Quasi Stochastic Approximation
American Control Conference
San Francisco, June 2011

Sean P. Meyn

Joint work with Darshan Shirodkar and Prashant Mehta

Coordinated Science Laboratory
and the Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, USA

Thanks to NSF & AFOSR
Outline

1. Background: Stochastic Approximation
2. Background: Q-learning
3. Quasi-Stochastic Approximation
4. Conclusions
Stochastic Approximation

Setting: Solve the equation $\bar{h}(\vartheta) = 0$, with

$$\bar{h}(\vartheta) = \mathbb{E}[h(\vartheta, \zeta)], \quad \text{where } \zeta \text{ is random.}$$

Robbins and Monro\textsuperscript{[5]}: Fixed point iteration with noisy measurements,

$$\vartheta_{n+1} = \vartheta_n + a_n h(\vartheta_n, \zeta_n), \quad n \geq 0, \quad \vartheta_0 \in \mathbb{R}^d \text{ given.}$$
Stochastic Approximation

Setting: Solve the equation $\overline{h}(\vartheta) = 0$, with

$$\overline{h}(\vartheta) = \mathbb{E}[h(\vartheta, \zeta)],$$
where $\zeta$ is random.

Robbins and Monro\[5]\: Fixed point iteration with noisy measurements,

$$\vartheta_{n+1} = \vartheta_n + a_n h(\vartheta_n, \zeta_n), \quad n \geq 0, \quad \vartheta_0 \in \mathbb{R}^d \text{ given.}$$

Typical assumptions for convergence:

- Step size $a_n = (1 + n)^{-1}$
- Random sequence $\zeta_n$ is identical to $\zeta$ in distribution and i.i.d..
- Global stability of ODE $\dot{\theta} = \overline{h}(\theta)$.

Excellent recent reference: Borkar 2008\[2].
Background: Stochastic Approximation

Linearization of Stochastic Approximation

Assuming convergence, write

$$\tilde{\vartheta}_{n+1} \approx a_n [A\tilde{\vartheta}_n + Z_n]$$

with $Z_n = h(\vartheta^*, \zeta_n)$ (zero mean), and $A = \nabla \bar{h}(\vartheta^*)$. 
Background: Stochastic Approximation

Variance

Linearization of Stochastic Approximation

Assuming convergence, write
\[ \tilde{\vartheta}_{n+1} \approx a_n [A \tilde{\vartheta}_n + Z_n] \]
with \( Z_n = h(\vartheta^*, \zeta_n) \) (zero mean), and \( A = \nabla \bar{h}(\vartheta^*) \).

Central Limit Theorem: \( \sqrt{n} \tilde{\vartheta}_n \sim N(0, \Sigma_{\vartheta}) \)
Background: Stochastic Approximation

Variance

Linearization of Stochastic Approximation

Assuming convergence, write

\[ \tilde{\theta}_{n+1} \approx a_n \left[ A \tilde{\theta}_n + Z_n \right] \]

with \( Z_n = h(\vartheta^*, \zeta_n) \) (zero mean), and \( A = \nabla \bar{h}(\vartheta^*) \).

Central Limit Theorem: \( \sqrt{n} \tilde{\theta}_n \sim \mathcal{N}(0, \Sigma_\vartheta) \)

Holds under mild conditions.

Lyapunov equation for variance, requires \( \text{eig}(A) < -\frac{1}{2} \):

\[
(A + \frac{1}{2}I)\Sigma_\vartheta + \Sigma_\vartheta(A + \frac{1}{2}I)^T + \Sigma_Z = 0
\]
Corollary to CLT

Question: What is the optimal matrix gain $\Gamma_n$?

$$\vartheta_{n+1} = \vartheta_n + a_n \Gamma_n h(\vartheta_n, \zeta_n), \quad n \geq 0, \quad \vartheta_0 \in \mathbb{R}^d \text{ given.}$$
Corollary to CLT

**Question**: What is the optimal matrix gain $\Gamma_n$?

$$\vartheta_{n+1} = \vartheta_n + a_n \Gamma_n h(\vartheta_n, \zeta_n), \quad n \geq 0, \quad \vartheta_0 \in \mathbb{R}^d \text{ given.}$$

