# Central Limit Theorem for Two-Timescale Stochastic Approximation with Markovian Noise: Theory and Applications

**Jie Hu**
North Carolina State University

**Vishwaraj Doshi**
IQVIA Inc.

**Do Young Eun**
North Carolina State University

## Abstract

Two-timescale stochastic approximation (TTSA) is among the most general frameworks for iterative stochastic algorithms. This includes well-known stochastic optimization methods such as SGD variants and those designed for bilevel or minimax problems, as well as reinforcement learning like the family of gradient-based temporal difference (GTD) algorithms. In this paper, we conduct an in-depth asymptotic analysis of TTSA under controlled Markovian noise via central limit theorem (CLT), uncovering the coupled dynamics of TTSA influenced by the underlying Markov chain, which has not been addressed by previous CLT results of TTSA only with Martingale difference noise. Building upon our CLT, we expand its application horizon of efficient sampling strategies from vanilla SGD to a wider TTSA context in distributed learning, thus broadening the scope of Hu et al. (2022). In addition, we leverage our CLT result to deduce the statistical properties of GTD algorithms with nonlinear function approximation using Markovian samples and show their identical asymptotic performance, a perspective not evident from current finite-time bounds.

## 1 INTRODUCTION

Two-timescale stochastic approximation (TTSA) serves as a cornerstone algorithm for identifying the root $(\mathbf{x}^*, \mathbf{y}^*)$ of two coupled functions, i.e.,

$$\begin{aligned} \bar{h}_1(\mathbf{x}^*, \mathbf{y}^*) &\triangleq \mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_1(\mathbf{x}^*, \mathbf{y}^*, \xi)] = 0, \\ \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*) &\triangleq \mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_2(\mathbf{x}^*, \mathbf{y}^*, \xi)] = 0, \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}$ is a probability vector and typically only noisy observations $h_1(\mathbf{x}, \mathbf{y}, \xi), h_2(\mathbf{x}, \mathbf{y}, \xi)$ are accessible (Kushner and Yin, 2003; Borkar, 2022). If either of the two functions $h_1, h_2$ is decoupled, e.g., $h_1(\mathbf{x}, \mathbf{y}, \xi) \equiv h_1(\mathbf{x}, \xi)$, TTSA degenerates into stochastic approximation (SA) as a special case, which itself has a wide range of applications, including, but not limited to, stochastic optimization (Bottou et al., 2018; Gower et al., 2019), reinforcement learning (RL) (Srikant and Ying, 2019; Dalal et al., 2020; Patil et al., 2023), and adaptive Markov chain Monte Carlo (MCMC) (Benaim et al., 2012; Avrachenkov et al., 2021; Doshi et al., 2023). In this paper, our primary focus is the analysis of the asymptotic behavior exhibited by a general nonlinear TTSA with Markovian noise, establishing a central limit theorem (CLT) to explore the effect of coupled variables $(\mathbf{x}, \mathbf{y})$. By leveraging this CLT, we address two applications: improvement of asymptotic performance in optimization algorithms, and the derivation of statistical property from a family of gradient-based TD (GTD) algorithms in RL.

The recursion of the TTSA algorithm considered in this work is described as follows:

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1} h_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}), \\ \mathbf{y}_{n+1} = \mathbf{y}_n + \gamma_{n+1} h_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}), \end{cases} \tag{2}$$

where $\beta_n, \gamma_n$ are decreasing step sizes at different rates,[1] $\{\xi_n\}_{n\geq 0}$ is a random sequence over a finite set $\Xi$. For instance, in stochastic bilevel optimizations, Hong et al. (2023) deploys TTSA to simultaneously optimize both primal and dual variables. Likewise, the work by Lin et al. (2020) highlights the applicability of TTSA in solving minimax problems for optimizing two competing objectives. In RL, a family of GTD algorithms utilize the two-timescale structure (Sutton et al., 2009; Dalal et al., 2018, 2020; Li

[1] For example, when the step size $\beta_n$ is much smaller than $\gamma_n$, i.e., $\beta_n = o(\gamma_n)$, iterates $\mathbf{x}_n$ converges slower than iterates $\mathbf{y}_n$, thereby $\mathbf{x}_n$ is on the slow timescale and $\mathbf{y}_n$ is on the fast timescale.

Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238. 

Table 1: Overview of TTSA literature. **Loc. Lipschitz**: locally Lipschitz; **high-prob. bound**: high-probability bound; **Mart. diff.**: Martingale difference noise; **exo. MC**: exogenous Markov chain, independent of TTSA iterates $(\mathbf{x}, \mathbf{y})$; **ctrl. MC**: controlled Markov chain, where the transition kernel is determined by iterates $(\mathbf{x}, \mathbf{y})$. Except for a.s. convergence, all other result types inherently include a.s. convergence.

| Existing Works | Result Type | Noise Type | Loc. Lipschitz | Nonlinear |
|---|---|---|---|---|
| Konda and Tsitsiklis (2004) | CLT | Mart. diff. | × | × |
| Mokkadem and Pelletier (2006) | CLT | Mart. diff. | ✓ | ✓ |
| Dalal et al. (2018) | high-prob. bound | Mart. diff. | × | × |
| Borkar and Pattathil (2018) | high-prob. bound | Mart. diff. | × | ✓ |
| Doan (2022, 2024); Hong et al. (2023) | finite-time bound | Mart. diff. | × | ✓ |
| Karmakar and Bhatnagar (2018) | a.s. convergence | ctrl. MC | × | ✓ |
| Yaji and Bhatnagar (2020) | a.s. convergence | ctrl. MC | ✓ | ✓ |
| Gupta et al. (2019); Haque et al. (2023) | finite-time bound | exo. MC | × | × |
| Doan (2021b) | finite-time bound | exo. MC | × | ✓ |
| Khodadadian et al. (2022) | finite-time bound | ctrl. MC | × | × |
| Barakat et al. (2022) | finite-time bound | ctrl. MC | × | × |
| Zeng et al. (2021) | finite-time bound | ctrl. MC | × | ✓ |
| Our Work | CLT | ctrl. MC | ✓ | ✓ |

et al., 2023a). Specifically, in these algorithms, the primary value function estimates update on slower timescale, while auxiliary variables or correction terms update on faster timescale. Furthermore, modern energy systems, such as power systems and smart grids, use TTSA for dynamic decision making (Lopez-Ramos et al., 2017; Yang et al., 2019). In the realm of game theory, a noteworthy application is Generative Adversarial Networks, where the game between a generator and a discriminator can be tackled using TTSA (Prasad et al., 2015; Heusel et al., 2017).

In this paper, we focus on the Markovian sequence $\{\xi_n\}$, which plays an important role in the TTSA algorithm and is inherent in many applications.[2] In distributed learning, token algorithms utilize a random walk, enabling tokens to traverse distributed agents over a graph, each possessing local datasets, and iteratively update model parameters, thus facilitating collaborative stochastic optimization across agents (Hu et al., 2022; Triastcyn et al., 2022; Hendrikx, 2023; Even, 2023). Apart from employing SGD iterates to minimize an objective function, such token algorithms of the form (2) can also address distributed bilevel or minimax problems that have been recently studied in Gao (2022); Gao et al. (2023). Meanwhile, in RL, the environment itself is modeled as a Markov Decision Process (MDP), which by design incorporates Markovian properties. When an agent interacts with this environment, the trajectory $\{\xi_n\}$ it follows, i.e., a sequence of states, actions, and rewards, is inherently Markovian. Notably, this Markovian sequence can be influenced by the agent's adaptive policy, as seen in actor-critic algorithms, yielding a *controlled* Markov chain dependent on the iterates $(\mathbf{x}_n, \mathbf{y}_n)$ (Karmakar and Bhatnagar, 2018; Yaji and Bhatnagar, 2020; Zeng et al., 2021). These examples underscore the importance of the Markovian sequence in the development of both theoretical frameworks and practical implementations of various learning algorithms.

[2] While the noise sequence $\{\xi_n\}$ is common to both recursions in the TTSA algorithm (2), it allows for two distinct Markov chains for each recursion. Further details can be found in Section 2.1.

### 1.1 Related Works

**Finite-time vs Asymptotic Analysis:** The convergence properties of SA have been studied extensively using both asymptotic (Kushner and Yin, 2003; Fort, 2015; Borkar, 2022; Li et al., 2023b) and finite-time (Srikant and Ying, 2019; Karimi et al., 2019; Chen et al., 2022) analyses. While recent trends have shown a preference for non-asymptotic analysis, discussions in Meyn (2022, Chapter 1.2) point out the often-underestimated significance of asymptotic statistics. This notion is highlighted in Mou et al. (2020); Chen et al. (2020); Srikant (2024), which demonstrate the broader applicability of CLT beyond purely asymptotic contexts. Specifically, the limiting covariance matrix, central to the CLT, finds its presence in high-probability bounds (Mou et al., 2020), and in finite-time bounds on mean square error (Chen et al., 2020) as well as 1-Wasserstein distance to measure the rate of convergence to normality (Srikant, 2024). Further underscoring its significance, Hu et al. (2022) showcases its accuracy in capturing the rate of conver-

gence compared to the mixing rate of the underlying Markov chain, frequently employed in finite-time analysis (Karimi et al., 2019; Chen et al., 2022).

**TTSA with Martingale Difference Noise:** For the TTSA algorithm (2), the stochastic sequence being an *i.i.d.* sequence $\{\xi_n\}$ allows for the decomposition of the noisy observation $h_1(\mathbf{x}, \mathbf{y}, \xi_n)$ into $\bar{h}_1(\mathbf{x}, \mathbf{y})$ and a Martingale difference noise term $h_1(\mathbf{x}, \mathbf{y}, \xi_n) – \bar{h}_1(\mathbf{x}, \mathbf{y})$; a similar decomposition applying to $h_2$. In the case of Martingale difference noise, an extensive body of research focuses on the analysis of CLT results (Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006), high-probability bounds (Dalal et al., 2018; Borkar and Pattathil, 2018), and finite-time bounds (Doan, 2022, 2024; Hong et al., 2023) for both linear and nonlinear TTSA, as shown in Table 1.

**TTSA with Markovian Noise — Asymptotic Results and Suboptimal Finite-Time Bounds:** Recently, increasing attention has been shifted towards analyzing TTSA with Markovian noise sequences $\{\xi_n\}$, which introduces technical challenges due to inherent bias in $h_i(\mathbf{x}, \mathbf{y}, \xi_n)$ as an estimator of $\bar{h}_i(\mathbf{x}, \mathbf{y})$ for $i = 1, 2$. Karmakar and Bhatnagar (2018); Yaji and Bhatnagar (2020) delve into the almost sure convergence of nonlinear TTSA with Markovian noise, showing that the two iterates $\mathbf{x}_n, \mathbf{y}_n$ asymptotically estimate the related differential inclusions, which are a generalized version of ordinary differential equations (ODEs). Yaji and Bhatnagar (2020) further relax to the locally Lipschitz functions $h_1, h_2$, which is commonly seen in the machine learning literature such as low-rank matrix recovery (Recht et al., 2010), tensor factorization problem (Kolda and Bader, 2009), and deep neural networks with unbounded Hessian matrices (Zhang and Hong, 2020, Appendix H).

Meanwhile, the mixing rate properties of Markov chains have been predominantly utilized for the finite-time analysis of both linear (Gupta et al., 2019; Kaledin et al., 2020; Doan, 2021a; Khodadadian et al., 2022; Barakat et al., 2022; Haque et al., 2023) and nonlinear TTSA (Doan, 2021b; Zeng et al., 2021) with Markovian noise.[3] Notably, the latter two works align closely with our TTSA settings. However, Doan (2021b) only provided a finite-time bound for the combined error of both iterations, i.e., $\mathbb{E}[\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \frac{\beta_n}{\gamma_n}\|\mathbf{y}_n - \mathbf{y}^*\|^2]$ at a suboptimal rate of $O(n^{-2/3})$ with a specific choice of step sizes $\beta_n = (n+1)^{-1}$ and $\gamma_n = (n+1)^{-2/3}$, while we show in Section 2.3 that for large $n$, the combined error should approximately decrease to zero at the speed of $O(\beta_n) = O(n^{-1})$. A similar bound for $\mathbb{E}[\|\mathbf{x}_n - \mathbf{x}^*\|^2]$ under the more general controlled Markov noise setting is provided in Zeng et al. (2021) at the suboptimal rate of $O(n^{-2/3})$ with the same choice of step sizes. Thus, even the state-of-the-art finite-time bounds in Doan (2021b); Zeng et al. (2021) do not preciously capture the leading term that determines the performance of each iterates $\mathbf{x}_n, \mathbf{y}_n$. A comprehensive non-asymptotic analysis with rate matching the CLT scale (i.e., $O(\beta_n), O(\gamma_n)$) has yet to be performed in the nonlinear TTSA with controlled Markovian noise under general decreasing step sizes $\beta_n, \gamma_n$.

[3]While nonlinear TTSA with Markovian noise is currently the most general framework, our emphasis is not solely on generalization. As we will demonstrate in Section 3, this setting has substantive implications in both stochastic optimization and RL.

### 1.2 Our Contributions

In this paper, we study the CLT of both iterates $\mathbf{x}_n$ and $\mathbf{y}_n$ in nonlinear TTSA with controlled Markovian noise, where $h_1, h_2$ are only locally Lipschitz continuous. Although Yaji and Bhatnagar (2020) considered more general set-valued functions $h_1, h_2$, they only obtained almost sure convergence. In contrast, we here target single-valued functions that are more common in the machine learning literature and extend the scope to include CLT results. Our work further generalizes the CLT analysis of the two-timescale framework in Mokkadem and Pelletier (2006) - still a state-of-the-art CLT result for Martingale difference noise - by necessitating a deeper exploration into the Markovian noise $\{\xi_n\}_{n\geq 0}$, given that $h_i(\mathbf{x}, \mathbf{y}, \xi_n) - \bar{h}_i(\mathbf{x}, \mathbf{y}), i = 1, 2$ are no longer Martingale difference.

Utilizing our CLT results, we demonstrate the impact of sampling strategies on the limiting covariance across a wide class of distributed optimization algorithms. Extending beyond the vanilla SGD setting studied in Hu et al. (2022), we show that improved sampling strategies lead to better performance for general TTSA including, but not restricted to, SGD variants and algorithms tailored for stochastic bilevel and minimax problems. Moreover, in the RL context, we introduce first of its kind statistical characterization of GTD2 and TDC algorithms with nonlinear function approximation (Maei et al., 2009) using Markovian samples. Using both theoretical and empirical results, we show that their asymptotic performance coincides, as evidenced by identical covariance matrix in our CLT. Such conclusions are not possible via current finite-time bounds (Doan, 2021b; Zeng et al., 2021).

**Notations.** We use $\|\cdot\|$ to denote both the Euclidean norm of vectors and the spectral norm of matrices. Two symmetric matrices $\mathbf{M}_1, \mathbf{M}_2$ follow Loewner ordering $\mathbf{M}_1 >_L \mathbf{M}_2$ (resp. '$\geq_L$') if $\mathbf{M}_1 - \mathbf{M}_2$ is positive definite (resp. positive semi-definite). A matrix is *Hurwitz* if all its eigenvalues possess strictly negative real parts. The function $\mathbb{1}_{(\cdot)}$ is an indicator func-

tion. $\nabla_{\mathbf{x}} h(\mathbf{x}, \mathbf{y})$ stands for the Jacobian matrix of the vector-valued function $h(\mathbf{x}, \mathbf{y})$ with respect to the variable $\mathbf{x}$. $C^1$ function $f$ means that function $f$ is both continuous and differentiable. We use '$\xrightarrow{d}$' for the convergence in distribution and $N(0, \mathbf{V})$ is the Gaussian random vector with covariance matrix $\mathbf{V}$.

## 2 MAIN RESULTS

In this section, we analyze the asymptotic behavior of the TTSA algorithm (2) with Markovian noise. First, we provide assumptions and the almost sure convergence result in Section 2.1. Before presenting our main CLT result in Section 2.3, we explain how our result is achieved by transforming the TTSA iteration into a single-timescale SA-like recursion, and introduce some key components related to asymptotic covariance of the iterates. This transformation resembles that in Konda and Tsitsiklis (2004); Mokkadem and Pelletier (2006) but with a fresh perspective by accounting for biased errors due to Markovian noise, as elaborated upon in Section 2.2.

### 2.1 Key Assumptions and a.s. Convergence

A1. The step sizes $\beta_n \triangleq (n+1)^{-b}$ and $\gamma_n \triangleq (n+1)^{-a}$, where $0.5 < a < b \leq 1$.

A2. For the $C^1$ function $h_1 : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \Xi \to \mathbb{R}^{d_1}$, there exists a positive constant $L_1$ such that $\|h_1(\mathbf{x}, \mathbf{y}, \xi)\| \leq L_1(1 + \|\mathbf{x}\| + \|\mathbf{y}\|)$ for every $\mathbf{x} \in \mathbb{R}^{d_1}, \mathbf{y} \in \mathbb{R}^{d_2}, \xi \in \Xi$. The same condition holds for the $C^1$ function $h_2$ as well.

A3. Consider a $C^1$ function $\lambda : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$. For every $\mathbf{x} \in \mathbb{R}^{d_1}$, the following three properties hold: (i) $\lambda(\mathbf{x})$ is the globally attracting point of the related ODE $\dot{\mathbf{y}} = \bar{h}_2(\mathbf{x}, \mathbf{y})$; (ii) $\nabla_{\mathbf{y}} \bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))$ is Hurwitz; (iii) $\|\lambda(\mathbf{x})\| \leq L_2(1 + \|\mathbf{x}\|)$ for some positive constant $L_2$. Additionally, let $\hat{h}_1(\mathbf{x}) \triangleq \bar{h}_1(\mathbf{x}, \lambda(\mathbf{x}))$, there exists a set of disjoint roots $\Lambda \triangleq \{\mathbf{x}^* : \hat{h}_1(\mathbf{x}^*) = 0, \nabla_{\mathbf{x}} \hat{h}_1(\mathbf{x}^*) + \frac{\mathbb{1}_{\{b=1\}}}{2} \mathbf{I} \text{ is Hurwitz}\}$, which is also the globally attracting set for trajectories of the related ODE $\dot{\mathbf{x}} = \hat{h}_1(\mathbf{x})$.

A4. $\{\xi_n\}_{n \geq 0}$ is an iterate-dependent Markov chain on finite state space $\Xi$. For every $n \geq 0$, $P(\xi_{n+1} = j | \mathbf{x}_m, \mathbf{y}_m, \xi_m, 0 \leq m \leq n) = P(\xi_{n+1} = j | \mathbf{x}_n, \mathbf{y}_n, \xi_n = i) = \mathbf{P}_{i,j}[\mathbf{x}_n, \mathbf{y}_n]$, where the transition kernel $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is continuous in $\mathbf{x}, \mathbf{y}$, and the Markov chain generated by $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is ergodic so that it admits a stationary distribution $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$, and $\boldsymbol{\pi}(\mathbf{x}^*, \lambda(\mathbf{x}^*)) = \boldsymbol{\mu}$.

A5. $\sup_{n \geq 0}(\|\mathbf{x}_n\| + \|\mathbf{y}_n\|) < \infty$ a.s.

In Assumption (A1), the step sizes $\beta_n, \gamma_n$ decay polynomially at distinct rates, i.e., $\beta_n = o(\gamma_n)$, which is standard in the TTSA literature (Zeng et al., 2021; Doan, 2021b; Hong et al., 2023). Assumption (A2) ensures that $C^1$ functions $h_1, h_2$ are locally Lipschitz and grow at most linearly with respect to the norms of their parameters, as also assumed in Yaji and Bhatnagar (2020). This is a far less stringent condition compared to the globally Lipschitz assumption used in most of the recent works, as listed in Table 1.

Assumption (A3) is crucial for the analysis of iterates $(\mathbf{x}_n, \mathbf{y}_n)$, which can be seen as a stochastic discretization of the ODEs $\dot{\mathbf{x}} = \hat{h}_1(\mathbf{x})$ and $\dot{\mathbf{y}} = \bar{h}_2(\mathbf{x}, \mathbf{y})$. This assumption guarantees the global asymptotic stability of these two ODEs, as demonstrated in Yaji and Bhatnagar (2020); Doan (2021b). The linear growth of $\lambda(\mathbf{x})$ is a milder condition than the globally Lipschitz assumption in Borkar and Pattathil (2018); Karmakar and Bhatnagar (2018); Zeng et al. (2021); Doan (2021b).

Assumption (A4) is standard to guarantee the asymptotic unbiasedness of $h_1, h_2$ in the existing literature on TTSA with Markovian noise (Karmakar and Bhatnagar, 2018; Yaji and Bhatnagar, 2020; Khodadadian et al., 2022; Barakat et al., 2022). It is worth noting that $\{\xi_n\}$ naturally allows for an augmentation of the form $\xi_n \triangleq (X_n, Y_n)$, with two independent Markovian noise sequences $\{X_n\}$, $\{Y_n\}$ corresponding to iterates $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$, respectively. In this case, the functions $h_1$ and $h_2$ act only on the entries of $\xi_n$ related to $X_n$ and $Y_n$.

Assumption (A5) assumes the a.s. boundedness of the coupled iterates $(\mathbf{x}_n, \mathbf{y}_n)$, which is commonly seen in the TTSA literature (Karmakar and Bhatnagar, 2018; Yaji and Bhatnagar, 2020). A similar stability condition is also found in the SA literature (Delyon et al., 1999; Borkar, 2022; Li et al., 2023b). In practice, to stabilize the TTSA algorithm (2) under Markovian noise, one could adopt algorithmic modifications from the SA literature, including the projection method onto (possibly expanding) compact sets (Chen, 2006; Andrieu and Vihola, 2014) or the truncation method with a restart process (Fort, 2015; Fort et al., 2016).

**Lemma 2.1** (Almost Sure Convergence)**.** *Under Assumptions (A1) - (A5), iterates* $(\mathbf{x}_n, \mathbf{y}_n)$ *in* (2) *almost surely converge to a set of roots, i.e.,* $(\mathbf{x}_n, \mathbf{y}_n) \to \bigcup_{\mathbf{x}^* \in \Lambda} (\mathbf{x}^*, \lambda(\mathbf{x}^*))$ *a.s.*

Lemma 2.1 follows from Yaji and Bhatnagar (2020, Theorem 4) by verifying the conditions therein and we defer the details to Appendix *A*.1. While they studied broader set-valued functions $h_1, h_2$ within the realm of stochastic recursive inclusion, they did not explore the CLT result. This is likely due to existing gaps in the

CLT analysis even for single-timescale stochastic recursive inclusion, as mentioned in Borkar (2022, Chapter 5). In contrast, we focus on single-valued functions $h_1, h_2$, as prevalent in the machine learning literature. This paves the way for the first CLT result, Theorem 2.2, for the general TTSA with *controlled* Markovian noise, as demonstrated in Table 1. In the following section, we will conduct a more detailed analysis of the asymptotic behavior of iterates $(\mathbf{x}_n, \mathbf{y}_n)$ near equilibrium $(\mathbf{x}^*, \lambda(\mathbf{x}^*))$ for some $\mathbf{x}^* \in \Lambda$.

### 2.2 Overview of the CLT Analysis for $(\mathbf{x}_n, \mathbf{y}_n)$

Assumption (A1) puts $\{\mathbf{y}_n\}_{n\geq 0}$ on a 'faster timescale' compared to $\{\mathbf{x}_n\}_{n\geq 0}$, and has implications on convergence rates of the two sequences. Under additional conditions on the function $\bar{h}_2(\cdot,\cdot)$ in Assumption (A3), the sequence $\{\mathbf{y}_n\}$ can be approximated by $\{\lambda(\mathbf{x}_n)\}$ for large time step $n$, where $\lambda(\mathbf{x})$ is an implicit function solving $\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) = 0$. Loosely speaking, when $n$ is large enough, the fast iterates $\mathbf{y}_n$ are nearly convergent to the root $\lambda(\mathbf{x}_n)$ of $\bar{h}_2(\mathbf{x}_n, \cdot)$. Iterates $\mathbf{x}_n$ on the slower timescale then guide the roots $\lambda(\mathbf{x}_n)$ of the iterates $\mathbf{y}_n$ until they reach $\mathbf{y}^* = \lambda(\mathbf{x}^*)$, which also satisfies $\bar{h}_1(\mathbf{x}^*, \lambda(\mathbf{x}^*)) = 0$. Consequently, resembling Konda and Tsitsiklis (2004, Section 2) and Mokkadem and Pelletier (2006, Section 2.3), we can show that $\{\mathbf{x}_n\}$ is now approximated by iterating a *single-timescale* SA update rule, independent of $\{\mathbf{y}_n\}$ but instead driven by $\{\lambda(\mathbf{x}_n)\}$, whose derivation we detail in what follows. For $i \in \{1, 2\}$, define

$$\begin{aligned}
&\mathbf{Q}_{i1} \triangleq \nabla_{\mathbf{x}} \bar{h}_i(\mathbf{x}, \mathbf{y})\big|_{(\mathbf{x},\mathbf{y})=(\mathbf{x}^*,\, \mathbf{y}^*)},\\
&\mathbf{Q}_{i2} \triangleq \nabla_{\mathbf{y}} \bar{h}_i(\mathbf{x}, \mathbf{y})\big|_{(\mathbf{x},\mathbf{y})=(\mathbf{x}^*,\, \mathbf{y}^*)},\\
&\Delta_n^{(i)} \triangleq h_i(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) - \bar{h}_i(\mathbf{x}_n, \mathbf{y}_n),\\
&\tilde{\Delta}_n^{(i)} \triangleq h_i(\mathbf{x}_n, \lambda(\mathbf{x}_n), \xi_{n+1}) - \bar{h}_i(\mathbf{x}_n, \lambda(\mathbf{x}_n)).
\end{aligned}$$

Adding and subtracting $\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n)$ and $\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n)$ to iterates $\mathbf{x}_n$ and $\mathbf{y}_n$ in (2) respectively, and taking their Taylor expansions at $(\mathbf{x}_n, \mathbf{y}_n) = (\mathbf{x}^*, \mathbf{y}^*)$, gives us

$$\begin{aligned}
\mathbf{x}_{n+1} = \mathbf{x}_n &+ \beta_{n+1}(\mathbf{Q}_{11}(\mathbf{x}_n - \mathbf{x}^*) + \mathbf{Q}_{12}(\mathbf{y}_n - \mathbf{y}^*))\\
&+ \beta_{n+1}\Delta_n^{(1)} + \beta_{n+1} O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2),
\end{aligned} \tag{3}$$

$$\begin{aligned}
\mathbf{y}_{n+1} = \mathbf{y}_n &+ \gamma_{n+1}(\mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*) + \mathbf{Q}_{22}(\mathbf{y}_n - \mathbf{y}^*))\\
&+ \gamma_{n+1}\Delta_n^{(2)} + \gamma_{n+1} O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2).
\end{aligned} \tag{4}$$

Re-arranging (4) by placing the $\mathbf{y}_n - \mathbf{y}^*$ on the left-hand side yields

$$\begin{aligned}
\mathbf{y}_n - \mathbf{y}^* = \gamma_{n+1}^{-1}\mathbf{Q}_{22}^{-1}(\mathbf{y}_{n+1} - \mathbf{y}_n) - \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*)\\
+ \mathbf{Q}_{22}^{-1}\Delta_n^{(2)} + O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2).
\end{aligned} \tag{5}$$

By substituting the above into (3), and then replacing (approximating) $\mathbf{y}_n$ with $\lambda(\mathbf{x}_n)$, we get

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1}\mathbf{K}_{\mathbf{x}}(\mathbf{x}_n - \mathbf{x}^*) + \beta_{n+1}\tilde{\Delta}_n^{\mathbf{x}} + \beta_{n+1}R_n, \tag{6}$$

where

$$\mathbf{K}_{\mathbf{x}} \triangleq \mathbf{Q}_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}, \;\; \tilde{\Delta}_n^{\mathbf{x}} \triangleq \tilde{\Delta}_n^{(1)} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\tilde{\Delta}_n^{(2)}, \tag{7}$$

and $R_n$ is comprised of residual errors from the earlier Taylor expansion and approximation of $\mathbf{y}_n$ by $\lambda(\mathbf{x}_n)$. The term $\tilde{\Delta}_n^{\mathbf{x}}$ can be further decomposed using the Poisson equation technique (Benveniste et al., 2012; Meyn, 2022) as

$$\begin{aligned}
\tilde{\Delta}_n^{\mathbf{x}} = &[M_{n+1}^{(1)} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}M_{n+1}^{(2)}]\\
&+ [\tilde{H}(\mathbf{x}_n, \lambda(\mathbf{x}_n), \xi_{n+1}) - \tilde{H}(\mathbf{x}_n, \lambda(\mathbf{x}_n), \xi_n)],
\end{aligned}$$

where $M_{n+1}^{(1)}$ and $M_{n+1}^{(2)}$ are Martingale difference terms adapted to filtration $\mathcal{F}_n \triangleq \sigma(\mathbf{x}_0, \mathbf{y}_0, \xi_0, \cdots, \xi_n)$. The exact expressions for the Martingale difference terms can be found in Appendix *A*.2.1, equations 9(a) and 9(b). The second summand including the $\tilde{H}$ terms, whose exact expression is provided in Appendix *A*.2.1, involves consecutive Markovian noise terms $\xi_{n+1}$ and $\xi_n$ which are responsible for biased errors in the iteration for $\mathbf{x}_n$. These additional terms are not present in existing works that focus only on *i.i.d.* stochastic inputs (Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006), even though their analysis leads to equations similar to (6). In Appendix *A*.2.2, we show that the $\tilde{H}$ terms along with residual errors $R_n$ at each step are $o(\sqrt{\beta_n})$, and thus do not influence the CLT result for iterates $\mathbf{x}_n$ of the slower timescale.

