Effectiveness of Constant Stepsize in Markovian LSA and Statistical Inference

Dongyan (Lucy) Huo[1], **Yudong Chen**[2], **Qiaomin Xie**[3]

[1]School of Operations Research and Information Engineering, Cornell University
[2]Department of Computer Sciences, University of Wisconsin-Madison
[3]Department of Industrial and Systems Engineering, University of Wisconsin-Madison
dh622@cornell.edu, yudong.chen@wisc.edu, qiaomin.xie@wisc.edu

Abstract

In this paper, we study the effectiveness of using a constant stepsize in statistical inference via linear stochastic approximation (LSA) algorithms with Markovian data. After establishing a Central Limit Theorem (CLT), we outline an inference procedure that uses averaged LSA iterates to construct confidence intervals (CIs). Our procedure leverages the fast mixing property of constant-stepsize LSA for better covariance estimation and employs Richardson-Romberg (RR) extrapolation to reduce the bias induced by constant stepsize and Markovian data. We develop theoretical results for guiding stepsize selection in RR extrapolation, and identify several important settings where the bias provably vanishes even without extrapolation. We conduct extensive numerical experiments and compare against classical inference approaches. Our results show that using a constant stepsize enjoys easy hyperparameter tuning, fast convergence, and consistently better CI coverage, especially when data is limited.

Introduction

Stochastic approximation (SA) algorithms use stochastic updates to iteratively approximate the solution to fixed-point equations. SA has wide applications, such as the stochastic gradient descent (SGD) algorithm for loss minimization and the Temporal Difference (TD) learning algorithm in reinforcement learning (RL) (Sutton 1988). Classical works on SA typically assume the stepsize sequence α_t is square-summable and diminishing, i.e., $\sum_t \alpha_t = \infty$, $\sum_t \alpha_t^2 < \infty$, under which asymptotic almost-sure convergence is well studied (Robbins and Monro 1951; Blum 1954; Borkar and Meyn 2000). Constant stepsize has gained popularity recently, particularly among practitioners, due to its fast initial convergence and easy hyperparameter tuning. A growing line of works studies the convergence properties of SA under constant stepsize, establishing upper bounds on the mean-squared error (MSE) (Lakshminarayanan and Szepesvári 2018; Srikant and Ying 2019; Mou et al. 2020, 2021) as well as weak convergence results (Dieuleveut, Durmus, and Bach 2020; Yu et al. 2021; Huo, Chen, and Xie 2023a).

Recent works have explored using SA and SGD iterates to perform statistical inference, e.g., constructing confidence intervals (CIs) around a point estimate (Li et al. 2018; Chen et al. 2020; Li et al. 2022; Lee et al. 2022; Liu, Chen, and Shang 2023). This approach is computationally cheap and scales well with the size and dimension of the dataset: SA updates are computed iteratively, without storing or multiple passes over the dataset. In comparison, classical inference techniques, such as bootstrapping, often require storing the entire dataset and repeating the estimation procedure, which may have prohibitive computational costs.

The aforementioned works on using SA for inference have focused on the diminishing stepsize paradigm, for which a mature convergence theory exists, ensuring the asymptotic correctness of the inference results. In contrast, constant-stepsize SA lacks last-iterate almost sure convergence: recent works have shown that the iterates converge only in distribution; moreover, the limit distribution may have a nonzero asymptotic bias due to the nonlinearity of the SA updates (Dieuleveut, Durmus, and Bach 2020) or the underlying Markovian data (Huo, Chen, and Xie 2023a), and this bias cannot be eliminated by iterate averaging. Partly due to these considerations, inference with constant stepsize SA iterates has been largely overlooked in the literature.

We study the effectiveness of statistical inference using constant-stepsize SA iterates. We focus on linear stochastic approximation (LSA), i.e., $\theta_{t+1} = \theta_t + \alpha(A(x_t)\theta_t + b(x_t))$, with a constant stepsize α and Markovian data $(x_t)_{t\geq 0}$. We first establish a Central Limit Theorem (CLT) for averaged Markovian LSA iterates. Built upon the CLT, we outline an inference procedure using averaged iterates and batch-mean covariance estimates. Our procedure leverages the fast mixing property of constant-stepsize updates for better covariance estimation and employs Richardson-Romberg (RR) extrapolation for bias reduction. We study two stepsize schemes in RR extrapolation. We further prove that with Markovian data, the asymptotic bias may vanish in several important settings, in which case inference with constant-stepsize LSA iterates is effective even without RR extrapolation. We conduct extensive experiments to benchmark our procedure against conventional inference approaches. Our results demonstrate superior and robust performance of the constant stepsize paradigm, which enjoys fast convergence, good coverage properties, and easy parameter tuning.[1]

[1]An extended paper with appendix can be found at Huo, Chen, and Xie (2023b).

Related Work

Dating back to Robbins and Monro (1951), classical works on stochastic approximation typically assume i.i.d. data and a square-summable and diminishing stepsize sequence. Subsequent works propose what is now known as the Polyak-Ruppert iterate averaging (Ruppert 1988; Polyak 1990) and establish a CLT for the averaged iterates (Polyak and Juditsky 1992). Recent work in Borkar et al. (2023) establishes a CLT for scaled iterates for general contractive SA under diminishing stepsize and Markovian data.

Constant-stepsize SA and SGD have attracted growing attention recently. Several works study constant-stepsize LSA with i.i.d. data and establish finite-time MSE bounds and a CLT for the average iterates (Lakshminarayanan and Szepesvári 2018; Mou et al. 2020). A parallel line of work studies SGD for both convex (Dieuleveut, Durmus, and Bach 2020) and nonconvex (Yu et al. 2021) functions with i.i.d. data and identifies an asymptotic bias arising from the nonlinearity of the function. More recent works study LSA with Markovian data. The works in Srikant and Ying (2019) and Durmus et al. (2023) study convergence bounds for MSE, while the work in Huo, Chen, and Xie (2023a) establishes weak convergence of the LSA iterates and characterizes its asymptotic bias due to Markovian data.

