Gaussian Approximation for Two-Timescale Linear Stochastic Approximation

Bogdan Butyrin[1,*], **Artemy Rubtsov**[1,*], **Alexey Naumov**[1], **Vladimir V. Ulyanov**[1,2], **Sergey Samsonov**[1]

[1]HSE University
[2]Lomonosov Moscow State University
bbutyrin@hse.ru, asrubtsov@hse.ru

Abstract

In this paper, we establish non-asymptotic bounds for accuracy of normal approximation for linear two-timescale stochastic approximation (TTSA) algorithms driven by martingale difference or Markov noise. Focusing on both the last iterate and Polyak–Ruppert averaging regimes, we derive bounds for normal approximation in terms of the convex distance between probability distributions. Our analysis reveals a non-trivial interaction between the fast and slow timescales: the normal approximation rate for the last iterate improves as the timescale separation increases, while it decreases in the Polyak–Ruppert averaged setting. We also provide the high-order moment bounds for the error of linear TTSA algorithm, which may be of independent interest. Finally, we demonstrate that our theoretical results are directly applicable to reinforcement learning algorithms such as GTD and TDC.

Extended version — https://arxiv.org/abs/2508.07928v2

1 Introduction

Stochastic approximation (SA) methods play an important role in the field of machine learning, especially due to their role in solving reinforcement learning (RL) problems (Sutton and Barto 2018). Recent studies cover both asymptotic (Nemirovskij and Yudin 1983; Polyak and Juditsky 1992) and non-asymptotic (Moulines and Bach 2011) properties of SA estimates. In particular, two-timescale stochastic approximation (TTSA) algorithms (Borkar 1997) refer to the class of methods that update two interdependent variables with separate step size sequences, one typically decreasing faster than the other. This class of methods is especially important in RL, where policy evaluation in the off-policy setting requires TTSA methods such as the Gradient Temporal Difference (GTD) method (Sutton, Maei, and Szepesvári 2008).

An important question for SA algorithms is related to the accuracy of Gaussian approximation (GAR) of the constructed estimates. Classical results on GAR for SA algorithms, such as (Polyak and Juditsky 1992; Konda and Tsitsiklis 2004), are asymptotic and do not provide convergence rates. At the same time, the latter results play an important role in statistical inference for optimization (Fan 2019), as they pave the way for non-asymptotic analysis of various procedures for constructing confidence intervals. We focus on the linear two-timescale SA problem, that is, we aim to find a solution $(\theta^\star, w^\star)$ that solves the system of linear equations:

$$A_{11}\theta + A_{12}w = b_1 \ , \quad A_{21}\theta + A_{22}w = b_2 \ ,$$

assuming that the solution $(\theta^\star, w^\star)$ is unique and is given by

$$\theta^\star = \Delta^{-1}(b_1 - A_{12}A_{22}^{-1}b_2), \quad w^\star = A_{22}^{-1}(b_2 - A_{21}\theta^\star) \ ,$$

with $\Delta := A_{11} - A_{12}A_{22}^{-1}A_{21}$. We consider the setting, where the underlying matrices A_{ij} and vectors b_i, $i,j \in \{1,2\}$, are not accessible. Instead, following (Borkar 1997), we assume that the learner has access to a sequence of random variables $\{X_k\}_{k\in\mathbb{N}}$ taking values in a measurable space $(\mathsf{X}, \mathcal{X})$, and vector/matrix-valued functions $\mathbf{b}_i(x)$, $\mathbf{A}_{ij}(x)$, $i,j \in \{1,2\}$, which serves as stochastic estimates of b_i and A_{ij}, respectively. The corresponding recurrence runs as

$$\begin{aligned} \theta_{k+1} &= \theta_k + \beta_k\{\mathbf{b}_1^{k+1} - \mathbf{A}_{11}^{k+1}\theta_k - \mathbf{A}_{12}^{k+1}w_k\} \ , \\ w_{k+1} &= w_k + \gamma_k\{\mathbf{b}_2^{k+1} - \mathbf{A}_{21}^{k+1}\theta_k - \mathbf{A}_{22}^{k+1}w_k\} \ , \end{aligned} \tag{1}$$

where $\theta_k \in \mathbb{R}^{d_\theta}$, $w_k \in \mathbb{R}^{d_w}$, and $\mathbf{b}_i^k$, $\mathbf{A}_{ij}^k$ are shorthand notations for $\mathbf{b}_i(X_k)$ and $\mathbf{A}_{ij}(X_k)$, respectively. The scalars $\gamma_k, \beta_k > 0$ in (1) are step sizes, and the underlying SA scheme is said to have two timescales as the step sizes satisfy $\lim_{k\to\infty}\beta_k/\gamma_k < 1$ such that w_k is updated at a faster timescale. In our paper we consider $\beta_k = c_{0,\beta}(k+k_0)^{-b}$ and $\gamma_k = c_{0,\gamma}(k+k_0)^{-a}$ with exponents a and b satisfying $1/2 < a < b < 1$. When $\{X_k\}_{k\in\mathbb{N}}$ are i.i.d., and under appropriate technical assumptions on the parameters of (1), it is known (see e.g. (Konda and Tsitsiklis 2004)), that the asymptotic normality of the "slow" timescale θ_k holds:

$$\beta_k^{-1/2}(\theta_k - \theta^\star) \to \mathcal{N}(0, \Sigma_\theta) \ , \tag{2}$$

with some covariance Σ_θ. The authors in (Mokkadem, Pelletier et al. 2006) generalized this result for the averaged iterates of non-linear SA:

$$\bar{\theta}_n := n^{-1}\textstyle\sum_{k=1}^n \theta_k \ , \quad \bar{w}_n := n^{-1}\sum_{k=1}^n w_k \ . \tag{3}$$

The latter estimates correspond to the Polyak-Ruppert averaging procedure introduced in (Ruppert 1988; Polyak and Juditsky 1992), a popular technique for stabilization of the SA algorithms. The authors of the recent paper (Kong et al.

*These authors contributed equally.

2025) obtained the non-asymptotic convergence rates for the averaged iterates $\bar{\theta}_n$ and $\bar{w}_n$ in Wasserstein distance of order 1, using the vector-valued versions of the Berry-Essen theorem for martingale-difference sequences due to (Srikant 2024). In this paper, we not only generalize these results for the setting of Markov noise, but also establish the corresponding convergence rates for the last iterate θ_k. The main contributions of this paper are the following:

- We derive non-asymptotic bounds for the accuracy of normal approximation for the Polyak–Ruppert-averaged TTSA $\sqrt{n}(\bar{\theta}_n - \theta^\star)$ and last iterate $\beta_n^{-1/2}(\theta_n - \theta^\star)$ in terms of convex distance under martingale-difference noise assumptions. Our results indicate that the normal approximation for the last iterate improves as the timescale separation increases and achieves a convergence rate of order up to $n^{-1/4}$, up to $\log n$ factors. We show that the Polyak–Ruppert averaged TTSA iterates achieve the same rate of normal approximation, but require that the timescales β_k and γ_k coincide up to a constant factor. While our analysis for the Polyak–Ruppert averaged TTSA generalizes recent results due to (Kong et al. 2025), we provide, to the best of our knowledge, the first fully non-asymptotic analysis of the normal approximation rates for the last iterate of TTSA.
- We generalize the obtained results for normal approximation for the averaged TTSA and the last iterate to the setting of Markov noise. Our results show a convergence rate of order up to $n^{-1/6}$, up to logarithmic factors, with the same conclusion regarding timescale separation as in the martingale noise case. This is the first result on the normal approximation rate for TTSA with Markov noise.

Notations. For a matrix $A \in \mathbb{R}^{d\times d}$ we denote by $\|A\|$ its operator norm. For symmetric positive-definite matrix $Q = Q^\top \succ 0$, $Q \in \mathbb{R}^{d\times d}$ and $x \in \mathbb{R}^d$ we define the corresponding norm $\|x\|_Q = \sqrt{x^\top Q x}$, and define the respective matrix Q-norm of the matrix $B \in \mathbb{R}^{d\times d}$ by $\|B\|_Q = \sup_{x\neq 0} \|Bx\|_Q/\|x\|_Q$. For sequences a_n and b_n, we write $a_n \lesssim_{\log_n} b_n$ if there exist $c, \alpha > 0$ (not depending upon n), such that $a_n \leq c(1+\log n)^\alpha b_n$. In the present text, the following abbreviations are used: "w.r.t." stands for "with respect to", "i.i.d. " - for "independent and identically distributed", "GAR" - for "Gaussian Approximation".