Consequently, the approximation $\mathbf{y}_n = \lambda(\mathbf{x}_n)$, together with the aforementioned analysis leading to $o(\sqrt{\beta_n})$, now allows us to analyze (6) as essentially a single-timescale SA with Markovian noise. We then apply Fort (2015, Proposition 4.1) to extract a CLT result, i.e., we prove that

$$\beta_n^{-1/2}(\mathbf{x}_n - \mathbf{x}^*) \xrightarrow{d} N(0, \mathbf{V}_{\mathbf{x}}), \tag{8}$$

where $\mathbf{V}_{\mathbf{x}}$ solves the Lyapunov equation $\mathbf{U}_{\mathbf{x}} + (\mathbf{K}_{\mathbf{x}} + \frac{\mathbb{1}_{\{\beta_n = O(1/n)\}}}{2}\mathbf{I})\mathbf{V}_{\mathbf{x}} + \mathbf{V}_{\mathbf{x}}(\mathbf{K}_{\mathbf{x}} + \frac{\mathbb{1}_{\{\beta_n = O(1/n)\}}}{2}\mathbf{I})^T = 0$,

$$\mathbf{U}_{\mathbf{x}} \triangleq \lim_{s\to\infty} \frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} \tilde{\Delta}_n^{\mathbf{x}^*}\right)\left(\sum_{n=1}^{s} \tilde{\Delta}_n^{\mathbf{x}^*}\right)^T\right], \tag{9}$$

and $\tilde{\Delta}_n^{\mathbf{x}^*}$ represents $\tilde{\Delta}_n^{\mathbf{x}}$ measured at $\mathbf{x}_n = \mathbf{x}^*$ for all $n$. Through $\mathbf{Q}_{22}$, $\mathbf{Q}_{21}$ and $\tilde{\Delta}_n^{(2)}$, both $\mathbf{K}_{\mathbf{x}}$ and $\mathbf{U}_{\mathbf{x}}$ capture the effect of deterministic field $\bar{h}_2(\cdot,\cdot)$ on the asymptotic behavior of $\mathbf{x}_n$. The matrix $\mathbf{U}_{\mathbf{x}}$ incorporates the effect of Markovian noise sequence $\{\xi_n\}_{n\geq 0}$ through $\tilde{\Delta}_n^{(1)}$ and $\tilde{\Delta}_n^{(2)}$, which will be utilized in Proposition 3.2 to identify the effect of the underlying Markov chain on the asymptotic behavior of iterates $\mathbf{x}_n$. We show in Appendix *A*.2.1 that $\mathbf{U}_{\mathbf{x}}$ can also be written as

$$\mathbf{U}_{\mathbf{x}} = \begin{bmatrix} \mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1} \end{bmatrix}^T, \tag{10}$$

where

$$\mathbf{U}_{ij} \triangleq \lim_{s\to\infty} \frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} \tilde{\Delta}_n^{(i)^\dagger}\right)\left(\sum_{n=1}^{s} \tilde{\Delta}_n^{(j)^\dagger}\right)^T\right], \tag{11}$$

$\tilde{\Delta}_n^{(i)^\dagger}$ denotes $\tilde{\Delta}_n^{(i)}$ measured at $\mathbf{x}_n = \mathbf{x}^*$ for all $n$ and $i$, and $\mathbf{U}_{12} = \mathbf{U}_{21}^T$. For an *i.i.d.* sequence $\{\xi_n\}$ with marginal $\boldsymbol{\mu}$, $\mathbf{U}_{ij} = \mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_i(\mathbf{x}^*, \mathbf{y}^*, \xi)h_j(\mathbf{x}^*, \mathbf{y}^*, \xi)^T]$ degenerates to the marginal covariance of functions $h_i(\mathbf{x}^*, \mathbf{y}^*, \cdot), h_j(\mathbf{x}^*, \mathbf{y}^*, \cdot)$, and (8) aligns with previously established CLT results for linear (Konda and Tsitsiklis, 2004) and nonlinear TTSA (Mokkadem and Pelletier, 2006), both with Martingale difference noise.

### 2.3 Central Limit Theorem of TTSA with Controlled Markovian Noise

Without loss of generality, our remaining results are stated while conditioning on the event that $\{\mathbf{x}_n \to \mathbf{x}^*, \mathbf{y}_n \to \mathbf{y}^*\}$, for some $\mathbf{x}^* \in \Lambda$ and $\mathbf{y}^* = \lambda(\mathbf{x}^*)$. Our main CLT result is as follows, with its proof deferred to Appendix *A*.2.

**Theorem 2.2** (Central Limit Theorem)**.** *Under Assumptions (A1) – (A5),*

$$\begin{pmatrix} \beta_n^{-1/2}(\mathbf{x}_n - \mathbf{x}^*) \\ \gamma_n^{-1/2}(\mathbf{y}_n - \mathbf{y}^*) \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} \mathbf{V}_{\mathbf{x}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\mathbf{y}} \end{pmatrix}\right), \tag{12}$$

*where the limiting covariance matrices* $\mathbf{V}_{\mathbf{x}} \in \mathbb{R}^{d_1\times d_1}, \mathbf{V}_{\mathbf{y}} \in \mathbb{R}^{d_2\times d_2}$ *are given by*

$$\begin{aligned} \mathbf{V}_{\mathbf{x}} &= \int_0^\infty e^{t\left(\mathbf{K}_{\mathbf{x}} + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)} \mathbf{U}_{\mathbf{x}} e^{t\left(\mathbf{K}_{\mathbf{x}} + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)^T} dt, \\ \mathbf{V}_{\mathbf{y}} &= \int_0^\infty e^{t\mathbf{Q}_{22}} \mathbf{U}_{22} e^{t\mathbf{Q}_{22}^T} dt, \end{aligned} \tag{13}$$

*with* $\mathbf{K}_{\mathbf{x}}$, $\mathbf{U}_{\mathbf{x}}$ *and* $\mathbf{U}_{22}$ *defined in* (7), (9) *and* (11), *respectively.*

Theorem 2.2 suggests that iterates $(\mathbf{x}_n, \mathbf{y}_n)$ evolve asymptotically independently, as evidenced by the zero covariance of off-diagonal terms in (12). This is due to the diminishing correlation between $(\mathbf{x}_n - \mathbf{x}^*)$ and $(\mathbf{y}_n - \mathbf{y}^*)$ at a rate of $O(\beta_n/\gamma_n)$, a characteristic of the two-timescale setup, aligning with existing CLT findings for TTSA with Martingale difference noise (Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006). The limiting covariance matrix $\mathbf{V}_{\mathbf{y}}$ is solely determined by the local function $h_2$ and $\mathbf{x}^*$ without an additional term $\frac{\mathbb{1}\{a=1\}}{2}\mathbf{I}$ due to $a < 1$ by assumption (A1), implying minimal effect of $\mathbf{x}_n$ on the asymptotic behavior of $\mathbf{y}_n$. In contrast, $\mathbf{V}_{\mathbf{x}}$ is significantly impacted by iterates $\mathbf{y}_n$ since matrices $\mathbf{K}_{\mathbf{x}}, \mathbf{U}_{\mathbf{x}}$ are comprised of functions $h_2$ and $\bar{h}_2$.

As a special case, when $h_1(\mathbf{x}, \mathbf{y}, \xi)$ in the TTSA algorithm is independent of the variable $\mathbf{y}$, i.e., $h_1(\mathbf{x}, \mathbf{y}, \xi) \equiv h_1(\mathbf{x}, \xi)$, then $\nabla_{\mathbf{y}} h_1(\mathbf{x}, \xi) = 0$ for any $\mathbf{y} \in \mathbb{R}^{d_2}$, implying $\mathbf{Q}_{12} = \mathbf{0}$. According to Theorem 2.2, $\mathbf{x}_n$ is decoupled from iterates $\mathbf{y}_n$ and reduces to the single-timescale SA with Markovian noise, where $\mathbf{V}_{\mathbf{x}}$ in (13), in view of (10) with $\mathbf{Q}_{12} = \mathbf{0}$, becomes

$$\mathbf{V}_{\mathbf{x}} = \int_0^\infty e^{t\left(\nabla\bar{h}_1(\mathbf{x}^*) + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)} \mathbf{U}_{11} e^{t\left(\nabla\bar{h}_1(\mathbf{x}^*) + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)^T} dt.$$

This $\mathbf{V}_{\mathbf{x}}$ is in line with the existing CLT result for the single-timescale SA with controlled Markovian noise (Delyon, 2000; Benveniste et al., 2012; Fort, 2015) under the same locally Lipschitz condition on $h_1(\mathbf{x}, \xi)$, as stated in Assumption (A2).

The limiting covariance matrices $\mathbf{V}_{\mathbf{x}}$ and $\mathbf{V}_{\mathbf{y}}$ are related to the mean square error (MSE) of their corresponding iterative errors $\mathbf{x}_n - \mathbf{x}^*$ and $\mathbf{y}_n - \mathbf{y}^*$. For large enough $n$, the diagonal entries of $\mathbf{V}_{\mathbf{x}}$ are approximated by $\mathbf{e}_i^T\mathbf{V}_{\mathbf{x}}\mathbf{e}_i \approx \mathbf{e}_i^T\mathbb{E}[(\mathbf{x}_n - \mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*)^T]\mathbf{e}_i/\beta_n$ for all $i \in \{1, \cdots, d_1\}$, where $\mathbf{e}_i$ is the $i$-th canonical vector. Then, the MSE of the iterate error $\mathbf{x}_n - \mathbf{x}^*$ can be estimated as $\mathbb{E}[\|\mathbf{x}_n - \mathbf{x}^*\|^2] = \sum_{i=1}^{d_1} \mathbf{e}_i^T\mathbb{E}[(\mathbf{x}_n - \mathbf{x}^*)(\mathbf{x}_n - \mathbf{x}^*)^T]\mathbf{e}_i \approx \beta_n \sum_{i=1}^{d_1} \mathbf{e}_i^T\mathbf{V}_{\mathbf{x}}\mathbf{e}_i$. This implies that $\mathbb{E}[\|\mathbf{x}_n - \mathbf{x}^*\|^2]$ resembles the trace[4] of $\mathbf{V}_{\mathbf{x}}$, and decreases at a rate of $\beta_n$. Similar arguments also hold for $\mathbb{E}[\|\mathbf{y}_n - \mathbf{y}^*\|^2]$ and $\mathbf{V}_{\mathbf{y}}$.

## 3 APPLICATIONS

### 3.1 Performance Ordering in TTSA

The limiting covariance matrices $\mathbf{V}_{\mathbf{x}}, \mathbf{V}_{\mathbf{y}}$ described in (13) for nonlinear TTSA with Markovian noise inherently incorporate the properties of the underlying Markov chain *completely* in terms of matrices $\mathbf{U}_{\mathbf{x}}$ and $\mathbf{U}_{22}$, as defined in (10) and (11). This raises an intuitive question: *If we can control the stochastic input sequence* $\{\xi_n\}_{n\geq 0}$*, how does it influence the performance of the TTSA algorithm?*

This question was originally studied by Hu et al. (2022), which introduces the notion of efficiency ordering of Markov chains, a metric prevalent in the MCMC literature, in the context of SGD algorithms, and proves that the presence of 'better', more efficient sampling strategy leads to improved SGD performance. Broadening this concept, we show that such performance improvements are applicable to the

[4]Sum of diagonal entries of a matrix.

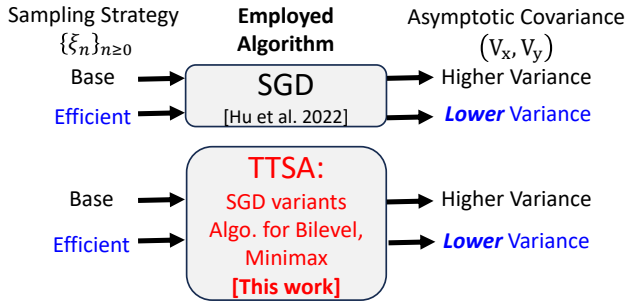
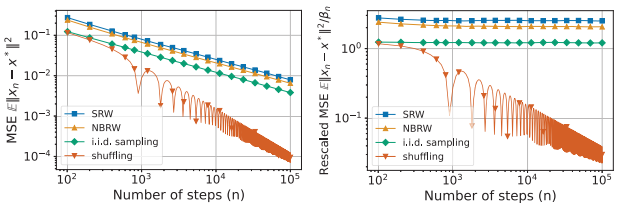



Figure 1: Efficiency Ordering: From SGD to TTSA.




(a) MSE (b) Rescaled MSE

Figure 2: Comparison of the performance among different sampling strategies in momentum SGD.

general TTSA framework, beyond mere SGD algorithms, as depicted in Figure 1. To better understand this, let $\mathbf{U}^Z(g) \triangleq \lim_{s\to\infty} \frac{1}{s}\mathbb{E}[\Delta_s \Delta_s^T]$ be the sampling covariance matrix for a vector-valued function $g : \Xi \to \mathbb{R}^d$ and stochastic process $\{Z_n\}$, where $\Delta_s = \sum_{n=1}^{s}(g(Z_n) - \mathbb{E}_{\boldsymbol{\mu}}[g])$ and $\mathbb{E}_{\boldsymbol{\mu}}[g] = \sum_{i\in\Xi} g(i)\mu_i$.

**Definition 3.1** (Efficiency Ordering, (Mira, 2001; Hu et al., 2022))**.** *For two Markov chains $\{W_n\}$ and $\{Z_n\}$ with identical stationary distribution $\boldsymbol{\mu}$, we say $\{Z_n\}$ is more sampling-efficient than $\{W_n\}$, denoted as $W \preceq Z$, if and only if $\mathbf{U}^W(g) \geq_L \mathbf{U}^Z(g)$ for any vector-valued function $g$.*

Examples of sampling strategies following Definition 3.1 include random and single shuffling paradigms (Ahn et al., 2020; Safran and Shamir, 2020), which are shown to be more sampling-efficient when compared to *i.i.d.* sampling. Another example, relevant in the context of token algorithms in distributed learning, is the so-called non-backtracking random walk (NBRW) (Alon et al., 2007; Lee et al., 2012; Ben-Hamou et al., 2018), which is more sampling-efficient than simple random walk (SRW). We point the reader to Hu et al. (2022, Section 4) for more detailed discussions, where more efficient sampling strategies employed in SGD algorithms lead to reduced asymptotic covariance of iterate errors. With two efficiency-ordered sampling strategies, we now extend the same performance ordering to TTSA, the proof of which can be found in Appendix *A*.3.1.

**Proposition 3.2.** *For the TTSA algorithm* (2)*, given two different underlying Markov chains $\{W_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$ that are efficiency ordered, i.e., $W \preceq Z$, we have $\mathbf{V}_{\mathbf{x}}^{(W)} \geq_L \mathbf{V}_{\mathbf{x}}^{(Z)}$ and $\mathbf{V}_{\mathbf{y}}^{(W)} \geq_L \mathbf{V}_{\mathbf{y}}^{(Z)}$.*

Proposition 3.2 enables us to expand the scope of Hu et al. (2022) by employing sampling-efficient strategies to a wider class of optimization problems within the TTSA framework. Specifically, our scope extends existing results as follows:

(i) *From vanilla SGD to its variants:* The TTSA structure accommodates many SGD variants for finite-sum minimization, including the Polyak-Ruppert averaging (Ruppert, 1988; Polyak and Juditsky, 1992) and momentum SGD (Gadat et al., 2018; Li et al., 2022). Other variants of SGD in the TTSA framework, e.g., signSGD and normalized SGD, are provided in Xiao et al. (2023, Section 4.3) with detailed expressions.

(ii) *From finite-sum minimization to bilievel and minimax problems:* Many algorithms within the TTSA framework can handle bilevel and minimax problems. For instance, Hong et al. (2023, Algorithm 1) effectively deals with both inner and outer objectives in bilevel optimization, while the stochastic gradient descent ascent algorithm (Lin et al., 2020, Algorithm 1) seeks saddle points in the minimax problem.

From Proposition 3.2, all the above algorithms enjoy improved asymptotic performance when driven by more efficient samples. For instance, in the token algorithm setting (Hu et al., 2022; Triastcyn et al., 2022; Hendrikx, 2023; Even, 2023), a token can employ NBRW over SRW to solve various optimization problems with these TTSA algorithms. When random access of each data point is possible, Hu et al. (2022, Lemma 4.2) highlights that through a state-space augmentation, shuffling algorithms – conceptualized as Markov chains – outperform *i.i.d.* sampling, achieving *zero* sampling covariance. Using Proposition 3.2, we can show that this leads to *zero* limiting covariance $\mathbf{V}_{\mathbf{x}}, \mathbf{V}_{\mathbf{y}}$ for all algorithms represented as TTSA. The superiority of shuffling techniques over *i.i.d.* sampling has indeed been studied for specific stochastic optimization settings, such as minimax optimization (Das et al., 2022; Cho and Yun, 2022) and SGD with momentum (Tran et al., 2021). However, Proposition 3.2 firmly establishes this at a much broader scope as described in (i) and (ii), such as bilevel optimization with shuffling methods, whose finite-time analysis remains an open problem.

**Simulations.** We present numerical experiments for different sampling strategies employed in the momentum SGD algorithm to solve the $L_2$-regularized binary classification problem using the dataset *a9a* (with 123 features) from LIBSVM (Chang and Lin, 2011).

Specifically, to simulate the token algorithm in distributed learning, we employ NBRW and SRW as the stochastic input to the momentum SGD on the wikiVote graph (Leskovec and Krevl, 2014), comprising 889 nodes and 2914 edges.[5] Each node on the wikiVote graph is assigned with one data point from the dataset *a9a*, thus 889 data points in total. We also assess the momentum SGD's performance under *i.i.d.* sampling and single shuffling using the same dataset of size 889. In Figure 2(a), we observe that NBRW has a smaller MSE than SRW across all time $n$, with a similar trend for single shuffling over *i.i.d.* sampling. Figure 2(b) demonstrates that the rescaled MSEs of NBRW, SRW and *i.i.d.* sampling approach some constants, while the curve for single shuffling still decreases in linear rate because eventually the limiting covariance matrix therein will be zero. We defer the detailed simulation settings and more simulation results to Appendix $A.5$.

### 3.2 Asymptotic Behavior of Nonlinear GTD Algorithms

The CLT result not only allows comparison of limiting covariance matrices of two efficiency-ordered stochastic inputs in distributed learning, but also offers insights into an algorithm's asymptotic performance, as showcased in Table 1. This is particularly relevant in RL where the stochastic sequence $\{\xi_n\}$ is generated by a given policy and thus uncontrollable. An important aspect in RL is policy evaluation in MDP with the primary goal of estimating the value function of a given policy, which is essential for further policy improvement (Sutton and Barto, 2018). In this part, we focus on a family of gradient-based TD learning (GTD) algorithms, which are instances of TTSA (Maei et al., 2009; Wang et al., 2021). We leverage Theorem 2.2 to derive the pioneering statistical properties of these algorithms when using *nonlinear* value function approximation and Markovian samples for policy evaluation.

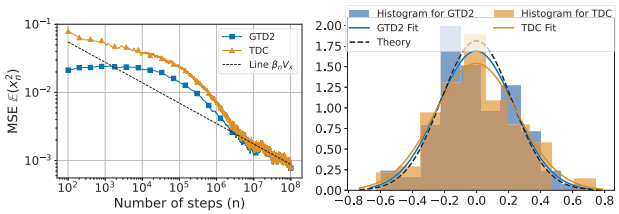
Tabular methods for estimating the value function, such as SARSA, have been widely used, but can be problematic when the state-action space is large (Sutton and Barto, 2018). TD learning with linear function approximation has been extensively studied (Srikant and Ying, 2019; Doan et al., 2019; Wang et al., 2020; Li et al., 2023a). In contrast to linear function approximation, nonlinear approaches, e.g. neural networks, are more practical choices known for their strong representation capabilities and eliminating the need for feature mapping (Wai et al., 2020; Wang et al., 2021). However, Tsitsiklis and Van Roy (1997) notes the potential divergence of TD learning with nonlinear function approximation. Addressing the divergence, Maei et al. (2009) introduces nonlinear GTD2 and TDC algorithms with almost sure convergence guarantees. These methods iterate over gradients of the mean-square projected Bellman error (MSPBE) in order to obtain the best estimate of the nonlinear value function that minimizes MSPBE (Maei et al., 2009; Xu and Liang, 2021; Wang et al., 2021).

While non-asymptotic analyses of GTD2 and TDC algorithms have been established for both *i.i.d.* and Markovian settings with linear approximations (Karmakar and Bhatnagar, 2018; Dalal et al., 2018, 2020; Kaledin et al., 2020; Li et al., 2023a), results for the nonlinear function approximation remain scarce since MSEPBE becomes nonconvex and the two-timescale update rule is nonlinear. For asymptotic analysis, Karmakar and Bhatnagar (2018) studies the almost sure convergence of general TTSA and applies it to nonlinear TDC algorithm, extending from *i.i.d.* (Maei et al., 2009) to Markovian samples. This analysis can also be applied to nonlinear GTD2 algorithm. Only a few works (Xu and Liang, 2021; Wang et al., 2021) provide non-asymptotic analysis specifically for nonlinear TDC algorithm with Markovian samples and constant step sizes while the results cannot be extrapolated to nonlinear GTD2 algorithm. Therefore, a comprehensive analysis of these algorithms with Markovian samples under decreasing step sizes remains lacking in RL.

We now summarize nonlinear GTD2 and TDC algorithms, followed by their asymptotic results in Proposition 3.3. An MDP is defined as a 5-tuple $(\mathcal{S}, \mathcal{A}, P, r, \alpha)$, where $\mathcal{S}$ and $\mathcal{A}$ are the finite state and action spaces, and $P$ and $r$ are transition kernel and reward function, with $\alpha$ being a discount factor. A policy $\boldsymbol{\pi}$ maps each state $s \in \mathcal{S}$ onto an action probability distribution $\boldsymbol{\pi}(\cdot|s)$, with $\boldsymbol{\mu}^{\boldsymbol{\pi}}$ being the corresponding stationary distribution. The Markovian samples $\{s_n\}$ then follow the transition probability $\mathbb{P}(s, s') = \sum_{a\in\mathcal{A}} P(s'|s, a)\boldsymbol{\pi}(a|s)$. The value function for policy $\boldsymbol{\pi}$ from initial state $s$ is $W^{\boldsymbol{\pi}}(s) = \mathbb{E}_{\boldsymbol{\pi}}[\sum_{n=0}^{\infty} \alpha^n r_n | s_0 = s]$, where $r_n \triangleq r(s_n, a_n, s_{n+1})$. The GTD2 and TDC algorithms estimate $W^{\boldsymbol{\pi}}(s)$ via nonlinear functions $W_{\mathbf{x}}(s)$ and its feature function $\phi_{\mathbf{x}}(s) = \nabla_{\mathbf{x}} W_{\mathbf{x}}(s)$ parameterized by $\mathbf{x}$. For linear approximation $W_{\mathbf{x}}(s) = \phi(s)^T\mathbf{x}$, $\phi(s)$ is independent of $\mathbf{x}$. However, with nonlinear $W_{\mathbf{x}}(s)$, $\phi_{\mathbf{x}}(s)$ depends on $\mathbf{x}$. Defining TD error as $\delta_n = r_n + \alpha W_{\mathbf{x}_n}(s_{n+1}) - W_{\mathbf{x}_n}(s_n)$, the iterates $(\mathbf{x}_n, \mathbf{y}_n)$ of the GTD2 and TDC algorithms admit an equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$, with $\mathbf{x}^*$ ensuring $\mathbb{E}_{s_n\sim\boldsymbol{\mu}}[\delta_n(\mathbf{x}^*)\phi_{\mathbf{x}^*}(s_n)] = \mathbf{0}$, and $\mathbf{y}^* = \mathbf{0}$. Details of these algorithms and conditions for the following CLT results are in Appendix $A.4.1$.

**Proposition 3.3.** *For both nonlinear GTD2 and TDC*

[5] We incorporate both NBRW and SRW with importance reweighting to achieve a uniform target distribution.




(a) MSE (b) Histograms of $\beta_n^{-1/2} x_n$

Figure 3: Comparison of nonlinear GTD2 and TDC algorithms in the 5-state random walk task.

*algorithms under Markovian samples, we have*

$$\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}^* \quad a.s. \quad and \quad \lim_{n\to\infty} \mathbf{y}_n = \mathbf{0} \quad a.s.$$

$$\frac{1}{\sqrt{\beta_n}}(\mathbf{x}_n - \mathbf{x}^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{V_x}), \quad \frac{1}{\sqrt{\gamma_n}}\mathbf{y}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{V_y}),$$

*where* $\mathbf{V_x}, \mathbf{V_y}$ *are identical for both algorithms.*

The proof of Proposition 3.3 and the exact forms of $\mathbf{V_x}, \mathbf{V_y}$ are in Appendix $A$.4.2. This proposition offers a state-of-the-art performance analysis of nonlinear GTD2 and TDC algorithms in RL, employing Markovian samples and general decreasing step sizes. While Doan (2021b); Zeng et al. (2021) provide finite-time bounds within the general TTSA framework, their applicability to nonlinear GTD2 and TDC algorithms is restricted by specific choice of the step sizes, as explained in Section 1.1. The usefulness of these finite-time results are further limited due to the lack of any definitive indication regarding the tightness of the bounds associated with these two algorithms. Moreover, empirical studies (Dann et al., 2014; Ghiassian et al., 2020) have not consistently favored either one of the two algorithms when compared across all tasks, leading to a lack of consensus regarding which one is the better performing overall. Proposition 3.3 clarifies that, in the long run, both GTD2 and TDC algorithms exhibit identical behaviors under the CLT scaling.

**Simulations.** We consider a 5-state random walk task (Dann et al., 2014; Sutton and Barto, 2018) for nonlinear GTD2 and TDC algorithms. Each state can transit to the right or left next state with probability 0.5, with reward $+0.5$ if turning right or $-0.5$ otherwise. Let discount factor $\alpha = 0.9$, we consider the nonlinear value function $W_x(s) = a(s)(e^{\kappa x} - 1)$ for a scalar parameter $x$, where $a = [-2, -6, -3, -4, -5], \kappa = 0.1$. The ground truth $W(s) = 0$ for $s \in [5]$ such that $x^* = 0$ for $W_{\mathbf{x}}(s)$ achieves the accurate approximation. Figure 3(a) illustrates the long-term performance of GTD2 and TDC algorithms. Starting from $n = 10^7$, they align with the line $\beta_n \mathbf{V_x}$, with $\mathbf{V_x}$ being a scalar from Proposition 3.3. This reaffirms the relationship between MSE and CLT, as detailed in Section 2.3. Figure 3(b) displays a histogram of $\beta_n^{-1/2} x_n$, generated from 100 independent trials at $n = 10^8$ for both algorithms. We show that their experimental density curves are close to the theoretical Gaussian density with zero mean and variance $\mathbf{V_x}$. We defer the detailed simulation settings, calculations related to Figure 3, and additional simulation results to Appendix $A$.5.

## 4 CONCLUSION

In this paper, we present the first CLT analysis of nonlinear TTSA in the context of controlled Markovian noise with general forms of decreasing step sizes. Our result greatly extends the scope of existing literature by allowing most general settings and performance ordering across nonlinear TTSA algorithms, notably in distributed optimization and RL. Our work highlights the potential of asymptotic analysis for algorithmic improvement and comparison, addressing areas where conventional finite-time analysis fall short, thus suggesting that more attention should be given to asymptotic statistics.

## 5 ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments. This work was supported in part by National Science Foundation under Grant Nos. CNS-2007423 and IIS-1910749.

### References

Ahn, K., Yun, C., and Sra, S. (2020). Sgd with shuffling: optimal rates without component convexity and large epoch requirements. In *Advances in Neural Information Processing Systems*, volume 33, pages 17526–17535.

Alon, N., Benjamini, I., Lubetzky, E., and Sodin, S. (2007). Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 9(04):585–603.

Andrieu, C. and Vihola, M. (2014). Markovian stochastic approximation with expanding projections. *Bernoulli*, pages 545–585.

Avrachenkov, K. E., Borkar, V. S., Moharir, S., and Shah, S. M. (2021). Dynamic social learning under graph constraints. *IEEE Transactions on Control of Network Systems*, 9(3):1435–1446.

Barakat, A., Bianchi, P., and Lehmann, J. (2022). Analysis of a target-based actor-critic algorithm with linear function approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 991–1040. PMLR.

Ben-Hamou, A., Lubetzky, E., and Peres, Y. (2018). Comparing mixing times on sparse random graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1734–1740. SIAM.

Benaim, M., Raimond, O., and Schapira, B. (2012). Strongly vertex-reinforced-random-walk on the complete graph. *arXiv preprint arXiv:1208.6375*.