Most related to this paper is a recent line of work on using SGD/SA iterates for statistical inference. The work in Chen et al. (2020) considers SGD with i.i.d. data and strongly-convex functions and proposes two covariance matrix estimators. This result is generalized to ϕ-mixing data in Liu, Chen, and Shang (2023). The work in Zhu, Chen, and Wu (2023) extends the batch-mean estimator in Chen et al. (2020) to a fully online version. The work in Lee et al. (2022) proposes a random scaling covariance estimator for robust online inference with SGD. Subsequent work in Li, Liang, and Zhang (2023) investigates online statistical inference using nonlinear SA with Markovian data. The above works all consider diminishing stepsizes. The work in Xie and Zhang (2022) extends the random scaling estimator for inference with SA under constant stepsizes but i.i.d. data. The work in Li et al. (2018) studies statistical inference with SGD and i.i.d. data, using a small stepsize whose value remains constant throughout the iterations but scales inversely with the total number of iterations. In contrast, we consider constant stepsizes whose values are independent of the total number of iterations and thus substantially larger than the typical stepsize values in Li et al. (2018).

RR extrapolation is a classical technique from numerical analysis to improve approximation errors. It has been used in Dieuleveut, Durmus, and Bach (2020) for SGD and Huo, Chen, and Xie (2023a) for LSA. See the survey by Bach (2021) for the use of RR extrapolation in other data science and machine learning problems.

Problem Setup

In this section, we formally set up the problem and introduce the assumptions. Let $(x_t)_{t\geq 0}$ be a time-homogeneous stochastic process on a Borel state space $\mathcal{X}$ with stationary distribution π. Define the target vector θ^* as the solution to the steady-state equation $\mathbb{E}_{x\sim\pi}[A(x)]\theta^* + \mathbb{E}_{x\sim\pi}[b(x)] = 0$, where $A : \mathcal{X} \to \mathbb{R}^{d\times d}$ and $b : \mathcal{X} \to \mathbb{R}^d$ are deterministic functions on $\mathcal{X}$. To approximate θ^*, we consider the following linear stochastic approximation iteraion:

$$\theta_{t+1}^{(\alpha)} = \theta_t^{(\alpha)} + \alpha_t\big(A(x_t)\theta_t^{(\alpha)} + b(x_t)\big), \quad t = 0, 1, \ldots, \tag{1}$$

where $(\alpha_t)_{t\geq 0}$ is the stepsize sequence. We focus on using a constant stepsize, i.e., $\alpha_t \equiv \alpha$ for all $t \geq 0$. (We omit the superscript (α) when the stepsize is clear from the context.) We emphasize that the constant stepsize considered here is independent of the *total* number of iterations. This is different from the work in Li et al. (2018), which pre-specifies the total number of iterations T and uses a small fixed stepsize of the form $\alpha_t = T^{-\beta}, \forall 0 \leq t \leq T$ for some $\beta > 0$.

We make the following standard assumptions.

Assumption 1. *$(x_t)_{t\geq 0}$ is a uniformly ergodic Markov chain with transition kernel P and a unique stationary distribution π.*

Assumpiton 1 is common in the literature on Markovian SA (Bhandari, Russo, and Singal 2021; Durmus et al. 2023; Huo, Chen, and Xie 2023a). Uniform ergodicity ensures that the distribution of x_t converges geometrically to π from any initial distribution. For example, all irreducible, aperiodic, and finite state space Markov chains are uniformly ergodic.

Assumption 2. *$A_{\max} := \sup_{x\in\mathcal{X}} \|A(x)\|_2 \leq 1$ and $b_{\max} := \sup_{x\in\mathcal{X}} \|b(x)\|_2 < \infty$. Moreover, $\bar{A} := \mathbb{E}_{x\sim\pi}[A(x)]$ is a Hurwitz matrix.*

The Hurwitz condition is again standard and ensures the stability of a dynamic system (Srikant and Ying 2019; Durmus et al. 2023; Huo, Chen, and Xie 2023a). This assumption is satisfied in, e.g., SGD for minimizing strongly convex quadratics and the linear TD algorithm in RL.

Central Limit Theorem

Our first result is a CLT for averaged LSA iterates, which lays the theoretical foundation for the inference procedure developed later. To state the CLT, we recall a known result for constant-stepsize Markovian LSA: the data-iterate pair (x_t, θ_t) converge weakly to a unique limit distribution, denoted by $(x_\infty, \theta_\infty) \sim \mu$ (Huo, Chen, and Xie 2023a).

Theorem 1 (CLT)**.** *Under Assumptions 1–2, there exists a threshold $\alpha_0 \in (0, 1)$ such that for all $\alpha \in (0, \alpha_0)$, we have*

$$\sqrt{T}(\bar{\theta}_T - \mathbb{E}[\theta_\infty]) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma^*), \quad \text{as } T \to \infty,$$

where $\bar{\theta}_T := \frac{1}{T}\sum_{t=0}^{T-1} \theta_t$ and $\Sigma^* := \lim_{T\to\infty} T \cdot \mathbb{E}_\mu[(\bar{\theta}_T - \mathbb{E}[\theta_\infty])(\bar{\theta}_T - \mathbb{E}[\theta_\infty])^\top]$.

The proof is deferred to the appendix. Theorem 1 extends existing CLT results, which focus on either LSA with i.i.d. data (Mou et al. 2020; Xie and Zhang 2022) or SA with diminishing stepsize (Borkar et al. 2023). When using Markovian data $(x_t)_{t\geq 0}$, the iterate sequence $(\theta_t)_{t\geq 0}$ is no longer a Markov chain on its own. Instead, we need to consider the joint process $(x_t, \theta_t)_{t\geq 0}$, which is a time-homogeneous Markov chain thanks to the use of a constant stepsize, and build the CLT accordingly. Moreover, a number of existing

Markov Chain CLT results require a one-step contraction property of the form $W_{p,q}(\mu Q, \nu Q) \leq \gamma W_{p,q}(\mu, \nu)$, where Q is the Markovian transition kernel, $\gamma < 1$ and $W_{p,q}$ is an appropriate Wasserstein distance between distributions (Xie and Zhang 2022). Our Markov chain $(x_t, \theta_t)_{t\geq 0}$ does not enjoy such a nice one-step contraction property, and hence proving our CLT requires additional work.

Statistical Inference Procedure Using LSA

We next present a statistical inference procedure using the averaged LSA iterates with constant stepsize and Markovian data. This procedure can be combined with RR extrapolation to construct confidence intervals for the target vector θ^*.

