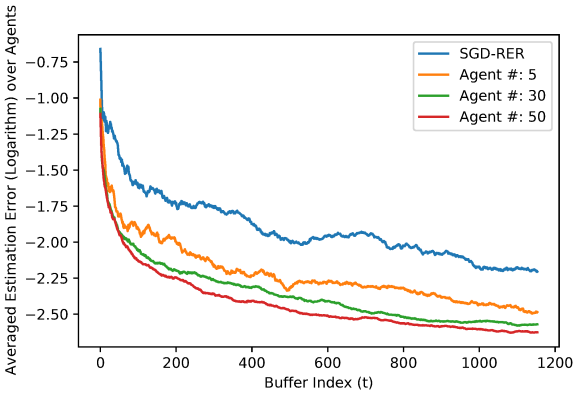

Fig. 2: The plot shows that the estimation error of DSGD-RER improves as the network size increases. Also, DSGD-RER outperforms it centralized counter part SGD-RER.

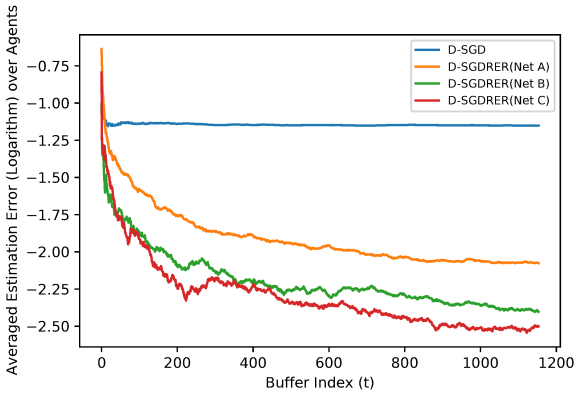

Fig. 3: The averaged estimation error over time for different network topologies. The better the network is connected, the lower the estimation error is. Vanilla D-SGD (even in a fully connected network) is not competitive due to the correlation of data points.

$(\frac{5m}{T^\rho}+\frac{3}{T^v})$ where ρ and v are some positive constants, the estimation error of the tail-averaged estimate of agent $k\in[m]$ over all buffers is bounded as follows

$$\begin{aligned}&\left\|\hat{\mathbf{A}}^k_{0,N-1}-\mathbf{A}\right\|\\ \le&2\sqrt{m}\beta^\tau\gamma RT^2+(16\gamma^2R^2T^2+8\gamma RT)\left\|\mathbf{A}^u\right\|\\ +&\sqrt{\frac{8\gamma C_\mu\sigma_{\max}(\Sigma)(c_6d+v\log T)(1+2\alpha)}{Nm}}+\frac{\zeta\left\|\mathbf{A}_0-\mathbf{A}\right\|}{N}\\ +&\frac{4\left\|\mathbf{A}_0-\mathbf{A}\right\|}{c_1\gamma\sigma_{\min}(\mathbf{G})NB}\exp\big(-\frac{c_1\gamma B\sigma_{\min}(\mathbf{G})}{2}(\lceil\zeta\rceil+1)\big),\end{aligned}\tag{1}$$

where $\zeta:=\max\{\frac{c_4d}{c_3}+\log(NT^v)/c_3-1,c_2\}$, $\alpha:=c_5(\zeta+\frac{1}{\gamma B\sigma_{\min}(\mathbf{G})})$ and $c_1,\dots,c_5$ are constants.

Remark 1: Comparing (1) and the estimation error upper bound of the centralized SGD-RER in [10], we can observe that for the dominant term in [10], the corresponding term in (1) is $\sqrt{\frac{8\gamma C_\mu\sigma_{\max}(\Sigma)(c_6d+v\log T)(1+2\alpha)}{Nm}}$, which shows that the contributed error can be improved by increasing the network size m.

IV. Numerical Experiments

Experiment Setup: We consider a network of LTI systems, where the transition matrix $\mathbf{A}\in\mathrm{R}^{d\times d}$ has the form $\mathbf{U}\Lambda\mathbf{U}^\top$. $\mathbf{U}$ is a randomly generated orthogonal matrix and Λ is a diagonal matrix with two of its diagonal entries equal to 0.9 and the rest equal to 0.3. For the step size, we set $\gamma_k=\frac{1}{2R_k}$, where R_k is estimated as the sum of the norms of the first $\lfloor 2\log T\rfloor$ samples of agent k. We also set the other hyperparameters as follows: $T=10^7$, $u=\sqrt{\frac{T}{\log T}}$, $B=10u$, $\tau=1$ and $d=5$. The starting state $\mathbf{x}^k_0=0$, and the noise follows the standard Gaussian distribution $\mathcal{N}(0,\mathbf{I})\ \forall k\in[m]$.

