Proceedings of Machine Learning Research vol 144:1–1, 2021

Nonlinear Two-Time-Scale Stochastic Approximation: Convergence and Finite-Time Performance

Thinh T. Doan THINHDOAN@VT.EDU
Department of Electrical and Computer Engineering
Virginia Tech, USA

Extended Abstract

We consider the so-called two-time-scale stochastic approximation (SA), a generalized variant of the classic SA, which is used to find the root of a system of two coupled nonlinear equations. Given two operators $F, G : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, we seek to find $x^\star$ and $y^\star$ satisfying

$$\begin{cases} F(x^\star, y^\star) = 0, \\ G(x^\star, y^\star) = 0. \end{cases} \tag{1}$$

We assume that we only have access to the noisy values of $F(x, y)$ and $G(x, y)$ for a given pair (x, y), that is, $F(x, y) + \xi$ and $G(x, y) + \psi$, where ξ and ψ are two random variables. For solving problem (1), the two-time-scale nonlinear SA iteratively updates the iterates x_k and y_k, the estimates of $x^\star$ and $y^\star$, respectively, given an arbitrarily initial conditions x_0 and y_0 as

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k \left(F(x_k, y_k) + \xi_k\right), \\ y_{k+1} &= y_k - \beta_k \left(G(x_k, y_k) + \psi_k\right), \end{aligned} \tag{2}$$

Here α_k, β_k are two step sizes chosen such that $\beta_k \ll \alpha_k$, i.e., the second iterate is updated using step sizes that are very small as compared to the ones used to update the first iterate. The update of x_k is referred to as the "fast-time scale" while the update of y_k is called the "slow-time scale". The time-scale difference is loosely defined as the ratio β_k/α_k. In (2), the update of the *fast iterate* depends on the *slow iterate* and vice versa, that is, they are coupled to each other. To handle this coupling, the two step sizes have to be properly chosen to guarantee the convergence of the method. Indeed, an important problem is to select the two step sizes so that the two iterates converge as fast as possible. Our main contribution is to address this problem.

Main contributions: Our main focus is to derive the asymptotic convergence and finite-time performance of the nonlinear two-time-scale SA. Under some proper choice of step sizes α_k and β_k, we show that the mean square error generated by the method converges at a rate

$$\mathbb{E}\left[\|y_k - y^\star\|^2\right] + (\beta_k/\alpha_k)\mathbb{E}\left[\|x_k - x^\star\|^2\right] \leq \mathcal{O}\left(\frac{1}{(k+1)^{2/3}}\right).$$

Our key technique is motivated by the classic control theory for singularly perturbed systems, that is, we properly choose the two step sizes to characterize the coupling between the fast and slow-time-scale iterates. More details can be found in Doan (2020).

References

T. T. Doan. Nonlinear two-time-scale stochastic approximation: Convergence and finite-time performance. Available at: https://arxiv.org/abs/2011.01868, 2020.