Concentration Bound for Stochastic Approximation with Markov Noise

Siddharth Chandak, 17D070019
Guide - Prof. Vivek S. Borkar
12^{th} May 2021

Electrical Engineering
IIT Bombay

Outline

- Introduction
    - An Example: Q-Learning
- Main Result
    - Setup
- Proof (Outline)
- Application: Asynchronous Q-Learning

Introduction

Stochastic Approximation

- Method to solve $h(x) = 0$ given noisy measurements of $h(\cdot)$
- Basic form:

$$x_{n+1} = x_n + a(n)\big(h(x_n) + M_{n+1}(x_n)\big), n \geq 0,$$

- Has applications in:
  - Reinforcement learning algorithms (will see soon)
  - Stochastic Gradient Descent

Fixed Point Schemes

- Contraction: $\|F(x-y)\| \leq \alpha\|(x-y)\|$ where $\alpha \in (0,1)$
- Fixed Point: $F(x^*) = x^*$
- Iteration:

$$x_{n+1} = x_n + a(n)\big(F(x_n) - x_n + M_{n+1}(x_n)\big), n \geq 0,$$

- Almost sure convergence[1] to x^*

[1] Under appropriate conditions

With Markov Noise

- Y_n: irreducible, finite state space Markov chain
- Iteration:

$$x_{n+1} = x_n + a(n)\big(F(x_n, Y_n) - x_n + M_{n+1}(x_n)\big), n \geq 0, \qquad (1)$$

- Contraction:

$$\|\sum_{i\in S} \pi(i)(F(x,i) - F(z,i))\| \leq \alpha\|x - z\|$$

- Fixed Point: $\sum_i \pi(i)F(x^*, i) = x^*$
- Almost sure convergence[2] of iterates x_n to x^*

[2] Under appropriate conditions

Our Work

- 'High probability' concentration bound
- $\|x_n - x^*\| \leq$ ____ for all $n \geq n_0$
  with probability exceeding $1-$____
- Extension of previous work[3] which considers contractive stochastic approximation

[3] V. S. Borkar, "A concentration bound for contractive stochastic approximation", *Systems and Control Letters*

Example: A very brief introduction to Q-learning

- Consider finite state space S and finite action space A
- At each time step n, agent chooses action $Z_n \in A$ when it is in state $X_n \in S$
- Markov control policy:

$$P(X_{n+1} = j | X_m, Z_m, m \leq n) = p(j | X_n, Z_n) \; \forall n,$$

- Objective: Minimize

$$E\left[\sum_{m=0}^{\infty} \gamma^m k(X_m, Z_m)\right]$$

Example: A very brief introduction to Q-learning

- Q-Learning Algorithm:

$$Q_{n+1}(i,u) = Q_n(i,u) + a(n) I\{X_n = i, Z_n = u\} \\ \times \Big(k(i,u) + \gamma \min_a Q_n(X_{n+1}, a) - Q_n(i,u) \Big)$$

- $Q_n \to Q^*$[4] where Q^* is a solution of

$$Q(i,u) = k(i,u) + \alpha \sum_j p(j|i,u) \min_a Q(j,a),$$

[4] Under appropriate conditions

Asynchronous vs Synchronous

- Synchronous - no Markov noise:

$$Q_{n+1}(i,u) = Q_n(i,u) + a(n) \times \left(k(i,u) + \gamma \min_a Q_n(Y_{n+1}(i,u),a) - Q_n(i,u)\right)$$

- Asynchronous - has Markov noise:

$$Q_{n+1}(i,u) = Q_n(i,u) + a(n) I\{X_n = i, Z_n = u\} \times \left(k(i,u) + \gamma \min_a Q_n(X_{n+1},a) - Q_n(i,u)\right)$$

- Can be rewritten in the form of (1) - gives us probabilistic bounds on $\|Q_n - Q^*\|$

Main Result

Setup

- For $x_n = [x_n(1), ..., x_n(d)]^T \in \mathcal{R}^d$,

$$x_{n+1} = x_n + a(n)\big(F(x_n, Y_n) - x_n + M_{n+1}(x_n)\big), n \geq 0,$$

- Y_n - irreducible Markov chain taking values in a finite state space S

$$\begin{aligned} P(Y_{n+1} = j | Y_n = i_n, ..., Y_0 = i_0) &= P(Y_{n+1} = j | Y_n = i_n) \\ &= p(j|i_n), i_0, ..., i_n, j \in S \end{aligned}$$