Networks: We examine the dependence of the estimation error to **1)** the network size and **2)** the network topology. For the former, we consider the centralized SGD-RER [10] and our distributed algorithm for 2-cyclic graph with various agent numbers $m\in\{5,30,50\}$. For the latter, we look at the performance of a 5-agent network with the following topologies: **a)** $\mathbf{P}=\mathbf{I}$ (Net A), fully disconnected; **b)** $\mathbf{P}$ captures a 2-cyclic graph, where each agent has a self-weight of 0.3 and assigns the weight of 0.35 to each of its two neighbors (Net B); **c)** $\mathbf{P}=\frac{1}{m}\mathbf{1}\mathbf{1}^\top$ (Net C), the fully connected network.

Performance: From Fig. 2, we verify the dependency of the estimation precision on the network size. For a larger network size, the error of the tail-averaged estimate is smaller. Furthermore, all decentralized schemes outperform centralized SGD-RER [10]. To see the influence of network topology, notice that the value of β for different networks is ordered as Net A $>$ Net B $>$ Net C. We can see in Fig. 3 that smaller β results in a better estimation error. Furthermore, to see the impact of reverse estimation, we also plot the estimation error of applying vanilla distributed SGD in a fully connected network. We can see that the error does not shrink over time as the estimator is biased, so the reverse update process in DSGD-RER is critical.

V. Conclusion

In this work, we considered the distributed online system identification problem for a network of identical LTI systems. We proposed a distributed online estimation algorithm and characterized the estimation error with respect to the network size and topology. We indeed observed in numerical experiments that larger network size as well as better connectivity can improve the estimation performance. For future directions, this online estimation method can be combined with decentralized control techniques to build decentralized, real-time adaptive controllers as explored in [23], [24]. Another potential direction is to extend the technique for identification of non-linear dynamical systems.

VI. Appendix

For the complete proof of claims, see the longer version of the paper [25], where all necessary lemmas are included.

A. Notations and Proof Sketch

To analyze the distributed estimation error, we decompose it into two parts: (I) The estimation error from applying

the fully-connected network $\mathbf{P}_{avg} = \frac{1}{m}\mathbf{1}\mathbf{1}^\top$ as the communication scheme; (II) The difference between estimates derived from $\mathbf{P}$ and $\mathbf{P}_{avg}$. To analyze (I), we apply the idea of coupled process: we consider $\tilde{\mathbf{A}}_{i,k}^{t,avg}$, agent k estimate based on the coupled process and $\mathbf{P}_{avg}$. From the estimate initialization and the update scheme, we know $\tilde{\mathbf{A}}_{i,1}^{t,avg} = \tilde{\mathbf{A}}_{i,2}^{t,avg} = \cdots = \tilde{\mathbf{A}}_{i,m}^{t,avg}$, which we denote as $\tilde{\mathbf{A}}_{i}^{t,avg}$, coming with the recursive update rule:

$$\tilde{\mathbf{A}}_{i+1}^{t,avg} = \tilde{\mathbf{A}}_{i}^{t,avg} - \frac{2\gamma}{m}\sum_{k=1}^{m}(\tilde{\mathbf{A}}_{i}^{t,avg}\tilde{\mathbf{x}}_{-i}^{k,t} - \tilde{\mathbf{x}}_{-(i-1)}^{k,t})\tilde{\mathbf{x}}_{-i}^{k,t\top}. \quad (2)$$

We use the following notations for the rest of the proof:

$$\tilde{\mathbf{P}}_{i}^{t,avg} := \mathbf{I} - 2\gamma\frac{\sum_{k=1}^{m}\tilde{\mathbf{x}}_{i}^{k,t}\tilde{\mathbf{x}}_{i}^{k,t\top}}{m},$$
$$\tilde{\mathbf{H}}_{i,j}^{t,avg} := \begin{cases}\prod_{s=i}^{j}\tilde{\mathbf{P}}_{-s}^{t,avg}, & i \le j\\ \mathbf{I}, & i > j\end{cases}$$

