Stochastic Approximation & the Need for Speed
Adithya M. Devraj1, Shuhang Chen1, Ana Bušić2 and Sean P. Meyn1
1University of Florida, 2Inria & École Normale Supérieure

Stochastic Approximation
Goal: Find \( \theta^* \in \mathbb{R}^d \) such that
\[
\frac{d}{dt} \theta(t) = -\nabla_{\theta} J(\theta(t), \Phi) = 0
\]
Algorithm [Robbins & Monro, 1951]:
\[
\theta_{n+1} = \theta_n + \alpha_n \nabla_{\theta} J(\theta_n, \Phi)
\]
Key assumption: associated ODE is stable
\[
\frac{d}{dt} \theta(t), \quad \text{stationary point: } \theta^*
\]

Central Limit Theorem & Rates of Convergence
Asymptotic covariance due to CLT
\[
\Sigma_n = \lim_{n \to \infty} \text{Var}(\theta_n) = \frac{1}{n} \langle (\theta_n - \theta^*)(\theta_n - \theta^*)^\top \rangle = \frac{1}{n} \Sigma_0
\]
Define incrementation matrix \( A(\theta) = \frac{d}{d\theta} J(\theta, \Phi) \)
• If \( \text{Real}(A(\theta)) < -\frac{1}{2} \), every eigenvalue \( \lambda \) of \( A(\theta) \), \( \lambda < 0 \)
  obtained as the solution to
  \[
  (\lambda I + A(\theta))^T \Sigma_n + \frac{1}{n} \langle J(\Phi)^\top A(\theta) \rangle^T \Sigma_n - \Sigma_0 = 0
  \]
• Implies 0(1/n) convergence rate for some \( \delta > 0 
  \Sigma_n = \text{O}((n^{-\delta})^2)
• Suppose \( \text{Real}(A(\theta)) \geq -\frac{1}{2} \), with left eigenvector \( \psi \neq 0 \), and \( \Sigma_n \neq 0 \),
  \( \lim \text{sup} E[(\theta_n - \theta^*)(\theta_n - \theta^*)^\top] = \infty, \quad \psi \neq 0 \)
  \( \lim \text{sup} E[(\theta_n - \theta^*)^2] = \infty, \quad \psi \neq 0 \)

Q-learning
Goal: Given a parameterized family of functions \( Q^\phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \),
discount factor \( \beta \in [0, 1] \), and \( \psi : \mathbb{R}^d \to \mathbb{R} \), find \( \phi \in \mathbb{R}^d \) such that
\[
\frac{d}{dt} \phi(t) = -\nabla_{\phi} J(\phi(t), \psi(t)) = 0
\]
• Stable ODE, even for Q-learning with non-linear function approximation
• Super fast convergence with optimal CLT covariance

ODE for tabular Q-learning:
\[
\frac{d}{dt} \phi(t) = -\nabla_{\phi} J(\phi(t), \psi(t)) = 0
\]
Application: Learn the shortest path via solving a Bellman equation

Matrix Momentum Newton-Raphson
Matrix gain SA: \( \theta_{n+1} = \theta_n + \alpha_n \nabla_{\theta} J(\theta_n, \Phi) \)
Associated ODE: \( \frac{d}{dt} \theta(t) = -\nabla_{\theta} J(\theta(t), \Phi) \)

Zap SNR: Choose \( \{\Phi_i\} \) such that \( J(\Phi_i) = -|I + A(\theta_i)^T A(\theta_i)|^{-1} A(\theta_i)^T f(\theta_i) \)

Associated ODEs:
\[
\frac{d}{dt} \theta(t) = -|I + A(\theta)^T A(\theta)|^{-1} A(\theta)^T f(\theta)
\]
CLT covariance: \( \Sigma_0 = \Sigma_n = |(\lambda I + A(\theta)|^{-1} \Sigma_n |(\lambda I + A(\theta)|^{-1} \Sigma_n = 0 \)
when \( \varepsilon = 0 \). It is optimal: \( \Sigma^* = \Sigma_n^* \) for any other \( \varphi \)

Main Results
• PoSA couples with \( \varepsilon = 0 \) Zap SNR at rate \( O(1/n^2) \)
  \( \sup_{t \geq 0} \text{sup} E[(\theta(t) - \theta^*)(\theta(t) - \theta^*)^\top] = \infty \)
  \( \lim \sup E[(\theta(t) - \theta^*)^2] = \infty \)
• Implies optimal CLT covariance for PoSA
• Expressions for CLT covariance of NeSA is also obtained: finite, but not optimal

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

References