**Answer**: Stochastic Newton-Raphson, so that $\Gamma_n^* \rightarrow -A^{-1}$. 
Background: Stochastic Approximation

Stochastic Newton Raphson

Corollary to CLT

**Question:** What is the optimal matrix gain $\Gamma_n$?

$$\vartheta_{n+1} = \vartheta_n + a_n \Gamma_n h(\vartheta_n, \zeta_n), \quad n \geq 0, \quad \vartheta_0 \in \mathbb{R}^d \text{ given.}$$

**Answer:** Stochastic Newton-Raphson, so that $\Gamma_n^* \rightarrow -A^{-1}$.

**Proof:** Linearization reduces to standard Monte-Carlo
Example: Root finding

\[ h(\vartheta, \zeta) = 1 - \tan(\vartheta) + \zeta. \]

\( \zeta \), normal random variable \( N(0, 9) \)

\[ h(\vartheta, \zeta) = 1 - \tan(\vartheta) + \zeta \]

\[ \bar{h}(\vartheta) = 1 - \tan(\vartheta) \quad \implies \vartheta^* = \pi/4. \]
Example: Root finding

\[ h(\vartheta, \zeta) = 1 - \tan(\vartheta) + \zeta. \quad \zeta, \text{ normal random variable } N(0, 9) \]

\[
\begin{align*}
  h(\vartheta, \zeta) &= 1 - \tan(\vartheta) + \zeta \\
  \overline{h}(\vartheta) &= 1 - \tan(\vartheta) \quad \implies \vartheta^{*} = \pi/4.
\end{align*}
\]

Estimates of \( \vartheta^{*} \) using i.i.d. and deterministic sequences (each \( \sim \zeta \)): 

\[
\begin{align*}
  \zeta &\sim \text{i.i.d. Gaussian} \\
  \zeta &= \phi^{-1}(\Delta) \quad \text{deterministic}
\end{align*}
\]
Example: Root finding

\[ h(\vartheta, \zeta) = 1 - \tan(\vartheta) + \zeta. \]

\( \zeta \), normal random variable \( N(0, 9) \)

\[ h(\vartheta, \zeta) = 1 - \tan(\vartheta) + \zeta \\
\overline{h}(\vartheta) = 1 - \tan(\vartheta) \quad \implies \quad \vartheta^* = \pi/4. \]

Estimates of \( \vartheta^* \) using i.i.d. and deterministic sequences (each \( \sim \zeta \)):

Analysis requires theory of quasi-stochastic approximation...
Background: Q-learning

M&M 2009 system: $\dot{x} = f(x, u)$  cost: $c(x, u)$

HJB equation for discounted-cost optimal control problem,

$$\min_u \left\{ c(x, u) + f(x, u) \cdot \nabla h(x) \right\} = \gamma h(x)$$

---

Watkins & Dayan, Q-learning, 1992; Bertsekas & Tsitsiklis, NDP, 1996; Mehta & Meyn, CDG 2009
Background: Q-learning

M&M 2009  

system: $\dot{x} = f(x, u)$  
cost: $c(x, u)$

HJB equation for discounted-cost optimal control problem,

$$
\min_u \{ c(x, u) + f(x, u) \cdot \nabla h(x) \} = \gamma h(x)
$$

Parameterization set of approximations,  

$\{ Q_\theta(x, u) : \theta \in \mathbb{R}^d \}$

Bellman error:  

$E_\theta(x, u) = \gamma (Q_\theta(x, u) - c(x, u)) - f(x, u) \cdot \nabla Q_\theta(x)$

$Q(x, u) = c(x, u) + f(x, u) \cdot \nabla h(x)$

Watkins & Dayan, Q-learning, 1992; Bertsekas & Tsitsiklis, NDP, 1996; Mehta & Meyn, CDG 2009
Background: Q-learning

M&M 2009 system: \( \dot{x} = f(x, u) \)  \( \text{cost: } c(x, u) \)

HJB equation for discounted-cost optimal control problem,

\[
\min_u \{ c(x, u) + f(x, u) \cdot \nabla h(x) \} = \gamma h(x)
\]