where the index $-j := (S-1)-j$, and

$$\mathbf{x}_{-i}^{k,t} := \mathbf{x}_{(S-1)-i}^{k,t}, \quad 0 \le i \le S-1,$$
$$\mathcal{D}_{-j}^{t} := \left\{\left\|\mathbf{x}_{-i}^{k,t}\right\| \le \sqrt{R} : \forall k,\ j \le i \le B-1\right\},$$
$$\tilde{\mathcal{D}}_{-j}^{t} := \left\{\left\|\tilde{\mathbf{x}}_{-i}^{k,t}\right\| \le \sqrt{R} : \forall k,\ j \le i \le B-1\right\},$$
$$\mathcal{D}^{s,t} := \cap_{r=s}^{t}\mathcal{D}_{-0}^{r},\ s \le t, \tilde{\mathcal{D}}^{s,t} := \cap_{r=s}^{t}\tilde{\mathcal{D}}_{-0}^{r},\ s \le t,$$
$$\hat{\mathcal{D}}^{s,t} := \mathcal{D}^{s,t} \cap \tilde{\mathcal{D}}^{s,t}.$$

The estimate $\tilde{\mathbf{A}}^{avg}$ consists of a bias term and a variance term, which have the following expressions (see equations 18-19 in [10]):

$$\tilde{\mathbf{A}}_{B}^{t,avg} - \mathbf{A} = (\tilde{\mathbf{A}}_{B,bias}^{t,avg} - \mathbf{A}) + \tilde{\mathbf{A}}_{B,var}^{t,avg},$$
$$(\tilde{\mathbf{A}}_{B,bias}^{t,avg} - \mathbf{A}) = (\mathbf{A}_0 - \mathbf{A})\prod_{s=0}^{t}\tilde{\mathbf{H}}_{0,B-1}^{s,avg},$$

$$\tilde{\mathbf{A}}_{B,var}^{t,avg} = 2\gamma\sum_{r=0}^{t}\sum_{j=0}^{B-1}\left(\frac{\sum_{k=1}^{m}\mathbf{w}_{-j}^{k,t-r}\tilde{\mathbf{x}}_{-j}^{k,t-r\top}}{m}\right)\tilde{\mathbf{H}}_{j+1,B-1}^{t-r,avg}\prod_{s=r-1}^{0}\tilde{\mathbf{H}}_{0,B-1}^{t-s,avg}.$$

B. Proof of the Main Theorem

First, we decompose the estimation error into several terms as follows:

$$\begin{aligned}&\left\|\hat{\mathbf{A}}_{0,N-1}^{k} - \mathbf{A}\right\|\\ \le&\left\|\hat{\mathbf{A}}_{0,N-1}^{k} - \hat{\mathbf{A}}_{0,N-1}^{avg}\right\| + \left\|\hat{\mathbf{A}}_{0,N-1}^{avg} - \hat{\tilde{\mathbf{A}}}_{0,N-1}^{avg}\right\| \quad (3)\\ +&\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{var,avg}\right\| + \left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{bias,avg} - \mathbf{A}\right\|.\end{aligned}$$

(I) For the term $\left\|\hat{\mathbf{A}}_{0,N-1}^{k} - \hat{\mathbf{A}}_{0,N-1}^{avg}\right\|$:
From Lemma 10, when the event of $\mathcal{D}^{0,N-1}$ holds, we have

$$\left\|\hat{\mathbf{A}}_{0,N-1}^{k} - \hat{\mathbf{A}}_{0,N-1}^{avg}\right\| \le \frac{1}{N}\sum_{t=0}^{N-1}\left\|\mathbf{A}_{B,k}^{t} - \mathbf{A}_{B}^{t,avg}\right\| \le 2\sqrt{m}\beta^{\tau}\gamma RT^2.$$

From Lemma 9 in [10], there exists a constant $\rho > 0$ such that if $R = \Omega(C_\mu\sigma_{\max}(\mathbf{G})\log T)$, $\mathcal{P}(\mathcal{D}^{0,N-1}) \ge 1 - \frac{m}{T^\rho}$, from which we have

$$\mathcal{P}\left(\left\|\hat{\mathbf{A}}_{0,N-1}^{k} - \hat{\mathbf{A}}_{0,N-1}^{avg}\right\| \ge 2\sqrt{m}\beta^{\tau}\gamma RT^2\right) \le \frac{m}{T^\rho}. \quad (4)$$