Related works Classical results in the stochastic approximation (Borkar 2008) study the asymptotic properties of the single timescale SA algorithms, with the properties of averaged estimated studied in (Polyak and Juditsky 1992). Two-timescale SA schemes were studied in (Borkar 1997; Tadić 2004; Tadic 2006) in terms of almost sure convergence. Asymptotic convergence rates of linear two-timescale SA were studied in (Konda and Tsitsiklis 2004), where the authors showed that asymptotically $\mathbb{E}[\|\theta_k - \theta^\star\|^2] = \mathcal{O}(\beta_k)$ and $\mathbb{E}[\|w_k - w^\star\|^2] = \mathcal{O}(\gamma_k)$.

Non-asymptotic error bounds for TTSA were first developed in (Dalal et al. 2018; Dalal, Szorenyi, and Thoppe 2020) under the martingale noise assumptions and additional projections used in the update scheme (1). These results were further improved in (Kaledin et al. 2020) for linear TTSA problems. (Haque, Khodadadian, and Maguluri 2023) refined the results of (Kaledin et al. 2020) obtaining the MSE bounds $\mathbb{E}[\|\theta_k - \theta^\star\|^2]$ and $\mathbb{E}[\|w_k - w^\star\|^2]$ with the leading terms given by $\beta_k \operatorname{Tr}\Sigma_\theta$ and $\gamma_k \operatorname{Tr}\Sigma_w$, where the covariances Σ_θ and Σ_w aligns with the CLT in (2). (Kwon et al. 2024) considered the version of (1) with constant step sizes and studied convergence to equilibrium for the corresponding Markov chain. Non-linear TTSA has been considered in (Doan 2024) under strong monotonicity assumptions, focusing on obtaining the MSE rate of order $\mathcal{O}(1/k)$ for k-th iterate.

Central limit theorem for TTSA iterates has been established in (Mokkadem, Pelletier et al. 2006), where the asymptotic version of the CLT was proved both for the last iterates (θ_k, w_k) and their Polyak–Ruppert averaged counterparts $(\bar{\theta}_n, \bar{w}_n)$. (Hu, Doshi, and Eun 2024) established an asymptotic CLT for general TTSA under Markov noise and controlled Markov chain dynamics, without quantifying the convergence rate. (Kong et al. 2025) studied the CLT for averaged iterates $(\bar{\theta}_n, \bar{w}_n)$ and provided a non-asymptotic CLT with the convergence rate studied in terms of Wasserstein distance of order 1.

2 Gaussian Approximation for SA Algorithms

We outline a general scheme for proving the normal approximation. We consider vector-valued nonlinear statistics $T(X_1, \ldots, X_n) \in \mathbb{R}^d$, which can be represented in the form

$$T = W + D\,, \tag{4}$$

where W is a linear statistic of the random variables $X_1, \ldots, X_n$, and D is a small perturbation. This approach is well studied when $X_1, \ldots, X_n$ are i.i.d. random variables (Chen and Shao 2007; Shao and Zhang 2022) or form a martingale-difference sequence (Shorack 2017). The case of Markov random variables can be reduced to the setting of martingale-difference sequences through the Poisson equation (Douc et al. 2018, Chapter 21). We consider the decomposition (4) and assume, without loss of generality, that $\mathbb{E}[WW^\top] = \mathrm{I}_d$. To measure the approximation quality, a common approach is to use the supremum of the difference between measures taken over some subclass $\mathcal{H} \subseteq \mathrm{Conv}(\mathbb{R}^d)$ of the collection of convex sets $\mathrm{Conv}(\mathbb{R}^d)$. Specifically, for probability measures μ, ν on $\mathbb{R}^d$, we write

$$\mathsf{d}_{\mathcal{H}}(\mu,\nu) = \sup_{B\in\mathcal{H}} |\mu(B) - \nu(B)|\ .$$

Examples of $\mathcal{H}$ include the class of all convex sets, halfspaces, rectangles, ellipsoids, etc. The choice of different collections of sets $\mathcal{H}$ may be motivated by the needs of a particular application and may introduce differences in the dependence of the results on the problem dimension d. Indeed, even this dimensional dependence for linear statistics W can vary; see (Bentkus 2003) and (Kojevnikov and Song 2022) for the respective results for i.i.d. sequences and martingale differences. In this paper, we focus on the convex distance ρ^{Conv}, defined as

$$\rho^{\mathrm{Conv}}(\mu,\nu) = \sup_{B\in\mathrm{Conv}(\mathbb{R}^d)} |\mu(B) - \nu(B)|\,,$$

and rely on the following proposition to reduce the problem of Gaussian approximation for the nonlinear statistic $W + D$ to that for the linear statistic W:

Proposition 1 (Proposition 2 in (Sheshukova et al. 2025))**.** *Let ν be a standard Gaussian measure in $\mathbb{R}^d$. Then for any random vectors W, D taking values in $\mathbb{R}^d$, and any $p \geq 1$,*

$$\rho^{\mathrm{Conv}}(W + D, \nu) \leq \rho^{\mathrm{Conv}}(W, \nu) + 2c_d^{p/(p+1)} \mathbb{E}^{1/(p+1)} \left[\|D\|^p\right],$$

where c_d is the isoperimetric constant of class $\mathrm{Conv}(\mathbb{R}^d)$.

Similar results can be derived for other classes of sets $\mathcal{H}$, with the constant c_d depending on the isoperimetric properties of the specific class $\mathcal{H}$; see, e.g., (Klivans, O'Donnell, and Servedio 2008). Proposition 1 shows that the estimation of $\rho^{\mathrm{Conv}}(W + D, \mathcal{N}(0, \mathrm{I}))$ can be reduced to:

1. Estimating $\rho^{\mathrm{Conv}}(W, \mathcal{N}(0, \mathrm{I}))$;
2. Estimating moments $\mathbb{E}[\|D\|^p]$ for some $p \geq 1$.

To bound $\rho^{\mathrm{Conv}}(W, \mathcal{N}(0, \mathrm{I}))$, one can apply a Berry–Esseen bound for the appropriate linear statistic, e.g., (Shao and Zhang 2022) for i.i.d. random variables or (Srikant 2024; Samsonov et al. 2025; Wu, Wei, and Rinaldo 2025) for the martingale-difference setting. The most involved part of the proof is the proper estimation of $\mathbb{E}[\|D\|^p]$.

3 GAR for TTSA with Martingale Noise

Assumptions and definitions. We investigate the linear TTSA algorithm given by the equivalent form of (1):

$$\theta_{k+1} = \theta_k + \beta_k(b_1 - A_{11}\theta_k - A_{12}w_k + V_{k+1}), \quad (5)$$
$$w_{k+1} = w_k + \gamma_k(b_2 - A_{21}\theta_k - A_{22}w_k + W_{k+1}) . \quad (6)$$

In this recurrence, the noise terms V_{k+1}, W_{k+1} are given by:

$$\begin{aligned} V_{k+1} &= \boldsymbol{\varepsilon}_V^{k+1} - \tilde{\mathbf{A}}_{11}^{k+1}(\theta_k - \theta^\star) - \tilde{\mathbf{A}}_{12}^{k+1}(w_k - w^\star), \\ W_{k+1} &= \boldsymbol{\varepsilon}_W^{k+1} - \tilde{\mathbf{A}}_{21}^{k+1}(\theta_k - \theta^\star) - \tilde{\mathbf{A}}_{22}^{k+1}(w_k - w^\star), \end{aligned} \quad (7)$$

where we used the notation $\tilde{\mathbf{A}}_{ij}^{k+1} := \mathbf{A}_{ij}^{k+1} - A_{ij}$ for $i, j \in \{1, 2\}$, and the random vectors $\boldsymbol{\varepsilon}_V^{k+1}, \boldsymbol{\varepsilon}_W^{k+1}$ are given by

$$\begin{aligned} \boldsymbol{\varepsilon}_V^{k+1} &= \mathbf{b}_1^{k+1} - \mathbf{A}_{11}^{k+1}\theta^\star - \mathbf{A}_{12}^{k+1}w^\star , \\ \boldsymbol{\varepsilon}_W^{k+1} &= \mathbf{b}_2^{k+1} - \mathbf{A}_{21}^{k+1}\theta^\star - \mathbf{A}_{22}^{k+1}w^\star . \end{aligned} \quad (8)$$

We consider a setting where the random elements V_{k+1} and W_{k+1} form a martingale-difference w.r.t. filtration $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$, $\mathcal{F}_0$ is trivial. We first consider the martingale noise setting. This setting covers the i.i.d. setting from (Konda and Tsitsiklis 2004) and also serves as a basis for subsequent analysis of the Markov noise setting.