Asymptotic Bias and RR Extrapolation

It has been shown in Huo, Chen, and Xie (2023a) that the asymptotic expectation $\mathbb{E}[\theta_\infty^{(\alpha)}]$ is biased with respect to θ^* and admits expansion: $\mathbb{E}[\theta_\infty^{(\alpha)}] = \theta^* + \sum_{i=1}^{\infty} \alpha^i B^{(i)}$, where $B^{(i)}$ are vectors independent from the stepsize α. Theorem 1 guarantees that the averaged iterates $\bar{\theta}_T$ are asymptotically normal, but $\mathbb{E}[\theta_\infty^{(\alpha)}] \neq \theta^*$. Moreover, the leading term in $\mathbb{E}[\theta_\infty^{(\alpha)}] - \theta^*$ scales with α. To construct CIs with good coverage properties, it is important to reduce the asymptotic bias.

In light of the bias expansion, we can employ Richardson-Romberg (RR) extrapolation to reduce the asymptotic bias to a higher order polynomial of the stepsize α (Dieuleveut, Durmus, and Bach 2020; Bach 2021; Huo, Chen, and Xie 2023a). Specifically, we run the LSA update (1) with M constant stepizes $\mathcal{A} = \{\alpha_1, \ldots, \alpha_M\}$ and compute a linear combination of the resulting iterates:

$$\widetilde{\theta}_t^{\mathcal{A}} = \sum_{m=1}^{M} h_m \theta_t^{(\alpha_m)}.$$

We carefully choose the coefficients $\{h_m\}$ to satisfy

$$\sum_{m=1}^{M} h_m = 1; \quad \sum_{m=1}^{M} h_m \alpha_m^l = 0, \ l = 1, 2, \ldots, M-1. \quad (2)$$

Using the aforementioned expansion, one sees that the bias $\mathbb{E}[\widetilde{\theta}_\infty^{\mathcal{A}}] - \theta^*$ is reduced exponentially in M and now scales with $(\max_{m=1,\ldots,M} \alpha_m)^M$ instead of α_m.

Inference Procedure

We now describe the inference procedure, which follows from the procedure in Li et al. (2018) originally designed for i.i.d. data and a small stepsize.

Point Estimation and Batching Given a trajectory of $(x_t)_{t\geq 0}$ sampled from a Markov chain, we run LSA with constant stepsize α and obtain iterates $(\theta_t^{(\alpha)})_{t\geq 0}$. The first b iterates $(\theta_t^{(\alpha)})_{t=0}^{b-1}$ are considered as initial burn-in and are not used in the inference procedure. For the remaining iterates, we divide them equally into K batches of size n. Within each batch, we discard the first $n_0 (\geq 0)$ iterates and compute the average of the remaining iterates:

$$\underbrace{\theta_0^{(\alpha)}, \ldots, \theta_{b-1}^{(\alpha)}}_{\text{burn in, discarded}}, \overbrace{\underbrace{\theta_b^{(\alpha)}, \ldots, \theta_{b+n_0-1}^{(\alpha)}}_{\text{discard}}, \underbrace{\theta_{b+n_0}^{(\alpha)}, \ldots, \theta_{b+n-1}^{(\alpha)}}_{\text{compute average } \bar{\theta}_1^{(\alpha)}}}^{\text{1st batch}},$$

$$\overbrace{\underbrace{\theta_{b+n}^{(\alpha)}, \ldots, \theta_{(b+n)+n_0-1}^{(\alpha)}}_{\text{discard}}, \underbrace{\theta_{(b+n)+n_0}^{(\alpha)}, \ldots, \theta_{b+2n-1}^{(\alpha)}}_{\text{compute average } \bar{\theta}_2^{(\alpha)}}}^{\text{2nd batch}}, \quad \ldots$$

Hence, for the k-th batch, we compute the point estimator $\bar{\theta}_k^{(\alpha)} = \frac{1}{n-n_0} \sum_{l=b+(n-1)k+n_0}^{b+nk-1} \theta_l^{(\alpha)}$. As such, we obtain a total of K batch-mean estimators $\{\bar{\theta}_k^{(\alpha)}\}$, which will be used for statistical inference. Note that we only need to save the running average of each batch-mean estimator without the necessity to store the entire trajectory $(\theta_t^{(\alpha)})_{t\geq 0}$.

Before delving into the construction of CIs, we briefly remark on several design choices. The initial b iterates are considered as burn-in and are omitted, as these iterates are far away from stationarity, and thus may have substantial optimization errors. The first n_0 iterates of each batch are also discarded, to reduce the correlation of the remaining iterates across batches, i.e., in the order of $\exp(-\alpha n_0)$.

Confidence Interval Construction Now that the CLT for the average iterates of Markovian LSA with constant stepsizes has been established, we construct estimators for $\mathbb{E}[\theta_\infty^{(\alpha)}]$ and Σ^* and subsequently build CIs for $\mathbb{E}[\theta_\infty]$.

With the K batch-mean estimators $\{\bar{\theta}_k^{(\alpha)}\}$ for $k = 1, \ldots, K$, we compute batch-mean estimators as

$$\bar{\theta}^{(\alpha)} = \frac{1}{K} \sum_{k=1}^{K} \bar{\theta}_k^{(\alpha)}.$$

For variance estimation, we adapt an estimator that has been considered in Flegal and Jones (2010); Chen et al. (2020), and Xie and Zhang (2022) to our problem. Given $\{\bar{\theta}_k^{(\alpha)}\}$, the variance estimator is computed as

$$\hat{\Sigma}^{(\alpha)} = \frac{(n-n_0)}{K} \sum_{k=1}^{K} \left(\bar{\theta}_k^{(\alpha)} - \bar{\theta}^{(\alpha)}\right)\left(\bar{\theta}_k^{(\alpha)} - \bar{\theta}^{(\alpha)}\right)^\top,$$

It has been shown in Flegal and Jones (2010) that $\hat{\Sigma}$ is a consistent estimator of Σ^* in Theorem 1 as $n, K \to \infty$,

Hence, for inference with LSA with stepsize α, for $q \in (0,1)$, we construct the $(1-q)100\%$-confidence interval for the i-th coordinate of $\mathbb{E}[\theta_\infty]$ as

$$\left[\bar{\theta}_i^{(\alpha)} - z_{1-\frac{q}{2}} \sqrt{\frac{\hat{\Sigma}_{i,i}^{(\alpha)}}{K(n-n_0)}}, \bar{\theta}_i^{(\alpha)} + z_{1-\frac{q}{2}} \sqrt{\frac{\hat{\Sigma}_{i,i}^{(\alpha)}}{K(n-n_0)}}\right].$$

In subsequent experiments, we focus on 95% CIs.

Combining With RR Extrapolation

Next, we apply RR extrapolation in addition to the above-delineated procedure to construct confidence intervals that have better coverage properties of θ^*.