(II) For the term $\left\|\hat{\mathbf{A}}_{0,N-1}^{avg} - \hat{\tilde{\mathbf{A}}}_{0,N-1}^{avg}\right\|$:
Based on the expression of the tail-averaged estimate and Lemma 9, under the event of $\hat{\mathcal{D}}^{0,N-1}$, we have

$$\begin{aligned}\left\|\hat{\mathbf{A}}_{0,N-1}^{avg} - \hat{\tilde{\mathbf{A}}}_{0,N-1}^{avg}\right\| \le&\frac{1}{N}\sum_{t=0}^{N-1}\left\|\mathbf{A}_{B}^{t,avg} - \tilde{\mathbf{A}}_{B}^{t,avg}\right\|\\ \le&(16\gamma^2R^2T^2 + 8\gamma RT)\left\|\mathbf{A}^u\right\|.\end{aligned}$$

Based on Lemma 9 in [10], there exists a constant $\rho > 0$ such that if $R = \Omega(C_\mu\sigma_{\max}(\mathbf{G})\log T)$, $\mathcal{P}(\hat{\mathcal{D}}^{0,N-1}) \ge 1 - \frac{2m}{T^\rho}$. Therefore, we conclude that

$$\mathcal{P}\left(\left\|\hat{\mathbf{A}}_{0,N-1}^{avg} - \hat{\tilde{\mathbf{A}}}_{0,N-1}^{avg}\right\| \ge (16\gamma^2R^2T^2 + 8\gamma RT)\left\|\mathbf{A}^u\right\|\right) \le \frac{2m}{T^\rho}. \quad (5)$$

(III) For the term $\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{var,avg}\right\|$:
By applying Theorem 8 with the probability upper bound equal to $\frac{1}{T^v}$ where v is some positive constant, we have that under the event of $\tilde{\mathcal{M}}^{0,N-1} \cap \tilde{\mathcal{D}}^{0,N-1}$ (the definition of $\tilde{\mathcal{M}}$ can be found in (16)), with probability at least $1 - \frac{1}{T^v}$,

$$\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{var,avg}\right\| \le \sqrt{\frac{8\gamma C_\mu\sigma_{\max}(\Sigma)(c_6 d + v\log T)(1+2\alpha)}{Nm}}. \quad (6)$$

Based on Lemma 9 in [10], there exists a constant $\rho > 0$ such that if $R = \Omega(C_\mu\sigma_{\max}(\mathbf{G})\log T)$, $\mathcal{P}(\tilde{\mathcal{D}}^{0,N-1}) \ge 1 - \frac{m}{T^\rho}$. Under the event of $\tilde{\mathcal{D}}^{0,N-1}$, by setting δ in Lemma 7 properly, we have

$$\mathcal{P}(\tilde{\mathcal{M}}^{0,N-1} \cap \tilde{\mathcal{D}}^{0,N-1}) \ge 1 - (\frac{1}{T^v} + \frac{m}{T^\rho}).$$

Combining above with (6), we conclude that

$$\begin{aligned}&\mathcal{P}\left(\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{var,avg}\right\| \ge \sqrt{\frac{8\gamma C_\mu\sigma_{\max}(\Sigma)(c_6 d + v\log T)(1+2\alpha)}{Nm}}\right)\\ \le&(\frac{2}{T^v} + \frac{m}{T^\rho}).\end{aligned} \quad (7)$$

(IV) For the term $\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{bias,avg} - \mathbf{A}\right\|$:
Based on the expression of the bias term in Section VI-A, we have that

$$\begin{aligned}\left\|\hat{\tilde{\mathbf{A}}}_{0,N-1}^{bias,avg} - \mathbf{A}\right\| =&\left\|\frac{1}{N}\sum_{t=0}^{N-1}(\tilde{\mathbf{A}}_{B,bias}^{t,avg} - \mathbf{A})\right\|\\ \le&\frac{\|\mathbf{A}_0 - \mathbf{A}\|}{N}\sum_{t=0}^{N-1}\left\|\prod_{s=0}^{t}\tilde{\mathbf{H}}_{0,B-1}^{s,avg}\right\|.\end{aligned} \quad (8)$$