A 1. *The noise terms are zero-mean given $\mathcal{F}_k$, i.e., $\mathbb{E}^{\mathcal{F}_k}[V_{k+1}] = 0$, and $\mathbb{E}^{\mathcal{F}_k}[W_{k+1}] = 0$.*

Next, for a given $p \geq 2$, we impose the following moment bound on V_{k+1}, W_{k+1}:

A 2 (p)**.** *There exist $m_W, m_V > 0$ such that for any $k \in \mathbb{N}$:*

$$\begin{aligned} \mathbb{E}^{1/p}[\|V_{k+1}\|^p] &\leq m_V\big(1 + \mathbb{E}^{1/p}[\|\theta_k - \theta^*\|^p] + \mathbb{E}^{1/p}[\|w_k - w^*\|^p]\big) \\ \mathbb{E}^{1/p}[\|W_{k+1}\|^p] &\leq m_W\big(1 + \mathbb{E}^{1/p}[\|\theta_k - \theta^*\|^p] + \mathbb{E}^{1/p}[\|w_k - w^*\|^p]\big) . \end{aligned}$$

The assumption A 2(p) appears in a similar form with $p = 2$ in (Kaledin et al. 2020, Assumption A4). Since our results require to control high-order moments of the TTSA iterates θ_k and w_k, it is natural to require that p-th moment of V_{k+1} and W_{k+1} are finite. Next, we present an assumption on the quadratic characteristic of V_k and W_k:

A3. *Noise variables $\boldsymbol{\varepsilon}_V^{k+1}$ and $\boldsymbol{\varepsilon}_W^{k+1}$ defined in* (8) *have zero conditional expectation given $\mathcal{F}_k$, that is, $\mathbb{E}^{\mathcal{F}_k}\left[\boldsymbol{\varepsilon}_V^{k+1}\right] = 0$ and $\mathbb{E}^{\mathcal{F}_k}\left[\boldsymbol{\varepsilon}_W^{k+1}\right] = 0$. Moreover, there exist matrices Σ_V, Σ_W, Σ_{VW} such that for any $k > 0$:*

$$\mathbb{E}^{\mathcal{F}_k}\left[\boldsymbol{\varepsilon}_V^{k+1}\{\boldsymbol{\varepsilon}_V^{k+1}\}^\top\right] = \Sigma_V , \mathbb{E}^{\mathcal{F}_k}\left[\boldsymbol{\varepsilon}_W^{k+1}\{\boldsymbol{\varepsilon}_W^{k+1}\}^\top\right] = \Sigma_W ,$$
$$\mathbb{E}^{\mathcal{F}_k}\left[\boldsymbol{\varepsilon}_V^{k+1}\{\boldsymbol{\varepsilon}_W^{k+1}\}^\top\right] = \Sigma_{VW} .$$

This assumption relaxes the one stated in (Kong et al. 2025), where the authors required the quadratic characteristic of the entire vectors V_{k+1} and W_{k+1} to be constant. However, this assumption is unlikely to hold due to the structure of these vectors outlined in (7). We also impose the following conditions on the problem matrices:

A4. *Matrices $-A_{22}$ and $-\Delta = -\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)$ are Hurwitz.*

A 4 is common for the analysis of both the linear two-timescale SA, see (Konda and Tsitsiklis 2004), and single-timescale SA, see (Durmus et al. 2025; Mou et al. 2020). A4 implies, due to the Lyapunov lemma (stated in the supplement paper for completeness), that there exist matrices $Q_{22}^\top = Q_{22} \succ 0$, $Q_\Delta^\top = Q_\Delta \succ 0$, such that

$$\begin{aligned} &\|\mathrm{I} - \gamma_k A_{22}\|_{Q_{22}} \leq 1 - a_{22}\gamma_k, \quad a_{22} := \tfrac{1}{4\|Q_{22}\|} , \\ &\|\mathrm{I} - \beta_k \Delta\|_{Q_\Delta} \leq 1 - a_\Delta \beta_k, \quad a_\Delta := \tfrac{1}{4\|Q_\Delta\|} , \end{aligned} \quad (9)$$

provided that the step sizes γ_k and β_k are small enough. Precisely, for $p \geq 2$, we impose the following assumption A 5(p) on the step sizes:

A5 (p)**.** *Step sizes $(\gamma_k)_{k\geq 1}$, $(\beta_k)_{k\geq 1}$ are non-increasing sequences of the form*

$$\beta_k = c_{0,\beta}(k + k_0)^{-b}, \quad \gamma_k = c_{0,\gamma}(k + k_0)^{-a} ,$$

where $1/2 < a < b < 1$, fraction $c_{0,\beta}/c_{0,\gamma}$ is small enough, and constant k_0 satisfies the bound $k_0 \geq C_{A5}p^{4/b}$, where the constant C_{A5} does not depend upon p.

In the subsequent main results, we set the parameter p of order $\log(n)$. Hence, the parameter k_0 will depend on the total number of iterations to be performed. The same effect appears in the single-timescale SA algorithms (Durmus et al. 2025; Wu et al. 2024). This effect is unavoidable at least in the setting of the constant step size algorithms, see (Durmus et al. 2021, Theorem 1).

A6. *There exist $\mathrm{C}_{\mathbf{A}}, \mathrm{C}_{\mathbf{b}} > 0$ such that for any $i, j \in \{1, 2\}$,*

$$\sup_{x\in\mathsf{X}} \|\mathbf{A}_{ij}(x)\| \vee \|\mathbf{A}_{ij}(x) - A_{ij}\| \leq \mathrm{C}_{\mathbf{A}} ,$$
$$\sup_{x\in\mathsf{X}} \|\mathbf{b}_i(x)\| \vee \|\mathbf{b}_i(x) - b_i\| \leq \mathrm{C}_{\mathbf{b}} .$$

We expect that A 6 can be replaced with an appropriate moment condition, at least in a setting where the noise variables V_k and W_k form a martingale difference. At the same time, our further generalizations to the Markov noise setting inherently rely on the boundedness of $\mathbf{A}_{ij}(x)$ and $\mathbf{b}_i(x)$.