Benveniste, A., Métivier, M., and Priouret, P. (2012). *Adaptive algorithms and stochastic approximations*, volume 22. Springer Science & Business Media.

Borkar, V. (2022). *Stochastic Approximation: A Dynamical Systems Viewpoint: Second Edition*. Texts and Readings in Mathematics. Hindustan Book Agency.

Borkar, V. S. and Pattathil, S. (2018). Concentration bounds for two time scale stochastic approximation. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 504–511. IEEE.

Bottou, L., Curtis, F. E., and Nocedal, J. (2018). Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311.

Brémaud, P. (2013). *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27.

Chellaboina, V. and Haddad, W. M. (2008). *Nonlinear dynamical systems and control: A Lyapunov-based approach*. Princeton University Press.

Chen, H.-F. (2006). *Stochastic approximation and its applications*, volume 64. Springer Science & Business Media.

Chen, S., Devraj, A., Busic, A., and Meyn, S. (2020). Explicit mean-square error bounds for monte-carlo and linear stochastic approximation. In *International Conference on Artificial Intelligence and Statistics*, pages 4173–4183. PMLR.

Chen, Z., Zhang, S., Doan, T. T., Clarke, J.-P., and Maguluri, S. T. (2022). Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning. *Automatica*, 146:110623.

Cho, H. and Yun, C. (2022). Sgda with shuffling: faster convergence for nonconvex-pł minimax optimization. In *The Eleventh International Conference on Learning Representations*.

Dalal, G., Szorenyi, B., and Thoppe, G. (2020). A tale of two-timescale reinforcement learning with the tightest finite-time bound. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3701–3708.

Dalal, G., Thoppe, G., Szörényi, B., and Mannor, S. (2018). Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In *Conference On Learning Theory*, pages 1199–1233. PMLR.

Dann, C., Neumann, G., Peters, J., et al. (2014). Policy evaluation with temporal differences: A survey and comparison. *Journal of Machine Learning Research*, 15:809–883.

Das, A., Schölkopf, B., and Muehlebach, M. (2022). Sampling without replacement leads to faster rates in finite-sum minimax optimization. In *Advances in Neural Information Processing Systems*.

Davis, B. (1970). On the intergrability of the martingale square function. *Israel Journal of Mathematics*, 8:187–190.

Delyon, B. (2000). Stochastic approximation with decreasing gain: Convergence and asymptotic theory. Technical report, Université de Rennes.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128.

Doan, T., Maguluri, S., and Romberg, J. (2019). Finite-time analysis of distributed td (0) with linear function approximation on multi-agent reinforcement learning. In *International Conference on Machine Learning*, pages 1626–1635. PMLR.

Doan, T. T. (2021a). Finite-time analysis and restarting scheme for linear two-time-scale stochastic approximation. *SIAM Journal on Control and Optimization*, 59(4):2798–2819.

Doan, T. T. (2021b). Finite-time convergence rates of nonlinear two-time-scale stochastic approximation under markovian noise. *arXiv preprint arXiv:2104.01627*.

Doan, T. T. (2022). Nonlinear two-time-scale stochastic approximation convergence and finite-time performance. *IEEE Transactions on Automatic Control*.

Doan, T. T. (2024). Fast nonlinear two-time-scale stochastic approximation: Achieving $\mathcal{O}(1/k)$ finite-sample complexity. *arXiv preprint arXiv:2401.12764*.

Doshi, V., Hu, J., and Eun, D. Y. (2023). Self-repellent random walks on general graphs–achieving minimal sampling variance via nonlinear markov chains. In *International Conference on Machine Learning*. PMLR.

Duflo, M. (1996). *Algorithmes stochastiques*, volume 23. Springer.

Even, M. (2023). Stochastic gradient descent under markovian sampling schemes. In *International Conference on Machine Learning*.

Fort, G. (2015). Central limit theorems for stochastic approximation with controlled markov chain dynamics. *ESAIM: Probability and Statistics*, 19:60–80.

Fort, G., Moulines, E., Schreck, A., and Vihola, M. (2016). Convergence of markovian stochastic approximation with discontinuous dynamics. *SIAM Journal on Control and Optimization*, 54(2):866–893.

Gadat, S., Panloup, F., and Saadane, S. (2018). Stochastic heavy ball. *Electronic Journal of Statistics*, 12:461–529.

Gao, H. (2022). Decentralized stochastic gradient descent ascent for finite-sum minimax problems. *arXiv preprint arXiv:2212.02724*.

Gao, H., Gu, B., and Thai, M. T. (2023). On the convergence of distributed stochastic bilevel optimization algorithms over a network. In *International Conference on Artificial Intelligence and Statistics*, pages 9238–9281. PMLR.

Ghiassian, S., Patterson, A., Garg, S., Gupta, D., White, A., and White, M. (2020). Gradient temporal-difference learning with regularized corrections. In *International Conference on Machine Learning*, pages 3524–3534. PMLR.

Gower, R. M., Loizou, N., Qian, X., Sailanbayev, A., Shulgin, E., and Richtárik, P. (2019). Sgd: General analysis and improved rates. In *International Conference on Machine Learning*, pages 5200–5209. PMLR.

Gupta, H., Srikant, R., and Ying, L. (2019). Finite-time performance bounds and adaptive learning rate selection for two time-scale reinforcement learning. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 4704–4713.

Hall, P., Heyde, C., Birnbauam, Z., and Lukacs, E. (2014). *Martingale Limit Theory and Its Application*. Communication and Behavior. Elsevier Science.

Haque, S. U., Khodadadian, S., and Maguluri, S. T. (2023). Tight finite time bounds of two-time-scale linear stochastic approximation with markovian noise. *arXiv preprint arXiv:2401.00364*.

Hendrikx, H. (2023). A principled framework for the design and analysis of token algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 470–489. PMLR.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Hong, M., Wai, H.-T., Wang, Z., and Yang, Z. (2023). A two-timescale stochastic algorithm framework for bilevel optimization: Complexity analysis and application to actor-critic. *SIAM Journal on Optimization*, 33(1):147–180.

Hu, J., Doshi, V., and Eun, D. Y. (2022). Efficiency ordering of stochastic gradient descent. In *Advances in Neural Information Processing Systems*.

Kaledin, M., Moulines, E., Naumov, A., Tadic, V., and Wai, H.-T. (2020). Finite time analysis of linear two-timescale stochastic approximation with markovian noise. In *Conference on Learning Theory*, pages 2144–2203. PMLR.

Karimi, B., Miasojedow, B., Moulines, E., and Wai, H.-T. (2019). Non-asymptotic analysis of biased stochastic approximation scheme. In *Conference on Learning Theory*, pages 1944–1974. PMLR.

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

Khodadadian, S., Doan, T. T., Romberg, J., and Maguluri, S. T. (2022). Finite sample analysis of two-time-scale natural actor-critic algorithm. *IEEE Transactions on Automatic Control*.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

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

Kushner, H. and Yin, G. G. (2003). *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer Science & Business Media.

Lee, C.-H., Xu, X., and Eun, D. Y. (2012). Beyond random walk and metropolis-hastings samplers: why you should not backtrack for unbiased graph sampling. *ACM SIGMETRICS Performance evaluation review*, 40(1):319–330.

Leskovec, J. and Krevl, A. (2014). Snap datasets: Stanford large network dataset collection.

Li, G., Wu, W., Chi, Y., Ma, C., Rinaldo, A., and Wei, Y. (2023a). Sharp high-probability sample complexities for policy evaluation with linear function approximation. *arXiv preprint arXiv:2305.19001*.

Li, T., Xiao, T., and Yang, G. (2022). Revisiting the central limit theorems for the sgd-type methods. *arXiv preprint arXiv:2207.11755*.

Li, X., Liang, J., and Zhang, Z. (2023b). Online statistical inference for nonlinear stochastic approximation with markovian data. *arXiv preprint arXiv:2302.07690*.

Lin, T., Jin, C., and Jordan, M. (2020). On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*, pages 6083–6093. PMLR.

Lopez-Ramos, L. M., Kekatos, V., Marques, A. G., and Giannakis, G. B. (2017). Two-timescale stochastic dispatch of smart distribution grids. *IEEE Transactions on Smart Grid*, 9(5):4282–4292.

Ma, S., Zhou, Y., and Zou, S. (2020). Variance-reduced off-policy tdc learning: non-asymptotic convergence analysis. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 14796–14806.

Maei, H. R., Szepesvári, C., Bhatnagar, S., Precup, D., Silver, D., and Sutton, R. S. (2009). Convergent temporal-difference learning with arbitrary smooth function approximation. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, pages 1204–1212.

Meyn, S. (2022). *Control systems and reinforcement learning*. Cambridge University Press.

Mira, A. (2001). Ordering and improving the performance of monte carlo markov chains. *Statistical Science*, pages 340–350.

Mokkadem, A. and Pelletier, M. (2005). The compact law of the iterated logarithm for multivariate stochastic approximation algorithms. *Stochastic analysis and applications*, 23(1):181–203.

Mokkadem, A. and Pelletier, M. (2006). Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *Annals of Applied Probability*, 16(3):1671–1702.

Mou, W., Li, C. J., Wainwright, M. J., Bartlett, P. L., and Jordan, M. I. (2020). On linear stochastic approximation: Fine-grained polyak-ruppert and non-asymptotic concentration. In *Conference on Learning Theory*, pages 2947–2997. PMLR.

Neal, R. M. (2004). Improving asymptotic variance of mcmc estimators: Non-reversible chains are better. Technical report.

Patil, G., Prashanth, L., Nagaraj, D., and Precup, D. (2023). Finite time analysis of temporal difference learning with linear function approximation: Tail averaging and regularisation. In *International Conference on Artificial Intelligence and Statistics*, pages 5438–5448. PMLR.

Pelletier, M. (1998). On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic processes and their applications*, 78(2):217–244.

Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855.

Prasad, H., LA, P., and Bhatnagar, S. (2015). Two-timescale algorithms for learning nash equilibria in general-sum stochastic games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1371–1379.

Recht, B., Fazel, M., and Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501.

Ruppert, D. (1988). Efficient estimations from a slowly convergent robbins-monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

Safran, I. and Shamir, O. (2020). How good is sgd with random shuffling? In *Conference on Learning Theory*, pages 3250–3284. PMLR.

Srikant, R. (2024). Rates of convergence in the central limit theorem for markov chains, with an application to td learning. *arXiv preprint arXiv:2401.15719*.

Srikant, R. and Ying, L. (2019). Finite-time error bounds for linear stochastic approximation and td learning. In *Conference on Learning Theory*, pages 2803–2830. PMLR.

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

Sutton, R. S., Maei, H. R., Precup, D., Bhatnagar, S., Silver, D., Szepesvári, C., and Wiewiora, E. (2009). Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *Proceedings of the 26th annual international conference on machine learning*, pages 993–1000.

Tarzanagh, D. A., Li, M., Thrampoulidis, C., and Oymak, S. (2022). Fednest: Federated bilevel, minimax, and compositional optimization. In *International Conference on Machine Learning*, pages 21146–21179. PMLR.

Tran, T. H., Nguyen, L. M., and Tran-Dinh, Q. (2021). Smg: A shuffling gradient-based method with momentum. In *International Conference on Machine Learning*, pages 10379–10389. PMLR.

Triastcyn, A., Reisser, M., and Louizos, C. (2022). Decentralized learning with random walks and communication-efficient adaptive optimization. In *Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*.

Tsitsiklis, J. and Van Roy, B. (1997). An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42(5):674–690.

Wai, H.-T., Yang, Z., Wang, Z., and Hong, M. (2020). Provably efficient neural gtd algorithm for off-policy learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 10431–10442.

Wang, G., Lu, S., Giannakis, G., Tesauro, G., and Sun, J. (2020). Decentralized td tracking with linear function approximation and its finite-time analysis. In *Advances in Neural Information Processing Systems*, volume 33, pages 13762–13772.

Wang, Y., Zou, S., and Zhou, Y. (2021). Non-asymptotic analysis for two time-scale tdc with general smooth function approximation. In *Advances in Neural Information Processing Systems*.

Xiao, N., Hu, X., and Toh, K.-C. (2023). Convergence guarantees for stochastic subgradient methods in nonsmooth nonconvex optimization. *arXiv preprint arXiv:2307.10053*.

Xu, T. and Liang, Y. (2021). Sample complexity bounds for two timescale value-based reinforcement learning algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 811–819. PMLR.

Yaji, V. G. and Bhatnagar, S. (2020). Stochastic recursive inclusions in two timescales with nonadditive iterate-dependent markov noise. *Mathematics of Operations Research*, 45(4):1405–1444.

Yang, Q., Wang, G., Sadeghi, A., Giannakis, G. B., and Sun, J. (2019). Two-timescale voltage control in distribution grids using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 11(3):2313–2323.

Zeng, S., Doan, T. T., and Romberg, J. (2021). A two-time-scale stochastic optimization framework with applications in control and reinforcement learning. *arXiv preprint arXiv:2109.14756*.

Zhang, J. and Hong, M. (2020). First-order algorithms without lipschitz gradient: A sequential local optimization approach. *arXiv preprint arXiv:2010.03194*.

## Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

# Central Limit Theorem for Two-Timescale Stochastic Approximation with Markovian Noise: Theory and Applications Supplementary Materials

## A.1 Proof of Lemma 2.1

The two-timescale SA form with iterate-dependent Markov chain we consider in this paper is as follows:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1} h_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}), \tag{14a}$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \gamma_{n+1} h_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}), \tag{14b}$$

with the goal of finding the root $(\mathbf{x}^*, \mathbf{y}^*)$ such that

$$\bar{h}_1(\mathbf{x}^*, \mathbf{y}^*) = \mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_1(\mathbf{x}^*, \mathbf{y}^*, \xi)] = 0, \quad \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*) = \mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_2(\mathbf{x}^*, \mathbf{y}^*, \xi)] = 0. \tag{15}$$

For self-contained purposes, we reproduce assumptions (A1) – (A5) for the TTSA algorithm (14) below.

A1. The step sizes $\beta_n \triangleq (n+1)^{-b}$ and $\gamma_n \triangleq (n+1)^{-a}$, where $0.5 < a < b \leq 1$.

A2. For the $C^1$ function $h_1 : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \Xi \to \mathbb{R}^{d_1}$, there exists a positive constant $L_1$ such that $\|h_1(\mathbf{x}, \mathbf{y}, \xi)\| \leq L_1(1 + \|\mathbf{x}\| + \|\mathbf{y}\|)$ for every $\mathbf{x} \in \mathbb{R}^{d_1}, \mathbf{y} \in \mathbb{R}^{d_2}, \xi \in \Xi$. The same condition holds for the $C^1$ function $h_2$ as well.

A3. Consider a $C^1$ function $\lambda : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$. For every $\mathbf{x} \in \mathbb{R}^{d_1}$, the following three properties hold: (i) $\lambda(\mathbf{x})$ is the globally attracting point of the related ODE $\dot{\mathbf{y}} = \bar{h}_2(\mathbf{x}, \mathbf{y})$; (ii) $\nabla_{\mathbf{y}} \bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))$ is Hurwitz; (iii) $\|\lambda(\mathbf{x})\| \leq L_2(1 + \|\mathbf{x}\|)$ for some positive constant $L_2$. Additionally, let $\hat{h}_1(\mathbf{x}) \triangleq \bar{h}_1(\mathbf{x}, \lambda(\mathbf{x}))$, there exists a set of disjoint roots $\Lambda \triangleq \{\mathbf{x}^* : \hat{h}_1(\mathbf{x}^*) = 0, \nabla_{\mathbf{x}} \hat{h}_1(\mathbf{x}^*) + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I} \text{ is Hurwitz}\}$, which is also the globally attracting set for trajectories of the related ODE $\dot{\mathbf{x}} = \hat{h}_1(\mathbf{x})$.

A4. $\{\xi_n\}_{n\geq 0}$ is an iterate-dependent Markov chain on finite state space $\Xi$. For every $n \geq 0$, $P(\xi_{n+1} = j | \mathbf{x}_m, \mathbf{y}_m, \xi_m, 0 \leq m \leq n) = P(\xi_{n+1} = j | \mathbf{x}_n, \mathbf{y}_n, \xi_n = i) = \mathbf{P}_{i,j}[\mathbf{x}_n, \mathbf{y}_n]$, where the transition kernel $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is continuous in $\mathbf{x}, \mathbf{y}$, and the Markov chain generated by $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is ergodic so that it admits a stationary distribution $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$, and $\boldsymbol{\pi}(\mathbf{x}^*, \lambda(\mathbf{x}^*)) = \boldsymbol{\mu}$.

A5. $\sup_{n\geq 0}(\|\mathbf{x}_n\| + \|\mathbf{y}_n\|) < \infty$ a.s.

Now, we translate the assumptions in Yaji and Bhatnagar (2020) below in our notations and TTSA algorithm (14) in order to apply the almost sure convergence result therein.

B1. The step sizes $\beta_n \triangleq n^{-b}$ and $\gamma_n \triangleq n^{-a}$, where $0.5 < a < b \leq 1$.

B2. Assume the function $h_1(\mathbf{x}, \mathbf{y}, \xi)$ is continuous and differentiable with respect to $\mathbf{x}, \mathbf{y}$. There exists a positive constant $L_1$ such that $\|h_1(\mathbf{x}, \mathbf{y}, \xi)\| \leq L_1(1 + \|\mathbf{x}\| + \|\mathbf{y}\|)$ for every $\mathbf{x} \in \mathbb{R}^{d_1}, \mathbf{y} \in \mathbb{R}^{d_2}, \xi \in \Xi$. The same condition holds for the function $h_2$ as well.

B3. Assume there exists a continuous function $\lambda : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ such that the following two properties hold for any $\mathbf{x} \in \mathbb{R}^{d_1}$: (i) $\|\lambda(\mathbf{x})\| \leq L_2(1 + \|\mathbf{x}\|)$ for some positive constant $L_2$; (ii) the ODE $\dot{\mathbf{y}} = \bar{h}_2(\mathbf{x}, \mathbf{y})$ has a globally asymptotically stable equilibrium $\lambda(\mathbf{x})$ such that $\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) = 0$. Additionally, let $\hat{h}_1(\mathbf{x}) \triangleq \bar{h}_1(\mathbf{x}, \lambda(\mathbf{x}))$, there exists a set of disjoint roots $\Lambda \triangleq \{\mathbf{x}^* : \hat{h}_1(\mathbf{x}^*) = 0\}$, which is the set of globally asymptotically stable equilibria of the ODE $\dot{\mathbf{x}} = \hat{h}_1(\mathbf{x})$.

B4. $\{\xi_n\}_{n\geq 0}$ is an iterate-dependent Markov process in finite state space $\Xi$. For every $n \geq 0$, $P(X_{n+1} = j|\mathbf{x}_m, \mathbf{y}_m, \xi_m, 0 \leq m \leq n) = P(X_{n+1} = j|\mathbf{x}_n, \mathbf{y}_n, \xi_n = i) = \mathbf{P}_{i,j}[\mathbf{x}_n, \mathbf{y}_n]$, where the transition kernel $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is continuous in $\mathbf{x}, \mathbf{y}$, and the Markov chain generated by $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ is ergodic so that it admits a stationary distribution $\boldsymbol{\pi}(\mathbf{x}, \mathbf{y})$, and $\boldsymbol{\pi}(\mathbf{x}^*, \lambda(\mathbf{x}^*)) = \boldsymbol{\mu}$.

B5. $\sup_{n\geq 0}(\|\mathbf{x}_n\| + \|\mathbf{y}_n\|) < \infty$ a.s.

Assumption (B1) is the standard condition on step sizes corresponding to Yaji and Bhatnagar (2020, Assumption A5). Assumption (B2) translates Yaji and Bhatnagar (2020, Assumptions A1, A2) from functions $h_1, h_2$ with set values to those with single values. Assumption (B3) is the condition on the relevant ODEs of the TTSA algorithm (14), which is derived from Yaji and Bhatnagar (2020, Assumption A9 – A11). Assumption (B4) simplifies Yaji and Bhatnagar (2020, Assumptions A3, A4) by using a single Markov sequence $\{\xi_n\}$ for both iterations in (14). This has a wide range of applications, such as constrained convex optimization in Yaji and Bhatnagar (2020, Section 7), performance ordering in distributed learning and reinforcement learning algorithms using Markovian samples, which is discussed in Section 3 of our paper. Assumption (B5) corresponds to Yaji and Bhatnagar (2020, Assumption A8). Besides, Yaji and Bhatnagar (2020, Assumptions A6, A7) are automatically satisfied since the noise terms therein are set to zero in (14). In the following, we provide the existing almost sure convergence result.

**Theorem A.1.1** (Yaji and Bhatnagar (2020) Theorem 4)**.** *Under Assumptions (B1) – (B5), for TTSA algorithm* (14)*, we have*

$$\begin{pmatrix}\mathbf{x}_n\\ \mathbf{y}_n\end{pmatrix} \xrightarrow[n\to\infty]{a.s.} \bigcup_{\mathbf{x}^*\in\Lambda} \begin{pmatrix}\mathbf{x}^*\\ \lambda(\mathbf{x}^*)\end{pmatrix}.$$

Our Assumptions (A1) – (A5) correspond to Assumptions (B1) – (B5). Consequently, Lemma 2.1 is a direct application of Yaji and Bhatnagar (2020, Theorem 4). Compared to assumption (B3), the additional conditions in our Assumption (A3), i.e., $\nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))$ is Hurwitz for every $\mathbf{x} \in \mathbb{R}^{d_1}$ and $\hat{h}_1(\mathbf{x}^*)$ is Hurwitz for $\mathbf{x}^* \in \Lambda$, are necessary for the proof of our CLT results in Theorem 2.2 and will be utilized in the following section.

## A.2 Proof of Theorem 2.2

Without loss of generality, the proof in this part is conditioned on the event that $\{\mathbf{x}_n \to \mathbf{x}^*, \mathbf{y}_n \to \mathbf{y}^* \triangleq \lambda(\mathbf{x}^*)\}$ for some $\mathbf{x}^* \in \Lambda$. The proof of Theorem 2.2 includes two parts: First, in Appendix A.2.1, we decompose the functions $h_1(\mathbf{x}, \mathbf{y}, \xi)$ and $h_2(\mathbf{x}, \mathbf{y}, \xi)$ of (14) into several terms and quantify the asymptotic behavior of each term. Second, in Appendix A.2.2, we partition those terms in each iteration into three parts (six sequences in total) and show that there is a sequence $L_n^{(\mathbf{x})}$ ($L_n^{(\mathbf{y})}$ resp.) in each iteration that contributes to weak convergence, while the remaining sequences diminish to zero when multiplied by the CLT scale $1/\sqrt{\beta_n}$ for iterates $\mathbf{x}_n$ ($1/\sqrt{\gamma_n}$ for iterates $\mathbf{y}_n$ resp.) so that they do not play a role in the final CLT result. The proofs of all technical lemmas used in Appendix A.2.1 and Appendix A.2.2 are deferred to Appendix A.2.3 – A.2.6 for better readability.

### A.2.1 Decomposition of Markovian Noise in the TTSA Algorithm

Throughout Appendix A.2, we define an operator $(\mathbf{P}m)(\mathbf{x}, \mathbf{y}, i)$ for any function $m : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \Xi \to \mathbb{R}^d$ as follows:

$$(\mathbf{P}m)(\mathbf{x}, \mathbf{y}, i) \triangleq \sum_{j\in\Xi} \mathbf{P}_{i,j}[\mathbf{x}, \mathbf{y}]m(\mathbf{x}, \mathbf{y}, j). \tag{16}$$

The ultimate goal in this subsection is to decompose (14) into

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1}\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n) + \beta_{n+1}M_{n+1}^{(\mathbf{x})} + \beta_{n+1}r_n^{(\mathbf{x},1)} + \beta_{n+1}r_n^{(\mathbf{x},2)}, \tag{17a}$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \gamma_{n+1}\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n) + \gamma_{n+1}M_{n+1}^{(\mathbf{y})} + \gamma_{n+1}r_n^{(\mathbf{y},1)} + \gamma_{n+1}r_n^{(\mathbf{y},2)}, \tag{17b}$$

where $M_{n+1}^{(\mathbf{x})}, M_{n+1}^{(\mathbf{y})}$ are two Martingale difference noise terms adapted to the filtration $\mathcal{F}_n \triangleq \{\mathbf{x}_0, \mathbf{y}_0, \xi_1, \cdots, \xi_n\}$. For $i = 1, 2$, $r_n^{(\mathbf{x},i)}, r_n^{(\mathbf{y},i)}$ are additional noise terms derived from Markovian noise and do not exist for *i.i.d.* sequence $\{\xi_n\}$. Therefore, these terms are absent from the previous CLT results for TTSA with Martingale

difference noise (Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006), where the *i.i.d.* sequence $\{\xi_n\}$ is their main focus.

We first rewrite (14) as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1}\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n) + \beta_{n+1}(h_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) - \bar{h}_1(\mathbf{x}_n, \mathbf{y}_n)), \tag{18a}$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \gamma_{n+1}\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n) + \gamma_{n+1}(h_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) - \bar{h}_2(\mathbf{x}_n, \mathbf{y}_n)). \tag{18b}$$

Then, given the underlying state-dependent Markov chain $\{\xi_n\}_{n\geq 0}$ with transition kernel $\mathbf{P}[\mathbf{x}, \mathbf{y}]$ that satisfies Assumption (A4), there exists a solution $m_1(\mathbf{x}, \mathbf{y}, \cdot) : \Xi \to \mathbb{R}^{d_1}$ to the following Poisson equation:

$$h_1(\mathbf{x}, \mathbf{y}, \xi) - \bar{h}_1(\mathbf{x}, \mathbf{y}) = m_1(\mathbf{x}, \mathbf{y}, \xi) - (\mathbf{P}m_1)(\mathbf{x}, \mathbf{y}, \xi). \tag{19}$$

Similarly, there exists a solution $m_2(\mathbf{x}, \mathbf{y}, \cdot) : \Xi \to \mathbb{R}^{d_2}$ to the following:

$$h_2(\mathbf{x}, \mathbf{y}, \xi) - \bar{h}_2(\mathbf{x}, \mathbf{y}) = m_2(\mathbf{x}, \mathbf{y}, \xi) - (\mathbf{P}m_2)(\mathbf{x}, \mathbf{y}, \xi). \tag{20}$$

This Poisson equation technique has been well discussed in Chen et al. (2020, Section 2) and Benveniste et al. (2012); Meyn (2022). For $l \in \{1, 2\}$, the explicit form of the solution $m_l$ to the corresponding Poisson equation (19) or (20) is given by

$$m_l(\mathbf{x}, \mathbf{y}, i) = \sum_{j\in\Xi}\sum_{k=0}^{\infty} \left(\mathbf{P}[\mathbf{x}, \mathbf{y}] - \mathbf{1}\boldsymbol{\pi}[\mathbf{x}, \mathbf{y}]^T\right)^k_{(i,j)} h_l(\mathbf{x}, \mathbf{y}, j) = \sum_{j\in\Xi} \left(\mathbf{I} - \mathbf{P}[\mathbf{x}, \mathbf{y}] + \mathbf{1}\boldsymbol{\pi}[\mathbf{x}, \mathbf{y}]^T\right)^{-1}_{(i,j)} h_l(\mathbf{x}, \mathbf{y}, j), \tag{21}$$

where $\boldsymbol{\pi}[\mathbf{x}, \mathbf{y}]$ is the stationary distribution of $\mathbf{P}[\mathbf{x}, \mathbf{y}]$, and $(i, j)$ represents the $(i, j)$-th entry of the corresponding matrix. The proof of (21) can be found in Delyon (2000, Appendix B.3.1 (B.16)) and Hu et al. (2022, Appendix B). Now, by (19), (20) and (21), we can further rewrite (18) as

$$\begin{aligned}
\mathbf{x}_{n+1} = \mathbf{x}_n &+ \beta_{n+1}\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n) + \beta_{n+1}\underbrace{(m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) - (\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n))}_{M^{(\mathbf{x})}_{n+1}} \\
&+ \beta_{n+1}\underbrace{((\mathbf{P}m_1)(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}, \xi_{n+1}) - (\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}))}_{r^{(\mathbf{x},1)}_n} \\
&+ \beta_{n+1}\underbrace{((\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) - (\mathbf{P}m_1)(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}, \xi_{n+1}))}_{r^{(\mathbf{x},2)}_n},
\end{aligned} \tag{22a}$$

and similarly,

$$\begin{aligned}
\mathbf{y}_{n+1} = \mathbf{y}_n &+ \gamma_{n+1}\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n) + \gamma_{n+1}\underbrace{(m_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) - (\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, \xi_n))}_{M^{(\mathbf{y})}_{n+1}} \\
&+ \gamma_{n+1}\underbrace{((\mathbf{P}m_2)(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}, \xi_{n+1}) - (\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}))}_{r^{(\mathbf{y},1)}_n} \\
&+ \gamma_{n+1}\underbrace{((\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) - (\mathbf{P}m_2)(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}, \xi_{n+1}))}_{r^{(\mathbf{y},2)}_n},
\end{aligned} \tag{22b}$$

which becomes (17). This kind of decomposition is well known for single-timescale SA with Markovian noise (Delyon, 2000; Benveniste et al., 2012; Fort, 2015; Fort et al., 2016), but now we need to deal with coupled variables $(\mathbf{x}_n, \mathbf{y}_n)$ for each iteration in (22).