- With stationary Distribution - $\pi(\cdot)$

Setup

$$x_{n+1} = x_n + a(n)\big(F(x_n, Y_n) - x_n + M_{n+1}(x_n)\big), n \geq 0,$$

- $\{M_n(x)\}$ - martingale difference sequence w.r.t. $\mathcal{F}_n \coloneqq \sigma(x_0, M_m(x), x \in \mathcal{R}^d, m \leq n)$, $n \geq 0$
- $E[M_{n+1}(x)|\mathcal{F}_n] = \theta$ a.s. $\forall x, n$
- $|M_n^l(x)| \leq K_0(1 + \|x\|)$ a.s., for some $K_0 > 0$

Setup

$$x_{n+1} = x_n + a(n)\big(F(x_n, Y_n) - x_n + M_{n+1}(x_n)\big), n \geq 0,$$

- Contraction:

$$\|\sum_{i \in S} \pi(i)(F(x,i) - F(z,i))\| \leq \alpha\|x - z\|, x, z \in \mathcal{R}^d$$

- $\widetilde{F}_n(x,i) := F(x,i) + M_{n+1}(x)$ satisfies

$$\|\widetilde{F}_n(x,i)\| \leq K + \alpha\|x\| \text{ a.s.}$$

Setup

$$x_{n+1} = x_n + a(n)\big(F(x_n, Y_n) - x_n + M_{n+1}(x_n)\big), n \geq 0,$$

- $a(n)$ - Non-negative stepsizes

$$\sum_n a(n) = \infty, \sum_n a(n)^2 < \infty$$

- Eventually non-increasing, i.e., there exists $n^* \geq 1$ such that $a(n+1) \leq a(n), \forall n \geq n^*$

Some Definitions

$$b_m(n) := \sum_{k=m}^{n} a(k), 0 \leq m \leq n < \infty$$

$$\beta(n) := \sup_{m \geq n} (1 - a(m+1))a(m)$$

$$\varphi(n) := \sup_{m \geq n} e^{a(m)}$$

$$\kappa(d) = \|\mathbb{1}\|$$

Theorem

Theorem 2.1

Let $n_0 \geq 0$ *satisfy* $\varphi(n_0) \leq \frac{1}{\alpha}$, $a(n_0) < 1$ *and* $a(n)$ *is non-increasing after* n_0*. Then there exist finite, positive constants* c_1, c_2 *and* D *such that for* $\delta > 0$ *and* $n \geq n_0$,

$$\|x_n - x^*\| \leq e^{-(1-\alpha)b_{n_0}(n)}\|x_{n_0} - x^*\| + \frac{\delta + (4a(n_0) + 2\varphi(n_0))c_1}{1 - \alpha\varphi(n_0)}$$

with probability exceeding

$$\begin{aligned} &1 \; - \; 2d(n-n_0)e^{-D\delta^2/\beta(n)}, \; 0 < \delta \leq C\varphi(n_0), \\ &1 \; - \; 2d(n-n_0)e^{-D\delta/\beta(n)}, \quad \delta > C\varphi(n_0), \end{aligned}$$

where $C = e^{\kappa(d)(K_0(1+\|x_{n_0}\|+\frac{K}{1-\alpha})+c_2)}$.

Theorem (cont.)

Theorem 2.2

Let $n_0 \geq 0$ *satisfy* $\varphi(n_0) \leq \frac{1}{\alpha}$, $a(n_0) < 1$ *and* $a(n)$ *is non-increasing after* n_0. *Then there exist finite, positive constants* c_1, c_2 *and* D *such that for* $\delta > 0$,

$$\|x_n - x^*\| \leq e^{-(1-\alpha)b_{n_0}(n)}\|x_{n_0} - x^*\| + \frac{\delta + (4a(n_0) + 2\varphi(n_0))c_1}{1 - \alpha\varphi(n_0)} \; \forall n \geq n_0,$$

with probability exceeding

$$\begin{aligned} &1 \; - \; 2d \sum_{n \geq n_0} (n - n_0) e^{-D\delta^2/\beta(n)}, \; 0 < \delta \leq C\varphi(n_0), \\ &1 \; - \; 2d \sum_{n \geq n_0} (n - n_0) e^{-D\delta/\beta(n)}, \quad \delta > C\varphi(n_0). \end{aligned}$$

Proof (Outline)

An Important Lemma

Lemma 3.1

$\sup_n \|x_n\| \leq \|x_{n_0}\| + \frac{K}{1-\alpha}$ *a.s. for* $n \geq n_0$

Proof (Outline): We use the fact that

$$\|\widetilde{F}_n(x,i)\| \leq K + \alpha\|x\| \text{ a.s..}$$

and then proceed inductively. □

Proof of the Main Result

Proof (Outline):

- Define z_n for $n \geq n_0$ by:

$$z_{n+1} = z_n + a(n)(\sum_i \pi(i)F(z_n, i) - z_n), \qquad (2)$$

  where $z_{n_0} = x_{n_0}$.

