Stochastic Approximation & the Need for Speed

Adithya M. Devraj[1], Shuhang Chen[1], Ana Bušić[2] and Sean P. Meyn[1]

[1]University of Florida, [2]Inria & École Normale Supérieure

Stochastic Approximation

Goal: Find $\theta^* \in \mathbb{R}^d$ such that

$$\bar f(\theta^*) := \mathsf{E}[f(\theta^*, \boldsymbol{\Phi})] = 0$$

Algorithm [Robbins & Monro, 1951]:

$$\theta_{n+1} = \theta_n + \alpha_{n+1} f(\theta_n, \Phi_{n+1})$$

Key assumption: *associated ODE* is stable

$$\frac{d}{dt}\vartheta_t = \bar f(\vartheta_t)\,, \qquad \text{stationary point: } \theta^*$$

Central Limit Theorem & Rates of Convergence

Asymptotic covariance due to CLT:

$$\Sigma^\theta := \lim_{n\to\infty} n\mathsf{E}[\tilde\theta_n \tilde\theta_n^\intercal]\,, \qquad \tilde\theta_n := \theta_n - \theta^*$$

Define *linearization matrix* $A(\theta) := \partial \bar f(\theta)$, and let $\alpha_n \equiv 1/n$

- If $\text{Real}(\lambda) < -\frac{1}{2}$ for every eigenvalue λ of $A(\theta^*)$, $\Sigma^\theta \geq 0$ is obtained as the solution to:

$$\big(\tfrac{1}{2}I + A(\theta^*)\big)\Sigma^\theta + \Sigma^\theta\big(\tfrac{1}{2}I + A(\theta^*)\big)^\intercal + \Sigma_\Delta = 0$$

- Implies $O(1/n)$ convergence rate: for some $\delta > 0$,

$$\Sigma_n := \mathsf{E}\big[\tilde\theta_n\tilde\theta_n^\intercal\big] = n^{-1}\Sigma^\theta + O(n^{-1-\delta})$$

- Suppose $\text{Real}(\lambda) \geq -\frac{1}{2}$ for some eigenvalue λ of $A(\theta^*)$, with left eigenvector $v \neq 0$, and $\Sigma_\Delta v \neq 0$; Defining $\bar\rho := 2|\text{Real}(\lambda)|$,

$$\lim_{n\to\infty} n^\rho \mathsf{E}\big[(v^\intercal\tilde\theta_n)^2\big] = 0 \ \ ,\rho < \bar\rho \qquad \text{and} \qquad \lim_{n\to\infty} n^\rho \mathsf{E}\big[(v^\intercal\tilde\theta_n)^2\big] = \infty \ \ ,\rho \geq \bar\rho$$

- Implies convergence rate (much) slower than $O(1/n)$

Q-learning

Goal: Given a parameterized family of functions $\{Q^\theta : \mathbb{R}^d \times \mathsf{X}\times\mathsf{U} \to \mathbb{R}\}$, discount factor $\beta \in [0,1)$, and $\psi : \mathsf{X}\times\mathsf{U} \to \mathbb{R}^d$, find $\theta^* \in \mathbb{R}^d$ such that

$$\bar f(\theta^*) = \mathsf{E}\Big[\Big(c(X_n,U_n) + \beta\min_u Q^{\theta^*}(X_{n+1},u) - Q^{\theta^*}(X_n,U_n)\Big)\psi(X_n,U_n)\Big] = 0$$

- *Tabular* Q-learning: ODE is stable, but largest $\lambda\big(A(\theta^*)\big)$ is like $(\beta - 1)$
- Infinite CLT covariance, and convergence rate $\approx O(n^{-2(1-\beta)})$, if $\beta > 0.5$
- Function approximation Q-learning: ODE need not be stable

Objectives

Propose new algorithms, so that:

- Their *associated ODE* is stable *and* converges to θ^*
- They have optimal CLT covariance

Zap Stochastic Newton-Raphson

Matrix gain SA: $\theta_{n+1} = \theta_n + G_n f(\theta_n, \Phi_{n+1})$ **Associated ODE:** $\frac{d}{dt}\vartheta_t = \mathcal{G}(\vartheta_t)\bar f(\vartheta_t)$

Zap SNR: Choose $\{G_n\}$ such that $\mathcal{G}(\vartheta_t) = -[\varepsilon I + A(\vartheta_t)^\intercal A(\vartheta_t)]^{-1}A(\vartheta_t)^\intercal$

Associated ODE's:

$$\boxed{\begin{aligned} &\frac{d}{dt}\vartheta_t = -[\varepsilon I + A(\vartheta_t)^\intercal A(\vartheta_t)]^{-1}A(\vartheta_t)^\intercal \bar f(\vartheta_t) \\ \Longrightarrow\ &\frac{d}{dt}\bar f(\vartheta_t) = -A(\vartheta_t)[\varepsilon I + A(\vartheta_t)^\intercal A(\vartheta_t)]^{-1}A(\vartheta_t)^\intercal \bar f(\vartheta_t)\end{aligned}}$$