Now, we further decompose the covariance of $M^{(\boldsymbol{\theta})}_{n+1}, M^{(\mathbf{x})}_{n+1}$ in Lemma A.2.1 and characterize the asymptotic behavior of each decomposed term in Lemma A.2.2, which will be used later in the proof of Lemma A.2.3 in the next subsection and is critical in quantifying the limiting covariance matrix in our main CLT result.

**Lemma A.2.1.** *For $M^{(\boldsymbol{\theta})}_{n+1}, M^{(\mathbf{x})}_{n+1}$ defined in (22), their covariance can be decomposed into the following forms:*

$$\mathbb{E}\left[M^{(\mathbf{x})}_{n+1}(M^{(\mathbf{x})}_{n+1})^T \middle| \mathcal{F}_n\right] = \mathbf{U}_{11} + \mathbf{D}^{(11)}_n + \mathbf{J}^{(11)}_n,$$

$$\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{y})})^T\middle|\mathcal{F}_n\right] = \mathbf{U}_{12} + \mathbf{D}_n^{(12)} + \mathbf{J}_n^{(12)}.$$
$$\mathbb{E}\left[M_{n+1}^{(\mathbf{y})}(M_{n+1}^{(\mathbf{x})})^T\middle|\mathcal{F}_n\right] = \mathbf{U}_{21} + \mathbf{D}_n^{(21)} + \mathbf{J}_n^{(21)}.$$
$$\mathbb{E}\left[M_{n+1}^{(\mathbf{y})}(M_{n+1}^{(\mathbf{y})})^T\middle|\mathcal{F}_n\right] = \mathbf{U}_{22} + \mathbf{D}_n^{(22)} + \mathbf{J}_n^{(22)}.$$

*where* $\mathbf{U}_{12} = \mathbf{U}_{21}^T$, $\mathbf{D}_n^{(12)} = (\mathbf{D}_n^{(21)})^T$, $\mathbf{J}_n^{(12)} = (\mathbf{J}_n^{(21)})^T$.

**Lemma A.2.2.** *For $i, j \in \{1, 2\}$, and with $\mathbf{U}_n^{(ij)}, \mathbf{D}_n^{(ij)}$ and $\mathbf{J}_n^{(ij)}$ as stated in Lemma A.2.1, we have*

$$\mathbf{U}_{ij} = \lim_{s\to\infty}\frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} h_i(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\right)\left(\sum_{n=1}^{s} h_j(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\right)^T\right],$$

$$\lim_{n\to\infty}\mathbf{D}_n^{(ij)} = 0 \quad a.s.$$

$$\lim_{n\to\infty}\gamma_n\mathbb{E}\left[\left\|\sum_{k=1}^{n}\mathbf{J}_k^{(ij)}\right\|\right] = 0 \quad a.s.$$

The proof of Lemmas A.2.1 and A.2.2 are located in Appendices A.2.3 and A.2.4.

*Remark* A.2.1. We note that for *i.i.d.* sequence $\{\xi_n\}$ with marginal $\boldsymbol{\mu}$, matrix $\mathbf{J}_n^{(ij)}$ in Lemma A.2.1 is *zero* for all time $n$, as indicated in the proof of Lemma A.2.2. Thus, the condition on $\mathbf{J}_n^{(ij)}$ in Lemma A.2.2 is automatically satisfied. Furthermore, the matrix $\mathbf{U}_{ij}$ in Lemma A.2.2 reduces to the marginal covariance $\mathbb{E}_{\xi\sim\boldsymbol{\mu}}[h_i(\mathbf{x}^*, \mathbf{y}^*, \xi)h_j(\mathbf{x}^*, \mathbf{y}^*, \xi)^T]$. Thus, Lemma A.2.1 and Lemma A.2.2 analyze the general Markovian sequence $\{\xi_n\}$ and include the *i.i.d.* sequence $\{\xi_n\}$ as a special case, whose resulting terms have been mainly analyzed in Konda and Tsitsiklis (2004); Mokkadem and Pelletier (2006).

*Remark* A.2.2. In Section 2.2, we originally obtain the matrix $\mathbf{U}_{\mathbf{x}}$ in the following form:

$$\mathbf{U}_{\mathbf{x}} = \lim_{s\to\infty}\frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s}\tilde{\Delta}_n^{\mathbf{x}}\right)\left(\sum_{n=1}^{s}\tilde{\Delta}_n^{\mathbf{x}}\right)^T\right],$$

where

$$\tilde{\Delta}_n^{\mathbf{x}} = h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n) - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n) = \begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}\begin{bmatrix}h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\\ h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\end{bmatrix}.$$

Then, $\mathbf{U}_{\mathbf{x}}$ can be rewritten as

$$\begin{aligned}\mathbf{U}_{\mathbf{x}} &= \lim_{s\to\infty}\frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s}\begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}\begin{bmatrix}h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\\ h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\end{bmatrix}\right)\left(\sum_{n=1}^{s}\begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}\begin{bmatrix}h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\\ h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\end{bmatrix}\right)^T\right]\\ &= \lim_{s\to\infty}\frac{1}{s}\begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}\mathbb{E}\left[\left(\sum_{n=1}^{s}\begin{bmatrix}h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\\ h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\end{bmatrix}\right)\left(\sum_{n=1}^{s}\begin{bmatrix}h_1(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\\ h_2(\mathbf{x}^*, \mathbf{y}^*, \xi_n)\end{bmatrix}\right)^T\right]\begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}^T \qquad (23)\\ &= \begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{U}_{21} & \mathbf{U}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\end{bmatrix}^T,\end{aligned}$$

where the last equality comes from the expression of $\mathbf{U}_{ij}$ in Lemma A.2.2.

### A.2.2 Analysis of Coupled Iterates $(\mathbf{x}_n, \mathbf{y}_n)$

In view of Lemma 2.1 (almost sure convergence) in Section 2.1, for large enough $n$, both iterates $\mathbf{x}_n, \mathbf{y}_n$ are close to the equilibrium $(\mathbf{x}^*, \mathbf{y}^*)$. In this part, we further decompose (17) in relation to the equilibrium. To do so, we first apply the Taylor expansion to functions $\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n)$ and $\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n)$ in (17) at $(\mathbf{x}^*, \mathbf{y}^*)$, which results in

$$\bar{h}_1(\mathbf{x}_n, \mathbf{y}_n) = \nabla_{\mathbf{x}}\bar{h}_1(\mathbf{x}^*, \mathbf{y}^*)(\mathbf{x}_n - \mathbf{x}^*) + \nabla_{\mathbf{y}}\bar{h}_1(\mathbf{x}^*, \mathbf{y}^*)(\mathbf{y}_n - \mathbf{y}^*) + O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2),$$

$$\bar{h}_2(\mathbf{x}_n, \mathbf{y}_n) = \nabla_{\mathbf{x}} \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*)(\mathbf{x}_n - \mathbf{x}^*) + \nabla_{\mathbf{y}} \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*)(\mathbf{y}_n - \mathbf{y}^*) + O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2).$$

Denote by

$$\mathbf{Q}_{11} \triangleq \nabla_{\mathbf{x}} \bar{h}_1(\mathbf{x}^*, \mathbf{y}^*), \quad \mathbf{Q}_{12} \triangleq \nabla_{\mathbf{y}} \bar{h}_1(\mathbf{x}^*, \mathbf{y}^*), \quad \mathbf{Q}_{21} \triangleq \nabla_{\mathbf{x}} \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*), \quad \mathbf{Q}_{22} \triangleq \nabla_{\mathbf{y}} \bar{h}_2(\mathbf{x}^*, \mathbf{y}^*),$$

we can rewrite (17) as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \beta_{n+1} \left( \mathbf{Q}_{11}(\mathbf{x}_n - \mathbf{x}^*) + \mathbf{Q}_{12}(\mathbf{y}_n - \mathbf{y}^*) + M_{n+1}^{(\mathbf{x})} + r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} \right), \tag{24a}$$

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \gamma_{n+1} \left( \mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*) + \mathbf{Q}_{22}(\mathbf{y}_n - \mathbf{y}^*) + M_{n+1}^{(\mathbf{y})} + r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})} \right), \tag{24b}$$

where $\rho_n^{(\mathbf{x})} = O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2)$, $\rho_n^{(\mathbf{y})} = O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2)$ are the error terms coming from Taylor expansion. Then, we rewrite iterates $\mathbf{y}_n - \mathbf{y}^*$ in (24b) as

$$\mathbf{y}_n - \mathbf{y}^* = \gamma_{n+1}^{-1} \mathbf{Q}_{22}^{-1}(\mathbf{y}_{n+1} - \mathbf{y}_n) - \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*) - \mathbf{Q}_{22}^{-1} \left( M_{n+1}^{(\mathbf{y})} + r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})} \right).$$

Substituting the above into the $(\mathbf{y}_n - \mathbf{y}^*)$ term in (24a) yields the following:

$$\begin{aligned}
\mathbf{x}_{n+1} - \mathbf{x}^* = \; & \mathbf{x}_n - \mathbf{x}^* + \beta_{n+1} \Big( \mathbf{Q}_{11}(\mathbf{x}_n - \mathbf{x}^*) + \gamma_{n+1}^{-1} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}(\mathbf{y}_{n+1} - \mathbf{y}_n) - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*) \\
& - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \left( M_{n+1}^{(\mathbf{y})} + r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})} \right) + M_{n+1}^{(\mathbf{x})} + r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} \Big) \\
= \; & (\mathbf{I} + \beta_{n+1} \mathbf{K}_{\mathbf{x}}) (\mathbf{x}_n - \mathbf{x}^*) + \beta_{n+1} \left( \gamma_{n+1}^{-1} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}(\mathbf{y}_{n+1} - \mathbf{y}_n) \right) + \beta_{n+1} \left( M_{n+1}^{(\mathbf{x})} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} M_{n+1}^{(\mathbf{y})} \right) \\
& + \beta_{n+1} \left( r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} (r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})}) \right),
\end{aligned} \tag{25}$$

where $\mathbf{K}_{\mathbf{x}} = \mathbf{Q}_{11} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21}$. As we observe from (25), iterates $\mathbf{x}_n$ naturally embed two sequences:

(i) $\beta_{n+1} \gamma_{n+1}^{-1} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}(\mathbf{y}_{n+1} - \mathbf{y}_n)$; (ii) $\beta_{n+1}(M_{n+1}^{(\mathbf{x})} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} M_{n+1}^{(\mathbf{y})})$.

These sequences can be expressed recursively by following similar steps as in Delyon (2000); Mokkadem and Pelletier (2006); Fort (2015). Specifically, let

$$u_n \triangleq \sum_{k=1}^{n} \beta_k, \quad s_n \triangleq \sum_{k=1}^{n} \gamma_k. \tag{26}$$

Then, the sequences (i) and (ii) can be used to drive the following iterations respectively:

$$R_n^{(\mathbf{x})} \triangleq e^{\beta_n \mathbf{K}_{\mathbf{x}}} R_{n-1}^{(\mathbf{x})} + \beta_n \gamma_n^{-1} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}(\mathbf{y}_n - \mathbf{y}_{n-1}) = \sum_{k=1}^{n} e^{(u_n - u_k)\mathbf{K}_{\mathbf{x}}} \beta_k \gamma_k^{-1} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}(\mathbf{y}_k - \mathbf{y}_{k-1}). \tag{27}$$

$$L_n^{(\mathbf{x})} \triangleq e^{\beta_n \mathbf{K}_{\mathbf{x}}} L_{n-1}^{(\mathbf{x})} + \beta_n (M_n^{(\mathbf{x})} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} M_n^{(\mathbf{y})}) = \sum_{k=1}^{n} e^{(u_n - u_k)\mathbf{K}_{\mathbf{x}}} \beta_k (M_k^{(\mathbf{x})} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} M_k^{(\mathbf{x})}), \tag{28}$$

The remaining noise in the iterates $\mathbf{x}_n$ is defined as $\Delta_{n+1}^{(\mathbf{x})} \triangleq \mathbf{x}_{n+1} - \mathbf{x}^* - L_{n+1}^{(\mathbf{x})} - R_{n+1}^{(\mathbf{x})}$.

Similarly, for iterates $\mathbf{y}_n$ in (24b), we define the following sequences:

$$R_n^{(\mathbf{y})} = e^{\gamma_n \mathbf{Q}_{22}} R_{n-1}^{(\mathbf{y})} + \gamma_n \mathbf{Q}_{21}(L_{n-1}^{(\mathbf{x})} + R_{n-1}^{(\mathbf{x})}) = \sum_{k=1}^{n} e^{(s_n - s_k)\mathbf{Q}_{22}} \gamma_k \mathbf{Q}_{21}(L_{k-1}^{(\mathbf{x})} + R_{k-1}^{(\mathbf{x})}). \tag{29}$$

$$L_n^{(\mathbf{y})} = e^{\gamma_n \mathbf{Q}_{22}} L_{n-1}^{(\mathbf{y})} + \gamma_n M_n^{(\mathbf{y})} = \sum_{k=1}^{n} e^{(s_n - s_k)\mathbf{Q}_{22}} \gamma_k M_k^{(\mathbf{y})}, \tag{30}$$

The remaining noise in iterates $\mathbf{y}_n$ is denoted as $\Delta_{n+1}^{(\mathbf{y})} \triangleq \mathbf{y}_{n+1} - \mathbf{y}^* - L_{n+1}^{(\mathbf{y})} - R_{n+1}^{(\mathbf{y})}$.

In view of Lemmas A.2.1 and A.2.2, we have the following results characterizing the weak convergence of sequences $\beta_n^{-1/2} L_n^{(\mathbf{x})}$ and $\gamma_n^{-1/2} L_n^{(\mathbf{y})}$, which are proved in Appendix A.2.5.

**Lemma A.2.3.** *For $L_n^{(\mathbf{x})}$ and $L_n^{(\mathbf{y})}$ defined in (28) and (30), we have*

$$\begin{pmatrix} \sqrt{\beta_n^{-1}} L_n^{(\mathbf{x})} \\ \sqrt{\gamma_n^{-1}} L_n^{(\mathbf{y})} \end{pmatrix} \xrightarrow[n\to\infty]{dist} N\left(0, \begin{pmatrix} \mathbf{V_x} & 0 \\ 0 & \mathbf{V_y} \end{pmatrix}\right),$$

*where*

$$\mathbf{V_x} = \int_0^\infty e^{t\left(\mathbf{K_x} + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right)} \mathbf{U_x} e^{t\left(\mathbf{K_x} + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right)^T} dt,$$
$$\mathbf{V_y} = \int_0^\infty e^{t\mathbf{Q}_{22}} \mathbf{U}_{22} e^{t\mathbf{Q}_{22}^T} dt,$$

*with $\mathbf{U_x}$ of the form (23) and $\mathbf{U}_{22}$ defined in Lemma A.2.2.*

Lemma A.2.3 confirms that $\beta_n^{-1/2} L_n^{(\mathbf{x})}$ and $\gamma_n^{-1/2} L_n^{(\mathbf{y})}$ weakly converge to Gaussian multivariate distributions, mirroring the weak convergence of $\beta_n^{-1/2}(\mathbf{x}_n - \mathbf{x}^*)$ and $\gamma_n^{-1/2}(\mathbf{y}_n - \mathbf{y}^*)$ as described in Theorem 2.2 (Section 2.3). Furthermore, we establish the following results that the sequences $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}, \Delta_n^{(\mathbf{x})}$, and $\Delta_n^{(\mathbf{y})}$ decay to zero at rates faster than their respective CLT scales. This is achieved by examining a bounded deterministic sequence that iteratively refines the upper bounds of $\mathbf{y}_n - \mathbf{y}^*$ and $\Delta_n^{(\mathbf{y})}$. Detailed insights are provided in Appendix A.2.6.

**Lemma A.2.4.** *We have*

1. *for some constant $c > b/2$, $\|R_n^{(\mathbf{x})}\| = O(n^{-c})$ a.s*
2. *$\|R_n^{(\mathbf{y})}\| = O(\sqrt{\beta_n \log u_n})$ a.s.*
3. *$\|\Delta_n^{(\mathbf{x})}\| = o(\sqrt{\beta_n})$ a.s.*
4. *$\|\Delta_n^{(\mathbf{y})}\| = o(\sqrt{\beta_n})$ a.s.*

Note that $\sqrt{\beta_n/\gamma_n \log u_n} = O(\sqrt{n^{a-b}\log n}) = o(1)$ since $a - b < 0$ by Assumption (A1). Therefore, from Lemma A.2.4, we have $\beta_n^{-1/2}(R_n^{(\mathbf{x})} + \Delta_n^{(\mathbf{x})}) \to 0$ and $\gamma_n^{-1/2}(R_n^{(\mathbf{y})} + \Delta_n^{(\mathbf{y})}) \to 0$ almost surely. By Lemma A.2.3, we establish the weak convergence for $\beta_n^{-1/2} L_n^{(\mathbf{x})}$ and $\gamma_n^{-1/2} L_n^{(\mathbf{y})}$. Together with $\mathbf{x}_n - \mathbf{x}^* = L_n^{(\mathbf{x})} + R_n^{(\mathbf{x})} + \Delta_n^{(\mathbf{x})}$ and $\mathbf{y}_n - \mathbf{y}^* = L_n^{(\mathbf{y})} + R_n^{(\mathbf{y})} + \Delta_n^{(\mathbf{y})}$, we show that

$$\beta_n^{-1/2}(\mathbf{x}_n - \mathbf{x}^*) \xrightarrow[n\to\infty]{dist} N(0, \mathbf{V_x}), \quad \gamma_n^{-1/2}(\mathbf{y}_n - \mathbf{y}^*) \xrightarrow[n\to\infty]{dist} N(0, \mathbf{V_y}),$$

which completes the proof.

### A.2.3 Proof of Lemma A.2.1

We decompose the covariance form of $M_{n+1}^{(\mathbf{x})}$ and $M_{n+1}^{(\mathbf{y})}$ into several terms using the Poisson equation method (Benveniste et al., 2012; Fort, 2015; Chen et al., 2020). The detailed steps are as follows.

$$\begin{aligned} &\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{x})})^T \middle| \mathcal{F}_n\right] \\ &= \mathbb{E}[m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1})^T | \mathcal{F}_n] + (\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) \left[(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)\right]^T \\ &\quad - \mathbb{E}[m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) | \mathcal{F}_n] \left[(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)\right]^T - (\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) \mathbb{E}[m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) | \mathcal{F}_n]^T \\ &= \mathbb{E}[m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1})^T | \mathcal{F}_n] - (\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) \left[(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)\right]^T. \end{aligned} \tag{31}$$

where the second equality is because $(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)$, as defined in (16), can also be written as

$$(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) = \mathbb{E}[m_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) | \mathcal{F}_n]. \tag{32}$$

Similarly, we have

$$\mathbb{E}\left[M_{n+1}^{(\mathbf{y})}(M_{n+1}^{(\mathbf{y})})^T \middle| \mathcal{F}_n\right] = \mathbb{E}[m_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) m_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1})^T | \mathcal{F}_n] - (\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, \xi_n) \left[(\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)\right]^T,$$

and

$$\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{y})})^T\middle|\mathcal{F}_n\right] = \mathbb{E}[m_1(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})m_2(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})^T|\mathcal{F}_n] - (\mathbf{P}m_1)(\mathbf{x}_n,\mathbf{y}_n,\xi_n)\left[(\mathbf{P}m_2)(\mathbf{x}_n,\mathbf{y}_n,\xi_n)\right]^T.$$

We now focus on $\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{x})})^T\middle|\mathcal{F}_n\right]$ as an example. The same steps of decomposition can be extrapolated to $\mathbb{E}\left[M_{n+1}^{(\mathbf{y})}(M_{n+1}^{(\mathbf{y})})^T\middle|\mathcal{F}_n\right]$ and $\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{y})})^T\middle|\mathcal{F}_n\right]$, and their proofs are omitted in this part to avoid repetition and maintain brevity. Define

$$\phi(\mathbf{x},\mathbf{y},i) \triangleq \sum_{j\in\Xi}\mathbf{P}_{i,j}[\mathbf{x},\mathbf{y}]m_1(\mathbf{x},\mathbf{y},j)m_1(\mathbf{x},\mathbf{y},j)^T - (\mathbf{P}m_1)(\mathbf{x},\mathbf{y},i)\left[(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},i)\right]^T, \tag{33}$$

and let its expectation w.r.t the stationary distribution $\boldsymbol{\pi}[\mathbf{x},\mathbf{y}]$ be $\bar{\phi}(\mathbf{x},\mathbf{y}) \triangleq \mathbb{E}_{i\sim\boldsymbol{\pi}[\mathbf{x},\mathbf{y}]}[\phi(\mathbf{x},\mathbf{y},i)]$. We can construct *another Poisson equation*, i.e.,

$$\begin{aligned}
&\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{x})})^T\middle|\mathcal{F}_n\right] - \sum_{\xi_n\in\Xi}\pi_{\xi_n}[\mathbf{x}_n,\mathbf{y}_n]\mathbb{E}\left[M_{n+1}^{(\mathbf{x})}(M_{n+1}^{(\mathbf{x})})^T\middle|\mathcal{F}_n\right]\\
&= \phi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) - \bar{\phi}(\mathbf{x}_n,\mathbf{y}_n)\\
&= \varphi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}),
\end{aligned}$$

with the matrix-valued function $\varphi:\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\times\Xi\to\mathbb{R}^{d_1\times d_1}$ as its solution. Then, we have

$$\begin{aligned}
\phi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) =& \underbrace{\bar{\phi}(\mathbf{x}^*,\mathbf{y}^*)}_{\mathbf{U}_{11}} + \underbrace{\bar{\phi}(\mathbf{x}_n,\mathbf{y}_n) - \bar{\phi}(\mathbf{x}^*,\mathbf{y}^*)}_{\mathbf{D}_n^{(11)}} + \underbrace{\varphi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_n)}_{\mathbf{J}_n^{(11,A)}}\\
&+ \underbrace{(\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_n) - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})}_{\mathbf{J}_n^{(11,B)}},
\end{aligned} \tag{34}$$

where matrix $\mathbf{J}_n^{(11)}$ defined in Lemma A.2.1 given by $\mathbf{J}_n^{(11)} \triangleq \mathbf{J}_n^{(11,A)} + \mathbf{J}_n^{(11,B)}$. This completes the proof.

### A.2.4 Proof of Lemma A.2.2

**Expression of matrices $\mathbf{U}_{ij}$.** We first give the exact expression of $\bar{\phi}(\mathbf{x}^*,\mathbf{y}^*)$ defined in Appendix A.2.3 in order to derive matrix $\mathbf{U}_{ij}$ for $i,j\in\{1,2\}$. Recall the explicit form of the solution to the Poisson equation as studied in Delyon (2000, Appendix B.3.1 (B.16)) and Hu et al. (2022, Appendix B), with $\boldsymbol{\pi}[\mathbf{x}^*,\mathbf{y}^*] = \boldsymbol{\mu}$, we have

$$\begin{aligned}
m_1(\mathbf{x}^*,\mathbf{y}^*,i) &= \sum_{j\in\Xi}\sum_{k=0}^{\infty}\left(\mathbf{P}[\mathbf{x}^*,\mathbf{y}^*] - \mathbf{1}\boldsymbol{\mu}^T\right)^k_{(i,j)}h_1(\mathbf{x}^*,\mathbf{y}^*,j)\\
&= \sum_{j\in\Xi}\sum_{k=0}^{\infty}\left(\mathbf{P}^k_{i,j}[\mathbf{x}^*,\mathbf{y}^*]h_1(\mathbf{x}^*,\mathbf{y}^*,j) - \underbrace{\bar{h}_1(\mathbf{x}^*,\mathbf{y}^*)}_{=\mathbf{0}}\right)\\
&= \mathbb{E}\left[\sum_{k=0}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\middle|\xi_0 = i\right],
\end{aligned}$$

where the second equality comes from $(\mathbf{P}[\mathbf{x}^*,\mathbf{y}^*] - \mathbf{1}\boldsymbol{\mu}^T)^k = \mathbf{P}[\mathbf{x}^*,\mathbf{y}^*]^k - \mathbf{1}\boldsymbol{\mu}^T$ for $k\geq 1$ by induction. The last equality is by rewriting $m_1(\mathbf{x}^*,\mathbf{y}^*,i)$ into a conditional expectation form on the Markov chain $\{\xi_n\}$ conditioned on $\xi_0 = i$. Similarly, for $(\mathbf{P}m_1)(\mathbf{x}^*,\mathbf{y}^*,i)$, we have

$$(\mathbf{P}m_1)(\mathbf{x}^*,\mathbf{y}^*,i) = \mathbb{E}\left[\sum_{k=1}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\middle|\xi_0 = i\right].$$

Hence, for function $\phi(\mathbf{x},\mathbf{y},i)$ defined in (33), taking the expectation over $\xi_0 \sim \boldsymbol{\mu}$, i.e., the underlying Markov chain is in the stationary regime from the beginning, we have

$$\begin{aligned}
\bar{\phi}(\mathbf{x}^*,\mathbf{y}^*) &= \sum_{i\in\Xi}\left[\mu_i m_1(\mathbf{x},\mathbf{y},i)m_1(\mathbf{x},\mathbf{y},i)^T - \mu_i(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},i)\left[(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},i)\right]^T\right],\\
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\left(\sum_{k=0}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)\left(\sum_{k=0}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)^T\right] - \mathbb{E}_{\boldsymbol{\mu}}\left[\left(\sum_{k=1}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)\left(\sum_{k=1}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)^T\right]\\
&= \mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)^T] + \mathbb{E}_{\boldsymbol{\mu}}\left[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)\left(\sum_{k=1}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)^T\right]\\
&\quad + \mathbb{E}_{\boldsymbol{\mu}}\left[\left(\sum_{k=1}^{\infty}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)^T\right]\\
&= \sum_{k=1}^{\infty}\left[\mathrm{Cov}(h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0),h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)) + \mathrm{Cov}(h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k),h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0))\right]\\
&\quad + \mathrm{Cov}(h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0),h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)),
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{Cov}(h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0),h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)) &\triangleq \mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)^T] - \mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)]\mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)]^T\\
&= \mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)^T] - \underbrace{\bar{h}_1(\mathbf{x}^*,\mathbf{y}^*)}_{\mathbf{0}}\bar{h}_1(\mathbf{x}^*,\mathbf{y}^*)^T\\
&= \mathbb{E}_{\boldsymbol{\mu}}[h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)^T]
\end{aligned}$$

is the covariance between $h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_0)$ and $h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)$ for the Markov chain $\{\xi_n\}$ in the stationary regime. By following steps similar to Brémaud (2013, Proof of Theorem 6.3.7) using $h_1(\mathbf{x}^*,\mathbf{y}^*,\xi)$ as the test function, we get

$$\bar{\phi}(\mathbf{x}^*,\mathbf{y}^*) = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\left(\sum_{k=0}^{n}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)\left(\sum_{k=0}^{n}h_1(\mathbf{x}^*,\mathbf{y}^*,\xi_k)\right)^T\right] = \mathbf{U}_{11}. \tag{35}$$

We can follow the same procedures as above to derive $\mathbf{U}_{12},\mathbf{U}_{21}$ and $\mathbf{U}_{22}$.

**Analysis of matrices $\mathbf{D}_n^{(ij)}$.** To analyze $\mathbf{D}_n^{(ij)}$, we first discuss the continuity of functions $m_1, m_2$ in (21), and $(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},\xi),(\mathbf{P}m_2)(\mathbf{x},\mathbf{y},\xi)$ defined in (16). By Assumption (A4), the transition kernel $\mathbf{P}[\mathbf{x},\mathbf{y}]$ and its corresponding stationary distribution $\boldsymbol{\pi}[\mathbf{x},\mathbf{y}]$ are continuous in $\mathbf{x},\mathbf{y}$, and the inverse $(\mathbf{I}-\mathbf{P}[\mathbf{x},\mathbf{y}]+\mathbf{1}\boldsymbol{\pi}[\mathbf{x},\mathbf{y}]^T)^{-1}$ is well defined and continuous in $\mathbf{x},\mathbf{y}$. This, together with the continuous functions $h_1,h_2$ assumed in Assumption (A2), leads to the continuity of functions $m_1(\mathbf{x},\mathbf{y},\xi), m_2(\mathbf{x},\mathbf{y},\xi)$ and $(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},\xi),(\mathbf{P}m_2)(\mathbf{x},\mathbf{y},\xi)$ in $\mathbf{x}\in\mathbb{R}^{d_1},\mathbf{y}\in\mathbb{R}^{d_2}$ for any $\xi\in\Xi$. Furthermore, this results in the conclusion that the function $\phi(\mathbf{x},\mathbf{y},i)$ defined in (33) is also continuous in $\mathbf{x}\in\mathbb{R}^{d_1},\mathbf{y}\in\mathbb{R}^{d_2}$ for any $\xi\in\Xi$. Thus, its mean field $\bar{\phi}(\mathbf{x},\mathbf{y})$ is continuous in $\mathbf{x}\in\mathbb{R}^{d_1},\mathbf{y}\in\mathbb{R}^{d_2}$ as well. By the almost sure convergence $\mathbf{x}_n\to\mathbf{x}^*$ and $\mathbf{y}_n\to\mathbf{y}^*$, along with the continuity of function $\bar{\phi}(\mathbf{x},\mathbf{y})$, we have

$$\lim_{n\to\infty}\mathbf{D}_n^{(11)} = \lim_{n\to\infty}\bar{\phi}(\mathbf{x}_n,\mathbf{y}_n) - \bar{\phi}(\mathbf{x}^*,\mathbf{y}^*) = 0 \quad a.s.$$

Similarly, we draw the same conclusion for $\mathbf{D}_n^{(12)},\mathbf{D}_n^{(21)}$ and $\mathbf{D}_n^{(22)}$.