To apply RR extrapolation, we first select a set of M distinct stepsizes, i.e., $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_M\}$ and run LSA iterates with these stepsizes simultaneously by using the same underlying data stream $(x_t)_{t\geq 0}$. For each trajectory of iterates $(\theta_t^{(\alpha_m)})_{t\geq 0}$, we follow the inference procedure and obtain K batch-mean estimators $\{\bar{\theta}_k^{(\alpha_m)}\}_{k=1}^K$. To obtain the RR extrapolated estimator, we linearly combine the k-th estimates across the M trajectories and obtain $\widetilde{\theta}_k^{\mathcal{A}} = \sum_{m=1}^M h_m \bar{\theta}_k^{(\alpha_m)}$, with $\{h_m\}$ computed according to (2). We then conduct statistical inference using the iterates $\{\widetilde{\theta}_k^{\mathcal{A}}\}$ and build confidence intervals following the similar methodology described above for a single trajectory.

Theoretical Guarantees

Next, we provide additional theoretical analysis that helps us achieve better statistical inference performance with RR extrapolated iterates. Please refer to the appendix for the proofs of propositions in this section.

Stepsize Selection in RR Extrapolation

As we discussed, one way to improve the coverage probability of θ^* is to reduce the bias via RR extrapolation. However, RR extrapolation does not come for free, as the coefficients $\{h_m\}$ solving (2) may blow up, thus resulting in large variance and offsetting the benefits of bias reduction. $\{h_m\}$ values are uniquely determined by inverting a Vandermonde matrix, which is infamous for being ill-conditioned when the "roots" $\{\alpha_m\}$ are positive real numbers. Therefore, when employing RR extrapolation, we need to carefully select M and $\{\alpha_m\}$ to maximize the benefits of bias reduction.

We study two stepsize selection schedules, namely geometric decaying and equidistant sequences. We assume that α_m decreases in value as m increases. We establish an upper bound to the variance of θ_∞ in each stepsize regime, which would offer some insight and guidance on stepsize selection.

The geometric decay schedule is not a conventional choice in numerical analysis, the field from which RR extrapolation originates, but it is frequently employed in machine learning.

Proposition 2. *Given unique stepsizes $\mathcal{A} = \{\alpha_1, \ldots, \alpha_M\}$. Assume $\alpha_1 < 1$ and $\alpha_m = \alpha_1/c^{m-1}$ with $c \geq 2$. We observe the following properties.*

1. $|h_m| \leq h_{\max}(c) = \exp\left(\frac{2}{c-1}\right)$.
2. $\mathrm{Var}(\widetilde{\theta}_\infty^{\mathcal{A}}) = \mathcal{O}\Big(c \cdot \exp(16\, c^{-1/2})\Big)$.

It is noteworthy that the variance upper bound for geometric decaying stepsizes does not depend on the number of stepsizes M used in the extrapolation, suggesting that the variance will not blow up as $M \to \infty$. Nonetheless, one problem with geometric decay is that the stepsize would decay too quickly to near zero as M increases. As such, RR iterates with a small constant stepsize mixes slowly and may need a much longer trajectory for the iterates to converge.

The equidistant decay schedule has been studied in numerical analysis (Gautschi 1990). We study the behavior of a generic equidistant decay schedule in RR exploration.

Proposition 3. *Given unique stepsizes $\mathcal{A} = \{\alpha_1, \ldots, \alpha_M\}$. Assume $a + b < 1$ and $\alpha_m = (a+b) - \frac{b(m-1)}{M-1}$ for $m = 1, \ldots, M$. We observe the following properties.*

1. $|h_m| = \binom{M}{m} \cdot \frac{bm}{a(M-1)+b(M-m)} \cdot \Big(\prod_{l=1}^M \frac{a(M-1)+b(l-1)}{bl}\Big)$.
2. $\mathrm{Var}(\widetilde{\theta}_\infty^{\mathcal{A}}) = \mathcal{O}((2M/b)^{2M})$.

When the stepsize sequence decays in equidistance, stepsizes do not decrease as quickly to zero based on the choice of a, b. However, the upper bound of variance now is of order M^M, which suggests that the variance could blow up quickly as M increases, potentially offsetting the benefits from reduced bias.

Zero Bias Special Cases

As discussed earlier, RR extrapolation reduces the asymptotic bias and hence increases the asymptotic coverage probability of θ^* based on the CI constructed for $\mathbb{E}[\theta_\infty]$. Hence, the Markovian underlying data and the presence of asymptotic bias should not discourage one from using constant stepsize LSA iterates for statistical inference.

In this section, we would like to highlight that Markovian data need not be a sufficient condition for the presence of asymptotic bias in LSA. That is, there are several commonly observed scenarios where no asymptotic bias is present, even when the underlying data is Markovian.

Independent Multiplicative Noise In this scenario, we expand our Markov chain state space to incorporate an independent bounded zero-mean random variable, i.e., $x_t = (s_t, \xi_t)$, where $(s_t)_{t\geq 0}$ is the uniformly ergodic Markov chain and $(\xi_t)_{t\geq 0}$ is uniformly bounded, i.e., $\|\xi_t\| \leq u$, and i.i.d. sampled from distribution ξ with $\mathbb{E}[\xi] = 0$. Consider the following Markovian LSA updates,

$$\theta_{t+1} = \theta_t + \alpha\Big((\bar{A} + \xi_t)\theta_t + b(s_t)\Big), \tag{3}$$

where $A(s_t, \xi_t) = \bar{A} + \xi_t$ and $b(s_t, \xi_t) = b(s_t)$. It can be easily verified that the above LSA iteration satisfies the required Assumptions 1–2.

Proposition 4. *Consider the LSA iteration of* (3)*, there exists some threshold $\alpha_0 \in (0, 1)$ such that $\forall \alpha \in [0, \alpha_0)$, $(x_t, \theta_t)_{t\geq 0}$ converges weakly to a unique stationary distribution. Moreover, $\mathbb{E}[\theta_\infty] = \theta^*$.*

The above proposition implies that when the Markovian noise is only additive, the limiting expectation of the iterates converges to θ^*. Hence, our CLT on $\mathbb{E}[\theta_\infty]$ allows us to construct CI and perform statistical inference directly on θ^*.

Independent Additive Noise in Linear Regression We consider an independent additive noise setting in linear regression, which has been previously discussed in Bresler et al. (2020) and Huo, Chen, and Xie (2023a).