Fixed point equation for Q-function: Writing \( Q = \min_u Q(x, u) = \gamma h(x) \),

\[
Q(x, u) = c(x, u) + f(x, u) \cdot \nabla h(x)
\]

Watkins & Dayan, Q-learning, 1992; Bertsekas & Tsitsiklis, NDP, 1996; Mehta & Meyn, CDG 2009
Background: Q-learning

M&M 2009  

system: \( \dot{x} = f(x,u) \)  
cost: \( c(x,u) \)

HJB equation for discounted-cost optimal control problem,

\[
\min_u \left\{ c(x,u) + f(x,u) \cdot \nabla h(x) \right\} = \gamma h(x)
\]

Fixed point equation for Q-function: Writing \( Q = \min_u Q(x,u) = \gamma h(x) \),

\[
Q(x,u) = c(x,u) + f(x,u) \cdot \nabla h(x) = c(x,u) + \gamma^{-1} f(x,u) \cdot \nabla Q(x)
\]
Background: Q-learning

M&M 2009 system: $\dot{x} = f(x, u)$  cost: $c(x, u)$

HJB equation for discounted-cost optimal control problem,

$$\min_u \left\{ c(x, u) + f(x, u) \cdot \nabla h(x) \right\} = \gamma h(x)$$

Fixed point equation for Q-function: Writing $Q = \min_u Q(x, u) = \gamma h(x)$,

$$Q(x, u) = c(x, u) + f(x, u) \cdot \nabla h(x) = c(x, u) + \gamma^{-1} f(x, u) \cdot \nabla Q(x)$$

Parameterization set of approximations, $\{Q^\vartheta(x, u) : \vartheta \in \mathbb{R}^d\}$

Bellman error: $\mathcal{E}^\vartheta(x, u) = \gamma(Q^\vartheta(x, u) - c(x, u)) - f(x, u) \cdot \nabla Q^\vartheta(x)$

Watkins & Dayan, Q-learning, 1992;  Bertsekas & Tsitsiklis, NDP, 1996;  Mehta & Meyn, CDG 2009
Background: Q-learning

M&M 2009 system: $\dot{x} = f(x, u) \quad \text{cost: } c(x, u)$

Parameterization, $\{Q^\vartheta(x, u) : \vartheta \in \mathbb{R}^d\}$

Bellman error: $E^\vartheta(x, u) = \gamma(Q^\vartheta(x, u) - c(x, u)) - f(x, u) \cdot \nabla Q^\vartheta(x)$
Background: Q-learning

Parameterization, \( \{ Q^\vartheta(x, u) : \vartheta \in \mathbb{R}^d \} \)

Bellman error:
\[
E^\vartheta(x, u) = \gamma(Q^\vartheta(x, u) - c(x, u)) - f(x, u) \cdot \nabla Q^\vartheta(x)
\]

Model-free form:
\[
E^\vartheta(x, u) = \gamma(Q^\vartheta(x, u) - c(x, u)) - \frac{d}{dt} Q^\vartheta(x(t)) \bigg|_{x=x(t), \ u=u(t)}
\]

M&M 2009 system:
\[
\dot{x} = f(x, u)
\]

cost:
\[
c(x, u)
\]

Watkins & Dayan, Q-learning, 1992; Bertsekas & Tsitsiklis, NDP, 1996; Mehta & Meyn, CDG 2009
Background: Q-learning

M&M 2009 system: \( \dot{x} = f(x, u) \)  cost: \( c(x, u) \)

Parameterization, \( \{Q^\theta(x, u) : \theta \in \mathbb{R}^d\} \)

Bellman error: \( \mathcal{E}^\theta(x, u) = \gamma(Q^\theta(x, u) - c(x, u)) - f(x, u) \cdot \nabla Q^\theta(x) \)

Model-free form:
\[
\mathcal{E}^\theta(x, u) = \gamma(Q^\theta(x, u) - c(x, u)) - \frac{d}{dt} Q^\theta(x(t)) \bigg|_{x=x(t), \; u=u(t)}
\]