- $\|x_n - x^*\| \leq \|x_n - z_n\| + \|z_n - x^*\|$.

Some Manipulation

Proof (cont.):

- With some manipulation:

$$\begin{aligned} x_{n+1} - z_{n+1} &= (1 - a(n))(x_n - z_n) \\ &+ a(n)M_{n+1} \\ &+ a(n)(\sum_i \pi(i)(F(x_n, i) - F(z_n, i))) \\ &+ a(n)(F(x_n, Y_n) - \sum_i \pi(i)F(x_n, i)). \end{aligned}$$

Further Manipulation

Proof (cont.):

- For $n, m \geq 0$, let $\phi(n,m) = \prod_{k=m}^{n}(1 - a(k))$ if $n \geq m$ and 1 otherwise. For some $n \geq n_0$, we iterate the above for $n_0 \leq m \leq n$,

$$
\begin{aligned}
x_{m+1} - z_{m+1} \quad = \quad & \sum_{k=n_0}^{m} \phi(m, k+1) a(k) M_{k+1} \\
+ \quad & \sum_{k=n_0}^{m} \phi(m, k+1) a(k) \Big(\sum_i \pi(i) (F(x_k, i) - F(z_k, i))\Big) \\
+ \quad & \sum_{k=n_0}^{m} \phi(m, k+1) a(k) \Big(F(x_k, Y_k) - \sum_i \pi(i) F(x_k, i)\Big).
\end{aligned}
$$

Proof (cont.):

- Poisson Equation:

$$V(x,i) = F(x,i) - \sum_j \pi(j)F(x,j) + \sum_j p(j|i)V(x,j). \quad (3)$$

- A possible solution:

$$V_1(x,i) = E_i[\sum_{m=0}^{\tau-1}(F(x,Y_m) - \sum_j \pi(j)F(x,j))], i \in S$$

- $V_{max} = \max_{x,i} \|V(x,i)\|$
- $V'_{max} = \max_{x,i,l} \|V^l(x,i)\|$

Proof (cont.):

$$\sum_{k=n_0}^{m} \phi(m, k+1)a(k)(F(x_k, Y_k) - \sum_i \pi(i)F(x_k, i))$$

$$= \sum_{k=n_0+1}^{m} \phi(m, k+1)a(k)(V(x_k, Y_k) - \sum_j p(j|Y_{k-1})V(x_k, j))$$

$$+ \sum_{k=n_0+1}^{m} ((\phi(m, k+1)a(k) - \phi(m, k)a(k-1)) \sum_j p(j|Y_{k-1})V(x_k, j))$$

$$+ \sum_{k=n_0+1}^{m} \phi(m, k)a(k-1)(\sum_j p(j|Y_{k-1})(V(x_k, j) - V(x_{k-1}, j)))$$

$$+ \phi(m, n_0+1)a(n_0)V(x_{n_0}, Y_{n_0}) - \phi(m, m+1)a(m) \sum_j p(j|Y_m)V(x_m, j)$$

Some Final Manipulations

Proof (cont.):

- Define

$$\zeta_m = \kappa(d) \max_l \max_{n_0 \le k \le m} |\sum_{r=n_0}^{k-1} \phi(k, r+1) a(r)(M_{r+1}^l(x_r) + V_r'^l(x_r))|$$

- Define $x'_m = \sup_{n_0 \le k \le m} \|x_k - z_k\|$
- Then

$$x'_{m+1} \le \alpha\varphi(n_0)x'_m + \zeta_n + V_c(n_0)$$

- And finally,

$$x'_m \le \frac{1}{1-\alpha\varphi(n_0)}(\zeta_n + V_c(n_0)), n_0 \le m \le n \qquad (4)$$

Using A Scalar Martingale Inequality[5]

Proof (cont.):

- Then for a suitable constant $D > 0$ and $\delta \in (0, C\gamma_1]$, we have

$$P(\zeta_n \geq \delta) \leq 2d(n - n_0)e^{-D\delta^2/\beta(n)} \tag{5}$$

  and for $\delta > C\gamma_1$,

$$P(\zeta_n \geq \delta) \leq 2d(n - n_0)e^{-D\delta/\beta(n)}. \tag{6}$$

- $C = e^{\kappa(d)(K_0(1+\|x_{n_0}\|+\frac{K}{1-\alpha})+2V'_{max})}$
- $\gamma_1 = \sup_{n\geq n_0} \varphi(n) = \varphi(n_0)$

[5] Appendix

The Other Term

Proof (cont.):

$$z_{n+1} - x^* = (1 - a(n))(z_n - x^*) + a(n) \sum_i \pi(i)(F(z_n, i) - F(x^*, i)),$$

$$\begin{aligned} \|z_{n+1} - x^*\| &\leq (1 - (1 - \alpha)a(n))\|z_n - x^*\| \\ &\leq e^{-(1-\alpha)b_{n_0}(n)}\|x_{n_0} - x^*\| \end{aligned} \tag{7}$$

The End

Proof (cont.):

- Combine (7) with (4) and use the fact that $\zeta_n < \delta$ holds with probabilities given by (5) and (6).