In this linear regression problem, independent observations $(s_t)_{t\geq 0}$ are sequentially generated from a uniformly ergodic Markov chain with stationary distribution ν and $\mathbb{E}_\nu[s_t s_t^\top]$ is positive definite. $y_t = s_t^\top w^* + \epsilon_t$, where ϵ_t is an i.i.d. zero-mean noise with variance σ_ϵ^2. The SGD iterates to estimate w^* are updated as $w_{t+1} = w_t - \alpha s_t\big(\langle w_t, s_t\rangle - y_t\big)$.

Casting the problem under the LSA framework, we have $w_{t+1} = w_t + \alpha(A(x_t)w_t + b(x_t))$ with $x_t = (s_t, \epsilon_t)$, $A(x_t) = -s_t s_t^\top$ and $b(x_t) = s_t(s_t^\top w^* + \epsilon_t)$. It has been shown in Huo, Chen, and Xie (2023a) that LSA under constant stepsizes has zero asymptotic bias in this setup.

Realizable Linear-TD Learning in RL In this section, we specialize our discussion on linear-TD learning in RL.

TD learning algorithm (Sutton 1988) is an important algorithm in RL for policy evaluation. Potentially equipped with linear function approximation, it is a special case of LSA.

We model the linear-TD problem with a Markov Reward Problem (MRP) $(\mathcal{S}, P^{\mathcal{S}}, r, \gamma)$. The value function $V : \mathcal{S} \to \mathbb{R}$ is approximated by a linear function $V(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r(s_t)|s_0 = s] \approx \phi(s)^\top \theta$, where $\phi : \mathcal{S} \to \mathbb{R}^d$ is a known feature map of finite rank d and θ is the unknown weight vector to be estimated. The feature map is normalized such that $\phi_{\max} := \sup_{s\in\mathcal{S}} \|\phi(s)\|_2 \le \frac{1}{\sqrt{1+\gamma}}$.

We consider the semi-simulator regime, in which the iterates are updated in the following fashion,

$$\theta_{t+1} = \theta_t + \alpha\Big(r(s_t) + \gamma\phi(s_t^{\text{next}})^\top \phi_t - \phi(s_t)^\top \theta_t\Big)\phi(s_t), \quad (4)$$

where $(s_t)_{t\ge 0}$ is a Markov chain and s_t^{next} is sampled independently from $P^{\mathcal{S}}(s_t, \cdot)$ conditioned on s_t.

Proposition 5. *Assuming $(s_t)_{t\ge 0}$ is a uniformly ergodic Markov chain on state space $\mathcal{S}$ with $s_0 \sim \pi^{\mathcal{S}}$. Assuming the linear-TD is realizable, i.e., $\exists v \in \mathbb{R}^d$ such that $V(s) = \phi(s)^\top v$ for all $s \in \mathcal{S}$. The linear-TD iterates of* (4) *converge without asymptotic bias, i.e., $\mathbb{E}[\theta_\infty] = \mathbb{E}[\theta^*]$.*

The realizability assumption is not particularly restricting, as there are a number of RL problems that satisfy this assumption, such as TD learning on tabular MDP or linear MDP. Hence, if we have such a structured linear-TD problem, we could use the iterates obtained in the semi-simulator setting to perform statistical inference on the weight vector θ estimation and subsequently the value function directly.

Numerical Experiments

We conduct extensive numerical experiments to examine our proposed inference procedure and stepsize selection guidelines in RR extrapolation. We present our main results in this section. Additional sets of experiments and detailed experiment designs are included in the appendix.

Inference Performance Comparison

We examine the empirical performance of our proposed inference procedure with constant stepsizes and RR extrapolation. We consider LSA problems in dimension $d = 5$ for a finite state, irreducible, and aperiodic Markov chain with $N = 10$ states. We generate the transition probability P, and the functions A and b randomly; see the appendix for details. We construct CIs using the LSA iterates. We examine the performance from three aspects, namely the ℓ_2 error of the point estimate to the target vector, i.e.,$\|\bar{\theta} - \theta^*\|_2$, the coordinate-wise CI width and coverage probabilities.

We mainly study under constant stepsizes $\alpha = 0.2$ and $\alpha = 0.02$. The two choices of constant stepsizes are of different scales, allowing us to demonstrate extreme effects in convergence and inference performance. RR extrapolation is conducted using the two constant stepsizes. For comparison, we also consider diminishing stepsizes with initial stepsize α_0 being 0.2 and 0.02 respectively, and a diminishing stepsize $\alpha_t = \alpha_0 t^{-0.5}$ for $t \ge 1$. The diminishing rate $t^{-0.5}$ is chosen as it is on the boundary of square-summable assumption and has often been observed with the best empirical performance among $t^{-\beta}$ with $\beta \in [0.5, 1]$.

Baseline Comparison for I.I.D. LSA We first conduct a cross-study to examine the performance of inference with constant and diminishing stepsize under i.i.d. data. We notice that constant stepsizes slightly outperform the $0.2t^{-0.5}$ diminishing stepsize. Please refer to the appendix for detailed experiment design and results.

Performance Comparison for Markovian LSA We examine 100 different LSA problems of the same dimension $|\mathcal{X}| = 10$ and $d = 5$. For each LSA setup, the parameters (P, A, b) are generated randomly, and we simulate 100 trajectories $(x_t)_{t\ge 0}$ of length 10^5, run LSA iterates with the above-described stepsize regimes, and perform inference. We summarize the distribution of the performance across the 100 cases in Table 1.

With Markovian data, the LSA with constant stepsize converges with nonzero asymptotic bias. The presence of bias is rather obvious when $\alpha = 0.2$, evident from the large ℓ_2 error. Hence, CIs constructed with $\alpha = 0.2$ iterates have the worst coverage probabilities. Reducing the stepsize to $\alpha = 0.02$ will reduce the asymptotic bias, and hence CIs constructed with $\alpha = 0.02$ iterates have significantly better performance.

When the RR extrapolation technique is employed, the confidence intervals constructed enjoy the best coverage properties, with the smallest ℓ_2 error, comparable CI width, and higher coverage probabilities of θ^*. Moreover, the median coverage probability is around the targeted 95%.