**Analysis of matrices $\mathbf{J}_n^{(ij)}$.** We still focus on the case where $i=j=1$, with other cases following the same steps. As demonstrated in (34), we can decompose $\mathbf{J}_n^{(11)} = \mathbf{J}_n^{(11,A)} + \mathbf{J}_n^{(11,B)}$, where

$$\mathbf{J}_n^{(11,A)} = \varphi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_n), \quad \mathbf{J}_n^{(11,B)} = (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_n) - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}).$$

$\mathbf{J}_n^{(11,A)}$ is a Martingale difference term adapted to $\mathcal{F}_n$, i.e.,

$$\mathbb{E}[\mathbf{J}_n^{(11,A)}|\mathcal{F}_n] = \mathbb{E}[\varphi(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})|\mathcal{F}_n] - (\mathbf{P}\varphi)(\mathbf{x}_n,\mathbf{y}_n,\xi_n) = 0$$

due to the definition of $(\mathbf{P}\varphi)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)$ as stated in (16). Using the Burkholder inequality from Lemma A.6.2 with $p = 1$, for some constant $C_1 > 0$ we get

$$\mathbb{E}\left[\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,A)}\right\|\right] \leq C_1 \mathbb{E}\left[\sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)}\right]. \tag{36}$$

By Assumption (A5), iterates $\mathbf{x}_n, \mathbf{y}_n$ are always within some compact set $\Omega$ such that $\sup_n \left\|\mathbf{J}_n^{(11,A)}\right\| \leq C_\Omega < \infty$ for a set-dependent constant $C_\Omega$, and thus

$$\gamma_n C_1 \sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)} \leq C_1 C_\Omega \gamma_n \sqrt{n}. \tag{37}$$

The last term of (37) decreases to zero in $n$ due to $a > 1/2$ in Assumption (A1) and is therefore uniformly bounded with respect to $n$.

For $\mathbf{J}_n^{(11,B)}$, we use the Abel transformation in Lemma A.6.1 to obtain

$$\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)} = \sum_{k=1}^{n} [(\mathbf{P}\varphi)(\mathbf{x}_k, \mathbf{y}_k, \xi_{k-1}) - (\mathbf{P}\varphi)(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}, \xi_{k-1})] + (\mathbf{P}\varphi)(\mathbf{x}_0, \mathbf{y}_0, \xi_0) - (\mathbf{P}\varphi)(\mathbf{x}_n, \mathbf{y}_n, \xi_n).$$

Following same steps to derive the continuity of function $m_1$ in the previous paragraph, we have that the matrix-valued function $(\mathbf{P}\varphi)(\mathbf{x}, \mathbf{y}, \xi)$ is continuous in $\mathbf{x} \in \mathbb{R}^{d_1}, \mathbf{y} \in \mathbb{R}^{d_2}$. Thus, by Assumption (A5) that $(\mathbf{x}_n, \mathbf{y}_n)$ are within some compact set $\Omega$, there exists a constant $L_\Omega$ such that

$$\|(\mathbf{P}\varphi)(\mathbf{x}_k, \mathbf{y}_k, \xi_{k-1}) - (\mathbf{P}\varphi)(\mathbf{x}_{k-1}, \mathbf{y}_{k-1}, \xi_{k-1})\| \leq L_\Omega(\|\mathbf{x}_k - \mathbf{x}_{k-1}\| + \|\mathbf{y}_k - \mathbf{y}_{k-1}\|) \leq C'_\Omega L_\Omega(\beta_k + \gamma_k),$$

where $C'_\Omega = \max_{(\mathbf{x},\mathbf{y})\in\Omega, \xi\in\Xi}\{h_1(\mathbf{x}, \mathbf{y}, \xi), h_2(\mathbf{x}, \mathbf{y}, \xi)\}$. Also, $\|(\mathbf{P}\varphi)(\mathbf{x}_0, \mathbf{y}_0, \xi_0)\| + \|(\mathbf{P}\varphi)(\mathbf{x}_n, \mathbf{y}_n, \xi_n)\|$ are upper-bounded by some positive constant $C''_\Omega$. From the above, we have

$$\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\| \leq C''_\Omega + C'_\Omega L_\Omega \sum_{k=1}^{n} (\beta_k + \gamma_k) \leq C''_\Omega + C'''_\Omega \sum_{k=1}^{n} \gamma_k \tag{38}$$

for some set-dependent constant $C'''_\Omega > 0$. Note that

$$\gamma_n \left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\| \leq \gamma_n C''_\Omega + C'''_\Omega \gamma_n \sum_{k=1}^{n} \gamma_k < \gamma_n C''_\Omega + \frac{C'''_\Omega}{a} n^{1-2a} \tag{39}$$

where the last inequality is from $\sum_{k=1}^{n} \gamma_k < \frac{1}{a} n^{1-a}$. We observe that (39) is decreasing to zero in $n$ due to $a > 1/2$ and is thus uniformly bounded with respect to $n$.

Note that $\mathbf{J}_k^{(11)} = \mathbf{J}_k^{(11,A)} + \mathbf{J}_k^{(11,B)}$, by triangular inequality we have

$$\begin{aligned}
\gamma_n \mathbb{E}\left[\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11)}\right\|\right] &\leq \gamma_n \mathbb{E}\left[\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,A)}\right\|\right] + \gamma_n \mathbb{E}\left[\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\|\right] \\
&\leq \gamma_n C_1 \mathbb{E}\left[\sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)}\right] + \gamma_n \mathbb{E}\left[\left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\|\right] \\
&= \mathbb{E}\left[\gamma_n C_1 \sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)} + \gamma_n \left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\|\right],
\end{aligned} \tag{40}$$

where the second inequality comes from (36). By (37) and (39) we know that both terms in the last line of (40) are bounded by constants that depend on the set $\Omega$. Therefore, by dominated convergence theorem, taking the limit over the last line of (40) gives

$$\lim_{n\to\infty} \mathbb{E}\left[\gamma_n C_1 \sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)} + \gamma_n \left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\|\right] = \mathbb{E}\left[\lim_{n\to\infty} \gamma_n C_1 \sqrt{\left(\sum_{k=1}^{n} \left\|\mathbf{J}_k^{(11,A)}\right\|^2\right)} + \gamma_n \left\|\sum_{k=1}^{n} \mathbf{J}_k^{(11,B)}\right\|\right] = 0.$$

Therefore, we have

$$\lim_{n\to\infty}\gamma_n\mathbb{E}\left[\left\|\sum_{k=1}^{n}\mathbf{J}_k^{(11)}\right\|\right]=0,$$

which completes the proof.

### A.2.5 Proof of Lemma A.2.3

Define a Martingale $Z^{(n)}=\{Z_k^{(n)}\}_{k\geq 1}$ such that

$$Z_k^{(n)}\triangleq\begin{pmatrix}\sqrt{\beta_n^{-1}}L_n^{(\mathbf{x})}\\ \sqrt{\gamma_n^{-1}}L_n^{(\mathbf{y})}\end{pmatrix}=\begin{pmatrix}\beta_n^{-1/2}e^{u_n\mathbf{K_x}} & 0\\ 0 & \gamma_n^{-1/2}e^{s_n\mathbf{Q}_{22}}\end{pmatrix}\times\sum_{j=1}^{k}\begin{pmatrix}e^{-u_k\mathbf{K_x}}\beta_k(M_k^{(\mathbf{x})}-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}M_k^{(\mathbf{y})})\\ e^{-s_k\mathbf{Q}_{22}}\gamma_k M_k^{(\mathbf{y})}\end{pmatrix}.$$

Then, the Martingale difference array $Z_k^{(n)}-Z_{k-1}^{(n)}$ becomes

$$Z_k^{(n)}-Z_{k-1}^{(n)}=\begin{pmatrix}\beta_n^{-1/2}e^{(u_n-u_k)\mathbf{K_x}}\beta_k(M_k^{(\mathbf{x})}-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}M_k^{(\mathbf{y})})\\ \gamma_n^{-1/2}e^{(s_n-s_k)\mathbf{Q}_{22}}\gamma_k M_k^{(\mathbf{y})}\end{pmatrix}$$

such that

$$\sum_{k=1}^{n}\mathbb{E}\left[(Z_k^{(n)}-Z_{k-1}^{(n)})(Z_k^{(n)}-Z_{k-1}^{(n)})^T|\mathcal{F}_{k-1}\right]=\begin{pmatrix}S_n^{(11)} & S_n^{(12)}\\ S_n^{(21)} & S_n^{(22)}\end{pmatrix},$$

where, in view of decomposition of $M_n^{(\mathbf{x})}$ and $M_n^{(\mathbf{y})}$ in Lemma A.2.1, we have

$$\begin{aligned}S_n^{(11)}=\beta_n^{-1}\sum_{k=1}^{n}\beta_k^2e^{(u_n-u_k)\mathbf{K_x}}\Big(&\mathbf{U}_{11}+\mathbf{D}_k^{(11)}+\mathbf{J}_k^{(11)}-(\mathbf{U}_{12}+\mathbf{D}_k^{(12)}+\mathbf{J}_k^{(12)})(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T\\&+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(\mathbf{U}_{22}+\mathbf{D}_k^{(22)}+\mathbf{J}_k^{(22)})(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T\\&-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(\mathbf{U}_{21}+\mathbf{D}_k^{(21)}+\mathbf{J}_k^{(21)})\Big)e^{(u_n-u_k)\mathbf{K}_{\mathbf{x}}^T},\end{aligned}\tag{41a}$$

$$S_n^{(12)}=\beta_n^{-1/2}\gamma_n^{-1/2}\sum_{k=1}^{n}\beta_k\gamma_k e^{(u_n-u_k)\mathbf{K_x}}(\mathbf{U}_{12}-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{U}_{22})e^{(s_n-s_k)\mathbf{Q}_{22}^T},\tag{41b}$$

$$S_n^{(22)}=\gamma_n^{-1}\sum_{k=1}^{n}\gamma_k^2e^{(s_n-s_k)\mathbf{Q}_{22}}(\mathbf{U}_{22}+\mathbf{D}_k^{(22)}+\mathbf{J}_k^{(22)})e^{(s_n-s_k)\mathbf{Q}_{22}^T},\tag{41c}$$

and $S_n^{(21)}=(S_n^{(12)})^T$.

We now focus on $S_n^{(11)}$, whose property is given by the following lemma.

**Lemma A.2.5.** $\lim_{n\to\infty}S_n^{(11)}=\mathbf{V_x}$, *where* $\mathbf{V_x}$ *is of the form in Lemma A.2.3.*

*Proof.* We rewrite $S_n^{(11)}$ into three parts:

$$\begin{aligned}
&S_n^{(11)}\\
&= \underbrace{\beta_n^{-1}\sum_{k=1}^{n}\left(\beta_k^2 e^{(u_n-u_k)\mathbf{K_x}}(\mathbf{U}_{11}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{U}_{22}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{U}_{12}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{U}_{21})e^{(u_n-u_k)\mathbf{K}_\mathbf{x}^T}\right)}_{S_n^{(11,A)}}\\
&+ \underbrace{\beta_n^{-1}\sum_{k=1}^{n}\left(\beta_k^2 e^{(u_n-u_k)\mathbf{K_x}}(\mathbf{D}_k^{(11)}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(22)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{D}_k^{(12)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(21)})e^{(u_n-u_k)\mathbf{K}_\mathbf{x}^T}\right)}_{S_n^{(11,B)}}\\
&+ \underbrace{\beta_n^{-1}\sum_{k=1}^{n}\left(\beta_k^2 e^{(u_n-u_k)\mathbf{K_x}}(\mathbf{J}_k^{(11)}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{J}_k^{(22)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{J}_k^{(12)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{J}_k^{(21)})e^{(u_n-u_k)\mathbf{K}_\mathbf{x}^T}\right)}_{S_n^{(11,C)}}.
\end{aligned} \tag{42}$$

We aim to demonstrate that

$$\lim_{n\to\infty} S_n^{(11,A)} = \mathbf{V_x}, \quad \lim_{n\to\infty}\|S_n^{(11,B)}\| = 0, \quad \lim_{n\to\infty}\|S_n^{(11,C)}\| = 0.$$

From Lemma A.6.4, we have for some $c, T > 0$ such that

$$\|S_n^{(11,B)}\| \le \beta_n^{-1}\sum_{k=1}^{n}\left\|\mathbf{D}_k^{(11)}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(22)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{D}_k^{(12)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(21)}\right\|\beta_k^2 c^2 e^{-2T(u_n-u_k)}.$$

Applying Lemma A.6.6, together with $\mathbf{D}_n^{(ij)} \to 0$ a.s. in Lemma A.2.2, gives

$$\limsup_n \|S_n^{(11,B)}\| \le \frac{1}{C(b,p)}\limsup_n \|\mathbf{D}_k^{(11)}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(22)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{D}_k^{(12)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{D}_k^{(21)}\| = 0,$$

for some constant $C(b,p) > 0$ defined in Lemma A.6.6.

We now consider $S_n^{(11,C)}$. Set

$$\psi_n \triangleq \sum_{k=1}^{n}\left(\mathbf{J}_k^{(11)}+\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{J}_k^{(22)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{J}_k^{(12)}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T-\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{J}_k^{(21)}\right),$$

we can rewrite $S_n^{(11,C)}$ as

$$S_n^{(11,C)} = \beta_n^{-1}\sum_{k=1}^{n}\beta_k^2 e^{(u_n-u_k)\mathbf{K_x}}(\psi_k-\psi_{k-1})e^{(u_n-u_k)(\mathbf{K_x})^T}.$$

By the Abel transformation in Lemma A.6.1, we have

$$S_n^{(11,C)} = \beta_n\psi_n + \beta_n^{-1}\sum_{k=1}^{n-1}\left[\beta_k^2 e^{(u_n-u_k)\mathbf{K_x}}\psi_k e^{(u_n-u_k)\mathbf{K}_\mathbf{x}^T} - \beta_{k+1}^2 e^{(u_n-u_{k+1})\mathbf{K_x}}\psi_k e^{(u_n-u_{k+1})\mathbf{K}_\mathbf{x}^T}\right]. \tag{43}$$

We know from Lemma A.2.2 that $\beta_n\psi_n \to 0$ a.s. because $\psi_n = o(\gamma_n^{-1})$ such that $\beta_n\psi_n = o(\beta_n/\gamma_n)$. Besides,

$$\begin{aligned}
\|\beta_k e^{(u_n-u_k)\mathbf{K_x}} - \beta_{k+1}e^{(u_n-u_{k+1})\mathbf{K_x}}\| &= \|(\beta_k-\beta_{k+1})e^{(u_n-u_k)\mathbf{K_x}} + \beta_{k+1}e^{(u_n-u_k)\mathbf{K_x}}(\mathbf{I}-e^{-\beta_{k+1}\mathbf{K_x}})\|\\
&\le C_1\beta_k^2 e^{-(u_n-u_k)T},
\end{aligned}$$

for some constant $C_1 > 0$ because $\beta_n - \beta_{n+1} \le C_2\beta_n^2$ and $\|\mathbf{I} - e^{-\beta_{k+1}\mathbf{K_x}}\| \le C_3\beta_{k+1}$ for some $C_2, C_3 > 0$. Moreover, for some $C_4 > 0$,

$$\begin{aligned}
\|\beta_k e^{(u_n-u_k)\mathbf{K_x}}\| + \|\beta_{k+1}e^{(u_n-u_{k+1})\mathbf{K_x}}\| &\le \beta_k\|e^{(u_n-u_k)\mathbf{K_x}}\| + \beta_k\|e^{(u_n-u_k)\mathbf{K_x}}\|\cdot\|e^{-\beta_{k+1}\mathbf{K_x}}\|\\
&\le C_4\beta_k e^{-(u_n-u_k)T}.
\end{aligned}$$

Using Lemma A.6.7 on (43) gives

$$\left\|S_n^{(11,C)}\right\| \leq C_1 C_4 \beta_n^{-1} \sum_{k=1}^{n-1} \beta_k^2 e^{-2(u_n-u_k)T} \|\beta_k \psi_k\| + \|\beta_n \psi_n\|.$$

Applying Lemma A.6.6 again gives

$$\limsup_n \left\|S_n^{(11,C)}\right\| \leq C_5 \limsup_n \|\beta_n \psi_n\| = 0,$$

for some constant $C_5 > 0$.

We provide an existing lemma below for the term $S_n^{(11,A)}$.

**Lemma A.2.6** (Mokkadem and Pelletier (2005) Lemma 4)**.** *For a sequence with decreasing step size $\beta_n = (n+1)^{-b}$ for $b \in (1/2, 1]$, $u_n = \sum_{k=1}^n \beta_k$, a positive semi-definite matrix $\mathbf{U}$ and a Hurwitz matrix $\mathbf{Q}$, which is given by*

$$\beta_n^{-1} \sum_{k=1}^n \beta_n^2 e^{(u_n-u_k)\mathbf{Q}} \mathbf{U} e^{(u_n-u_k)\mathbf{Q}^T},$$

*we have*

$$\lim_{n\to\infty} \beta_n^{-1} \sum_{k=1}^n \beta_n^2 e^{(u_n-u_k)\mathbf{Q}} \mathbf{U} e^{(u_n-u_k)\mathbf{Q}^T} = \mathbf{V}$$

*where $\mathbf{V}$ is the solution of the Lyapunov equation*

$$\left(\mathbf{Q} + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right)\mathbf{V} + \mathbf{V}\left(\mathbf{Q}^T + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right) + \mathbf{U} = 0.$$

Recall in (23) that $\mathbf{U_x} = \mathbf{U}_{11} + \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{U}_{22}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T - \mathbf{U}_{12}(\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})^T - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{U}_{21}$. Then, we rewrite $S_n^{(11,A)}$, defined in (42), as

$$S_n^{(11,A)} = \beta_n^{-1} \sum_{k=1}^n \beta_k^2 e^{(u_n-u_k)\mathbf{K_x}} \mathbf{U_x} e^{(u_n-u_k)\mathbf{K_x}^T}$$

Therefore, $\lim_{n\to\infty} S_n^{(11,A)} = \mathbf{V_x}$ is a direct application of Lemma A.2.6, where $\mathbf{V_x}$ is the solution to the following Lyapunov equation

$$\left(\mathbf{K_x} + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right)\mathbf{V_x} + \mathbf{V_x}\left(\mathbf{K_x}^T + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}\right) + \mathbf{U_x} = 0.$$

Together with Lemma A.6.8, we show the closed form $\mathbf{V_x}$ in Lemma A.2.3. □

By repeating the same process as in the demonstration of Lemma A.2.5, we conclude that $\lim_{n\to\infty} S_n^{(22)} = \mathbf{V_y}$.

Moreover, the property of $S_n^{(12)}$ is given as follows.

**Lemma A.2.7.** $\lim_{n\to\infty} S_n^{(12)} = 0$.

*Proof.* Note that

$$\begin{aligned}
\left\|S_n^{(12)}\right\| &= O\left(\beta_n^{-1/2}\gamma_n^{-1/2} \sum_{k=1}^n \beta_k\gamma_k \|e^{(u_n-u_k)\mathbf{K_x}}\| \|e^{(s_n-s_k)\mathbf{Q}_{22}^T}\|\right) \\
&= O\left(\beta_n^{-1/2}\gamma_n^{-1/2} \sum_{k=1}^n \beta_k\gamma_k e^{-(u_n-u_k)T} e^{-(s_n-s_k)T'}\right) \\
&= O\left(\beta_n^{-1/2}\gamma_n^{-1/2} \sum_{k=1}^n \beta_k\gamma_k e^{-(s_n-s_k)T'}\right)
\end{aligned}$$

for some $T, T' > 0$, where the second equality is from Lemma A.6.4 and the third equality comes from $e^{-(u_n-u_k)T} \leq 1$. Then, we use Lemma A.6.6 with $p = 0$ to obtain

$$\sum_{k=1}^{n} \beta_k \gamma_k e^{-(s_n-s_k)T'} = O(\beta_n) \tag{44}$$

where $\beta_n/\beta_{n+1} = (1+1/n)^b = 1 + O(1/n) = 1 + o(\gamma_n)$ satisfies the condition in Lemma A.6.6. Additionally, since $\beta_n = o(\gamma_n)$, we have

$$\beta_n^{-1/2}\gamma_n^{-1/2} \sum_{k=1}^{n} \beta_k \gamma_k^{-1/2} \gamma_k^{3/2} e^{-(s_n-s_k)T'} = O(\beta_n^{1/2}\gamma_n^{-1/2}) = o(1).$$

Then, it follows that $\lim_{n\to\infty} S_n^{(12)} = 0$. □

Consequently, we obtain

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[(Z_k^{(n)} - Z_{k-1}^{(n)})(Z_k^{(n)} - Z_{k-1}^{(n)})^T | \mathcal{F}_{k-1}\right] = \begin{pmatrix} \mathbf{V_x} & 0 \\ 0 & \mathbf{V_y} \end{pmatrix}.$$

The last part of this proof is to verify the conditions of the Martingale CLT in Theorem A.6.3. For some $\tau > 0$, we have

$$\begin{aligned} &\sum_{k=1}^{n} \mathbb{E}\left[\|Z_k^{(n)} - Z_{k-1}^{(n)}\|^{2+\tau} | \mathcal{F}_{k-1}\right] \\ &= O\left(\beta_n^{-(1+\frac{\tau}{2})} \sum_{k=1}^{n} \beta_k^{2+\frac{\tau}{2}} \beta_k^{\frac{\tau}{2}} e^{-(2+\tau)(u_n-u_k)T} + \gamma_n^{-(1+\frac{\tau}{2})} \sum_{k=1}^{n} \gamma_k^{2+\frac{\tau}{2}} \gamma_k^{\frac{\tau}{2}} e^{-(2+\tau)(s_n-s_k)T'}\right) \\ &= O\left(\beta_n^{\frac{\tau}{2}} + \gamma_n^{\frac{\tau}{2}}\right) \end{aligned} \tag{45}$$

where the last equality comes from Lemma A.6.6. Since (45) also holds for $\tau = 0$, we have

$$\sum_{k=1}^{n} \mathbb{E}\left[\|Z_k^{(n)} - Z_{k-1}^{(n)}\|^2 | \mathcal{F}_{k-1}\right] = O(1) < \infty.$$

Therefore, all the conditions in Theorem A.6.3 are satisfied and its application then gives

$$Z^{(n)} = \begin{pmatrix} \sqrt{\beta_n^{-1}} L_n^{(\boldsymbol{\theta})} \\ \sqrt{\gamma_n^{-1}} L_n^{(\mathbf{x})} \end{pmatrix} \xrightarrow[n\to\infty]{dist} N\left(0, \begin{pmatrix} \mathbf{V_x} & 0 \\ 0 & \mathbf{V_y} \end{pmatrix}\right). \tag{46}$$

This completes the proof.

### A.2.6 Upper Bounds of $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}, \Delta_n^{(\mathbf{x})}$, and $\Delta_n^{(\mathbf{y})}$ Towards Lemma A.2.4

In this part, we aim to show that $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}, \Delta_n^{(\mathbf{x})}$, and $\Delta_n^{(\mathbf{y})}$ decrease to zero faster than the CLT scaling factor, and are thus not present in the final CLT results. To proceed with the analysis, we provide the additional lemma to get the tighter upper bounds of $\|L_n^{(\mathbf{x})}\|$ and $\|L_n^{(\mathbf{y})}\|$ as follows, which is useful in deriving the tight upper bounds of $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}, \Delta_n^{(\mathbf{x})}$, and $\Delta_n^{(\mathbf{y})}$ later in Lemmas A.2.10 and A.2.12.

**Lemma A.2.8.** *For $L_n^{(\mathbf{x})}$ and $L_n^{(\mathbf{y})}$ defined in* (28) *and* (30)*, we further have*

$$\|L_n^{(\mathbf{x})}\| = O\left(\sqrt{\beta_n \log(u_n)}\right) \quad a.s.$$

$$\|L_n^{(\mathbf{y})}\| = O\left(\sqrt{\gamma_n \log(s_n)}\right) \quad a.s.$$

*Proof.* This proof follows from Pelletier (1998, Lemma 1). We only need the special case of Pelletier (1998, Lemma 1) that fits our scenario, i.e., we let the two types of step sizes therein be the same. For self-contained purposes, we attach the following lemma.

**Lemma A.2.9** (Pelletier (1998) Lemma 1)**.** *Consider a sequence*

$$L_{n+1} = e^{u_n \mathbf{Q}} \sum_{k=1}^{n} e^{-u_k \mathbf{Q}} \beta_k M_{k+1},$$

*where* $\beta_n = (n+1)^{-b}$, $1/2 < b \leq 1$, $u_n = \sum_{k=1}^{n} \beta_k$, *matrix* $\mathbf{Q}$ *is Hurwitz, and* $\{M_n\}$ *is a Martingale difference sequence adapted to the filtration* $\mathcal{F}$. *Almost surely,* $\limsup_n \mathbb{E}[\|M_{n+1}\|^2|\mathcal{F}_n] \leq M^2$ *and there exists* $\tau \in (0,2)$, $b(2+\tau) > 2$, *such that* $\sup_n \mathbb{E}[\|M_{n+1}\|^{2+\tau}|\mathcal{F}_n] < \infty$. *Then, almost surely,*

$$\limsup_n \frac{\|L_n\|}{\sqrt{\beta_n \log(u_n)}} \leq C_M, \tag{47}$$

*where* $C_M$ *is a constant dependent on* $M$.

By Assumption (A5), the iterates $(\boldsymbol{\theta}_n, \mathbf{x}_n)$ are bounded within a compact subset $\Omega$. Recall the forms of $M_{n+1}^{(\mathbf{x})}, M_{n+1}^{(\mathbf{y})}$ defined in (22), they comprise the functions $m_1(\mathbf{x}_n, \mathbf{y}_n, i), m_2(\mathbf{x}_n, \mathbf{y}_n, i)$ and $(\mathbf{P}m_1)(\mathbf{x}_n, \mathbf{y}_n, i), (\mathbf{P}m_2)(\mathbf{x}_n, \mathbf{y}_n, i)$, which in turn include the function $h_1(\mathbf{x}_n, \mathbf{y}_n, i), h_2(\mathbf{x}_n, \mathbf{y}_n, i)$. We know that $h_1(\mathbf{x}_n, \mathbf{y}_n, i), h_2(\mathbf{x}_n, \mathbf{y}_n, i)$ are bounded for $(\mathbf{x}_n, \mathbf{y}_n)$ within some compact set $\Omega$. Thus, $M_{n+1}^{(\mathbf{x})}, M_{n+1}^{(\mathbf{y})}$ are bounded (because of finite state space $\Xi$) and we denote by $C_\Omega^{(\mathbf{x})}$ and $C_\Omega^{(\mathbf{y})}$ as their corresponding upper bounds, i.e., $\mathbb{E}[\|M_{n+1}^{(\mathbf{x})}\|^2|\mathcal{F}_n] \leq C_\Omega^{(\mathbf{x})}$ and $\mathbb{E}[\|M_{n+1}^{(\mathbf{x})}\|^2|\mathcal{F}_n] \leq C_\Omega^{(\mathbf{y})}$. Thus, by the application of Lemma A.2.9, we have

$$\limsup_n \frac{\|L_n^{(\mathbf{x})}\|}{\sqrt{\beta_n \log(u_n)}} \leq C_\Omega^{(\mathbf{x})}, \quad \limsup_n \frac{\|L_n^{(\mathbf{y})}\|}{\sqrt{\gamma_n \log(s_n)}} \leq C_\Omega^{(\mathbf{y})}, \tag{48}$$

such that almost surely, $\|L_n^{(\mathbf{x})}\| = O(\sqrt{\beta_n \log(u_n)})$ and $\|L_n^{(\mathbf{y})}\| = O(\sqrt{\gamma_n \log(s_n)})$, which completes the proof. □

Now, we present the following condition on any given real-valued deterministic sequence $\{\omega_n\}$ in a similar vein as in Mokkadem and Pelletier (2006, Definition 2).

(C). Let $\{\omega_n\}$ be a positive real-valued and uniformly bounded deterministic sequence. Moreover, $\{\omega_n\}$ satisfies $\dfrac{\omega_n}{\omega_{n+1}} = 1 + o(\gamma_n)$.

In what follows, choices of sequences satisfying Condition C will be employed towards proving Lemma A.2.4. We explain how to derive the upper bounds for $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}, \Delta_n^{(\mathbf{x})}$, and $\Delta_n^{(\mathbf{y})}$, while the main difficulty in the procedure is to derive the upper bounds of $\Delta_n^{(\mathbf{x})}, \Delta_n^{(\mathbf{y})}$ because we have to deal with the additional noise terms $r_n^{(\mathbf{x},2)}, r_n^{(\mathbf{y},2)}$ therein, which arise from the decomposition of Markovian noise.

