Optimal and instance-dependent guarantees for Markovian linear stochastic approximation

Wenlong Mou WMOU@BERKELEY.EDU
Department of EECS, UC Berkeley
Ashwin Pananjady ASHWINPM@GATECH.EDU
School of ISyE and School of ECE, Georgia Tech
Martin J. Wainwright WAINWRIG@BERKELEY.EDU
Peter L. Bartlett PETER@BERKELEY.EDU
Department of EECS and Department of Statistics, UC Berkeley

Editors: Po-Ling Loh and Maxim Raginsky

Keywords: Markov chains, stochastic approximation, reinforcement learning, temporal difference methods, minimax lower bound

Let $s_1, s_2, \cdots, s_n$ denote a trajectory of length n from an ergodic Markov chain with stationary distribution ξ, and let $\{L_t\}_{t=1}^n$ and $\{b_t\}_{t=1}^n$ denote sequences of random functions from the state space, taking values in $\mathbb{R}^{d\times d}$ and $\mathbb{R}^d$, respectively. We study stochastic approximation (SA) procedures for approximately solving the d-dimensional linear fixed point equation

$$\bar{\theta} = \bar{L}\bar{\theta} + \bar{b}, \quad \text{where} \quad \bar{L} = \mathbb{E}_{s\sim\xi}[L_t(s)] \quad \text{and} \quad \bar{b} = \mathbb{E}_{s\sim\xi}[b_t(s)], \tag{1}$$

In particular, we consider the classical stochastic approximation iterate sequence (with constant stepsize η), as well as its Polyak–Ruppert averaged analog (with burn-in period n_0), given by

$$\theta_{t+1} = \theta_t + \eta \cdot (L_{t+1}(s_{t+1})\theta_t - b_t(s_{t+1})) \;\text{ and }\; \widehat{\theta}_n = (\theta_{n_0} + \ldots + \theta_{n-1})/(n - n_0), \quad \text{respectively.}$$

The random observations typically satisfy $|||L_t(s)|||_{\text{op}} = \Theta(d)$ and $\|b_t(s)\|_2 = \Theta(\sqrt{d})$ almost surely, and our main goal is to establish that the estimators converge to $\bar{\theta}$ at a rate that depends optimally on the dimension and mixing time. Accordingly, we first establish the MSE bound on the SA iterates

$$\mathbb{E}[\|\theta_n - \bar{\theta}\|_2^2] \lesssim \eta d t_{\text{mix}}, \quad \text{for constant stepsize choice } \eta \in \big(\log n/n, (t_{\text{mix}}d)^{-1}\big) \tag{2}$$

With the optimal choice of stepsize and burn-in period, we then prove a non-asymptotic, instance-dependent bound on the averaged iterate $\widehat{\theta}_n$:

$$\mathbb{E}\big[\|\widehat{\theta}_n - \bar{\theta}\|_2^2\big] \leq n^{-1}\mathrm{Tr}\big((I_d - \bar{L})^{-1}\Sigma^*(I_d - \bar{L})^{-\top}\big) + O\big((dt_{\text{mix}}/n)^{4/3}\big), \tag{3}$$

where the matrix Σ^* is the covariance matrix in the Markovian central limit theorem satisfied by the appropriately defined noise process at $\bar{\theta}$. Both the leading-order (first) term and high-order (second) term exhibit sharp dependence on the parameters (d, t_{mix}). We complement these upper bounds with a non-asymptotic local minimax lower bound over a small neighborhood of a given Markovian transition kernel, and this matches the leading-order term in Eq. (3). Taken together, these results establish the instance-optimality of the averaged SA estimator $\widehat{\theta}_n$ in the Markovian setting.

We derive corollaries of these results for policy evaluation with Markov noise—covering the TD(λ) family of algorithms for all $\lambda \in [0, 1)$—and parameter estimation in linear autoregressive models. Our instance-dependent characterizations open the door to designing fine-grained model selection procedures for hyperparameter tuning (e.g., choosing the value of λ when running the TD(λ) algorithm).

. Extended abstract. Full version appears as [arXiv:2112.12770]

Acknowledgments

We gratefully acknowledge the support of the ONR through MURI award N000142112431 to PLB. This work was also supported in part by NSF-DMS grant 2015454, and DOD-ONR Office of Naval Research N00014-21-1-2842 to MJW. This work was also supported by NSF-IIS grant 1909365 and NSF FODSI grant 202350 to PLB and MJW. AP was supported in part by the NSF grant CCF-2107455 and an Adobe Data Science Research Award. He is thankful to the Simons Institute for the Theory of Computing for their hospitality when part of this work was performed.