Q-learning of M&M

- Find zeros of \( \overline{h}(\theta) = \nabla \mathbb{E}[\mathcal{E}^\theta(x, u)^2] \)
- \( \zeta = (x_\infty, u_\infty) \) ergodic steady-state.
- Choose input: stable feedback + mixture of sinusoids,
  \( u(t) = -k(x(t)) + \omega(t) \),

Watkins & Dayan, Q-learning, 1992; Bertsekas & Tsitsiklis, NDP, 1996; Mehta & Meyn, CDG 2009
Example: Q-Learning

\[ \dot{x} = -x^3 + u, \quad c(x, u) = \frac{1}{2}(x^2 + u^2) \]

HJB Equation:

\[ \min_u \left[ c(x, u) + (-x^3 + u) \cdot \nabla h(x) \right] = \gamma h(x) \]
Example: Q-Learning

\[ \dot{x} = -x^3 + u, \quad c(x, u) = \frac{1}{2}(x^2 + u^2) \]

HJB Equation: \[ \min_{u} \left[ c(x, u) + (-x^3 + u) \cdot \nabla h(x) \right] = \gamma h(x) \]

Basis: \[ Q^\vartheta(x, u) = c(x, u) + \vartheta_1 x^2 + \vartheta_2 \frac{xu}{1 + 2x^2} \]
Example: Q-Learning

\[ \dot{x} = -x^3 + u, \quad c(x, u) = \frac{1}{2}(x^2 + u^2) \]

HJB Equation:  \[
\min_u \left[ c(x, u) + (-x^3 + u) \cdot \nabla h(x) \right] = \gamma h(x)
\]

Basis:  \[
Q^\vartheta(x, u) = c(x, u) + \vartheta_1 x^2 + \vartheta_2 \frac{xu}{1 + 2x^2}
\]

Control:  \[
u(t) = A[\sin(t) + \sin(\pi t) + \sin(\epsilon t)]
\]
Example: Q-Learning

\[ \dot{x} = -x^3 + u, \quad c(x, u) = \frac{1}{2} (x^2 + u^2) \]

**HJB Equation:**

\[
\min_u \left[ c(x, u) + (-x^3 + u) \cdot \nabla h(x) \right] = \gamma h(x)
\]

**Basis:**

\[ Q^\vartheta(x, u) = c(x, u) + \vartheta_1 x^2 + \vartheta_2 \frac{xu}{1 + 2x^2} \]

**Control:**

\[ u(t) = A[\sin(t) + \sin(\pi t) + \sin(et)] \]

Analysis requires theory of quasi-stochastic approximation...
Example: Q-Learning

\[ \dot{x} = -x^3 + u, \quad c(x, u) = \frac{1}{2}(x^2 + u^2) \]

HJB Equation:

\[ \min_u \left[ c(x, u) + (-x^3 + u) \cdot \nabla h(x) \right] = \gamma h(x) \]

Basis:

\[ Q^\theta(x, u) = c(x, u) + \theta_1 x^2 + \theta_2 \frac{xu}{1 + 2x^2} \]

Control:

\[ u(t) = A[\sin(t) + \sin(\pi t) + \sin(et)] \]

Analysis requires theory of quasi-stochastic approximation...
Quasi-Stochastic Approximation

Continuous time, deterministic version of SA

\[
\frac{d}{dt} \vartheta(t) = a(t) h(\vartheta(t), \zeta(t))
\]
Quasi-Stochastic Approximation

Continuous time, deterministic version of SA

Assumptions

- **Ergodicity**: $\zeta$ satisfies

  \[
  \bar{h}(\theta) = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} h(\theta, \zeta(t)) \, dt, \quad \text{for all } \theta \in \mathbb{R}^d.
  \]

- **Decreasing gain**: $a(t) \downarrow 0$, with

  \[
  \int_{0}^{\infty} a(t) \, dt = \infty, \quad \int_{0}^{\infty} a(t)^2 \, dt < \infty
  \]

    
    usually, take $a(t) = (1 + t)^{-1}$
Quasi-Stochastic Approximation

Continuous time, deterministic version of SA

Assumptions

- **Ergodicity**: $\zeta$ satisfies

$$\bar{h}(\theta) = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} h(\theta, \zeta(t)) \, dt, \quad \text{for all } \theta \in \mathbb{R}^d.$$