3.1 Moment Bounds for Martingale TTSA

Given the assumptions A 1 - A 6, we present the classical reformulation of the two-timescale SA scheme (5)-(6), which is due to (Konda and Tsitsiklis 2004), see also (Kaledin et al. 2020). We define recursively the following sequence of matrices $\{L_k\}_{k\in\mathbb{N}}$, with $L_0 = 0$, and

$$L_{k+1} := \big(L_k - \gamma_k A_{22} L_k + \beta_k A_{22}^{-1} A_{21} U_k\big) \times \big(\mathrm{I} - \beta_k U_k\big)^{-1} , \; U_k := \Delta - A_{12} L_k .$$

and define $\mathrm{L}_\infty = a_\Delta \lambda_{\mathsf{max}}(Q_\Delta)/(\lambda_{\mathsf{min}}(Q_{22}) 2\|A_{12}\|)$. As shown in (Kaledin et al. 2020, Lemma 18), under A5 above recursion on L_k is well-defined, and every L_k satisfies the relation $\|L_k\| \le \mathrm{L}_\infty$. In addition, define the matrices:

$$B_{11}^k := \Delta - A_{12} L_k , \quad D_k := L_{k+1} + A_{22}^{-1} A_{21} ,$$
$$B_{22}^k := (\beta_k/\gamma_k)\big(L_{k+1} + A_{22}^{-1} A_{21}\big) A_{12} + A_{22} .$$

In a similar vein as performing Gaussian elimination, we obtain a simplified two-timescale SA recursions:

Proposition 2 (Observation 1 in (Kaledin et al. 2020))**.** *Consider the following change of variables:*

$$\tilde\theta_k := \theta_k - \theta^\star, \quad \tilde w_k = w_k - w^\star + D_{k-1}\tilde\theta_k.$$

Then the two-timescale SA (5)-(6) *is equivalent to:*

$$\begin{aligned} \tilde\theta_{k+1} &= (\mathrm{I} - \beta_k B_{11}^k)\tilde\theta_k - \beta_k A_{12}\tilde w_k - \beta_k V_{k+1} , \\ \tilde w_{k+1} &= (\mathrm{I} - \gamma_k B_{22}^k)\tilde w_k - \beta_k D_k V_{k+1} - \gamma_k W_{k+1} . \end{aligned} \tag{10}$$

Our further analysis, both for martingale and Markov noise, will essentially rely on the decoupled TTSA updates (10). We refer to this dynamics as to the "decoupled" one, since the update of the scale $\tilde w_{k+1}$ no longer depends directly on $\tilde\theta_k$, only through the noise variables V_{k+1} and W_{k+1}. Now we aim to upper bound the quantities

$$M_{k,p}^{\tilde w} := \mathbb{E}^{1/p}[\|\tilde w_k\|^p], \quad M_{k,p}^{\tilde\theta} := \mathbb{E}^{1/p}[\|\tilde\theta_k\|^p] .$$

Similarly to (9), we show in the supplement paper, that

$$\begin{aligned} \|\mathrm{I} - \beta_k B_{11}^k\|_{Q_\Delta} &\le 1 - (1/2)\beta_k a_\Delta , \\ \|\mathrm{I} - \gamma_k B_{22}^k\|_{Q_{22}} &\le 1 - (1/2)\gamma_k a_{22} . \end{aligned} \tag{11}$$

The result (11) together with the structure of the updates (10) enables us to expand the recurrence and to show that the error component, associated with the initial error $\theta_0 - \theta^\star$ and $w_0 - w^\star$ decay at the exponential rate. Precisely, the following bound holds:

Proposition 3. *Let $p \ge 2$ and assume A1,A2(p), A3, A4, A 5(p), and A6. Then for any $k \in \mathbb{N}$ it holds*

$$M_{k+1,p}^{\tilde\theta} \lesssim \textstyle\prod_{j=0}^k \big(1 - \beta_j a_\Delta/8\big) + p^2 \beta_k^{1/2} , \tag{12}$$
$$M_{k+1,p}^{\tilde w} \lesssim \textstyle\prod_{j=0}^k \big(1 - \gamma_j a_{22}/8\big) + p^3 \gamma_k^{1/2} , \tag{13}$$

where $\lesssim$ stands for inequality up to constants not depending upon k and p.

Discussion. Proposition 3 provides, to best of our knowledge, the first high-order moment bounds in the linear TTSA with martingale noise. The scaling of the r.h.s. with $\beta_k^{1/2}$ for $M_{k+1,p}^{\tilde\theta}$ and $\gamma_k^{1/2}$ for $M_{k+1,p}^{\tilde w}$ coincides with the one previously obtained for the particular case $p = 2$ in (Kaledin et al. 2020). Similar asymptotic results were previously obtained in (Konda and Tsitsiklis 2004). We expect that the dependence of the r.h.s. of (12) and (13) upon p can be improved based on applying the Pinelis version of Rosenthal inequality (Pinelis 1994, Theorem 4.1) instead of Burkholder's inequality (Osekowski 2012, Theorem 8.6), that was used in the current proof, yet we expect that this approach introduces additional technical difficulties.

3.2 GAR for Polyak-Ruppert Averaged TTSA

Based on the results of the previous section, we can now quantify the Gaussian approximation rates for $\sqrt{n}(\bar\theta_n - \theta^\star)$ for the Polyak-Ruppert averaged estimator $\bar\theta_n$ from (3). Now we present the key decomposition:

$$\Delta(\theta_k - \theta^\star) = \frac{\theta_k - \theta_{k+1}}{\beta_k} - \frac{A_{12}A_{22}^{-1}(w_k - w_{k+1})}{\gamma_k} + (V_{k+1} - A_{12}A_{22}^{-1}W_{k+1}) . \tag{14}$$

The proof of the above identity is given in the supplement paper. Taking sum in (14) for $k = 1$ to n, and using the definition of V_{k+1}, W_{k+1} in (7), we get:

$$\sqrt{n}\Delta(\bar\theta_n - \theta^\star) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{k=1}^n \psi_{k+1} + R_n^{\mathrm{pr}} , \tag{15}$$

where we set $\psi_{k+1} = \boldsymbol{\varepsilon}_V^{k+1} - A_{12}A_{22}^{-1}\boldsymbol{\varepsilon}_W^{k+1}$, and R_n^{pr} is a residual term defined in the supplement paper. Assumption A 3 implies that the variance $\mathrm{Var}[\boldsymbol{\varepsilon}_V^{k+1} - A_{12}A_{22}^{-1}\boldsymbol{\varepsilon}_W^{k+1}]$ is constant for any k, so we can define

$$\boldsymbol{\Sigma}_\varepsilon := \mathrm{Var}[\boldsymbol{\varepsilon}_V^1 - A_{12}A_{22}^{-1}\boldsymbol{\varepsilon}_W^1] \in \mathbb{R}^{d_\theta \times d_\theta} . \tag{16}$$

The following theorem holds:

Theorem 1. *Assume A1,A2($\log n$), A3, A4, A5($\log n$), and A6. Then, it holds that*

$$\rho^{\mathrm{Conv}}\big(\sqrt{n}\Delta(\bar\theta_n - \theta^\star), \mathcal{N}(0, \boldsymbol{\Sigma}_\varepsilon)\big) \lesssim_{\log n} \frac{1}{n^{a/2}} + \frac{1}{n^{(1-b)/2}} .$$

Proof sketch. We apply Proposition 1 to the decomposition (15) and obtain, with $\nu \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\varepsilon)$, that

$$\rho^{\mathrm{Conv}}\big(\sqrt{n}\Delta(\bar\theta_n - \theta^\star), \nu\big) \le \underbrace{\rho^{\mathrm{Conv}}\big(n^{-1/2}\sum_{k=1}^n \psi_{k+1}, \nu\big)}_{T_1} + \underbrace{2c_d^{p/p+1}\mathbb{E}^{1/(p+1)}\big[\|\boldsymbol{\Sigma}_\varepsilon^{-1/2} R_n^{\mathrm{pr}}\|^p\big]}_{T_2} .$$

Due to A 1 and A 6, sequence $\{\psi_{k+1}\}_{k\in\mathbb{N}}$ is a bounded martingale-difference sequence w.r.t. $\mathcal{F}_k$ with constant quadratic characteristic. Hence, T_1 can be estimated applying a slight modification of (Wu, Wei, and Rinaldo 2025, Theorem 1). It remains to bound the moments of T_2, which is done using Proposition 3.

Discussion. In the theorem above, the coefficients before the terms depend upon the initial errors $\|\theta_0-\theta^\star\|, \|w_0-w^\star\|$, and upon the factors $1/(1-a)$ and $1/(1-b)$. That is why the result in its current form does not apply directly if $b=1$. We expect that the result holds in this case as well, perhaps at a price of introducing additional logarithmic factors. The same remark applies to Theorem 2-Theorem 4 stated below.