From the result of Lemma 6 with $\zeta := \max\{\frac{c_4 d}{c_3} + \log(\frac{N}{\delta})/c_3 - 1, c_2\}$, we have, for all $t \ge \zeta$

$$\left\|\prod_{s=0}^{t}\tilde{\mathbf{H}}_{0,B-1}^{s,avg}\right\| \le 2\big(1 - \gamma B\sigma_{\min}(\mathbf{G})\big)^{\frac{c_1}{2}(t+1)}, \quad (9)$$

with probability at least $(1-\delta)$ under the event of $\tilde{\mathcal{D}}^{0,N-1}$. Based on Lemma 6, under the event of $\tilde{\mathcal{D}}^{0,N-1}$ we also have for all $t < \zeta$,

$$\left\| \prod_{s=0}^{t} \tilde{\mathbf{H}}_{0,B-1}^{s,avg} \right\| \leq 1 \text{ almost surely.} \tag{10}$$

Based on (8), (9) and (10), we have, under the event of $\tilde{\mathcal{D}}^{0,N-1}$ with probability at least $(1-\delta)$,

$$\begin{aligned}
&\left\| \hat{\tilde{\mathbf{A}}}_{0,N-1}^{bias,avg} - \mathbf{A} \right\| \\
\leq & \frac{\|\mathbf{A}_0 - \mathbf{A}\|}{N} \left(\sum_{t=0}^{\lceil\zeta\rceil-1} \left\| \prod_{s=0}^{t} \tilde{\mathbf{H}}_{0,B-1}^{s,avg} \right\| + \sum_{t=\lceil\zeta\rceil}^{N-1} \left\| \prod_{s=0}^{t} \tilde{\mathbf{H}}_{0,B-1}^{s,avg} \right\| \right) \\
\leq & \frac{\|\mathbf{A}_0 - \mathbf{A}\|}{N} \left(\zeta + \sum_{t=\lceil\zeta\rceil}^{N-1} 2\big(1 - \gamma B \sigma_{\min}(\mathbf{G})\big)^{\frac{c_1}{2}(t+1)} \right) \\
\leq & \frac{\|\mathbf{A}_0 - \mathbf{A}\|}{N} 2\big(1 - \gamma B \sigma_{\min}(\mathbf{G})\big)^{\frac{c_1}{2}} \sum_{t=\lceil\zeta\rceil}^{N-1} \big(1 - \frac{c_1 \gamma B \sigma_{\min}(\mathbf{G})}{2}\big)^t \\
& + \frac{\zeta \|\mathbf{A}_0 - \mathbf{A}\|}{N} \\
\leq & \frac{2(1 - \frac{c_1 \gamma B \sigma_{\min}(\mathbf{G})}{2})^{\lceil\zeta\rceil}}{c_1 \gamma B \sigma_{\min}(\mathbf{G})} \frac{\|\mathbf{A}_0 - \mathbf{A}\|}{N} 2\big(1 - \gamma B \sigma_{\min}(\mathbf{G})\big)^{\frac{c_1}{2}} \\
& + \frac{\zeta \|\mathbf{A}_0 - \mathbf{A}\|}{N} \\
\leq & \frac{4 \|\mathbf{A}_0 - \mathbf{A}\|}{c_1 \gamma \sigma_{\min}(\mathbf{G}) NB} \exp\big(- \frac{c_1 \gamma B \sigma_{\min}(\mathbf{G})}{2} (\lceil\zeta\rceil + 1)\big) \\
& + \frac{\zeta \|\mathbf{A}_0 - \mathbf{A}\|}{N}
\end{aligned} \tag{11}$$

By setting δ as $\frac{1}{T^v}$, based on (11) and the fact that $\mathcal{P}(\tilde{\mathcal{D}}^{0,N-1}) \geq 1 - \frac{m}{T^\rho}$, we have

$$\begin{aligned}
&\mathcal{P}\left(\left\| \hat{\tilde{\mathbf{A}}}_{0,N-1}^{bias,avg} - \mathbf{A} \right\| \geq \frac{4 \|\mathbf{A}_0 - \mathbf{A}\|}{c_1 \gamma \sigma_{\min}(\mathbf{G}) NB} \exp\big(- \frac{c_1 \gamma B \sigma_{\min}(\mathbf{G})}{2} (\lceil\zeta\rceil + 1)\big) + \frac{\zeta \|\mathbf{A}_0 - \mathbf{A}\|}{N} \right) \\
&\leq (\frac{1}{T^v} + \frac{m}{T^\rho}).
\end{aligned} \tag{12}$$