First of all, by almost sure convergence, we have $\|\mathbf{y}_n - \mathbf{y}^*\| = o(1)$, the only upper bound of $\mathbf{y}_n - \mathbf{y}^*$ for us. Letting $\omega_n \equiv 1$ is obviously one of the choices of $\{\omega_n\}$ that satisfy Condition C. Thus, setting $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$ allows us to present the initial upper bounds for $R_n^{(\mathbf{x})}, R_n^{(\mathbf{y})}$, which involve $\omega_n$, as indicated in Lemma A.2.10.

**Lemma A.2.10.** *Suppose there exists a nonrandom sequence* $\{\omega_n\}$ *satisfying Condition C such that* $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$ *a.s. Then, with* $R_n^{(\mathbf{x})}$, $R_n^{(\mathbf{y})}$ *defined in* (27) *and* (29), *for some* $s > 1/2$, *we have*

$$\|R_n^{(\mathbf{x})}\| = O(\beta_n \gamma_n^{-1} w_n + n^{-s}) \quad a.s.$$

$$\|R_n^{(\mathbf{y})}\| = O\left(\beta_n \gamma_n^{-1} w_n + \sqrt{\beta_n \log(u_n)}\right) \quad a.s.$$

For the proof of Lemma A.2.10 we refer the reader to Mokkadem and Pelletier (2006, Lemma 5), which applies directly since both sequences $R_n^{(\mathbf{x})}$ and $R_n^{(\mathbf{y})}$, as defined in (27) and (29), have the same form as in Mokkadem

and Pelletier (2006) and do not include additional terms $r_n^{(\mathbf{x},1)}, r_n^{(\mathbf{x},2)}$, $r_n^{(\mathbf{y},1)}, r_n^{(\mathbf{y},2)}$ arising from the Markovian noise.

Lemma A.2.10 implies that $\|R_n^{(\mathbf{x})}\| = o(1)$ and $\|R_n^{(\mathbf{y})}\| = o(1)$ since $\beta_n\gamma_n^{-1} \to 0$ and $\omega_n$ is uniformly bounded in Condition C. Since $\|\mathbf{y}_n - \mathbf{y}^*\| = o(1)$ by almost sure convergence, Lemma A.2.8 and Lemma A.2.10 indicate that $L_n^{(\mathbf{y})} = o(1), R_n^{(\mathbf{y})} = o(1)$, we thus have $\Delta_n^{(\mathbf{y})} = o(1)$. So, in addition to set $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$, we also let $\Delta_n^{(\mathbf{y})} = O(\eta_n)$ for some sequence $\{\eta_n\}$ satisfying Condition C. Then, we need to characterize the exact forms of $\Delta_n^{(\mathbf{x})}$, $\Delta_n^{(\mathbf{y})}$.

By substituting (27) and (28) in (25), along with the definition $\Delta_{n+1}^{(\mathbf{x})} \triangleq \mathbf{x}_{n+1} - \mathbf{x}^* - L_{n+1}^{(\mathbf{x})} - R_{n+1}^{(\mathbf{x})}$, we obtain $\Delta_{n+1}^{(\mathbf{x})}$ as follows.

$$\begin{aligned}
\Delta_{n+1}^{(\mathbf{x})} &= (\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}})(\mathbf{x}_n - \mathbf{x}^*) + \beta_{n+1}\left(r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})})\right) \\
&\quad - e^{\beta_{n+1}\mathbf{K}_{\mathbf{x}}}L_n^{(\mathbf{x})} - e^{\beta_{n+1}\mathbf{K}_{\mathbf{x}}}R_n^{(\mathbf{x})} \\
&= (\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}})(\mathbf{x}_n - \mathbf{x}^*) + \beta_{n+1}\left(r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})})\right) \\
&\quad - (\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}} + O(\beta_{n+1}^2))L_n^{(\mathbf{x})} - (\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}} + O(\beta_{n+1}^2))R_n^{(\mathbf{x})} \\
&= (\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}})\Delta_n^{(\mathbf{x})} + \beta_{n+1}\left(r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})})\right) \\
&\quad + O(\beta_{n+1}^2)\left(L_n^{(\mathbf{x})} + R_n^{(\mathbf{x})}\right),
\end{aligned} \tag{49}$$

where the third equality is by using the Taylor expansion $e^{\beta_{n+1}\mathbf{K}_{\mathbf{x}}} = \mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}} + O(\beta_{n+1}^2)$, and the fourth equality stems from the definition $\Delta_n^{(\mathbf{x})} = \mathbf{x}_n - \mathbf{x}^* - L_n^{(\mathbf{x})} - R_n^{(\mathbf{x})}$.

Similarly, for $\Delta_{n+1}^{(\mathbf{y})}$, we have

$$\begin{aligned}
\Delta_{n+1}^{(\mathbf{y})} &\triangleq \mathbf{y}_{n+1} - \mathbf{y}^* - L_{n+1}^{(\mathbf{y})} - R_{n+1}^{(\mathbf{y})} \\
&= (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22})(\mathbf{y}_n - \mathbf{y}^*) + \gamma_{n+1}\left(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})}\right) + \gamma_{n+1}\mathbf{Q}_{21}(\mathbf{x}_n - \mathbf{x}^*) \\
&\quad - e^{\gamma_{n+1}\mathbf{Q}_{22}}L_n^{(\mathbf{y})} - e^{\gamma_{n+1}\mathbf{Q}_{22}}R_n^{(\mathbf{y})} - \gamma_{n+1}\mathbf{Q}_{21}(L_n^{(\mathbf{x})} + R_n^{(\mathbf{x})}) \\
&= (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22})(\mathbf{y}_n - \mathbf{y}^*) + \gamma_{n+1}\left(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})}\right) + \gamma_{n+1}\mathbf{Q}_{21}\Delta_n^{(\mathbf{x})} \\
&\quad - (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22} + O(\gamma_{n+1}^2))L_n^{(\mathbf{y})} - (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22} + O(\gamma_{n+1}^2))R_n^{(\mathbf{y})} \\
&= (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22})\Delta_n^{(\mathbf{y})} + \gamma_{n+1}\left(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})}\right) + \gamma_{n+1}\mathbf{Q}_{21}\Delta_n^{(\mathbf{x})} + O(\gamma_{n+1}^2)\left(L_n^{(\mathbf{y})} + R_n^{(\mathbf{y})}\right),
\end{aligned} \tag{50}$$

where the third equality is from $e^{\gamma_{n+1}\mathbf{Q}_{22}} = \mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22} + O(\gamma_{n+1}^2)$ and $\Delta_n^{(\mathbf{x})} = \mathbf{x}_n - \mathbf{x}^* - L_n^{(\mathbf{x})} - R_n^{(\mathbf{x})}$, and the fourth equality is because the definition $\Delta_n^{(\mathbf{y})} = \mathbf{y}_n - \mathbf{y}^* - L_n^{(\mathbf{y})} - R_n^{(\mathbf{y})}$.

In what follows, we investigate the asymptotic behavior of the terms $r_n^{(\mathbf{x},1)}, r_n^{(\mathbf{y},2)}$ and $r_n^{(\mathbf{y},1)}, r_n^{(\mathbf{y},2)}$ that are part of the sequences $\Delta_n^{(\mathbf{x})}$ and $\Delta_n^{(\mathbf{y})}$, respectively. The results of this analysis will be used in Lemma A.2.12 to show the upper bounds of $\Delta_n^{(\mathbf{x})}, \Delta_n^{(\mathbf{y})}$.

**Lemma A.2.11.** *For $r_n^{(\mathbf{x},1)}, r_n^{(\mathbf{x},2)}, r_n^{(\mathbf{y},1)}, r_n^{(\mathbf{y},2)}$ defined in (22), the following holds almost surely:*

$$\|r_n^{(\mathbf{x},1)}\| = O(\gamma_n) = o(\sqrt{\beta_n}), \quad \sup_n \left\|\sum_{k=1}^n r_k^{(\mathbf{x},2)}\right\| < \infty,$$

$$\|r_n^{(\mathbf{y},1)}\| = O(\gamma_n) = o(\sqrt{\beta_n}), \quad \sup_n \left\|\sum_{k=1}^n r_k^{(\mathbf{y},2)}\right\| < \infty.$$

*Proof.* We only prove the result for $r_n^{(\mathbf{x},1)}$ and $r_n^{(\mathbf{x},2)}$, since the result for $r_n^{(\mathbf{y},1)}$ and $r_n^{(\mathbf{y},2)}$ follows similar arguments.

Observe that for any compact set $\Omega$ satisfying Assumption (A5), we have

$$\begin{aligned} r_n^{(\mathbf{x},1)} &= (\mathbf{P}m_1)(\mathbf{x}_{n+1},\mathbf{y}_{n+1},\xi_{n+1}) - (\mathbf{P}m_1)(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1}) \\ &\le \sum_{j\in\Xi} L_\Omega(\|\mathbf{x}_{n+1}-\mathbf{x}_n\| + \|\mathbf{y}_{n+1}-\mathbf{y}_n\|) \\ &\le |\Xi| L_\Omega C_\Omega(\beta_{n+1}+\gamma_{n+1}) \end{aligned}$$

where the first inequality is because $(\mathbf{P}m_1)(\mathbf{x},\mathbf{y},\xi)$ is continuous in $\mathbf{x},\mathbf{y}$ for any $\xi\in\Xi$, and the fact that any continuous function is locally Lipschitz with a set-dependent Lipschitz constant $L_\Omega$. The second inequality is from update rule (14), Assumption (A2), and $(\mathbf{x}_n,\mathbf{y}_n)\in\Omega$ for some compact subset $\Omega$ (by Assumption (A5)) such that

$$\max_{(\mathbf{x},\mathbf{y})\in\Omega,\xi\in\Xi}\{\|h_1(\mathbf{x},\mathbf{y},\xi)\|, \|h_2(\mathbf{x},\mathbf{y},\xi)\|\} \le C_\Omega,$$

Then, because $a > 1/2 \ge b/2$ by Assumption (A1), we have $\|r_n^{(\mathbf{x},1)}\| = O(\gamma_n) = o(\sqrt{\beta_n})$.

Now, let $\nu_n \triangleq (\mathbf{P}m_1)(\mathbf{x}_n,\mathbf{y}_n,\xi_n)$ such that $r_n^{(\mathbf{x},2)} = \nu_n - \nu_{n+1}$. Note that

$$\sum_{k=1}^n r_k^{(\mathbf{x},2)} = \nu_1 - \nu_{n+1},$$

and by Assumption (A5), $\|\nu_n\|$ is upper bounded by a constant dependent on the compact set $\Omega$, which leads to

$$\sup_n \left\|\sum_{k=1}^n r_k^{(\mathbf{x},2)}\right\| = \sup_n \|\nu_1 - \nu_{n+1}\| < \infty \quad \text{a.s.}$$

This completes the proof. □

We are now ready to state the lemma for the sequences $\Delta_n^{(\mathbf{x})}, \Delta_n^{(\mathbf{y})}$.

**Lemma A.2.12.** *Suppose that there exist two sequences $\{\omega_n\}$ and $\{\eta_n\}$ satisfying Condition C such that $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$ a.s. and $\|\Delta_n^{(\mathbf{y})}\| = O(\eta_n)$ a.s. We have*

$$\|\Delta_n^{(\mathbf{x})}\| = O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + o(\sqrt{\beta_n}) \quad a.s.$$

$$\|\Delta_n^{(\mathbf{y})}\| = O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + o(\sqrt{\beta_n}) \quad a.s.$$

*Proof.* The sequences in $\Delta_n^{(\mathbf{x})}, \Delta_n^{(\mathbf{y})}$ involve additional noise terms $r_n^{(\mathbf{x},2)}, r_n^{(\mathbf{y},2)}$ arising from Markovian noise, causing the challenge of deriving their upper bounds. In this proof we separate out these terms $r_n^{(\mathbf{x},2)}, r_n^{(\mathbf{y},2)}$ from other noise terms, and specifically analyze their asymptotic rates, which contribute to the $o(\sqrt{\beta_n})$ term.

Recall from (49), (50), we have

$$\begin{aligned} \Delta_{n+1}^{(\mathbf{x})} &= (\mathbf{I}+\beta_{n+1}\mathbf{K}_\mathbf{x})\Delta_n^{(\mathbf{x})} + \beta_{n+1}\left(r_n^{(\mathbf{x},1)} + r_n^{(\mathbf{x},2)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})})\right) + O(\beta_{n+1}^2)\left(L_n^{(\mathbf{x})} + R_n^{(\mathbf{x})}\right), \\ \Delta_{n+1}^{(\mathbf{y})} &= (\mathbf{I}+\gamma_{n+1}\mathbf{Q}_{22})\Delta_n^{(\mathbf{y})} + \gamma_{n+1}\left(r_n^{(\mathbf{y},1)} + r_n^{(\mathbf{y},2)} + \rho_n^{(\mathbf{y})}\right) + \gamma_{n+1}\mathbf{Q}_{21}\Delta_n^{(\mathbf{x})} + O(\gamma_{n+1}^2)\left(L_n^{(\mathbf{y})} + R_n^{(\mathbf{y})}\right). \end{aligned}$$

We observe that they include the additional terms $r_n^{(\mathbf{x},2)}, r_n^{(\mathbf{y},2)}$ arising from the decomposition of the Markovian noise, which are missing in Mokkadem and Pelletier (2006, equations (20), (21)). To deal with this issue, we can decompose $\Delta_{n+1}^{(\mathbf{x})}$ into two parts, i.e., $\Delta_{n+1}^{(\mathbf{x})} = \Delta_{n+1}^{(\mathbf{x},1)} + \Delta_{n+1}^{(\mathbf{x},2)}$, where

$$\Delta_{n+1}^{(\mathbf{x},1)} \triangleq (\mathbf{I}+\beta_{n+1}\mathbf{K}_\mathbf{x})\Delta_n^{(\mathbf{x},1)} + \beta_{n+1}\left(r_n^{(\mathbf{x},1)} + \rho_n^{(\mathbf{x})} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + \rho_n^{(\mathbf{y})})\right) + O(\beta_{n+1}^2)\left(L_n^{(\mathbf{x})} + R_n^{(\mathbf{x})}\right), \quad (51)$$

$$\Delta_{n+1}^{(\mathbf{x},2)} \triangleq (\mathbf{I}+\beta_{n+1}\mathbf{K}_\mathbf{x})\Delta_n^{(\mathbf{x},2)} + \beta_{n+1}\left(r_n^{(\mathbf{x},2)} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}r_n^{(\mathbf{y},2)}\right). \quad (52)$$

Similarly, we can decompose $\Delta_{n+1}^{(\mathbf{y})}$ into two parts, i.e., $\Delta_{n+1}^{(\mathbf{y})} = \Delta_{n+1}^{(\mathbf{y},1)} + \Delta_{n+1}^{(\mathbf{y},2)}$, where

$$\Delta_{n+1}^{(\mathbf{y},1)} \triangleq (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22})\Delta_n^{(\mathbf{y},1)} + \gamma_{n+1}\left(r_n^{(\mathbf{y},1)} + \rho_n^{(\mathbf{y})}\right) + \gamma_{n+1}\mathbf{Q}_{21}\Delta_n^{(\mathbf{x})} + O(\gamma_{n+1}^2)\left(L_n^{(\mathbf{y})} + R_n^{(\mathbf{y})}\right), \tag{53}$$

$$\Delta_{n+1}^{(\mathbf{y},2)} \triangleq (\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22})\Delta_n^{(\mathbf{y},2)} + \gamma_{n+1}r_n^{(\mathbf{y},2)}. \tag{54}$$

Let us first focus on the terms $\Delta_{n+1}^{(\mathbf{x},2)}$ and $\Delta_{n+1}^{(\mathbf{y},2)}$. In the following, we show that

$$\|\Delta_{n+1}^{(\mathbf{x},2)}\| = o(\sqrt{\beta_n}), \quad \|\Delta_{n+1}^{(\mathbf{y},2)}\| = o(\sqrt{\beta_n}). \tag{55}$$

Denote by

$$\Phi_{k,n} \triangleq \prod_{j=k+1}^{n}(\mathbf{I} + \beta_j\mathbf{K}_{\mathbf{x}}), \quad \theta_n \triangleq \sum_{k=1}^{n} r_k^{(\mathbf{x},2)} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}r_k^{(\mathbf{y},2)},$$

and by convention $\Phi_{n+1,n} = \mathbf{I}$, we can rewrite $\Delta_{n+1}^{(\mathbf{x},2)}$ in (52) as

$$\Delta_{n+1}^{(\mathbf{x},2)} = \sum_{k=1}^{n}\Phi_{k,n}\beta_{k+1}\left(\theta_k - \theta_{k-1}\right)$$

because $\theta_k - \theta_{k-1} = r_k^{(\mathbf{x},2)} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}r_k^{(\mathbf{y},2)}$. By Abel transformation in Lemma A.6.1, we have

$$\Delta_{n+1}^{(\mathbf{x},2)} = \beta_{n+1}\theta_n + \sum_{k=1}^{n-1}(\beta_k\Phi_{k,n} - \beta_{k+1}\Phi_{k+1,n})\theta_k.$$

Note that $\beta_{n+1}\theta_n = \beta_{n+1}\left(\sum_{k=1}^{n} r_k^{(\mathbf{x},2)}\right) - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\beta_{n+1}\left(\sum_{k=1}^{n} r_k^{(\mathbf{y},2)}\right)$. By Lemma A.2.11, we have

$$\|\beta_{n+1}\theta_n\| = O(\beta_n) = o(\sqrt{\beta_n}) \quad a.s.$$

such that $\beta_{n+1}\theta_n \to 0$ almost surely. Furthermore,

$$\begin{aligned}\|\beta_k\Phi_{k,n} - \beta_{k+1}\Phi_{k+1,n}\| &\le \beta_{k+1}\|\Phi_{k,n} - \Phi_{k+1,n}\| + (\beta_k - \beta_{k+1})\|\Phi_{k,n}\| \\ &\le \beta_{k+1}\|\Phi_{k+1,n}\|\beta_k\|\mathbf{K}_{\mathbf{x}}\| + C_6\beta_k^2\|\Phi_{k,n}\| \\ &\le C_7\beta_k^2 e^{-(u_n - u_k)T}\end{aligned}$$

for some constant $T, C_6, C_7 > 0$, where the last inequality is from Lemma A.6.4 and $\|\Phi_{k+1,n}\| \le C_8\|\Phi_{k,n}\|$ for some constant $C_8 > 0$ that depends on $e^T$. Then,

$$\sum_{k=1}^{n}\|\beta_k\Phi_{k,n} - \beta_{k+1}\Phi_{k+1,n}\|\|\theta_k\| \le C_8\sum_{k=1}^{n}\beta_k e^{-(u_n - u_k)T}\|\beta_k\Psi_k\| = O(\|\beta_n\theta_n\|),$$

where the last equality is the application of Lemma A.6.6. Thus, we have $\|\Delta_{n+1}^{(\mathbf{x},2)}\| = o(\sqrt{\beta_n})$. Repeating the same steps, we get $\|\Delta_{n+1}^{(\mathbf{y},2)}\| = o(\sqrt{\beta_n})$.

We now turn to $\Delta_{n+1}^{(\mathbf{x},1)}$ and $\Delta_{n+1}^{(\mathbf{y},1)}$. As shown in Mokkadem and Pelletier (2006, p.11), there exist two matrix norms $\|\cdot\|_T$ and $\|\cdot\|_M$ such that for large enough $n$,

$$\|\mathbf{I} + \beta_{n+1}\mathbf{K}_{\mathbf{x}}\|_T \le 1 - \beta_{n+1}T, \quad \|\mathbf{I} + \gamma_{n+1}\mathbf{Q}_{22}\|_M \le 1 - \gamma_{n+1}M,$$

for some $T, M > 0$. The corresponding vector norm $\|\mathbf{v}\|_T \triangleq \|[\mathbf{v}\cdots\mathbf{v}]\|_T$, where $[\mathbf{v}\cdots\mathbf{v}] \in \mathbb{R}^{d_1\times d_1}$, and $\|\mathbf{u}\|_M \triangleq \|[\mathbf{u}\cdots\mathbf{u}]\|_M$, where $[\mathbf{u}\cdots\mathbf{u}] \in \mathbb{R}^{d_2\times d_2}$.

Then, we have

$$\begin{aligned}\|\Delta_{n+1}^{(\mathbf{x},1)}\|_T \le{}& (1 - \beta_{n+1}T)\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}\left(\|r_n^{(\mathbf{x},1)}\|_T + \|\rho_n^{(\mathbf{x})}\|_T + \|\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}(r_n^{(\mathbf{y},1)} + \rho_n^{(\mathbf{y})})\|_T\right) \\ &+ O(\beta_{n+1}^2)\left(\|L_n^{(\mathbf{x})}\|_T + \|R_n^{(\mathbf{x})}\|_T\right), \\ \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M \le{}& (1 - \gamma_{n+1}M)\|\Delta_n^{(\mathbf{y},1)}\|_M + \gamma_{n+1}\left(\|r_n^{(\mathbf{y},1)}\|_M + \|\rho_n^{(\mathbf{y})}\|_M\right) + \gamma_{n+1}\|\mathbf{Q}_{21}\Delta_n^{(\mathbf{x})}\|_M \\ &+ O(\gamma_{n+1}^2)\left(\|L_n^{(\mathbf{y})}\|_M + \|R_n^{(\mathbf{y})}\|_M\right).\end{aligned}$$

We note that

$$\begin{aligned}\|\rho_n^{(\mathbf{x})}\|_T + \|\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\rho_n^{(\mathbf{y})}\|_T &= O(\|\rho_n^{(\mathbf{x})}\| + \|\rho_n^{(\mathbf{y})}\|)\\ &= O(\|\mathbf{x}_n - \mathbf{x}^*\|^2 + \|\mathbf{y}_n - \mathbf{y}^*\|^2)\\ &= O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x})}\|_T^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2 + \|\Delta_n^{(\mathbf{y})}\|_M^2).\end{aligned}$$

where the first and the last equalities are from the equivalence of norms. Similarly,

$$\|\rho_n^{(\mathbf{y})}\|_M = O(\|\rho_n^{(\mathbf{x})}\| + \|\rho_n^{(\mathbf{y})}\|) = O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x})}\|_T^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2 + \|\Delta_n^{(\mathbf{y})}\|_M^2).$$

Also note that $\|\Delta_n^{(\mathbf{x})}\|_T^2 = O(\|\Delta_n^{(\mathbf{x},1)}\|_T^2) + o(\beta_n)$ and $\|\Delta_n^{(\mathbf{y})}\|_M^2 = O(\|\Delta_n^{(\mathbf{y},1)}\|_M^2) + o(\beta_n)$ by (55). It then follows that

$$\begin{aligned}\|\Delta_{n+1}^{(\mathbf{x},1)}\|_T \le{}& (1-\beta_{n+1}T)\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}O(\beta_{n+1}\|L_n^{(\mathbf{x})}\| + \beta_{n+1}\|R_n^{(\mathbf{x})}\| + \|r_n^{(\mathbf{x},1)}\|)\\ &+ \beta_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2 + \|\Delta_n^{(\mathbf{y},1)}\|_M^2 + o(\beta_n)), \qquad (56)\\ \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M \le{}& (1-\gamma_{n+1}M)\|\Delta_n^{(\mathbf{y},1)}\|_M + \gamma_{n+1}O(\gamma_{n+1}\|L_n^{(\mathbf{y})}\| + \gamma_{n+1}\|R_n^{(\mathbf{y})}\| + \|r_n^{(\mathbf{y},1)}\|)\\ &+ \gamma_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2 + \|\Delta_n^{(\mathbf{y},1)}\|_M^2 + o(\sqrt{\beta_n})). \qquad (57)\end{aligned}$$

From Lemma A.2.11 we know that $\|r_n^{(\mathbf{x},1)}\| = o(\sqrt{\beta_n})$ and $\|r_n^{(\mathbf{y},1)}\| = o(\sqrt{\beta_n})$, thus we can omit $o(\beta_n)$ and $o(\sqrt{\beta_n})$ terms in (56), (57). Lemma A.2.8 suggests that $\lim_{n\to\infty} L_n^{(\mathbf{x})} = 0$ and $\lim_{n\to\infty} L_n^{(\mathbf{y})} = 0$. Moreover, Lemma A.2.10 implies that $\lim_{n\to\infty} R_n^{(\mathbf{x})} = 0$ and $\lim_{n\to\infty} R_n^{(\mathbf{y})} = 0$ since $\omega_n$ is a bounded sequence by Condition C and $\beta_n\gamma_n^{-1} = o(1)$. By almost sure convergence $\lim_{n\to\infty}\mathbf{x}_n = \mathbf{x}^*$ and $\lim_{n\to\infty}\mathbf{y}_n = \mathbf{y}^*$, we have $\lim_{n\to\infty}\Delta_n^{(\mathbf{x})} = 0$ and $\lim_{n\to\infty}\Delta_n^{(\mathbf{y})} = 0$. This implies that for large enough $n$, there exists some $0 < T' < T, 0 < M' < M$ such that

$$-T\|\Delta_n^{(\mathbf{x},1)}\|_T + O(\|\Delta_n^{(\mathbf{x},1)}\|_T^2) \le T'\|\Delta_n^{(\mathbf{x},1)}\|_T, \quad -M\|\Delta_n^{(\mathbf{y},1)}\|_M + O(\|\Delta_n^{(\mathbf{y},1)}\|_M^2) \le -M'\|\Delta_n^{(\mathbf{y},1)}\|_T.$$

Bringing the above inequalities back to (56) and (57) (and omit $o(\beta_n)$ and $o(\sqrt{\beta_n})$ terms therein) leads to

$$\begin{aligned}\|\Delta_{n+1}^{(\mathbf{x},1)}\|_T \le{}& (1-\beta_{n+1}T')\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}O(\beta_{n+1}\|L_n^{(\mathbf{x})}\| + \beta_{n+1}\|R_n^{(\mathbf{x})}\| + \|r_n^{(\mathbf{x},1)}\|)\\ &+ \beta_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2 + \|\Delta_n^{(\mathbf{y},1)}\|_M^2), \qquad (58)\\ \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M \le{}& (1-\gamma_{n+1}M')\|\Delta_n^{(\mathbf{y},1)}\|_M + \gamma_{n+1}O(\gamma_{n+1}\|L_n^{(\mathbf{y})}\| + \gamma_{n+1}\|R_n^{(\mathbf{y})}\| + \|r_n^{(\mathbf{y},1)}\|)\\ &+ \gamma_{n+1}C_9(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2). \qquad (59)\end{aligned}$$

for some $C_9 > 0$. Since $\lim_{n\to\infty}\Delta_n^{(\mathbf{y})} = 0$, we have $O(\|\Delta_n^{(\mathbf{y},1)}\|_M^2) \le C_{10}\|\Delta_n^{(\mathbf{y},1)}\|_M$ for some $0 < C_{10} < T'/C_9$ for large enough $n$ so that we can further modify (58) as

$$\begin{aligned}\|\Delta_{n+1}^{(\mathbf{x},1)}\|_T \le{}& (1-\beta_{n+1}T')\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}O(\beta_{n+1}\|L_n^{(\mathbf{x})}\| + \beta_{n+1}\|R_n^{(\mathbf{x})}\| + \|r_n^{(\mathbf{x},1)}\|)\\ &+ \beta_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2) + \beta_{n+1}C_{10}\|\Delta_n^{(\mathbf{y},1)}\|_M.\end{aligned} \qquad (60)$$

Rewriting (59) gives

$$\begin{aligned}\|\Delta_{n+1}^{(\mathbf{y},1)}\|_M \le{}& \frac{1}{\gamma_{n+1}M'}\left[\|\Delta_n^{(\mathbf{y},1)}\|_M - \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M\right] + O(\gamma_{n+1}\|L_n^{(\mathbf{y})}\| + \gamma_{n+1}\|R_n^{(\mathbf{y})}\| + \|r_n^{(\mathbf{y},1)}\|)\\ &+ C_9(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2).\end{aligned}$$

Taking it back to (60) induces

$$\begin{aligned}
\|\Delta_{n+1}^{(\mathbf{x},1)}\|_T &\le (1-\beta_{n+1}T')\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}O(\beta_{n+1}\|L_n^{(\mathbf{x})}\| + \beta_{n+1}\|R_n^{(\mathbf{x})}\| + \|r_n^{(\mathbf{x},1)}\|) \\
&\quad + \beta_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2) \\
&\quad + \frac{\beta_{n+1}C_{10}}{\gamma_{n+1}M'}\left[\|\Delta_n^{(\mathbf{y},1)}\|_M - \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M\right] + \beta_{n+1}O(\gamma_{n+1}\|L_n^{(\mathbf{y})}\| + \gamma_{n+1}\|R_n^{(\mathbf{y})}\| + \|r_n^{(\mathbf{y},1)}\|) \\
&\quad + \beta_{n+1}C_9C_{10}(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|\Delta_n^{(\mathbf{x},1)}\|_T + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2) \qquad (61) \\
&\le (1-\beta_{n+1}T'')\|\Delta_n^{(\mathbf{x},1)}\|_T + \beta_{n+1}O(\|L_n^{(\mathbf{x})}\|^2 + \|R_n^{(\mathbf{x})}\|^2 + \|L_n^{(\mathbf{y})}\|^2 + \|R_n^{(\mathbf{y})}\|^2) \\
&\quad + \beta_{n+1}O(\beta_{n+1}\|L_n^{(\mathbf{x})}\| + \beta_{n+1}\|R_n^{(\mathbf{x})}\| + \|r_n^{(\mathbf{x},1)}\| + \gamma_{n+1}\|L_n^{(\mathbf{y})}\| + \gamma_{n+1}\|R_n^{(\mathbf{y})}\| + \|r_n^{(\mathbf{y},1)}\|) \\
&\quad + \frac{\beta_{n+1}C_{10}}{\gamma_{n+1}M'}\left[\|\Delta_n^{(\mathbf{y},1)}\|_M - \|\Delta_{n+1}^{(\mathbf{y},1)}\|_M\right],
\end{aligned}$$

where the last inequality is by setting $0 < T'' < T' - C_9C_{10}$. Now, (59) and (61) correspond to Mokkadem and Pelletier (2006, equations (27), (28)). Thus, we can leverage the result therein for $\Delta_{n+1}^{(\mathbf{x},1)}$ and $\Delta_{n+1}^{(\mathbf{y},1)}$, which is given below.