Since $1/2 < a < b < 1$, the bound of Theorem 1 is optimized when setting $a = 1/2 + 1/\log n$ and $b = a + 1/\log n$, yielding the final rate of convergence of order

$$\rho^{\mathrm{Conv}}\big(\sqrt{n}\Delta(\bar\theta_n - \theta^\star), \mathcal{N}(0, \mathbf{\Sigma}_\varepsilon)\big) \lesssim_{\log n} n^{-1/4} \ . \quad (17)$$

The result of (17) improves upon the previously established results of (Kong et al. 2025). The authors of that paper obtained a rate of $n^{-1/4}$, up to $\log n$ factors, in terms of Wasserstein distance. This implies convergence rate $n^{-1/8}$ in the convex distance, which is slower than (17). The choice of a and b in (17) corresponds to nearly the same scales for β_k and γ_k, effectively reducing the problem to a single-scale LSA. The obtained $n^{-1/4}$ rate aligns with the one established for this problem with i.i.d. noise in (Samsonov et al. 2024).

3.3 GAR for the Last Iterate

In this section, we derive the normal approximation rates for the last iterate $\beta_n^{-1/2}\tilde\theta_{n+1}$. Following (Konda and Tsitsiklis 2004) and using (10), equations for $\tilde\theta_k$ and $\tilde w_k$ writes as

$\tilde\theta_{k+1} = (\mathrm{I} - \beta_k\Delta)\tilde\theta_k - \beta_k A_{12}\tilde w_k - \beta_k V_{k+1} + \beta_k\delta_k^{(1)}$,
$\tilde w_{k+1} = (\mathrm{I} - \gamma_k A_{22})\tilde w_k - \beta_k D_k V_{k+1} - \gamma_k W_{k+1} - \beta_k\delta_k^{(2)}$,

where we set

$$\delta_k^{(1)} = A_{12}L_k\tilde\theta_k \ , \quad \delta_k^{(2)} = -(L_{k+1} + A_{22}^{-1}A_{21})A_{12}\tilde w_k \ .$$

Throughout the analysis we use the following convention:

$$G_{m:k}^{(1)} := \textstyle\prod_{i=m}^k(\mathrm{I} - \beta_i\Delta), \ G_{m:k}^{(2)} := \prod_{i=m}^k(\mathrm{I} - \gamma_i A_{22}) \ .$$

Enrolling the above recurrence and following (Konda and Tsitsiklis 2004), we get from the previous recurrence that

$$\tilde\theta_{n+1} = -\textstyle\sum_{j=0}^n \beta_j G_{j+1:n}^{(1)}\psi_{j+1} + R_n^{\mathrm{last}} \ , \quad (18)$$

where R_n^{last} is a remainder term defined in the supplement paper. The leading term in representation (18) is a linear statistics of $\varepsilon_V, \varepsilon_W$ which are martingale difference sequences with constant quadratic characteristics due to A3. Now we define

$$\mathbf{\Sigma}_n^{\mathrm{last}} = \mathrm{Var}\big[\textstyle\sum_{j=0}^n \beta_j G_{j+1:n}^{(1)}\psi_{j+1}\big] \ .$$

It is known that $\beta_n^{-1}\mathbf{\Sigma}_n^{\mathrm{last}}$ converges to a fixed matrix $\mathbf{\Sigma}_\infty^{\mathrm{last}}$ which is a solution of the Ricatti equation

$$\mathbf{\Sigma}_\infty^{\mathrm{last}} = \beta_0(\Delta\mathbf{\Sigma}_\infty^{\mathrm{last}} + \mathbf{\Sigma}_\infty^{\mathrm{last}}\Delta^\top - \mathbf{\Sigma}_\varepsilon) \ ,$$

where $\mathbf{\Sigma}_\varepsilon$ is defined in (16). Moreover, the convergence rate is proportional to β_n, i.e.

$$\|\beta_n^{-1}\mathbf{\Sigma}_n^{\mathrm{last}} - \mathbf{\Sigma}_\infty^{\mathrm{last}}\| \lesssim n^{-b} \ .$$

The proof of the above result is given in the supplement paper. The following assumption guarantees that the covariance matrix $\beta_n^{-1}\mathbf{\Sigma}_n^{\mathrm{last}}$ is non-degenerate, which is important for the further applications of Proposition 1.

A 7. *Step size exponents a, b satisfy $2b > 1 + a$. Moreover, assume that the total number of iterations n satisfies $n^b \geq C_{A7}$, where C_{A7} does not depend on a, b, and can be traced following the supplement paper.*

Theorem 2. *Assume A1,A2($\log n$), A3, A4, A5($\log n$), A6, A 7. Then, it holds that*

$$\rho^{\mathrm{Conv}}\big(\beta_n^{-1/2}\tilde\theta_{n+1}, \mathcal{N}(0, \mathbf{\Sigma}_\infty^{\mathrm{last}})\big) \lesssim_{\log n} n^{b/2}\prod_{j=0}^n(1 - \frac{a_\Delta}{8}\beta_j) + \frac{1}{n^{(3b-a-2)/2}} \ . \quad (19)$$

Discussion The proof of Theorem 2 is similar to the one of Theorem 1, but relies on the decomposition (18) instead of (15) used in the averaged setting. Additional technical difficulties arises when controlling the moments of the term R_n^{last}. Bounding the latter term requires additional constraint $2b > 1 + a$ imposed in A7.

Since $1/2 < a < b < 1$, the bound of Theorem 2 is optimized when setting $a = 1/2 + 1/\log n$ and $b = 1 - 1/\log n$, yielding the final rate

$$\rho^{\mathrm{Conv}}\big(\beta_n^{-1/2}\tilde\theta_{n+1}, \mathcal{N}(0, \mathbf{\Sigma}_\infty^{\mathrm{last}})\big) \lesssim_{\log n} n^{-1/4} \ ,$$

provided that n is large enough. To the best of our knowledge, this is the first result concerning the Gaussian approximation rate for the TTSA last iterate.

Note that Theorem 2 reveals phenomenon, which is completely different from what was previously observed for the Polyak-Ruppert averaged iterates in Theorem 1. Indeed, the right-hand side of the bound (19) contains the term $n^{-(3b-a-2)/2}$, which favors separation between β_k and γ_k, and vanishes when the scale exponents are close.

4 GAR for TTSA with Markov Noise

In this section we generalize the results obtained in Section 3 to the more practical scenario when $\{X_k\}_{k\in\mathbb{N}}$ form a Markov chain. Namely, we impose the following assumption:

B 1. *The sequence $\{X_k\}_{k\in\mathbb{N}}$ is a Markov chain taking values in a Polish space $(\mathsf{X}, \mathcal{X})$ with the Markov kernel P. Moreover, P admits π as a unique invariant distribution and is uniformly geometrically ergodic, that is, there exists $t_{\mathrm{mix}} \in \mathbb{N}$, such that for any $k \in \mathbb{N}$, it holds that*

$$\Delta(\mathrm{P}^k) := \sup_{x,x'\in\mathsf{X}} \mathsf{d}_{\mathrm{tv}}(\mathrm{P}^k(x,\cdot), \mathrm{P}^k(x',\cdot)) \leq (1/4)^{\lceil k/t_{\mathrm{mix}}\rceil} \ .$$

Moreover, for all $k \in \mathbb{N}$ and $i, j \in \{1, 2\}$ it holds that

$$\mathbb{E}_\pi[\mathbf{A}_{ij}^k] = A_{ij} \text{ and } \mathbb{E}_\pi[\mathbf{b}_i^k] = b_i \ .$$

Parameter t_{mix} in B 1 is referred to as a *mixing time*, see e.g. (Paulin 2015), and controls the rate of convergence of the iterates P^k to π as k increases.

4.1 Moment Bounds for TTSA with Markov Noise

First, we introduce a counterpart to A 5 that is needed to derive moment bounds for the setting of Markov noise.