- **Decreasing gain**: $a(t) \downarrow 0$, with

$$\int_{0}^{\infty} a(t) \, dt = \infty, \quad \int_{0}^{\infty} a(t)^2 \, dt < \infty$$

usually, take $a(t) = (1 + t)^{-1}$

- **Stable ODE**: $\bar{h}$ is globally Lipschitz, and $\dot{\vartheta} = \bar{h}(\vartheta)$ is globally asymptotically stable, with globally Lipschitz Lyapunov function
Quasi-Stochastic Approximation

Stability & Convergence

Assumptions

- **Ergodicity**: $\zeta$ satisfies

  $$\bar{h}(\theta) = \lim_{T \to \infty} \frac{1}{T} \int_0^T h(\theta, \zeta(t)) \, dt, \text{ for all } \theta \in \mathbb{R}^d.$$ 

- **Decreasing gain**: $a(t) \downarrow 0$, with

  $$\int_0^\infty a(t) \, dt = \infty, \quad \int_0^\infty a(t)^2 \, dt < \infty$$

- **Stable ODE**: $\bar{h}$ is globally Lipschitz, and $\dot{\vartheta} = \bar{h}(\vartheta)$ is globally asymptotically stable, with globally Lipschitz Lyapunov function

**Theorem**: $\vartheta(t) \to \vartheta^*$ for any initial condition.
Quasi-Stochastic Approximation

Variance

- \( a(t) = 1/(1 + t) \)
- The model is linear: \( h(\theta, \zeta) = A\theta + \zeta \), and each \( \lambda(A) \) satisfies \( \text{Re}(\lambda) < -1 \).
- \( \zeta \) has zero mean: \( \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} \zeta(t) \, dt = 0 \), at rate \( 1/T \).
Quasi-Stochastic Approximation

Variance

- $a(t) = 1/(1 + t)$
- The model is linear: $h(\theta, \zeta) = A\theta + \zeta$, and each $\lambda(A)$ satisfies $\text{Re}(\lambda) < -1$.
- $\zeta$ has zero mean: \[
    \lim_{T \to \infty} \frac{1}{T} \int_0^T \zeta(t) \, dt = 0, \text{ at rate } 1/T.
\]

Rate of convergence is $t^{-1}$, and not $t^{-1/2}$, under these assumptions:
Quasi-Stochastic Approximation

Variance

- $a(t) = 1/(1 + t)$
- The model is linear: $h(\theta, \zeta) = A\theta + \zeta$, and each $\lambda(A)$ satisfies $\text{Re}(\lambda) < -1$.
- $\zeta$ has zero mean: $\lim_{T \to \infty} \frac{1}{T} \int_0^T \zeta(t) \, dt = 0$, at rate $1/T$.

Rate of convergence is $t^{-1}$, and not $t^{-\frac{1}{2}}$, under these assumptions:

**Theorem:** For some constant $\bar{\sigma} < \infty$,

$$\limsup_{t \to \infty} t \| \vartheta(t) - \vartheta^* \| \leq \bar{\sigma}$$
Polyak and Juditsky

Polyak and Juditsky\cite{4,2} obtain \textit{optimal variance} for SA using a very simple approach: \textit{High gain, and averaging.}
Polyak and Juditsky

Polyak and Juditsky\cite{polyak1992stochastic, polyak1995stochastic} obtain optimal variance for SA using a very simple approach: \textit{High gain, and averaging}.  

Deterministic P&J: Choose \textit{high-gain}, $a(t) = 1/(1 + t)^\delta$, with $\delta \in (0, 1)$.  

$$\frac{d}{dt} \gamma(t) = \frac{1}{(1 + t)^\delta} h(\gamma(t), \zeta(t))$$
Polyak and Juditsky

Polyak and Juditsky\cite{4,2} obtain \textit{optimal variance} for SA using a very simple approach: \textit{High gain, and averaging}.