**Lemma A.2.13** (Mokkadem and Pelletier (2006) Appendix A.4.2 and Appendix A.4.3)**.** *For $\Delta_{n+1}^{(\mathbf{x},1)}$ and $\Delta_{n+1}^{(\mathbf{y},1)}$ with inequalities in* (61) *and* (59)*, and assume that $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$, $\|\Delta_n^{(\mathbf{y},1)}\| = O(\eta_n')$ where $\{\omega_n\}$ and $\{\eta_n'\}$ satisfy Condition C, we have*

$$\|\Delta_n^{(\mathbf{x},1)}\| = O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n') + o(\sqrt{\beta_n}) \quad a.s.$$

$$\|\Delta_n^{(\mathbf{y},1)}\| = O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n') + o(\sqrt{\beta_n}) \quad a.s.$$

Since $\|\Delta_n^{(\mathbf{y})}\| = O(\eta_n)$ in Lemma A.2.12, and $\|\Delta_n^{(\mathbf{y},1)}\| \le \|\Delta_n^{(\mathbf{y})}\| + \|\Delta_n^{(\mathbf{y},2)}\| = O(\eta_n + \sqrt{\beta_n})$ by (55), we set $\eta_n' = \eta_n + \sqrt{\beta_n}$. With Lemma A.2.13, it follows that, almost surely,

$$\begin{aligned}
\|\Delta_n^{(\mathbf{x})}\| &\le \|\Delta_n^{(\mathbf{x},1)}\| + \|\Delta_n^{(\mathbf{x},2)}\| \\
&= O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + O(\beta_n\gamma_n^{-1}\sqrt{\beta_n}) + o(\sqrt{\beta_n}) \\
&= O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + o(\sqrt{\beta_n}),
\end{aligned}$$

where the last equality is from $O(\beta_n\gamma_n^{-1}\sqrt{\beta_n}) = o(\sqrt{\beta_n})$ because $\beta_n\gamma_n^{-1} = o(1)$. Similarly,

$$\begin{aligned}
\|\Delta_n^{(\mathbf{y})}\| &\le \|\Delta_n^{(\mathbf{y},1)}\| + \|\Delta_n^{(\mathbf{y},2)}\| \\
&= O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + O(\beta_n\gamma_n^{-1}\sqrt{\beta_n}) + o(\sqrt{\beta_n}) \\
&= O(\beta_n^2\gamma_n^{-2}w_n^2 + \beta_n\gamma_n^{-1}\eta_n) + o(\sqrt{\beta_n}).
\end{aligned}$$

This completes the proof of Lemma A.2.12. □

By Lemma A.2.10 and Lemma A.2.12, we can iteratively fine-tune the expression of $\omega_n$ and $\eta_n$ (in other words, tightening the upper bounds of $\omega_n$ and $\eta_n$). We are now ready to prove lemma A.2.4.

***Proof of lemma A.2.4.*** Since $\lim_{n\to\infty}\Delta_n^{(\mathbf{y})} = 0$ almost surely by Lemma A.2.12, we can set $\eta_n \equiv 1$ such that

$$\|\Delta_n^{(\mathbf{y})}\| = O(\beta_n^2\gamma_n^{-2}\omega_n^2 + \beta_n\gamma_n^{-1}) + o(\sqrt{\beta_n}). \qquad (62)$$

According to (62), we set $\eta_n = O(\beta_n^2\gamma_n^{-2}\omega_n^2 + [\beta_n\gamma_n^{-1}]^k) + o(\sqrt{\beta_n})$ for integer $k \geq 1$. Then, we have

$$\begin{aligned}\frac{\eta_n}{\eta_{n+1}} &\leq \frac{\beta_n^2\gamma_n^{-2}\omega_n^2 + [\beta_n\gamma_n^{-1}]^k + \sqrt{\beta_n}}{\beta_{n+1}^2\gamma_{n+1}^{-2}\omega_{n+1}^2 + [\beta_{n+1}\gamma_{n+1}^{-1}]^k + \sqrt{\beta_{n+1}}} \\ &= 1 + \frac{(\beta_n^2\gamma_n^{-2}\omega_n^2 - \beta_{n+1}^2\gamma_{n+1}^{-2}\omega_{n+1}^2) + ([\beta_n\gamma_n^{-1}]^k - [\beta_{n+1}\gamma_{n+1}^{-1}]^k) + (\sqrt{\beta_n} - \sqrt{\beta_{n+1}})}{\beta_{n+1}^2\gamma_{n+1}^{-2}\omega_{n+1}^2 + [\beta_{n+1}\gamma_{n+1}^{-1}]^k + \sqrt{\beta_{n+1}}} \\ &\leq 1 + \frac{\beta_n^2\gamma_n^{-2}\omega_n^2 - \beta_{n+1}^2\gamma_{n+1}^{-2}\omega_{n+1}^2}{\beta_{n+1}^2\gamma_{n+1}^{-2}\omega_{n+1}^2} + \frac{[\beta_n\gamma_n^{-1}]^k - [\beta_{n+1}\gamma_{n+1}^{-1}]^k}{[\beta_{n+1}\gamma_{n+1}^{-1}]^k} + \frac{\sqrt{\beta_n} - \sqrt{\beta_{n+1}}}{\sqrt{\beta_{n+1}}}.\end{aligned}$$

Since $[\beta_n\gamma_n^{-1}]^k/[\beta_{n+1}\gamma_{n+1}^{-1}]^k = 1 + O(1/n) = 1 + o(\gamma_n)$ for $k \geq 1$ and $\sqrt{\beta_n}/\sqrt{\beta_{n+1}} = 1 + O(1/n) = 1 + o(\gamma_n)$, together with $\omega_n/\omega_{n+1} = 1 + o(\gamma_n)$ in Condition C, we have

$$\frac{\eta_n}{\eta_{n+1}} = 1 + o(\gamma_n).$$

Thus, the new expression of $\eta_n$ also satisfies Condition C for *all* $k \geq 1$. There exists some integer $k_0$ such that for all $k \geq k_0$, we have $[\beta_n\gamma_n^{-1}]^k = n^{ka-kb} = o(\sqrt{\beta_n})$ such that

$$\|\Delta_n^{(\mathbf{y})}\| = O(\eta_n) = O(\beta_n^2\gamma_n^{-2}\omega_n^2) + o(\sqrt{\beta_n}).$$

Similarly, we have

$$\|\Delta_n^{(\mathbf{x})}\| = O(\beta_n^2\gamma_n^{-2}\omega_n^2) + o(\sqrt{\beta_n}).$$

For $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\omega_n)$, we set $\omega_n = \sqrt{\gamma_n \log s_n} + [\beta_n\gamma_n^{-1}]^k$ for integer $k \geq 1$ and check that

$$\frac{\omega_n}{\omega_{n+1}} = \frac{\sqrt{\gamma_n \log s_n} + [\beta_n\gamma_n^{-1}]^k}{\sqrt{\gamma_{n+1} \log s_{n+1}} + [\beta_{n+1}\gamma_{n+1}^{-1}]^k} \leq 1 + \frac{\sqrt{\gamma_n \log s_n} - \sqrt{\gamma_{n+1} \log s_{n+1}}}{\sqrt{\gamma_{n+1} \log s_{n+1}}} + \frac{[\beta_n\gamma_n^{-1}]^k - [\beta_{n+1}\gamma_{n+1}^{-1}]^k}{[\beta_{n+1}\gamma_{n+1}^{-1}]^k}.$$

After algebraic calculation, we have $\sqrt{\gamma_n \log s_n}/\sqrt{\gamma_{n+1} \log s_{n+1}} < \sqrt{\gamma_n}/\sqrt{\gamma_{n+1}} = 1 + O(1/n) = 1 + o(\gamma_n)$. Along with $[\beta_n\gamma_n^{-1}]^k/[\beta_{n+1}\gamma_{n+1}^{-1}]^k = 1 + o(\gamma_n)$, it then follows that

$$\frac{\omega_n}{\omega_{n+1}} = 1 + o(\gamma_n).$$

Therefore, $\omega_n = \sqrt{\gamma_n \log s_n} + [\beta_n\gamma_n^{-1}]^k$ satisfies Condition C for all $k \geq 1$. There exists some integer $k_1$ such that for all $k \geq k_1$, $[\beta_n\gamma_n^{-1}]^k = o(\sqrt{\gamma_n})$, which in turn implies that $[\beta_n\gamma_n^{-1}]^k = o(\sqrt{\gamma_n \log s_n})$. Thus, $\omega_n = O(\sqrt{\gamma_n \log s_n})$ and $\|\mathbf{y}_n - \mathbf{y}^*\| = O(\sqrt{\gamma_n \log s_n})$ almost surely.

Consequently, with $\omega_n = O(\sqrt{\gamma_n \log s_n})$, we have

$$\|\Delta_n^{(\mathbf{x})}\| = O(\beta_n^2\gamma_n^{-1} \log s_n) + o(\sqrt{\beta_n}) = o(\sqrt{\beta_n}), \quad \|\Delta_n^{(\mathbf{y})}\| = O(\beta_n^2\gamma_n^{-1} \log s_n) + o(\sqrt{\beta_n}) = o(\sqrt{\beta_n}).$$

For $\|R_n^{(\mathbf{x})}\|, \|R_n^{(\mathbf{y})}\|$ of the forms in Lemma A.2.10, we have the following:

$$\|R_n^{(\mathbf{x})}\| = O(\beta_n\gamma_n^{-1/2} \log s_n + n^{-s}) = O(n^{a/2-b} \log s_n + n^{-s}) = O(n^{-c}),$$

where $c \triangleq \min\{b - a/2 + \epsilon, s\} > b/2$ for some small enough $\epsilon > 0$. In addition, we have

$$\|R_n^{(\mathbf{y})}\| = O(\beta_n\gamma_n^{-1/2} \log s_n + \sqrt{\beta_n \log u_n}) = O(\sqrt{\beta_n \log u_n}),$$

which completes the proof. □

## A.3 Performance Ordering in TTSA

### A.3.1 Proof of Proposition 3.2

By definition of efficiency ordering in Definition 3.1 (in Section 3.1), we recall that two efficiency-ordered Markov chains $\{W_n\}$ and $\{Z_n\}$ with $W \preceq Z$ and same stationary distribution $\boldsymbol{\mu}$ obey the following Loewner ordering:

$$\mathbf{U}^{(W)}(g) \geq_L \mathbf{U}^{(Z)}(g)$$

for any vector-valued function $g : \Xi \to \mathbb{R}^d$, where

$$\mathbf{U}^{(W)}(g) = \lim_{s\to\infty} \frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} g(W_n) - \mathbb{E}_{\boldsymbol{\mu}}[g]\right)\left(\sum_{n=1}^{s} g(W_n) - \mathbb{E}_{\boldsymbol{\mu}}[g]\right)^T\right]$$

and $\mathbb{E}_{\boldsymbol{\mu}}[g] = \mathbb{E}_{W\sim\boldsymbol{\mu}}[g(W)]$.

Now, we turn to the explicit form of $\mathbf{V_x}$ and $\mathbf{V_y}$ in our Theorem 2.2 in Section 2.3 and give them below for completeness.

$$\mathbf{V_x} = \int_0^\infty e^{t\left(\mathbf{K_x} + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)}\mathbf{U_x} e^{t\left(\mathbf{K_x} + \frac{\mathbb{1}\{b=1\}}{2}\mathbf{I}\right)^T} dt,$$
$$\mathbf{V_y} = \int_0^\infty e^{t\mathbf{Q}_{22}}\mathbf{U}_{22} e^{t\mathbf{Q}_{22}^T} dt,$$

where the expression of $\mathbf{U_x}$ and $\mathbf{U}_{22}$ can be found in Remark A.2.2 in Appendix A.2.1. The only components in $\mathbf{V_x}, \mathbf{V_y}$ that are associated with the underlying Markov chain are $\mathbf{U_x}$ and $\mathbf{U}_{22}$. Let the function $g(W_n) \equiv h_2(\mathbf{x}^*, \mathbf{y}^*, W_n)$ for the TTSA algorithm (14) driven by the Markov chain $\{W_n\}$, then replacing $\{W_n\}$ with a more efficient chain $\{Z_n\}$ leads to $\mathbf{U}^{(W)}(g) \geq_L \mathbf{U}^{(Z)}(g)$, or equivalently, $\mathbf{U}_{22}^{(W)} \geq_L \mathbf{U}_{22}^{(Z)}$, where the superscript $(W)$ indicates that the TTSA algorithm (14) is driven by the Markov chain $\{W_n\}$.

Similarly, let $g(W_n) \equiv \begin{bmatrix} h_1(\mathbf{x}^*, \mathbf{y}^*, W_n) \\ h_2(\mathbf{x}^*, \mathbf{y}^*, W_n) \end{bmatrix}$, then we have

$$\begin{aligned}\begin{bmatrix} \mathbf{U}_{11}^{(W)} & \mathbf{U}_{12}^{(W)} \\ \mathbf{U}_{21}^{(W)} & \mathbf{U}_{22}^{(W)} \end{bmatrix} &= \lim_{s\to\infty} \frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} g(W_n)\right)\left(\sum_{n=1}^{s} g(W_n)\right)^T\right] \\ &\geq_L \lim_{s\to\infty} \frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^{s} g(Z_n)\right)\left(\sum_{n=1}^{s} g(Z_n)\right)^T\right] = \begin{bmatrix} \mathbf{U}_{11}^{(Z)} & \mathbf{U}_{12}^{(Z)} \\ \mathbf{U}_{21}^{(Z)} & \mathbf{U}_{22}^{(Z)} \end{bmatrix}.\end{aligned} \tag{63}$$

We then focus on the matrix $\mathbf{U_x}$ in the form of (23). From the definition of Loewner ordering, for any matrix $\mathbf{C}$ with suitable dimension, $\mathbf{A} \geq \mathbf{B}$ leads to $\mathbf{CAC}^T \geq_L \mathbf{CBC}^T$. Let $\mathbf{C} \equiv [\mathbf{I} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}]$ such that $\mathbf{U_x} = \mathbf{C}\begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{bmatrix}\mathbf{C}^T$. Then, with (63), we have $\mathbf{U}_{\mathbf{x}}^{(W)} \geq_L \mathbf{U}_{\mathbf{x}}^{(Z)}$.

Then, for $\mathbf{V}_{\mathbf{x}}^{(W)}$ and $\mathbf{V}_{\mathbf{x}}^{(Z)}$, by the fact that $\mathbf{A}_1 \geq_L \mathbf{B}_1$ and $\mathbf{A}_2 \geq \mathbf{B}_2$ leads to $\mathbf{A}_1 + \mathbf{A}_2 \geq \mathbf{B}_1 + \mathbf{B}_2$, i.e., Loewner ordering is closed under addition, together with $\mathbf{CAC}^T \geq_L \mathbf{CBC}^T$, we have $\mathbf{V}_{\mathbf{x}}^{(W)} \geq_L \mathbf{V}_{\mathbf{x}}^{(Z)}$ since $\mathbf{U}_{\mathbf{x}}^{(W)} \geq_L \mathbf{U}_{\mathbf{x}}^{(Z)}$. Following similar steps above gives $\mathbf{V}_{\mathbf{y}}^{(W)} \geq_L \mathbf{V}_{\mathbf{y}}^{(Z)}$ because $\mathbf{U}_{22}^{(W)} \geq_L \mathbf{U}_{22}^{(Z)}$. This completes the proof.

## A.4 Asymptotic Behavior of Nonlinear GTD Algorithms

### A.4.1 Introduction to GTD2 and TDC Algorithms with Nonlinear Function Approximation

Before starting the proof, for self-contained purposes, we here present the GTD2 and TDC algorithm with nonlinear function approximation, first proposed by Maei et al. (2009). In particular, both algorithms can be represented as TTSA in (14), where

$$h_2(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) \equiv \delta_n(\mathbf{x}_n)\phi_{\mathbf{x}_n}(s_n) - \phi_{\mathbf{x}_n}(s_n)\phi_{\mathbf{x}_n}(s_n)^T\mathbf{y}_n,$$

and

(i) For TDC algorithm:

$$h_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) \equiv \delta_n(\mathbf{x}_n)\phi_{\mathbf{x}_n}(s_n) - f_n(\mathbf{x}_n, \mathbf{y}_n) - \alpha\phi_{\mathbf{x}_n}(s_{n+1})\phi_{\mathbf{x}_n}(s_n)^T\mathbf{y}_n,$$

(ii) For GTD2 algorithm:

$$h_1(\mathbf{x}_n, \mathbf{y}_n, \xi_{n+1}) \equiv (\phi_{\mathbf{x}_n}(s_n) - \alpha\phi_{\mathbf{x}_n}(s_{n+1}))\phi_{\mathbf{x}_n}(s_n)^T\mathbf{y}_n - f_n(\mathbf{x}_n, \mathbf{y}_n),$$

where $\xi_{n+1} \triangleq (s_n, s_{n+1})$, the feature vector $\phi_{\mathbf{x}}(s) \triangleq \nabla_{\mathbf{x}} V_{\mathbf{x}}(s) \in \mathbb{R}^d$ and the TD error $\delta_n(\mathbf{x}) \triangleq r(s_n, a_n, s_{n+1}) + \alpha V_{\mathbf{x}}(s_{n+1}) - V_{\mathbf{x}}(s_n) \in \mathbb{R}$. Lastly, we introduce $f_n(\mathbf{x}, \mathbf{y}) \triangleq (\delta_n(\mathbf{x}) - \phi_{\mathbf{x}}(s_n)^T\mathbf{y})\nabla_{\mathbf{x}}\phi_{\mathbf{x}}(s_n)\mathbf{y} \in \mathbb{R}^d$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. The conditions on step sizes $\beta_n$ and $\gamma_n$ are in Assumption (A1) in Section 2.1. For both algorithms, the iterates $\mathbf{x}_n$ evolve the parameter for $V_{\mathbf{x}}$ to accurately estimate the value function $V^{\boldsymbol{\pi}}$, and iterates $\mathbf{y}_n$ aim to approximate $\mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\delta_n(\mathbf{x})\phi_{\mathbf{x}}(s_n)]$ for each $\mathbf{x}$ value from iterates $\mathbf{x}_n$. As demonstrated in Maei et al. (2009, Corollary 1), iterates $(\mathbf{x}_n, \mathbf{y}_n)$ admit a root $(\mathbf{x}^*, \mathbf{y}^*)$, where $\mathbf{x}^*$ satisfies $\mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\delta_n(\mathbf{x}^*)\phi_{\mathbf{x}^*}(s_n)] = \mathbf{0}$, and $\mathbf{y}^* = \mathbf{0}$.

In the following, we list the conditions (C1) - (C4) commonly assumed in RL.

C1. For any $s \in \mathcal{S}$ and $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$, $|V_{\mathbf{x}}(s)| \le C_v$, $\|\phi_{\mathbf{x}}(s)\| \le C_\phi$, and $\|\nabla_{\mathbf{x}}\phi_{\mathbf{x}}(s)\| \le D_v$ for some positive constants $C_v, C_\phi, D_v$. Besides, we assume $f_n(\mathbf{x}, \mathbf{y})$ is Lipschitz continuous in $\mathbf{y}$ for any $\mathbf{x} \in \mathbb{R}^d$;

C2. The point $(\mathbf{x}^*, \mathbf{y}^*)$, where $\mathbf{x}^*$ satisfies $\mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\delta_n(\mathbf{x}^*)\phi_{\mathbf{x}^*}(s_n)] = \mathbf{0}$ and $\mathbf{y}^* = \mathbf{0}$, is the globally asymptotically stable equilibrium of the related ODE $\dot{\mathbf{x}} = \bar{h}_1(\mathbf{x}, \mathbf{C}(\mathbf{x})^{-1}\mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\delta_n(\mathbf{x})\phi_{\mathbf{x}}(s_n)])$, where $\mathbf{C}(\mathbf{x}) \triangleq \mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\phi_{\mathbf{x}}(s)\phi_{\mathbf{x}}(s)^T]$;

C3. Matrix $\mathbf{C}(\mathbf{x}) \ge_L \zeta\mathbf{I} >_L \mathbf{0}, \forall \mathbf{x} \in \mathbb{R}^d$ for some $\zeta > 0$, and we use the shorthand notation for $\mathbf{C}_* \triangleq \mathbf{C}(\mathbf{x}^*)$. We also assume that $\mathbf{A}_*$ is full rank, where $\mathbf{A}_* \triangleq \mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\phi_{\mathbf{x}^*}(s_n)(\phi_{\mathbf{x}^*}(s_n) - \alpha\phi_{\mathbf{x}^*}(s_{n+1}))^T + \delta_n(\mathbf{x}^*)\nabla_{\mathbf{x}}\phi_{\mathbf{x}^*}(s_n)]$;[6]

C4. The Markov chain with transition kernel $\mathbb{P}(s_{n+1} = s' | s_n = s) \triangleq \sum_{a \in \mathcal{A}} \boldsymbol{\pi}(a|s)P(s'|s, a)$ is ergodic;

C5. $\sup_{n \ge 0}(\|\mathbf{x}_n\| + \|\mathbf{y}_n\|) < \infty$ a.s.

The boundedness assumption imposed on $\phi_{\mathbf{x}}(s), \nabla_{\mathbf{x}}\phi_{\mathbf{x}}(s)$ and $V_{\mathbf{x}}(s)$ in Condition (C1), as well as the Lipschitz condition on $f_n(\mathbf{x}, \mathbf{y})$, are in line with the assumptions made in the state-of-the-art work for nonlinear TDC algorithm (Xu and Liang, 2021; Wang et al., 2021). Condition (C2) is to ensure the globally asymptotically stability of the related ODE $\dot{\mathbf{x}} = \bar{h}_1(\mathbf{x}, \mathbf{C}(\mathbf{x})^{-1}\mathbb{E}_{\boldsymbol{\mu}^{\boldsymbol{\pi}}}[\delta_n(\mathbf{x})\phi_{\mathbf{x}}(s_n)])$. A similar assumption has been made in Maei et al. (2009, Section 5) where they consider the asymptotically stable equilibrium for any trajectory of the aforementioned ODE in a compact set because their algorithms project iterates $(\mathbf{x}_n, \mathbf{y}_n)$ onto that set. Condition (C3) aligns with current work on the nonlinear function approximation Maei et al. (2009); Wang et al. (2021). In the special case of linear function approximation, Condition (C3) is also widely used in the literature (Sutton et al., 2009; Dalal et al., 2018, 2020; Li et al., 2023a). Condition (C4) is typical for Markovian samples (Ma et al., 2020; Xu and Liang, 2021; Wang et al., 2021). Condition (C5) ensures the stability of $(\mathbf{x}_n, \mathbf{y}_n)$, which serves the same purpose as the projection operator for the iterates $(\mathbf{x}_n, \mathbf{y}_n)$ in the original GTD2 and TDC algorithms in Maei et al. (2009), where $\|\mathbf{x}_n\|$, $\|\mathbf{y}_n\|$ remain constrained by an upper bound.

### A.4.2 Proof of Proposition 3.3

We now explain how conditions (C1) – (C5) correspond to assumptions (A2) – (A5) in order to apply our main CLT result in Theorem 2.2. By Condition (C1), we have that

$$\|h_2(\mathbf{x}, \mathbf{y}, \xi)\| \le (r_{\max} + (1+\alpha)C_v)C_\phi + C_\phi^2\|\mathbf{y}\| = O(1 + \|\mathbf{y}\|),$$

where $r_{\max} \triangleq \max_{(s,a,s') \in \mathcal{S}\times\mathcal{A}\times\mathcal{S}} r(s, a, s') < \infty$. Moreover, for GTD2 algorithm,

$$\|h_1(\mathbf{x}, \mathbf{y}, \xi)\| \le (1+\alpha)C_\phi^2\|\mathbf{y}\| + \|f_n(\mathbf{x}, \mathbf{y})\| = O(1 + \|\mathbf{y}\|).$$

For TDC algorithm,

$$\|h_1(\mathbf{x}, \mathbf{y}, \xi)\| \le (r_{\max} + (1+\alpha)C_v)C_\phi + \|f_n(\mathbf{x}, \mathbf{y})\| + \alpha C_\phi^2\|\mathbf{y}\| = O(1 + \|\mathbf{y}\|).$$

Therefore, Assumption (A2) is satisfied.

[6] Only when we consider the CLT result and the fastest decaying step size $\beta_n = 1/(n+1)$, we need an extra assumption, i.e., $-\mathbf{A}_*^T\mathbf{C}_*^{-1}\mathbf{A}_* + \frac{1}{2}\mathbf{I}$ be to Hurwitz. This condition is not needed for almost sure convergence even when $\beta_n = 1/(n+1)$.

Then, we turn to verifying Assumption (A3). For any $\mathbf{x} \in \mathbb{R}^d$, to ensure $\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) = 0$, we can set $\lambda(\mathbf{x}) = \mathbf{C}(\mathbf{x})^{-1}\mathbb{E}_{\boldsymbol{\mu}\boldsymbol{\pi}}[\delta_n(\mathbf{x})\phi_{\mathbf{x}}(s_n)]$, where $\mathbf{C}(\mathbf{x})^{-1}$ is well defined by Condition (C3). Clearly, $\lambda(\mathbf{x})$ is the globally asymptotically stable equilibrium of $\dot{\mathbf{y}} = \bar{h}_2(\mathbf{x}, \mathbf{y}) = \mathbb{E}_{\boldsymbol{\mu}}[\delta_n(\mathbf{x}), \phi_{\mathbf{x}}(s_n)] - \mathbf{C}(\mathbf{x})\mathbf{y}$ because this ODE is linear in $\mathbf{y}$, and $\nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) = -\mathbf{C}(\mathbf{x})$ is Hurwitz for any $\mathbf{x} \in \mathbb{R}^d$, as stated in Condition (C3). By Condition (C1), we have

$$\|\lambda(\mathbf{x})\| \leq \|\mathbf{C}(\mathbf{x})^{-1}\|(r_{\max} + (1+\alpha)C_v)C_\phi \leq \zeta^{-1}(r_{\max} + (1+\alpha)C_v)C_\phi.$$

When $\mathbf{x} = \mathbf{x}^*$, $\mathbf{y}^* = \lambda(\mathbf{x}^*) = 0$ such that $f_n(\mathbf{x}^*, \mathbf{y}^*) = 0$ and $\hat{h}_1(\mathbf{x}^*) = \bar{h}_1(\mathbf{x}^*, \lambda(\mathbf{x}^*)) = 0$ for both GTD2 and TDC algorithms. Finally, we deal with $\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*)$. By chain rule, we have $\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}) = \nabla_{\mathbf{x}}\bar{h}_1(\mathbf{x}, \lambda(\mathbf{x})) + \nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))\nabla_{\mathbf{x}}\lambda(\mathbf{x})$. Although $\lambda(\mathbf{x})$ is an implicit function, after taking the derivative on both sides of $\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) = 0$ with respective to $\mathbf{x}$, we have

$$\nabla_{\mathbf{x}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})) + \nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))\nabla_{\mathbf{x}}\lambda(\mathbf{x}) = 0.$$

This further leads to

$$\nabla_{\mathbf{x}}\lambda(\mathbf{x}) = -(\nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})))^{-1}\nabla_{\mathbf{x}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x})),$$

where the matrix inverse exists because $\nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}, \lambda(\mathbf{x}))$ is Hurwitz for any $\mathbf{x} \in \mathbb{R}^d$. Thus, we obtain the following form for $\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*)$:

$$\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*) = \nabla_{\mathbf{x}}\bar{h}_1(\mathbf{x}^*, \lambda(\mathbf{x}^*)) - \nabla_{\mathbf{y}}\bar{h}_1(\mathbf{x}^*, \lambda(\mathbf{x}^*))(\nabla_{\mathbf{y}}\bar{h}_2(\mathbf{x}^*, \lambda(\mathbf{x}^*)))^{-1}\nabla_{\mathbf{x}}\bar{h}_2(\mathbf{x}^*, \lambda(\mathbf{x}^*)). \tag{64}$$

After algebraic calculation, for both GTD2 and TDC algorithms, $\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*) = -\mathbf{A}_*^T\mathbf{C}_*^{-1}\mathbf{A}_*$. Now that $\mathbf{A}_*$ is full rank, for any non-zero vector $\mathbf{v} \in \mathbb{R}^d$, we have $\mathbf{v}^T\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*)\mathbf{v} = -\mathbf{v}^T\mathbf{A}_*\mathbf{C}_*^{-1}\mathbf{A}_*\mathbf{v} = -\mathbf{u}^T\mathbf{C}_*\mathbf{u} < 0$ by letting $\mathbf{u} = \mathbf{A}_*\mathbf{v} \neq \mathbf{0}$. Thus, $\nabla_{\mathbf{x}}\hat{h}_1(\mathbf{x}^*)$ is negative definite and thus Hurwitz. Additionally, Condition (C2) indicates that $(\mathbf{x}^*, \lambda(\mathbf{x}^*))$ is the globally asymptotically stable equilibrium of the related ODE $\dot{\mathbf{x}} = \bar{h}_1(\mathbf{x}, \lambda(\mathbf{x}))$. Therefore, Assumption (A3) is verified.