B 2 (p)**.** $(\gamma_k)_{k\geq 1}$, $(\beta_k)_{k\geq 1}$ *are non-increasing sequences of the form*

$$\beta_k = c_{0,\beta}(k+k_0)^{-b}, \quad \gamma_k = c_{0,\gamma}(k+k_0)^{-a} ,$$

where $1/2 < a < b < 1$*, fraction* $c_{0,\beta}/c_{0,\gamma}$ *is small enough, and constant* k_0 *satisfies the bound* $k_0 \geq C_{B\,2}p^{4/b}$*, where the constant* $C_{B\,2}$ *does not depend upon* p*.*

The proof of moment bounds is more involved compared to the martingale noise case. Following the decomposition outlined in (Kaledin et al. 2020), we first represent the noise variables (V_{k+1}, W_{k+1}) as a sum of their martingale $(V_{k+1}^{(0)}, W_{k+1}^{(0)})$ and Markovian components $(V_{k+1}^{(1)}, W_{k+1}^{(1)})$ in a way that

$$V_{k+1} = V_{k+1}^{(0)} + V_{k+1}^{(1)} , \quad W_{k+1} = W_{k+1}^{(0)} + W_{k+1}^{(1)} .$$

Here $\mathbb{E}^{\mathcal{F}_k}\left[V_{k+1}^{(0)}\right] = 0$ and $\mathbb{E}^{\mathcal{F}_k}\left[W_{k+1}^{(0)}\right] = 0$. This representation is obtained using the decomposition associated with the Poisson equation, see (Douc et al. 2018, Chapter 21) and additional summation by parts. Then we define a pair of coupled recursions, which form exact counterparts of (10):

$$\tilde{\theta}_{k+1}^{(i)} = (\mathrm{I} - \beta_k B_{11}^k)\tilde{\theta}_k^{(i)} - \beta_k A_{12}\tilde{w}_k^{(i)} - \beta_k V_{k+1}^{(i)} ,$$
$$\tilde{w}_{k+1}^{(i)} = (\mathrm{I} - \gamma_k B_{22}^k)\tilde{w}_k^{(i)} - \beta_k D_k V_{k+1}^{(i)} - \gamma_k W_{k+1}^{(i)} ,$$

where $i \in \{0,1\}$. Then it is easy to see that $\tilde{\theta}_k = \tilde{\theta}_k^{(0)} + \tilde{\theta}_k^{(1)}$ and $\tilde{w}_k = \tilde{w}_k^{(0)} + \tilde{w}_k^{(1)}$. Precise expressions for $\tilde{\theta}_k^{(i)}, \tilde{w}_k^{(i)}, V_k^{(i)}, W_k^{(i)}$ can be found in the supplement paper.

Proposition 4. *Let* $p \geq 2$*. Assume A4, A6, B 1, B 2(p). Thus, it holds for any* $k \geq 0$ *that*

$$M_{k+1,p}^{\tilde{\theta}} \lesssim \textstyle\prod_{j=0}^k (1 - \frac{a_\Delta \beta_j}{8}) + p^2\sqrt{\beta_k} ,$$
$$M_{k+1,p}^{\tilde{w}} \lesssim \textstyle\prod_{j=0}^k (1 - \frac{a_{22}\gamma_j}{8}) + p^3\sqrt{\gamma_k} .$$

Proof sketch. The idea of the proof is to bound martingale and Markov parts separately using the techniques from Section 3. Note that Proposition 4 directly mimics the similar result obtained under the martingale noise setting in Proposition 3. The only difference is that the constants hidden under $\lesssim$ additionally depends upon the parameter t_{mix}.

4.2 GAR for Polyak-Ruppert Averaged TTSA

To proceed with Gaussian approximation for Polyak-Ruppert averaging, we use the decomposition (15) to transform the linear statistic $\sum_{k=1}^n \psi_{k+1}$ to a sum of martingale-increments. This transformation is done through the Poisson equation, see (Douc et al. 2018, Chapter 21). Under A 6, function $\psi(x) = \boldsymbol{\varepsilon}_V(x) - A_{12}A_{22}^{-1}\boldsymbol{\varepsilon}_W(x)$ is a.s. bounded, which implies that there exists a function $\boldsymbol{g}^\psi : \mathsf{X} \to \mathbb{R}^{d_\theta}$, such that

$$\boldsymbol{g}^\psi(x) - \mathrm{P}\boldsymbol{g}^\psi(x) = \psi(x) .$$

We set $\boldsymbol{g}_{k+1}^\psi := \boldsymbol{g}^\psi(X_{k+1})$ and define

$$\mathrm{M}_k = \boldsymbol{g}_{k+1}^\psi - \mathrm{P}\boldsymbol{g}_k^\psi ,$$

which form a martingale-increment w.r.t. $\mathcal{F}_k$. Then we can rewrite (15) as

$$\sqrt{n}\Delta(\bar{\theta}_n - \theta^\star) = \tfrac{1}{\sqrt{n}}\textstyle\sum_{k=1}^n \mathrm{M}_k + R_n^{\mathrm{pr,m}} , \qquad (20)$$

where $R_n^{\mathrm{pr,m}}$ is a residual term defined in the supplement. Under B 1 there exists a matrix $\boldsymbol{\Sigma}_\infty^{\mathsf{mark}} \in \mathbb{R}^{d_\theta \times d_\theta}$ such that

$$n^{-1/2}\textstyle\sum_{k=1}^n \{\psi_{k+1} - \pi(\psi)\} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_\infty^{\mathsf{mark}}) . \qquad (21)$$

Due to (Douc et al. 2018, Theorem 21.2.5), we get that

$$\mathrm{Var}[\mathrm{M}_k] = \boldsymbol{\Sigma}_\infty^{\mathsf{mark}} .$$

Now we state the counterpart to Theorem 1:

Theorem 3. *Assume A4, A6, B 1, B 2(*$\log n$*). Then it holds that*

$$\rho^{\mathrm{Conv}}(\sqrt{n}\Delta(\bar{\theta}_n - \theta^\star), \mathcal{N}(0, \boldsymbol{\Sigma}_\infty^{\mathsf{mark}})) \qquad (22)$$
$$\lesssim_{\log n} \frac{1}{n^{1/4}} + \frac{1}{n^{(1-b)/2}} + \frac{1}{n^{a-\frac{1}{2}}} + \sqrt{n}\prod_{j=0}^{n-1}(1 - \frac{a_\Delta\beta_j}{16}) .$$

Proof sketch. The proof of Theorem 3 consists of two main parts. First, we derive a Gaussian approximation rate for the linear statistic $\frac{1}{\sqrt{n}}\sum_{k=1}^n \mathrm{M}_k$ using an appropriate martingale CLT. It is especially non-trivial, since $\mathbb{E}^{\mathcal{F}_k}\left[\mathrm{M}_k\{\mathrm{M}_k\}^\top\right]$ is not constant. We circumvent this problem using an appropriate modification of the argument due to (Fan 2019). Next, we estimate the moments of $R_n^{\mathrm{pr,m}}$ using the techniques established in Proposition 3 for $\tilde{\theta}_k^{(0)}, \tilde{w}_k^{(0)}$ and then combining this with a separate bounds for the Markov part $\tilde{\theta}_k^{(1)}, \tilde{w}_k^{(1)}$.

Discussion. It is easy to see that, given that $b > a$, the right-hand side of (22) is optimized when setting $a = 2/3$ and $b = 2/3 + 1/(\log n)$. This yields the final rate of order $n^{-1/6}$ up to logarithmic factors:

$$\rho^{\mathrm{Conv}}(\sqrt{n}\Delta(\bar{\theta}_n - \theta^\star), \mathcal{N}(0, \boldsymbol{\Sigma}_\infty^{\mathsf{mark}})) \lesssim_{\log n} n^{-1/6} . \qquad (23)$$

To the best of our knowledge, (23) provides the first result concerning the Gaussian approximation rates for the TTSA problems with Markov noise. The suggested step size schedule mimics the one predicted by Theorem 1 and essentially reduces the TTSA scheme to a single-timescale one.

