E1 396: Topics in Stochastic Approximation Algorithms

Naman Mishra

August 2025

Contents

Lectures

The course

Instructor: Shalabh Bhatnagar

Lectu
Wedne
Augus

Textbook: Stochastic Approximation: A Dynamical Systems Viewpoint by Vivek S. Borkar. (2022, Hindustan Book Agency)

Evaluation:

(40%) Two midterms

(25%) Final exam

(35%) Project

H. Robbins and S. Monro published a landmark paper in the Annals of Mathematical Statistic in 1951. The subject of the paper was the following. Suppose there is a function $f\colon \mathbb{R} \to \mathbb{R}$ and iid noises $M_1, M_2, \ldots$ with zero mean. We are to find a θ^* such that $f(\theta^*) = 0$.

They proposed what is now known as the Robbins-Monro algorithm. Set

$$\theta_{n+1} = \theta_n + a_n(f(\theta_n) + M_{n+1})$$

a_n is called the *step size* or the *learning rate*. They showed that $\theta_n \to \theta^*$ with $f(\theta^*) = 0$.

Assume that the domain is a bounded box $[0,1]^d$ One could take many samples at each mesh point in a dense mesh, and hope that the law of large numbers allows you to sketch out f. This is infeasible in practice. The method of stochastic approximation *does not yield* much information about f, other than that it has a root at the limit θ^*.

Examples.

- To find the fixed point of a given function h, set $f = h - \mathrm{id}$.
- Suppose $h\colon \mathbb{R}^d \to \mathbb{R}$ and we wish to find a local minimum. Set f to be $-\nabla h$, where of course the gradients we compute will be noisy.

> There are known simple techniques to obtain estimates of gradients that at least have zero mean at minima.

Roadmap

- Introduce stochastic approximation (chapter 1)
- Background in probability, analysis and ODEs
- Asymptoptic analysis of stochastic approximation, assuming stability of the iterates (chapter 2)
- Sufficient conditions for stability (chapter 3)
- Lock-in probability and sample complexity of stochastic approximation (chapter 4)
- Multi-timescale stochastic approximation
- If time permits, stochastic approximation with Markov noise
- Stochastic approximation with set-valued maps

Stability in this context means that $\sup_n \|\theta_n\| < \infty$ almost surely.

Chapter I

First chapter?

The strong law of large number states that the empirical mean of iid random variables with finite mean converges to the mean almost surely. Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Writing

$$\begin{aligned}
S_{n+1} &= \frac{n}{n+1} S_n + \frac{1}{n+1} X_{n+1} \\
&= \left(1 - \frac{1}{n+1}\right) S_n + \frac{1}{n+1} X_{n+1} \\
&= S_n + \frac{1}{n+1}(X_{n+1} - S_n) \\
&= S_n + a_n(h(S_n) + X_{n+1})
\end{aligned}$$

where $a_n = \frac{1}{n+1}$ and $h = -\operatorname{id}$. This is now a stochastic approximation scheme if one assumes $\mathbf{E}[X_1] = 0$.

The corresponding ODE is

$$\dot{s}(t) = -s(t),$$

which has an equilibrium at $s(t) = 0$.

Suppose instead that $a_n = \frac{1}{(n+1)^{2/3}}$. We would still recover a strong law-like statement even though the mean is not uniformly weighted.