Percentile		Comparison table				
		0.2	0.02	RR	$\frac{0.2}{\sqrt{t}}$	$\frac{0.02}{\sqrt{t}}$
10	ℓ_2	3.60	1.01	0.92	0.87	0.90
	CI	1.44	1.30	1.31	1.21	0.87
	Cov	0	72	90	86	62
25	ℓ_2	6.05	1.25	1.08	1.06	1.11
	CI	1.87	1.68	1.70	1.52	1.20
	Cov	0	82	91	88	70
50	ℓ_2	8.12	1.59	1.32	1.32	1.42
	CI	2.70	2.38	2.41	2.14	1.51
	Cov	11	90	94	91	76
75	ℓ_2	14.82	2.39	1.90	1.85	2.13
	CI	3.95	3.40	3.47	3.05	2.50
	Cov	66	93	95	94	83
90	ℓ_2	25.53	4.20	4.14	3.44	6.92
	CI	10.49	6.31	8.91	5.21	4.82
	Cov	92	95	97	96	90

Table 1: Inference comparison of different stepsize regimes. ℓ_2 denotes the ℓ_2 error of point estimates and is of unit 10^{-3}. "CI" refers to the CI width and is also of unit 10^{-3}. "Cov" represents the coverage probability. Both the CI width and coverage probability are for the 1-st coordinate estimate.

Batch No.		Comparison table				
		0.2	0.02	RR	$\frac{0.2}{\sqrt{t}}$	$\frac{0.02}{\sqrt{t}}$
100	ℓ_2	7.80 (0.02)	1.01 (0.02)	0.70 (0.02)	0.69 (0.02)	0.72 (0.02)
	CI	1.53 (0.01)	1.37 (0.00)	1.38 (0.00)	1.34 (0.00)	0.81 (0.00)
	Cov	0 (0)	80.6 (1.8)	94.4 (1.0)	95.0 (1.0)	71.2 (2.0)
500	ℓ_2	7.80 (0.02)	1.03 (0.02)	0.72 (0.02)	0.70 (0.02)	0.70 (0.02)
	CI	1.54 (0.00)	1.38 (0.00)	1.39 (0.00)	1.06 (0.00)	0.40 (0.00)
	Cov	0 (0)	80.8 (1.8)	94.2 (1.0)	88.8 (1.4)	42.2 (2.2)
1000	ℓ_2	7.79 (0.02)	1.01 (0.02)	0.73 (0.02)	0.69 (0.02)	0.74 (0.03)
	CI	1.54 (0.00)	1.38 (0.00)	1.39 (0.00)	0.82 (0.00)	0.29 (0.00)
	Cov	0 (0)	83.0 (1.7)	94.2 (1.0)	75.4 (1.9)	30.4 (2.1)

Table 2: Inference comparison of different batch numbers. ℓ_2 and "CI" values are of unit 10^{-3}. Both CI width and coverage probability are for the 1-st coordinate estimate.

Batch Number Selection We further inspect the impact of the number of batches used in inference, i.e., the value K in the confidence interval construction procedure. We focus on one specific LSA problem with $|\mathcal{X}| = 10$ and $d = 5$ Markovian data. The trajectory length is set at 10^6 and the number of batches varies. For each combination of stepsize and number of batches used, we run 500 independent runs and record the mean and the standard error in Table 2.

For constant stepsizes, as we compare across different rows, the mean estimates are at the same scale, and not influenced much by the choice of batch number. However, for diminishing stepsizes, as the batch number increases, the CI widths decrease quickly, which drastically reduces the coverage probability and implies an underestimation of the variance. As the stepsize decays in the diminishing stepsize sequence, the iterates become increasingly correlated, and hence a longer batch size is needed to overcome the strong correlation. Therefore, the number of batches used in diminishing stepsize regime cannot arbitrarily increase. In contrast, iterates under constant stepsize mixes at the same rate and is more robust to the number of batches used.

Trajectory Length We next investigate the impact of increasing trajectory length with a range of stepsizes to better visualize the trend. For each trajectory length, we fix the number of batches as $T^{0.3}$ as recommended in Chen et al. (2020). The statistics are in Table 3.

We observe that inference with RR iterates consistently presenting the best results. When the trajectory length is at 10^3, the iterates under various stepsize regimes would still be distant from θ^*, which explains the mediocre coverage of θ^*. As the trajectory length increases, the iterates gradually approach stationarity. Hence, the inference performance deteriorates for $\alpha = 0.2$ with 0 coverage probability, as the iterates converge to $\mathbb{E}[\theta_\infty]$, which is asymptotically biased from θ^*. On the other hand, as the trajectory length is longer, the iterates under diminishing stepsize converge closer to θ^*. Therefore, the inference performance with diminishing stepsize improves. Inference with RR extrapolated iterates generally gives satisfactory performance. Hence, we would like to note that inference with constant stepsize might be especially useful when the simulation trajectory length budget is limited, as diminishing stepsize iterates struggle to output good inference results.

Comparison Against Bootsrapping In all previous experiments, we are comparing inference using SA iterates under different stepsize regimes. Here, we compare to bootstrapping, a more classical approach to statistical inference. Detailed experimental design can be found in the appendix.

While constant stepsize together with RR obtains 95.2% coverage with ℓ_2 error 1.0×10^{-5} and CI width 0.00134, bootstrapping achieves 93.4% coverage with ℓ_2 error 2.0×10^{-3} and CI width 0.00411. The large ℓ_2 error and wide confidence interval of bootstrapping suggest that it may not be able to handle correlated data efficiently.

Bootstrapping requires the entire data set to be stored, requiring $\mathcal{O}(T^d)$ memory, whereas inference with SA iterates can accommodate online data, hence $\mathcal{O}(d)$ memory. Inference with SA iterates only needs first-order information, while bootstrapping may need higher-order information, making it potentially computationally prohibitive.

Extending to Nonlinear SA Lastly, we examine the performance of our proposed technique in an example of logistic regression with unbounded Markovian data, to demonstrate that our proposed technique is robust in a wide range of settings, even if not currently addressed by our theory.

The categorical data (x_t, y_t) arrives online, with x_t sequentially sampled from a 2-dimensional Gaussian AR(1) process, and $y_t \sim \text{Bernoulli}(\frac{1}{1+e^{-w_* x_t}})$ with $w_* \in \mathbb{R}^2$ being a random unit vector. This setting is non-Hurwitz, nonlinear, and with an unbounded underlying Markov chain, but our proposed inference procedure still exhibits similar performance as seen in Table 4.

RR Stepsize Selection

In this section, we numerically examine the impacts of different stepsize selection decisions in RR extrapolation.

Comparing Different Regimes We first compare the geometric and equidistant regimes. To ensure a fair comparison, we keep the range of the stepsize selection the same across the two regimes, so the difference would come solely from how the stepsize decays within the range. The range is determined by a dyadic decaying stepsize sequence, i.e. the smallest stepsize across the two regimes is fixed at $\alpha_M = \alpha_1/2^{(M-1)}$. Under this setup, when the RR order is 2, there is no difference between these two schedules, so we start the comparison from the extrapolation with 3 stepsizes.