4.3 GAR for Last Iterate of TTSA

We start this section by introducing a counterpart to (18) based on the idea of the decomposition (20) for Polyak-Ruppert averaging:

$$\beta_n^{-1/2}\tilde{\theta}_{n+1} = -\textstyle\sum_{j=0}^n \beta_j G_{j+1:n}^{(1)}\mathrm{M}_j + R_n^{\mathrm{last,m}} , \qquad (24)$$

where $R_n^{\mathrm{last,m}}$ is a residual term that is given in the supplement paper. Note that the leading term in representation (24) is martingale difference sequence. Now we define

$$\boldsymbol{\Sigma}_n^{\mathrm{last,m}} = \mathrm{Var}\left[\textstyle\sum_{j=0}^n \beta_j G_{j+1:n}^{(1)}\mathrm{M}_j\right] .$$

It is known that $\beta_n^{-1}\boldsymbol{\Sigma}_n^{\mathrm{last,m}}$ converges to a fixed matrix $\boldsymbol{\Sigma}_\infty^{\mathrm{last,m}}$ which is a solution of the Ricatti equation

$$\boldsymbol{\Sigma}_\infty^{\mathrm{last,m}} = \beta_0(\Delta\boldsymbol{\Sigma}_\infty^{\mathrm{last,m}} + \boldsymbol{\Sigma}_\infty^{\mathrm{last,m}}\Delta^\top - \boldsymbol{\Sigma}_\infty^{\mathsf{mark}}) ,$$

where $\mathbf{\Sigma}_\infty^{\mathrm{mark}}$ is defined in (21). Moreover, the convergence rate is proportional to β_n, i.e.

$$\|\beta_n^{-1}\mathbf{\Sigma}_n^{\mathrm{last,m}} - \mathbf{\Sigma}_\infty^{\mathrm{last,m}}\| \lesssim n^{-b} \ .$$

The proof of the above result is given in the supplement paper. Now we formulate a counterpart to A7:

B 3. *Step size exponents a, b satisfy $2b > 1 + a$. Moreover, assume that the total number of iterations n satisfies $n^b \geq C_{B\,3}$, where $C_{B\,3}$ does not depend on a, b, and can be traced from the supplement paper.*

Theorem 4. *Assume A 4, A 6, B 1, B 2($\log n$), B 3. Then it holds that*

$$\rho^{\mathrm{Conv}}(\beta_n^{-1/2}\tilde{\theta}_{n+1}, \mathcal{N}(0, \mathbf{\Sigma}_\infty^{\mathrm{last,m}})) \tag{25}$$
$$\lesssim_{\log n} n^{\frac{b}{2}} \prod_{j=0}^{n}(1 - \frac{a_\Delta \beta_j}{8}) + \frac{1}{n^{\frac{2b-1}{4}}} + \frac{1}{n^{\frac{2a-b}{2}}} + \frac{1}{n^{\frac{3b-a-2}{2}}} \ .$$

Proof sketch. The proof of Theorem 4 uses the same machinery of Gaussian approximation for non-linear statistics based on representation (24). In this setting control of the moments of the term $R_n^{\mathrm{last,m}}$ is a delicated problem, which requires the additional constraint $2b > 1 + a$ imposed in B 3.

Discussion. It is easy to see that, given that $b \geq a$, the right-hand side of (25) is optimized when setting $a = 2/3$ and $b = 1 - 1/(\log n)$, and yields the final rate in terms of n of order up to $n^{-1/6}$ up to logarithmic factors:

$$\rho^{\mathrm{Conv}}(\beta_n^{-1/2}\tilde{\theta}_{n+1}, \mathcal{N}(0, \mathbf{\Sigma}_\infty^{\mathrm{last,m}})) \lesssim_{\log n} n^{-1/6} \ .$$

This rate, to the best of our knowledge, is the first one obtained for the last iterate of TTSA with Markov noise.

5 Applications to TDC and GTD

In this section, we show that the results derived in Section 3 and Section 4 apply to the Gradient Temporal Difference (GTD) (Sutton, Maei, and Szepesvári 2008) and Temporal Difference with Gradient Correction (TDC) (Sutton et al. 2009) methods. These methods address the problem of classical TD learning, which is based on single-timescale stochastic approximation and is known to fail in off-policy RL settings where data are drawn from a *behavior policy* different from the target policy (Baird 1995; Tsitsiklis and Van Roy 1997). We consider a discounted MDP (Markov Decision Process) given by a tuple $(\mathcal{S}, \mathcal{A}, \mathrm{P}, r, \lambda)$. Here $\mathcal{S}$ and $\mathcal{A}$ denote state and action spaces, which are assumed to be complete separable metric spaces with their Borel σ-algebras $\mathcal{B}(\mathcal{S})$ and $\mathcal{B}(\mathcal{A})$, and $\lambda \in (0, 1)$ is a discount factor. Let $\mathrm{P}(\cdot|s, a)$ be a state-action transition kernel, which determines the probability of moving from (s, a) to a set $B \in \mathcal{B}(\mathcal{S})$. Reward function $r\colon \mathcal{S} \times \mathcal{A} \to [0, 1]$ is deterministic. A *Policy* $\pi(\cdot|s)$ is the distribution over action space $\mathcal{A}$ corresponding to agent's action preferences in state $s \in \mathcal{S}$. We aim to estimate *value function*

$$V^\pi(s) = \mathbb{E}\big[\textstyle\sum_{k=0}^{\infty} \lambda^k r(S_k, A_k) | S_0 = s\big] \ ,$$

where $A_k \sim \pi(\cdot|s_k)$, and $S_{k+1} \sim \mathrm{P}(\cdot|S_k, A_k)$. Define the transition kernel under π,

$$\mathrm{P}_\pi(B|s) = \textstyle\int_{\mathcal{A}} \mathrm{P}(B|s, a)\pi(\mathrm{d}a|s) \ . \tag{26}$$

We consider the *linear function approximation* for $V^\pi(s)$, defined for $s \in \mathcal{S}, \theta \in \mathbb{R}^d$, and a feature mapping $\varphi\colon \mathcal{S} \to \mathbb{R}^d$ as $V_\theta^\pi(s) = \varphi^\top(s)\theta$. Our goal is to find a parameter $\theta^\star$, which defines the best linear approximation to V^π. We denote by μ the invariant distribution over the state space $\mathcal{S}$ induced by $\mathrm{P}^\pi(\cdot|s)$ in (26). Consider the following assumptions on the generative mechanism and on the feature mapping $\varphi(\cdot)$:

TD 1. *Tuples (s_k, a_k, s_k') are generated i.i.d.with $s_k \sim \mu$, $a_k \sim \pi(\cdot|s_k)$, $s_k' \sim \mathrm{P}(\cdot|s_k, a_k)$.*

TD 2. *Feature mapping $\varphi(\cdot)$ satisfies $\sup_{s\in\mathcal{S}} \|\varphi(s)\| \leq 1$.*

As an alternative to the generative model setting **TD** 1, our analysis covers the Markov noise setting:

TD 3. *Suppose that we obtain a Markovian sample trajectory $\{(s_k, a_k, r_k)\}_{k=0}^\infty$ which is generated when a stationary behavior policy π is employed. Assume that the Markov kernel P_π admits a unique invariant distribution μ and is uniformly geometrically ergodic, that is, there exist $t_{\mathrm{mix}} \in \mathbb{N}$, such that for any $k \in \mathbb{N}$, it holds that*

$$\sup_{s,s'\in\mathcal{S}} \mathsf{d}_{\mathrm{tv}}(\mathrm{P}_\pi^k(\cdot|s), \mathrm{P}_\pi^k(\cdot|s')) \leq (1/4)^{\lceil k/t_{\mathrm{mix}}\rceil} \ .$$

We introduce the k-th step TD error for the linear setting:

$$\delta_k = r_k + \lambda\theta_k^\top\varphi_{k+1} - \theta_k^\top\varphi_k \ ,$$

where we have defined

$$\varphi_k = \varphi(s_k) \ , \quad r_k = r(s_k, a_k) \ .$$

Generalized Temporal Difference learning. The GTD(0) algorithm is defined by the following recurrence for $k \geq 1$:

$$\begin{cases} \theta_{k+1} = \theta_k + \beta_k(\varphi_k - \lambda\varphi_{k+1})(\varphi_k)^\top w_k \ , \quad \theta_0 \in \mathbb{R}^d \ , \\ w_{k+1} = w_k + \gamma_k(\delta_k\varphi_k - w_k) \ , \quad w_0 = 0 \ . \end{cases} \tag{27}$$

It is clear that the GTD(0) recurrence (27) is a particular case of the linear TTSA given in (5)-(6).