Applying the results of (4), (5), (7) and (12) on (3), with probability at least $1 - (\frac{5m}{T^\rho} + \frac{3}{T^v})$ we have

$$\begin{aligned}
&\left\| \hat{\mathbf{A}}_{0,N-1}^{k} - \mathbf{A} \right\| \\
\leq & 2\sqrt{m} \beta^\tau \gamma R T^2 + (16\gamma^2 R^2 T^2 + 8\gamma R T) \|\mathbf{A}^u\| \\
& + \sqrt{\frac{8\gamma C_\mu \sigma_{\max}(\Sigma)(c_6 d + v \log T)(1+2\alpha)}{Nm}} + \frac{\zeta \|\mathbf{A}_0 - \mathbf{A}\|}{N} \\
& + \frac{4 \|\mathbf{A}_0 - \mathbf{A}\|}{c_1 \gamma \sigma_{\min}(\mathbf{G}) NB} \exp\big(- \frac{c_1 \gamma B \sigma_{\min}(\mathbf{G})}{2} (\lceil\zeta\rceil + 1)\big).
\end{aligned} \tag{13}$$

REFERENCES

[1] K. J. Åström and P. Eykhoff, "System identification—a survey," *Automatica*, vol. 7, no. 2, pp. 123–162, 1971.

[2] L. Ljung, "System identification," *Wiley encyclopedia of electrical and electronics engineering*, pp. 1–19, 1999.

[3] H.-F. Chen and L. Guo, *Identification and stochastic adaptive control*. Springer Science & Business Media, 2012.

[4] G. C. Goodwin, G. GC, and P. RL, "Dynamic system identification. experiment design and data analysis." 1977.

[5] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *Foundations of Computational Mathematics*, pp. 1–47, 2019.

[6] Y. Jedra and A. Proutiere, "Finite-time identification of stable linear systems optimality of the least-squares estimator," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 996–1001.

[7] M. K. S. Faradonbeh, A. Tewari, and G. Michailidis, "Finite time identification in unstable linear systems," *Automatica*, vol. 96, pp. 342–353, 2018.

[8] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Conference On Learning Theory*. PMLR, 2018, pp. 439–473.

[9] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," in *International Conference on Machine Learning*. PMLR, 2019, pp. 5610–5618.

[10] S. Kowshik, D. Nagaraj, P. Jain, and P. Netrapalli, "Streaming linear system identification with reverse experience replay," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[11] Y. Jedra and A. Proutiere, "Sample complexity lower bounds for linear system identification," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2676–2681.

[12] S. Oymak and N. Ozay, "Non-asymptotic identification of lti systems from a single trajectory," in *American control conference (ACC)*, 2019, pp. 5655–5661.

[13] T. Sarkar, A. Rakhlin, and M. A. Dahleh, "Finite time lti system identification," *Journal of Machine Learning Research*, vol. 22, pp. 1–61, 2021.

[14] A. Tsiamis and G. J. Pappas, "Finite sample analysis of stochastic system identification," in *IEEE Conference on Decision and Control (CDC)*, 2019, pp. 3648–3654.

[15] M. Simchowitz, R. Boczar, and B. Recht, "Learning linear dynamical systems with semi-parametric least squares," in *Conference on Learning Theory (COLT)*. PMLR, 2019, pp. 2714–2802.

[16] Y. Zheng and N. Li, "Non-asymptotic identification of linear dynamical systems using multiple trajectories," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1693–1698, 2020.

[17] S. Fattahi, "Learning partially observed linear dynamical systems from logarithmic number of samples," in *Learning for Dynamics and Control (L4DC)*. PMLR, 2021, pp. 60–72.

[18] Y. Sattar and S. Oymak, "Non-asymptotic and accurate learning of nonlinear dynamical systems," *arXiv preprint arXiv:2002.08538*, 2020.

[19] D. Foster, T. Sarkar, and A. Rakhlin, "Learning nonlinear dynamical systems from a single trajectory," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 851–861.

[20] S. Kowshik, D. Nagaraj, P. Jain, and P. Netrapalli, "Near-optimal offline and streaming algorithms for learning non-linear dynamical systems," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[21] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Online least squares estimation with self-normalized processes: An application to bandit problems," *arXiv preprint arXiv:1102.2670*, 2011.

[22] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[23] T.-J. Chang and S. Shahrampour, "Distributed online linear quadratic control for linear time-invariant systems," in *American Control Conference (ACC)*, 2021, pp. 923–928.

[24] ——, "Regret analysis of distributed online LQR control for unknown LTI systems," *arXiv preprint arXiv:2105.07310*, 2021.

[25] ——, "Distributed online system identification for lti systems using reverse experience replay," *arXiv preprint arXiv:2207.01062*, 2022.