As depicted in Figure 1, when the order of extrapolation increases to around 5, the benefits of RR extrapolation have been mostly exploited. The difference between the two decaying schedules is more significant when the order is low.

Length		Comparison table							
		0.2	0.15	0.1	0.05	0.02	RR	$\frac{0.2}{\sqrt{t}}$	$\frac{0.02}{\sqrt{t}}$
10^3	ℓ_2	2.84 (0.06)	2.55 (0.06)	2.48 (0.05)	2.35 (0.05)	2.28 (0.05)	2.28 (0.05)	2.36 (0.05)	4.28 (0.07)
	CI	4.16 (0.07)	3.91 (0.06)	3.81 (0.06)	3.68 (0.06)	3.71 (0.06)	3.76 (0.06)	3.10 (0.05)	1.01 (0.02)
	Cov	82.2 (1.7)	84.0 (1.6)	83.0 (1.7)	83.6 (1.7)	83.6 (1.7)	83.4 (1.7)	76.8 (1.9)	24.8 (1.9)
10^4	ℓ_2	1.15 (0.02)	0.99 (0.02)	0.89 (0.02)	0.79 (0.02)	0.73 (0.02)	0.72 (0.02)	0.73 (0.02)	1.14 (0.02)
	CI	1.44 (0.02)	1.38 (0.01)	1.34 (0.01)	1.31 (0.01)	1.29 (0.01)	1.30 (0.01)	1.12 (0.01)	0.67 (0.01)
	Cov	75.2 (1.9)	81.8 (1.7)	85.2 (1.6)	88.8 (1.4)	90.0 (1.3)	89.2 (1.4)	85.4 (1.6)	56.0 (2.2)
10^5	ℓ_2	0.82 (0.01)	0.61 (0.01)	0.44 (0.01)	0.29 (0.01)	0.24 (0.01)	0.23 (0.01)	0.22 (0.01)	0.24 (0.01)
	CI	0.46 (0.00)	0.45 (0.01)	0.43 (0.01)	0.42 (0.01)	0.41 (0.00)	0.41 (0.00)	0.38 (0.00)	0.29 (0.00)
	Cov	0.04 (0.9)	22.2 (1.8)	56.4 (2.2)	83.2 (1.7)	88.2 (1.4)	91.2 (1.3)	90.0 (1.3)	79.2 (1.8)
10^6	ℓ_2	0.81 (0.00)	0.57 (0.00)	0.38 (0.00)	0.20 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	CI	0.15 (0.00)	0.15 (0.00)	0.14 (0.00)	0.14 (0.00)	0.13 (0.00)	0.13 (0.00)	0.13 (0.00)	0.11 (0.00)
	Cov	0 (0)	0 (0)	0 (0)	19.0 (1.8)	80.2 (1.8)	95.2 (1.0)	92.8 (1.2)	90.2 (1.3)

Table 3: Inference comparison of different trajectory lengths. ℓ_2 and "CI" values are of unit 10^{-2}. Both CI width and coverage probability are for the 1-st coordinate estimate.

Length		Comparison table				
		0.2	0.02	RR	$\frac{0.2}{\sqrt{t}}$	$\frac{0.02}{\sqrt{t}}$
10^3	ℓ_2	11.63 (0.28)	21.53 (0.26)	24.4 (0.25)	20.89 (0.28)	41.14 (0.25)
	CI	24.01 (0.40)	34.87 (0.27)	38.07 (0.26)	15.17 (0.20)	16.31 (0.12)
	Cov	70.2 (2.0)	34.0 (2.1)	26.4 (2.0)	3.6 (0.8)	0 (0)
10^4	ℓ_2	8.35 (0.11)	2.87 (0.07)	3.27 (0.08)	3.65 (0.09)	10.64 (0.11)
	CI	9.27 (0.10)	9.37 (0.09)	9.79 (0.09)	5.84 (0.07)	10.23 (0.07)
	Cov	12.0 (1.5)	89.8 (1.4)	84.8 (1.6)	53.6 (2.2)	1.4 (0.5)
10^5	ℓ_2	8.14 (0.04)	1.10 (0.03)	0.88 (0.02)	0.90 (0.02)	1.40 (0.03)
	CI	3.09 (0.03)	2.83 (0.02)	2.82 (0.02)	1.83 (0.02)	2.74 (0.02)
	Cov	0 (0)	80.2 (1.8)	90.2 (1.3)	73.0 (2.0)	59.6 (2.2)

Table 4: Inference comparison of different trajectory lengths. ℓ_2 and "CI" values are of unit 10^{-2}. Both CI width and coverage probability are for the 1-st coordinate estimate.

Comparing Spacing in the Equidistant Regime We study the impact of decay distance in the equidistant decaying scheme, i.e., varying the value of b in $\alpha_m = (a+b) - b\frac{(m-1)}{M-1}$. Specifically, we keep the $a+b$ value constant, and test with different $b/(M-1) \in \{0.01, 0.02, 0.03, 0.04\}$.

As shown in Figure 2, the smaller $b/(M-1)$ is, the worse the RR extrapolation performance is. This phenomenon is in line with Proposition 3, as $b/(M-1)$ inversely scales with the variance upper bound. What is surprising is that when using small $b/(M-1)$, increasing the number of the stepsizes used does not always reduce the ℓ_2 error of the extrapolated iterates. This may be due to the proximity of stepsizes, leading to large coefficients $\{h_m\}$, increasing the variance and offsetting the reduction in bias. Hence, this suggests that for RR extrapolation to be effective, stepsizes should not be too close to each other, but should explore a range of values.

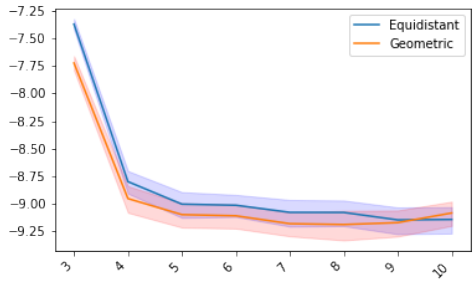

Figure 1: ℓ_2 errors of averaged iterates with different order of RR extrapolation. The Y-axis is of logarithmic scale.