CLT covariance: $\Sigma^\theta = \Sigma^* := A(\theta^*)^{-1}\Sigma_\Delta(A(\theta^*)^{-1})^\intercal$ when $\varepsilon = 0$; It is optimal: $\Sigma^{\mathcal{G}} \geq \Sigma^*$ for any other $\mathcal{G}$

Zap Q-learning

Zap Q-learning $\equiv$ Zap Stochastic Newton-Raphson for Q-learning

- Stable ODE, even for Q-learning with ***non-linear*** function approximation
- Super fast convergence with optimal CLT covariance
- **ODE for tabular Q-learning:** $\boxed{Q^{\theta_t} := \mathcal{Q}(c_t)\,, \qquad \frac{d}{dt}c_t = -c_t + c}$

Where $\mathcal{Q}$ is a mapping from cost functions to corresponding *optimal* Q-functions

Application: Learn the shortest path via solving a Bellman equation

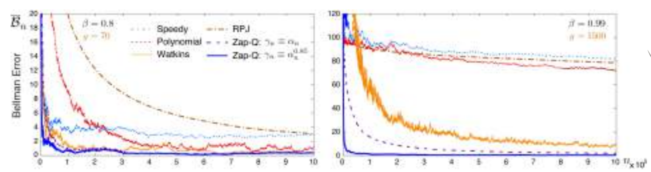

Figure: Zap Q-learning applied to a stochastic shortest-path problem

Application: Q-learning for Optimal Stopping Time in Finance

Objective: Maximize pay-off from a single stock by learning the optimal time to exercise using Q-learning.

Figure: Histograms of pay-off obtained using different Q-learning algorithms

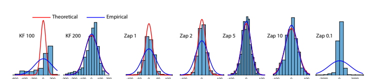

Figure: Theoritical and experimental CLT covariance for different Q-learning algorithms

Matrix Momentum Stochastic Approximation

Consider a linear model:

$$f(\theta_n, \Phi_{n+1}) = A_{n+1}\theta_n + b_{n+1}\,, \qquad \bar f(\theta_n) = A\theta_n + b$$

Polyak's momentum technique [Polyak, 1964]: $\gamma > 0$,

$$\theta_{n+1} = \theta_n + \gamma(\theta_n - \theta_{n-1}) + \alpha_{n+1}f(\theta_n, \Phi_{n+1})$$

Matrix Momentum SA: $\{M_{n+1}\}$ and $\{G_{n+1}\} \in \mathbb{R}^{d\times d}$

$$\theta_{n+1} = \theta_n + M_{n+1}(\theta_n - \theta_{n-1}) + \alpha_{n+1}G_{n+1}f(\theta_n, \Phi_{n+1})$$

PolSA: $\theta_{n+1} = \theta_n + [I + \zeta A](\theta_n - \theta_{n-1}) + \alpha_{n+1}\zeta f(\theta_n, \Phi_{n+1})$

NeSA: $\theta_{n+1} = \theta_n + [I + \zeta A_{n+1}](\theta_n - \theta_{n-1}) + \alpha_{n+1}\zeta f(\theta_n, \Phi_{n+1})$

Main Results

- PolSA couples with ($\varepsilon = 0$) Zap SNR at rate $O(1/n^2)$:

$$\sup_{n\geq 0}\ n^2\mathsf{E}\big[\|\theta_n^{\text{PolSA}} - \theta_n^{\text{Zap}}\|^2\big] < \infty$$

- Implies optimal CLT covariance for PolSA
- Expression for CLT covariance of NeSA is also obtained: finite, but not optimal

Application to Q-learning

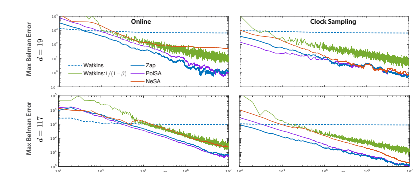

Figure: Matrix momentum & Zap Q-learning applied to a stochastic shortest-path problem

References

- A. M. Devraj and S. P. Meyn, *Zap Q-learning*. *NIPS*. Dec. 2017.
- A. M. Devraj and S. P. Meyn, *Fastest convergence for Q-learning*. Available on *ArXiv*. Jul. 2017.
- A. M. Devraj, A. Bušić, and S. P. Meyn. *Optimal Matrix Momentum Stochastic Approximation and Applications to Q-learning*. *ArXiv e-prints*, Feb. 2019.
- S. Chen, A. M. Devraj, A. Bušić, and S. P. Meyn. *Zap Q-learning for Optimal Stopping Time Problems*. *ArXiv e-prints*, Apr. 2019.
- S. Chen, A. M. Devraj, A. Bušić, and S. P. Meyn. *Zap Q-learning with Nonlinear Function Approximation*. *ArXiv e-prints*, June. 2019.