**Deterministic P\&J:** Choose \textit{high-gain}, \(a(t) = 1/(1 + t)^\delta\), with \(\delta \in (0, 1)\).

\[
\frac{d}{dt} \gamma(t) = \frac{1}{(1 + t)^\delta} h(\gamma(t), \zeta(t)) \]

The output of this algorithm is then averaged:

\[
\frac{d}{dt} \vartheta(t) = \frac{1}{1 + t} (-\vartheta(t) + \gamma(t)) \]
Polyak and Juditsky

Polyak and Juditsky\cite{4,2} obtain optimal variance for SA using a very simple approach: High gain, and averaging.

Deterministic P&J: Choose high-gain, $a(t) = 1/(1 + t)^\delta$, with $\delta \in (0, 1)$.

$$ \frac{d}{dt} \gamma(t) = \frac{1}{(1 + t)^\delta} h(\gamma(t), \zeta(t)) $$

The output of this algorithm is then averaged:

$$ \frac{d}{dt} \vartheta(t) = \frac{1}{1 + t} (-\vartheta(t) + \gamma(t)) $$

Linear convergence under stability alone

There is a finite constant $\bar{\sigma}$ satisfying,

$$ \limsup_{t \to \infty} t \| \vartheta(t) - \vartheta^* \| \leq \bar{\sigma} $$
Polyak and Juditsky

Polyak and Juditsky\cite{4, 2} obtain optimal variance for SA using a very simple approach: High gain, and averaging.

Deterministic P&J: Choose high-gain, \( a(t) = 1/(1 + t)^\delta \), with \( \delta \in (0, 1) \).

\[
\frac{d}{dt} \gamma(t) = \frac{1}{(1 + t)^\delta} h(\gamma(t), \zeta(t))
\]

The output of this algorithm is then averaged:

\[
\frac{d}{dt} \vartheta(t) = \frac{1}{1 + t} \left( -\vartheta(t) + \gamma(t) \right)
\]

Linear convergence under stability alone

1. There is a finite constant \( \bar{\sigma} \) satisfying,

\[
\limsup_{t \to \infty} t \| \vartheta(t) - \vartheta^* \| \leq \bar{\sigma}
\]

2. Question: Is \( \bar{\sigma} \) minimal?
Conclusions

Summary:

- QSA is the most natural approach to approximate dynamic programming for deterministic systems, such as Q-learning\cite{3}.
- Stability is established using standard techniques.
- Linear convergence is obtained under mild stability assumptions.
Conclusions

Summary:

- QSA is the most natural approach to approximate dynamic programming for deterministic systems, such as Q-learning\cite{3}.
- Stability is established using standard techniques
- Linear convergence is obtained under mild stability assumptions.

Current research:

1. Open question: Can we extend the approach of Polyak and Juditsky to obtain optimal convergence?
Conclusions

Summary:

- QSA is the most natural approach to approximate dynamic programming for deterministic systems, such as Q-learning\cite{3}.
- Stability is established using standard techniques.
- Linear convergence is obtained under mild stability assumptions.

Current research:

1. Open question: Can we extend the approach of Polyak and Juditsky to obtain optimal convergence?
2. First we must answer, what is the optimal value of $\bar{\sigma}$?
Conclusions

Summary:

- QSA is the most natural approach to approximate dynamic programming for deterministic systems, such as Q-learning\(^3\).
- Stability is established using standard techniques.
- Linear convergence is obtained under mild stability assumptions.

Current research:

1. Open question: Can we extend the approach of Polyak and Juditsky to obtain optimal convergence?
2. First we must answer, what is the optimal value of \(\bar{\sigma}\)?
3. Concentration on applications:
   - Further development of Q-learning
   - Applications to nonlinear filtering.
References

D.P. Bertsekas and J. N. Tsitsiklis.
Neuro-Dynamic Programming.

V. S. Borkar.
Stochastic Approximation: A Dynamical Systems Viewpoint.

P. G. Mehta and S. P. Meyn.
Q-learning and Pontryagin’s minimum principle.

B. T. Polyak and A. B. Juditsky.
Acceleration of stochastic approximation by averaging.

H. Robbins and S. Monro.
A stochastic approximation method.

Q-learning.