Regarding $\{\xi_n\}$ with $\xi_{n+1} = (s_n, s_{n+1})$, it can seen as the Markov chain $\{s_n\}$ on the augmented state space. Then, we have the following result.

**Theorem A.4.1** (Neal (2004) Theorem 2)**.** *Suppose that $\{s_n\}$ is an irreducible, reversible Markov chain on the finite state space $\mathcal{V}$ with transition matrix $\mathbf{P} = \{P(i,j)\}$ and stationary distribution $\boldsymbol{\pi}$. Construct a Markov chain $\{\xi_n\}$ on the augmented state space $\mathcal{E} \triangleq \{(i,j) : i,j \in \mathcal{V} \ \ s.t. \ \ P(i,j) > 0\} \subseteq \mathcal{V} \times \mathcal{V}$ with transition matrix $\boldsymbol{P'} = \{P'(e_{ij}, e_{lk})\}$ in which $e_{ij} \triangleq (s = i, s' = j)$ and the transition probabilities $P'(e_{ij}, e_{lk})$ satisfy the following two conditions: for all $e_{ij}, e_{jk} \in \mathcal{E}$,*

$$P'(e_{ij}, e_{jk}) = P(j,k). \tag{65}$$

*Then, the Markov chain $\{\xi_n\}_{n\geq 0}$ is irreducible with a unique stationary distribution $\boldsymbol{\pi}'$ in which*

$$\pi'(e_{ij}) = \pi_i P(i,j) = \pi_j P(j,i), \quad e_{ij} \in \mathcal{E}. \tag{66}$$

By Theorem A.4.1, $\{\xi_n\}$ is an ergodic Markov chain and thus satisfies Assumption (A4). Condition (C4) matches Assumption (A5). Therefore, we can apply Lemma 2.1 (almost sure convergence) and Theorem 2.2 (CLT result) to GTD2 and TDC algorithms. The results are given as follows.

$$\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}^* \quad \text{a.s. and} \quad \lim_{n\to\infty} \mathbf{y}_n = \mathbf{0} \quad \text{a.s.}$$
$$\frac{1}{\sqrt{\beta_n}}(\mathbf{x}_n - \mathbf{x}^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{\mathbf{x}}), \ \frac{1}{\sqrt{\gamma_n}}\mathbf{y}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{\mathbf{y}}),$$

where $\mathbf{V}_{\mathbf{x}}, \mathbf{V}_{\mathbf{y}}$ are identical for both algorithms, and

$$\mathbf{V}_{\mathbf{x}} = \int_0^\infty e^{t\mathbf{K}'}\mathbf{U}_{\mathbf{x}}e^{t(\mathbf{K}')^T}, \quad \mathbf{V}_{\mathbf{y}} = \int_0^\infty e^{-t\mathbf{C}_*}\mathbf{U}_{\mathbf{y}}e^{-t\mathbf{C}_*^T},$$

where

$$\mathbf{K}'_{\mathbf{x}} = -\mathbf{A}_*^T\mathbf{C}_*^{-1}\mathbf{A}_* + \frac{\mathbb{1}_{\{b=1\}}}{2}\mathbf{I}, \tag{67}$$

$$\mathbf{U}_{\mathbf{y}} = \lim_{s\to\infty}\frac{1}{s}\mathbb{E}\left[\left(\sum_{n=1}^s \delta_n(\mathbf{x}^*)\phi_{\mathbf{x}^*}(s_n)\right)\left(\sum_{n=1}^s \delta_n(\mathbf{x}^*)\phi_{\mathbf{x}^*}(s_n)\right)^T\right], \tag{68}$$

$$\mathbf{U}_{\mathbf{x}} = \mathbf{A}_*^T\mathbf{C}_*^{-1}\mathbf{U}_{\mathbf{y}}\mathbf{C}_*^{-1}\mathbf{A}_*. \tag{69}$$

## A.5 Simulation Setups and Additional Numerical Results

In this appendix, we provide a more detailed illustration of the numerical results. The simulations are conducted on a PC with AMD Ryzen R9 5950X, 128GB RAM and RTX 3080.

### A.5.1 Distributed Learning in Section 3.1

#### A.5.1.1 Simulation Setup for L2-regularized Binary Classification

In Section 3.1, we perform the L2-regularized binary classification problem using the momentum SGD algorithm on the wikiVote graph (Leskovec and Krevl, 2014). Specifically, the problem has the following objective function:

$$\min_{\mathbf{x}\in\mathbb{R}^d}\left\{f(\mathbf{x})=\frac{1}{N}\sum_{i=1}^{N}F(\mathbf{x},i)\triangleq\frac{1}{N}\sum_{i=1}^{N}\log\left(1+e^{\mathbf{x}^T\mathbf{s}_i}\right)-z_i\left(\mathbf{x}^T\mathbf{s}_i\right)+\frac{\kappa}{2}\|\mathbf{x}\|^2\right\},\tag{70}$$

where $\{(\mathbf{s}_i,z_i)\}_{i=1}^N$ is the *a9a* dataset (with 123 features, i.e., $\mathbf{s}_i\in\mathbb{R}^{123}$) from LIBSVM (Chang and Lin, 2011), and penalty parameter $\kappa=1$. The momentum SGD algorithm (Gadat et al., 2018; Li et al., 2022) employed in this simulation is given below.

**Algorithm 1:** Momentum SGD

Initial parameters $\mathbf{x}_0,\mathbf{y}_0=\mathbf{0}$, data point $\xi_0$, step sizes $\beta_n=(n+1)^{-1}$ and $\gamma_n=(n+1)^{-0.501}$, number of iterations $T$;
**for** $n=0$ **to** $T$ **do**
  Sample new data point: $\xi_{n+1}\leftarrow$ Sampling Strategy
  Compute gradient: $\mathbf{g}_n\leftarrow\nabla F(\mathbf{x}_n,\xi_{n+1})$
  Update momentum: $\mathbf{y}_{n+1}\leftarrow\mathbf{y}_n-\gamma_{n+1}(\mathbf{g}_n+\mathbf{y}_n)$
  Update parameter: $\mathbf{x}_{n+1}\leftarrow\mathbf{x}_n+\beta_{n+1}\mathbf{y}_n$
**end**

For *i.i.d.* sampling and single shuffling in the simulation, where the whole dataset is available in each iteration, we have the following schemes: at $n$-th iteration,

- ***i.i.d.* sampling**: $\xi_{n+1}$ is sampled from $[N]$ uniformly at random;
- **single shuffling**: At the beginning of the simulation, we shuffle the sequence $\{1,2,\cdots,N\}$ by the permutation operator $\sigma:[N]\to[N]$ and then sample the data point according to $\xi_{n+1}=\sigma(n+1 \bmod N)$.

When dataset is distributed over the wikiVote graph, i.e., each node on the graph is assigned a data point, we use simple random walk (SRW) and its sampling-efficient counterpart non-backtracking random walk (NBRW) (Alon et al., 2007; Lee et al., 2012; Ben-Hamou et al., 2018) in the simulation. Specifically, denote by $\xi_n$ the index of the node in the $n$-th iteration and $\mathcal{N}(\xi_n)$ the list of neighboring nodes of node $\xi_n$, we have

- **SRW**: $\xi_{n+1}$ is sampled from $\mathcal{N}(\xi_n)$ uniformly at random;
- **NBRW**: $\xi_{n+1}$ is sampled from $\mathcal{N}(\xi_n)\backslash\{\xi_{n-1}\}$ uniformly at random. If $\mathcal{N}(\xi_n)\backslash\{\xi_{n-1}\}=\emptyset$, then $\xi_{n+1}=\xi_{n-1}$.

Note that SRW and NBRW both have a stationary distribution proportional to degree distribution, while the objective function indicates that each node is treated equally, which results in the bias from SRW and NBRW. To overcome this problem, we employ importance reweighting, e.g., by modifying the momentum update step in Algorithm 1 in the following form:

$$\mathbf{y}_{n+1}\leftarrow\mathbf{y}_n-\gamma_{n+1}\left(\mathbf{g}_n\cdot\frac{1}{d_{\xi_{n+1}}}-\mathbf{y}_n\right),$$

where $d_i$ is defined as the degree of node $i$.

#### A.5.1.2 Additional Simulation on Distributed Minimax Problem

In this part, we consider the following minimax problem

$$\min_{\mathbf{x}\in\mathbb{R}^d}\max_{\mathbf{y}\in\mathbb{R}^d}\left\{f(\mathbf{x},\mathbf{y})=\frac{1}{N}\sum_{i=1}^{N}F(\mathbf{x},\mathbf{y},i)\triangleq\frac{1}{N}\sum_{i=1}^{N}-\left[\frac{1}{2}\|\mathbf{y}\|^2-\mathbf{b}(i)^T\mathbf{y}+\mathbf{y}^T\mathbf{A}(i)\mathbf{x}\right]+\frac{\kappa}{2}\|\mathbf{x}\|^2\right\}, \tag{71}$$

and follow the same setup as in Tarzanagh et al. (2022) to generate the dataset. In particular, we let $\kappa=10$, $d=10$, $\mathbf{b}(i)=\mathbf{b}'(i)-\frac{1}{N}\sum_{i=1}^{N}\mathbf{b}'(i)$ and $\mathbf{A}(i)=t_i\mathbf{I}$, where $\mathbf{b}'(i)\sim N(0,\mathbf{I})$ and $t_i$ are drawn from $(0,0.1)$ uniformly at random. We test the distributed minimax problem over the WikiVote graph (Leskovec and Krevl, 2014), where each node is assigned a data point, thus 889 data points in total. To solve the minimax problem (71), we leverage the stochastic gradient descent ascent (SGDA) algorithm as follows, which is regarded as a special case of TTSA (Lin et al., 2020).

**Algorithm 2:** SGDA

Initial parameters $\mathbf{x}_0,\mathbf{y}_0=\mathbf{0}$, data point $\xi_0$, step sizes $\beta_n=(n+1)^{-1}$ and $\gamma_n=(n+1)^{-0.8}$, number of iterations $T$;
**for** $n=0$ **to** $T$ **do**
  Sample new data point: $\xi_{n+1}\leftarrow$ Sampling Strategy
  Compute gradient w.r.t $\mathbf{x}$: $\mathbf{g}_n\leftarrow\nabla_{\mathbf{x}}F(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})$
  Compute gradient w.r.t $\mathbf{y}$: $\mathbf{h}_n\leftarrow\nabla_{\mathbf{y}}F(\mathbf{x}_n,\mathbf{y}_n,\xi_{n+1})$
  Update inner parameter: $\mathbf{y}_{n+1}\leftarrow\mathbf{y}_n+\gamma_{n+1}\mathbf{h}_n$
  Update outer parameter: $\mathbf{x}_{n+1}\leftarrow\mathbf{x}_n-\beta_{n+1}\mathbf{g}_n$
**end**

In this simulation, we compare two pairs of sampling strategies for the performance ordering, i.e., SRW versus NBRW, *i.i.d.* sampling versus single shuffling, which have been introduced in Appendix A.5.1.1. Especially, for SRW and NBRW, we reweight the gradient in the update of both inner and outer parameters, i.e.,

$$\mathbf{y}_{n+1}\leftarrow\mathbf{y}_n+\gamma_{n+1}\mathbf{h}_n\cdot\frac{1}{d_{\xi_{n+1}}},$$
$$\mathbf{x}_{n+1}\leftarrow\mathbf{x}_n-\beta_{n+1}\mathbf{g}_n\cdot\frac{1}{d_{\xi_{n+1}}}.$$

In both Figure 4(a) and Figure 5(a), we observe that NBRW has a smaller MSE than SRW across all time $n$ in iterates $\mathbf{x}_n$ and iterates $\mathbf{y}_n$, with a similar trend for single shuffling over *i.i.d.* sampling. Figure 4(b) and Figure 5(b) demonstrate that for both iterates $\mathbf{x}_n$ and iterates $\mathbf{y}_n$, the rescaled MSEs of NBRW, SRW and *i.i.d.* sampling approach some constants, while the curve for single shuffling still decreases in linear rate because eventually the limiting covariance matrix therein will be zero. This simulation result, along with the one in Section 3.1, demonstrates the effectiveness of Proposition 3.2 in Section 3.1, even under the finite-time regime.

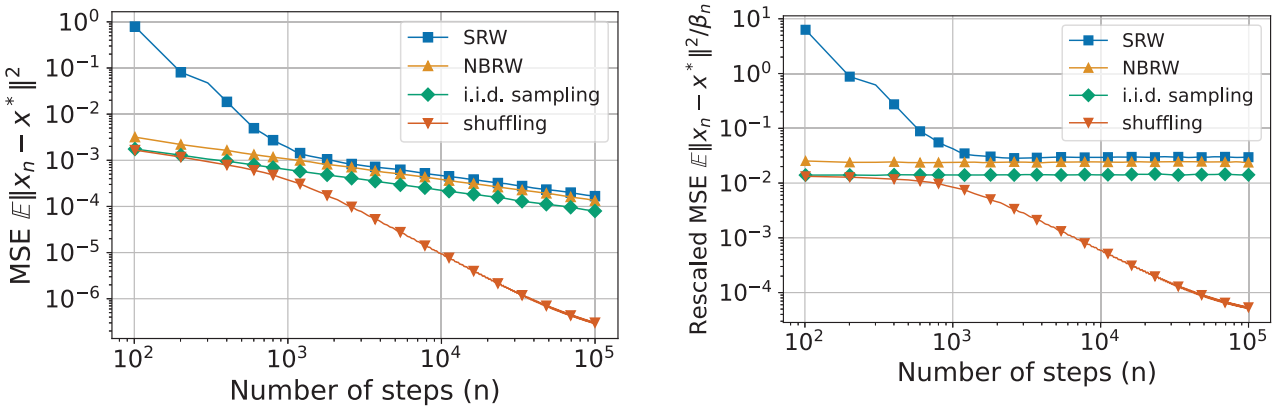
### A.5.2 Random Walk Task for GTD2 and TDC algorithms in Section 3.2

#### A.5.2.1 Simulation Setups and Computation of $\mathbf{V}_{\mathbf{x}}$

For the 5-state random walk task, the problem setting is given in Figure 6. Using the value iteration algorithm, we obtain the true value function $W(s)=0$ for $s=\{1,2,3,4,5\}$. In the simulation in Section 3.2, we consider the discount factor $\alpha=0.9$, and the nonlinear function approximation $W_x(s)=a(s)(e^{0.1x}-1)$, where $a=[-2,-6,-3,-4,-5]$, and the optimal parameter $x^*=0$. Then, $\phi_x(s)=0.1\cdot a(s)e^{0.1x}$ and $\nabla_x\phi_x(s)=0.01\cdot a(s)e^{0.1x}$. In both GTD2 and TDC algorithms, we set the step sizes $\beta_n=(n+1)^{-0.6},\gamma_n=(n+1)^{-0.501}$.

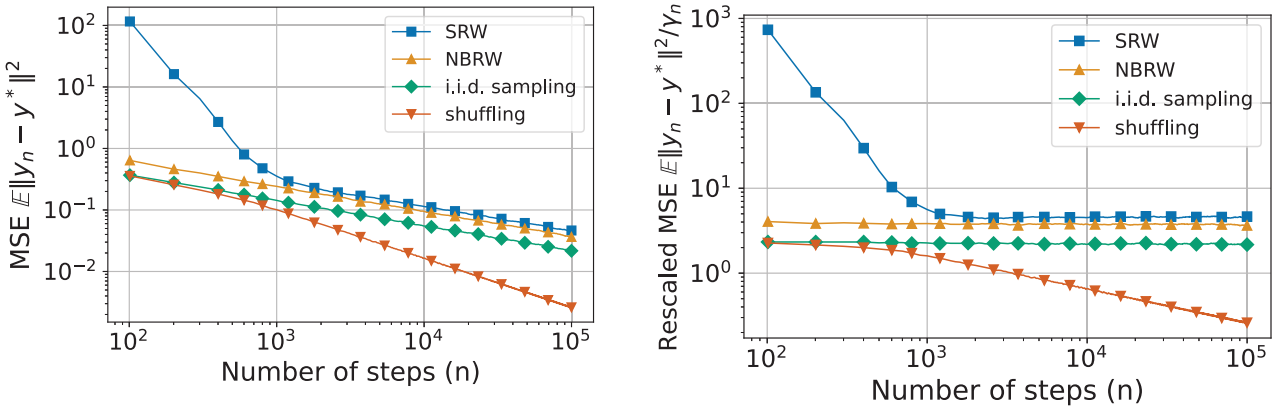
Now, we leverage the expression in Appendix A.4.2 to calculate the theoretical value of $\mathbf{V}_{\mathbf{x}}$,[7] which is used for

[7] In this task, all matrices, e.g., $\mathbf{V}_{\mathbf{x}},\mathbf{K}_{\mathbf{x}},\mathbf{C}_*,\mathbf{A}_*,\mathbf{U}_{\mathbf{x}},\mathbf{U}_{\mathbf{y}}$, degenerate to scalars.




(a) MSE (b) Rescaled MSE

Figure 4: Comparison of the performance ordering in SGDA in terms of iterates $\mathbf{x}_n$.




(a) MSE (b) Rescaled MSE

Figure 5: Comparison of the performance ordering in SGDA in terms of iterates $\mathbf{y}_n$.

Figure 3 in Section 3.2. In particular, $\mu_i = 0.2$ for $i \in \{1, 2, \cdots, 5\}$, and

$$\mathbf{C}_* = \mathbb{E}_{\boldsymbol{\mu}}\left[\phi_{x^*}(s_n)^2\right] = \sum_{i=1}^{5} \frac{1}{5} \cdot 0.01 \cdot a(i)^2 = 0.18.$$

Note that $\mathbf{V}_\mathbf{x}$ now becomes $\mathbf{V}_\mathbf{x} = 2\mathbf{K}_\mathbf{x}^{-1} \cdot \mathbf{U}_\mathbf{x} = \frac{1}{2}\mathbf{C}_*^{-1}\mathbf{U}_\mathbf{y} = \frac{25}{9}\mathbf{U}_\mathbf{y}$. Then, we run the simulation to estimate $\mathbf{U}_\mathbf{y}$ of the form in (67), which gives $\mathbf{U}_\mathbf{y} \approx 0.0174$ with 100 independent trials. Thus, we have $\mathbf{V}_\mathbf{x} \approx 0.0484$.

#### A.5.2.2 Additional Choice of Nonlinear Function Approximation

In this part, we conduct the 5-state random walk task for GTD2 and TDC algorithms with step sizes $\beta_n = (n+1)^{-0.6}, \gamma_n = (n+1)^{-0.501}$, and another choice of the nonlinear function approximation, i.e., $W_x(s) = 0.1 \cdot (x + \sin(x))$, which becomes the ground truth $W(s) = 0$ by setting $x^* = 0$. Then, we have

$$\mathbf{C}_* = \mathbb{E}_{\boldsymbol{\mu}}\left[\phi_{x^*}(s_n)^2\right] = \sum_{i=1}^{5} \frac{1}{5} \cdot 0.04 \cdot a(i)^2 = 0.72,$$

such that $\mathbf{V}_\mathbf{x} = \frac{1}{2}\mathbf{C}_*\mathbf{U}_\mathbf{y} = \frac{25}{36}\mathbf{U}_\mathbf{y}$. Similarly, we run the simulation to compute $\mathbf{U}_\mathbf{y} \approx 0.1384$ so that $\mathbf{V}_\mathbf{x} \approx 0.0961$.

Figure 7 shows the long-term performance of both the GTD2 and TDC algorithms, as well as the deviation from the optimal value $x^*$ at time $n = 10^8$. This is in agreement with our Proposition 3.3 in Section 3.2. Specifically,

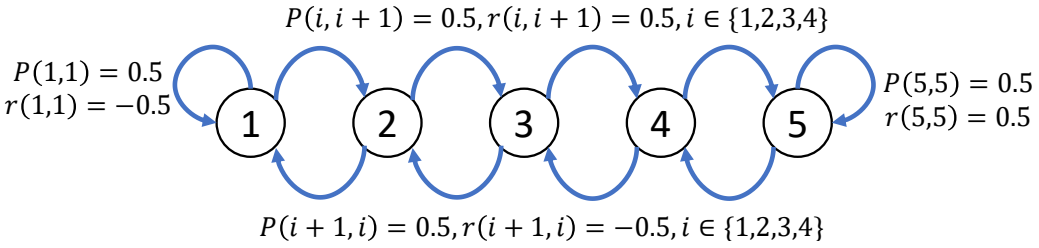
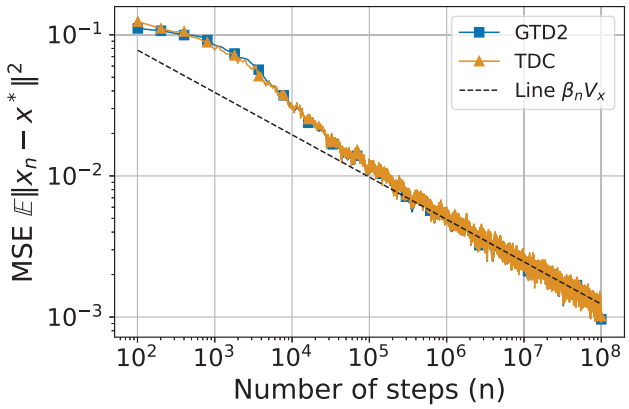
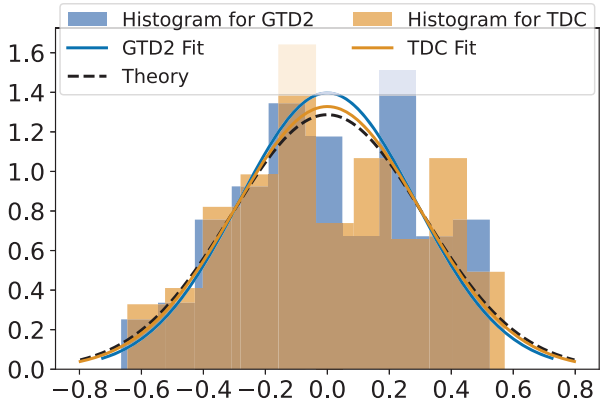



Figure 6: 5-state random walk problem.

we show in 7(a) that for this choice of nonlinear function approximation, the GTD2 and TDC algorithms achieve almost the same performance starting from $n = 10^4$. This is because $W_x(s) = 0.1 \cdot (x + \sin(x))$ acts more like a linear function in $x$ than the selection $W_x(s) = a(s)(e^{0.1x} - 1)$ in Section 3.2, which means the effect of $f_n(\mathbf{x}, \mathbf{y})$ introduced by the nonlinear approximation is reduced in the iterates $\mathbf{x}_n$ for GTD2 and TDC algorithms, thus diminishing the performance gap between these two algorithms. Figure 7(b) represents the histogram of $\beta_n^{-1/2} x_n$ for both algorithms from 100 independent experiments. Their experimental density curves approach the theoretical Gaussian curve with zero mean and variance $\mathbf{V_x} = 0.0961$.




(a) MSE




(b) Histograms of $\beta_n^{-1/2} x_n$ at $n = 10^8$

Figure 7: Comparison of nonlinear GTD2 and TDC algorithms in the 5-state random walk task with the nonlinear function approximation $W_x(s) = 0.1 \cdot (x + \sin(x))$.

## A.6 Useful Theoretical Results

**Lemma A.6.1** (Abel Transformation). *Suppose $\{f_k\}$ and $\{g_k\}$ are two sequences. Then,*

$$\sum_{k=m}^{n} f_k(g_{k+1-g_k}) = f_n g_{n+1} - f_m g_m - \sum_{k=m+1}^{n} g_k(f_k - f_{k-1}).$$

**Lemma A.6.2** (Burkholder Inequality, Davis (1970), Hall et al. (2014) Theorem 2.10). *Given a Martingale difference sequence $\{M_{i,n}\}_{i=1}^n$, for $p \geq 1$ and some positive constant $C_p$, we have*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{n} M_{i,n}\right\|^p\right] \leq C_p \mathbb{E}\left[\left(\sum_{i=1}^{n} \|M_{i,n}\|^2\right)^{p/2}\right] \tag{72}$$

**Theorem A.6.3** (Martingale CLT, Delyon (2000) Theorem 30). *If a Martingale difference array $\{X_{n,i}\}$ satisfies the following condition: for some $\tau > 0$,*

$$\sum_{k=1}^{n} \mathbb{E}\left[\|X_{n,k}\|^{2+\tau} | \mathcal{F}_{k-1}\right] \xrightarrow{\mathbb{P}} 0, \quad \sup_n \sum_{k=1}^{n} \mathbb{E}\left[\|X_{n,k}\|^2 | \mathcal{F}_{k-1}\right] < \infty, \quad \sum_{k=1}^{n} \mathbb{E}\left[X_{n,k} X_{n,k}^T | \mathcal{F}_{k-1}\right] \xrightarrow{\mathbb{P}} \boldsymbol{V}, \tag{73}$$

*then*

$$\sum_{i=1}^{n} X_{n,i} \xrightarrow{d} N(0, \boldsymbol{V}). \tag{74}$$

**Lemma A.6.4** (Duflo (1996) Proposition 3.I.2)**.** *For a Hurwitz matrix* $\mathbf{H}$*, there exist some positive constants* $C, b$ *such that for any* $n$*,*

$$\left\|e^{\mathbf{H}n}\right\| \leq Ce^{-bn}. \tag{75}$$

**Lemma A.6.5** (Fort (2015) Lemma 5.8)**.** *For a Hurwitz matrix* $\mathbf{A}$*, denote by* $-r$*,* $r > 0$*, the largest real part of its eigenvalues. Let a positive sequence* $\{\gamma_n\}$ *such that* $\lim_n \gamma_n = 0$*. Then for any* $0 < r' < r$*, there exists a positive constant* $C$ *such that for any* $k < n$*,*

$$\left\|\prod_{j=k}^{n}(\mathbf{I} + \gamma_j \mathbf{A})\right\| \leq Ce^{-r'\sum_{j=k}^{n}\gamma_j}. \tag{76}$$

**Lemma A.6.6** (Fort (2015) Lemma 5.9, Mokkadem and Pelletier (2006) Lemma 10)**.** *Let* $\{\gamma_n\}$ *be a positive sequence such that* $\lim_n \gamma_n = 0$ *and* $\sum_n \gamma_n = \infty$*. Let* $\{\epsilon_n, n \geq 0\}$ *be a nonnegative sequence. Then, for* $b > 0$*,* $p \geq 0$*,*

$$\limsup_n \gamma_n^{-p} \sum_{k=1}^{n} \gamma_k^{p+1} e^{-b\sum_{j=k+1}^{n}\gamma_j}\epsilon_k \leq \frac{1}{C(b,p)} \limsup_n \epsilon_n \tag{77}$$

*for some constant* $C(b, p) > 0$*.*

*When* $p = 0$ *and define a positive sequence* $\{w_n\}$ *satisfying* $w_{n-1}/w_n = 1 + o(\gamma_n)$*, we have*

$$\sum_{k=1}^{n} \gamma_k e^{-b\sum_{j=k+1}^{n}\gamma_j}\epsilon_k = \begin{cases} O(w_n), & \text{if } \epsilon_n = O(w_n), \\ o(w_n), & \text{if } \epsilon_n = o(w_n). \end{cases} \tag{78}$$

**Lemma A.6.7** (Fort (2015) Lemma 5.10)**.** *For any matrices* $A, B, C$*,*

$$\|ABA^T - CBC^T\| \leq \|A - C\|\|B\|(\|A\| + \|C\|). \tag{79}$$

**Lemma A.6.8.** *Suppose* $\mathbf{K}$ *is a Hurwitz matrix. Then, for any positive semi-definite matrix* $\mathbf{U}$*, there exists a unique positive semi-definite matrix* $\mathbf{V}$ *such that* $\mathbf{KV} + \mathbf{VK}^T + \mathbf{U} = 0$ *(Lyapunov equation), where the closed form of* $\mathbf{V}$ *is given by*

$$\mathbf{V} = \int_0^\infty e^{t\mathbf{K}}\mathbf{U}e^{t\mathbf{K}^T}dt. \tag{80}$$

Lemma A.6.8 come from Theorem 3.16 Chellaboina and Haddad (2008). Nevertheless, they necessitate a positive definite matrix $\mathbf{U}$ so that the solution $\mathbf{V}$ is also positive definite. Throughout this paper, we do not require the solution $\mathbf{V}$ to be positive definite so that the matrix $\mathbf{U}$ can be relaxed to be positive semi-definite. This relaxation does not change any steps as in the proof of Theorem 3.16 Chellaboina and Haddad (2008), and is thus omitted here.