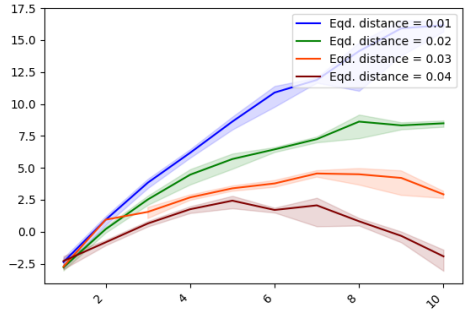

Figure 2: ℓ_2 errors of averaged iterates with different order of RR extrapolation. The Y-axis is of logarithmic scale.

Conclusion

In this paper, we demonstrate the effectiveness of statistical inference with LSA iterates under Markovian data, constant stepsizes, and RR extrapolation. An immediate next step is to provide theoretical results to justify the validity of the CI constructed with Markovian LSA iterates with constant stepsize, as the CI is non-consistent. Further directions include: (a) develop an anytime variance estimator so that the proposed inference procedure can be fully online; (b) given a fixed simulation length, how one should decide the order of RR extrapolation to achieve decent inference performance.

Acknowledgments

Y. Chen is supported in part by NSF CAREER Award CCF-2233152. Q. Xie is supported in part by NSF grant CNS-1955997.

References

Bach, F. 2021. On the Effectiveness of Richardson Extrapolation in Data Science. *SIAM Journal on Mathematics of Data Science*, 3(4): 1251–1277.

Bhandari, J.; Russo, D.; and Singal, R. 2021. A Finite Time Analysis of Temporal Difference Learning with Linear Function Approximation. *Operations Research*, 69(3): 950–973.

Blum, J. R. 1954. Approximation Methods Which Converge with Probability One. *The Annals of Mathematical Statistics*, 25(2): 382 – 386.

Borkar, V.; Chen, S.; Devraj, A.; Kontoyiannis, I.; and Meyn, S. 2023. The ODE Method for Asymptotic Statistics in Stochastic Approximation and Reinforcement Learning. arXiv:2110.14427.

Borkar, V. S.; and Meyn, S. P. 2000. The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning. *SIAM Journal on Control and Optimization*, 38(2): 447–469.

Bresler, G.; Jain, P.; Nagaraj, D.; Netrapalli, P.; and Wu, X. 2020. Least Squares Regression with Markovian Data: Fundamental Limits and Algorithms. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.

Chen, X.; Lee, J. D.; Tong, X. T.; and Zhang, Y. 2020. Statistical inference for model parameters in stochastic gradient descent. *The Annals of Statistics*, 48(1): 251 – 273.

Dieuleveut, A.; Durmus, A.; and Bach, F. 2020. Bridging the gap between constant step size stochastic gradient descent and Markov chains. *The Annals of Statistics*, 48(3): 1348 – 1382.

Durmus, A.; Moulines, E.; Naumov, A.; and Samsonov, S. 2023. Finite-time High-probability Bounds for Polyak-Ruppert Averaged Iterates of Linear Stochastic Approximation. arXiv:2207.04475.

Flegal, J. M.; and Jones, G. L. 2010. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2): 1034 – 1070.

Gautschi, W. 1990. How (Un)stable Are Vandermonde Systems? In Wong, R. S. C., ed., *Asymptotic and Computational Analysis*, chapter 10, 266–290. CRC Press.

Huo, D. L.; Chen, Y.; and Xie, Q. 2023a. Bias and Extrapolation in Markovian Linear Stochastic Approximation with Constant Stepsizes. arXiv:2210.00953.

Huo, D. L.; Chen, Y.; and Xie, Q. 2023b. Effectiveness of Constant Stepsize in Markovian LSA and Statistical Inference. arXiv:2312.10894.

Lakshminarayanan, C.; and Szepesvári, C. 2018. Linear Stochastic Approximation: How Far Does Constant Step-Size and Iterate Averaging Go? In Storkey, A.; and Perez-Cruz, F., eds., *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, 1347–1355. PMLR.

Lee, S.; Liao, Y.; Seo, M. H.; and Shin, Y. 2022. Fast and Robust Online Inference with Stochastic Gradient Descent via Random Scaling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7): 7381–7389.

Li, T.; Liu, L.; Kyrillidis, A.; and Caramanis, C. 2018. Statistical Inference Using SGD. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Li, X.; Liang, J.; Chang, X.; and Zhang, Z. 2022. Statistical Estimation and Online Inference via Local SGD. In Loh, P.-L.; and Raginsky, M., eds., *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, 1613–1661. PMLR.

Li, X.; Liang, J.; and Zhang, Z. 2023. Online Statistical Inference for Nonlinear Stochastic Approximation with Markovian Data. arXiv:2302.07690.

Liu, R.; Chen, X.; and Shang, Z. 2023. Statistical Inference with Stochastic Gradient Methods under ϕ-mixing Data. arXiv:2302.12717.

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 Abernethy, J.; and Agarwal, S., eds., *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, 2947–2997. PMLR.

Mou, W.; Pananjady, A.; Wainwright, M. J.; and Bartlett, P. L. 2021. Optimal and instance-dependent guarantees for Markovian linear stochastic approximation. arXiv:2112.12770.

Polyak, B. T. 1990. New stochastic approximation type procedures. *Automation and Remote Control*, 51(7): 98–107.

Polyak, B. T.; and Juditsky, A. B. 1992. Acceleration of Stochastic Approximation by Averaging. *SIAM Journal on Control and Optimization*, 30(4): 838–855.

Robbins, H.; and Monro, S. 1951. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3): 400 – 407.

Ruppert, D. 1988. Efficient Estimations from a Slowly Convergent Robbins-Monro Process. Technical report, Cornell University.

Srikant, R.; and Ying, L. 2019. Finite-Time Error Bounds For Linear Stochastic Approximation and TD Learning. In Beygelzimer, A.; and Hsu, D., eds., *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, 2803–2830. PMLR.

Sutton, R. S. 1988. Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1): 9–44.

Xie, C.; and Zhang, Z. 2022. A Statistical Online Inference Approach in Averaged Stochastic Approximation. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 8998–9009. Curran Associates, Inc.

Yu, L.; Balasubramanian, K.; Volgushev, S.; and Erdogdu, M. A. 2021. An Analysis of Constant Step Size SGD in the Non-convex Regime: Asymptotic Normality and Bias. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 4234–4248. Curran Associates, Inc.

Zhu, W.; Chen, X.; and Wu, W. B. 2023. Online Covariance Matrix Estimation in Stochastic Gradient Descent. *Journal of the American Statistical Association*, 118(541): 393–404.