Temporal-difference learning with gradient correction. The TDC algorithm employs dual updates for the primary parameter vector θ_k and the auxiliary weight vector w_k. Its update rule is given by

$$\begin{cases} \theta_{k+1} = \theta_k + \beta_k\delta_k\varphi_k - \beta_k\lambda\varphi_{k+1}(\varphi_k^\top w_k) \ , \\ w_{k+1} = w_k + \gamma_k(\delta_k - \varphi_k^\top w_k)\varphi_k \ . \end{cases} \tag{28}$$

It is possible to check that both updates schemes (27) and (28) satisfy the general assumptions A 1-A 4 and A 6 under **TD** 1 and **TD** 2. Similar, B 1 holds under **TD** 3. Thus, all the results from Section 3 and Section 4 applies for both algorithms. We provide details in the supplemental paper.

6 Conclusion

In this paper, we provided, to the best of our knowledge, the first rate of normal approximation for the last iterate and Polyak-Ruppert averaged TTSA iterates in a sense of convex distance, covering both the martingale-difference and Markov noise settings. A natural further research direction is to consider the problem of constructing confidence intervals for the TTSA solution $(\theta^\star, w^\star)$ based on bootstrap approach or asymptotic covariance matrix estimation, and perform a fully non-asymptotic analysis of the suggested procedure. Another important direction is the construction of lower bounds to ensure tightness of the rates obtained in Theorem 1-4.

Acknowledgments

The work was supported by the grant for research centers in the field of AI provided by the Ministry of Economic Development of the Russian Federation in accordance with the agreement 000000C313925P4E0002 and the agreement with HSE University № 139-15-2025-009.

References

Baird, L. 1995. Residual Algorithms: Reinforcement Learning with Function Approximation. In *International Conference on Machine Learning*, 30–37.

Bentkus, V. 2003. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2): 385–402.

Borkar, V. S. 1997. Stochastic approximation with two time scales. *Systems & Control Letters*, 29(5): 291–294.

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

Chen, L. H.; and Shao, Q.-M. 2007. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli*, 13(2): 581 – 599.

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, 3701–3708.

Dalal, G.; Szörényi, B.; Thoppe, G.; and Mannor, S. 2018. Finite Sample Analysis of Two-Timescale Stochastic Approximation with Applications to Reinforcement Learning. In *Conference on Learning Theory (COLT)*, 1199–1233. PMLR.

Doan, T. T. 2024. Fast Nonlinear Two-Time-Scale Stochastic Approximation: Achieving $O(1/k)$ Finite-Sample Complexity. *arXiv preprint arXiv:2401.12764*.

Douc, R.; Moulines, E.; Priouret, P.; and Soulier, P. 2018. *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer. ISBN 978-3-319-97703-4.

Durmus, A.; Moulines, E.; Naumov, A.; and Samsonov, S. 2025. Finite-Time High-Probability Bounds for Polyak–Ruppert Averaged Iterates of Linear Stochastic Approximation. *Mathematics of Operations Research*, 50(2): 935–964.

Durmus, A.; Moulines, E.; Naumov, A.; Samsonov, S.; Scaman, K.; and Wai, H.-T. 2021. Tight High Probability Bounds for Linear Stochastic Approximation with Fixed Stepsize. In *Advances in Neural Information Processing Systems*, volume 34, 30063–30074. Curran Associates, Inc.

Fan, X. 2019. Exact rates of convergence in some martingale central limit theorems. *Journal of Mathematical Analysis and Applications*, 469(2): 1028–1044.

Haque, M. S.; Khodadadian, A.; and Maguluri, S. T. 2023. Tight Finite-Time Bounds for Two-Timescale Stochastic Approximation and Applications to Reinforcement Learning. *arXiv preprint arXiv:2312.13613*.

Hu, M.; Doshi, P.; and Eun, C. 2024. A Central Limit Theorem for Two-Timescale Stochastic Approximation under Controlled Markov Noise. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

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*.

Klivans, A. R.; O'Donnell, R.; and Servedio, R. A. 2008. Learning Geometric Concepts via Gaussian Surface Area. In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '08, 541–550. USA: IEEE Computer Society. ISBN 9780769534367.

Kojevnikov, D.; and Song, K. 2022. A Berry–Esseen bound for vector-valued martingales. *Statistics & Probability Letters*, 186: 109448.

Konda, V. R.; and Tsitsiklis, J. N. 2004. Convergence rate of linear two-time-scale stochastic approximation. *Ann. Appl. Probab.*, 14(2): 796–819.

Kong, S. T.; Zeng, S.; Doan, T. T.; and Srikant, R. 2025. Nonasymptotic CLT and Error Bounds for Two-Time-Scale Stochastic Approximation. *arXiv preprint arXiv:2502.09884*.

Kwon, J.; Dotson, L.; Chen, Y.; and Xie, Q. 2024. Two-Timescale Linear Stochastic Approximation: Constant Stepsizes Go a Long Way. *arXiv preprint arXiv:2410.13067*.

Mokkadem, A.; Pelletier, M.; et al. 2006. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *The 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*, 2947–2997. PMLR.

Moulines, E.; and Bach, F. 2011. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. *Advances in neural information processing systems*, 24: 451–459.

Nemirovskij, A. S.; and Yudin, D. B. 1983. *Problem complexity and method efficiency in optimization*. Wiley-Interscience.

Osekowski, A. 2012. *Sharp Martingale and Semimartingale Inequalities*. Monografie Matematyczne 72. Birkhäuser Basel, 1 edition. ISBN 3034803699,9783034803694.

Paulin, D. 2015. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electronic Journal of Probability*, 20(none): 1 – 32.

Pinelis, I. 1994. Optimum Bounds for the Distributions of Martingales in Banach Spaces. *The Annals of Probability*, 22(4): 1679 – 1706.

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

Ruppert, D. 1988. Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

Samsonov, S.; Moulines, E.; Shao, Q.-M.; Zhang, Z.-S.; and Naumov, A. 2024. Gaussian Approximation and Multiplier Bootstrap for Polyak-Ruppert Averaged Linear Stochastic Approximation with Applications to TD Learning. In *Advances in Neural Information Processing Systems*, volume 37, 12408–12460. Curran Associates, Inc.

Samsonov, S.; Sheshukova, M.; Moulines, E.; and Naumov, A. 2025. Statistical Inference for Linear Stochastic Approximation with Markovian Noise. *arXiv preprint arXiv:2505.19102*.

Shao, Q.-M.; and Zhang, Z.-S. 2022. Berry–Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms. *Bernoulli*, 28(3): 1548–1576.

Sheshukova, M.; Samsonov, S.; Belomestny, D.; Moulines, E.; Shao, Q.-M.; Zhang, Z.-S.; and Naumov, A. 2025. Gaussian Approximation and Multiplier Bootstrap for Stochastic Gradient Descent. *arXiv preprint arXiv:2502.06719*.

Shorack, G. R. 2017. *Probability for Statisticians*. Springer Texts in Statistics. Springer, 2 edition.

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*.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. The MIT Press, second edition.

Sutton, R. S.; Maei, H.; and Szepesvári, C. 2008. A Convergent O(n) Temporal-difference Algorithm for Off-policy Learning with Linear Function Approximation. In Koller, D.; Schuurmans, D.; Bengio, Y.; and Bottou, L., eds., *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc.

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 *International Conference on Machine Learning*, 993–1000.

Tadić, V. 2004. Almost Sure Convergence of Two Time-Scale Stochastic Approximation Algorithms. *IEEE Transactions on Automatic Control*, 49(9): 1469–1474.

Tadic, V. 2006. Asymptotic analysis of temporal-difference learning algorithms with constant step-sizes. *Machine Learning*, 63: 107–133.

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

Wu, W.; Li, G.; Wei, Y.; and Rinaldo, A. 2024. Statistical Inference for Temporal Difference Learning with Linear Function Approximation. *arXiv preprint arXiv:2410.16106*.

Wu, W.; Wei, Y.; and Rinaldo, A. 2025. Uncertainty quantification for Markov chains with application to temporal difference learning. *arXiv preprint arXiv:2502.13822*.