□

Asynchronous Q-Learning

A Brief Introduction (Again)

- Consider finite state space S and finite action space A
- At each time step n, agent chooses action $Z_n \in A$ when it is in state $X_n \in S$
- Markov control policy:

$$P(X_{n+1} = j | X_m, Z_m, m \leq n) = p(j | X_n, Z_n) \ \forall n,$$

- Objective: Minimize

$$E\left[\sum_{m=0}^{\infty} \gamma^m k(X_m, Z_m)\right]$$

A Brief Introduction (Again)

- Q-Learning Algorithm:

$$Q_{n+1}(i,u) = Q_n(i,u) + a(n)I\{X_n = i, Z_n = u\} \\ \times \Big(k(i,u) + \gamma \min_a Q_n(X_{n+1}, a) - Q_n(i,u)\Big)$$

- $Q_n \to Q^*$[6] where Q^* is a solution of

$$Q(i,u) = k(i,u) + \alpha \sum_j p(j|i,u) \min_a Q(j,a),$$

[6] Under appropriate conditions

Applying our Result

- Assume off-line simulation with a fixed randomized stationary policy. For application of our theorem
- (X_n, Z_n) together forms the Markov chain with the transition probabilities as:

$$P(j, v|i, u) = p(j|i, u)\Phi(v|j)$$

- $\Phi(v|j)$ is the randomized policy
- Stationary distribution for this Markov chain - $\pi(i, u) = \pi_\Phi(i)\Phi(u|i)$
- Under the norm $\|\cdot\|_\infty$

Applying our Result

- Can be rewritten as:

$$Q_{n+1}(i,u) = Q_n(i,u) + a(n)\Big(F^{(i,u)}(Q_n, Y_n) - Q_n(i,u) + M^{(i,u)}_{n+1}(Q_n)\Big)$$

- where

$$F^{i,u}(Q,X,Y) = I\{X = i, Z = u\}\times \Big(k(i,u) + \gamma \sum_j p(j|i,u) \min_a Q(j,a) - Q(i,u)\Big) + Q(i,u)$$

- and

$$M^{i,u}_{n+1}(Q) = \gamma I\{X_n = i, Z_n = u\}\times (\min_a Q(X_{n+1},a) - \sum_j p(j|i,u) \min_a Q(j,a))\Big).$$

Applying our Result

- Most assumptions of the theorem can be easily verified
- The map $\sum_{i,u} \pi(i,u) F(\cdot, i, u)$ is a contraction with $\alpha = (1 - (1 - \gamma)\pi_{min})$
- The result can be used on iterates Q_n.

Applying our Result

Corollary 4.1

Let $n_0 \geq 0$ satisfy $\varphi(n_0) \leq \frac{1}{\alpha}$, $a(n_0) < 1$ and $a(n)$ is non-increasing after n_0. Then there exist finite positive constants c_1, c_2 and D such that for $\delta > 0$ and $n \geq n_0$,

$$\|Q_n - Q^*\| \leq e^{-(1-\alpha)b_{n_0}(n)}\|Q_{n_0} - Q^*\| + \frac{\delta + (4a(n_0) + 2\varphi(n_0))c_1}{1 - \alpha\varphi(n_0)}$$

with probability exceeding

$$\begin{aligned} &1 \; - \; 2d(n - n_0)e^{-D\delta^2/\beta(n)}, \; 0 < \delta \leq C\varphi(n_0), \\ &1 \; - \; 2d(n - n_0)e^{-D\delta/\beta(n)}, \quad \delta > C\varphi(n_0), \end{aligned}$$

where $C = e^{\kappa(d)(2(1+\|Q_{n_0}\|_\infty + \frac{\|k\|_\infty}{1-\alpha}) + c_2)}$.

Applying our Result

Corollary 4.2

Let $n_0 \geq 0$ *satisfy* $\varphi(n_0) \leq \frac{1}{\alpha}$, $a(n_0) < 1$ *and* $a(n)$ *is non-increasing after* n_0. *Then there exist finite positive constants* c_1, c_2 *and* D *such that for* $\delta > 0$ *and for all* $n \geq n_0$,

$$\|Q_n - Q^*\| \leq e^{-(1-\alpha)b_{n_0}(n)}\|Q_{n_0} - Q^*\| + \frac{\delta + (4a(n_0) + 2\varphi(n_0))c_1}{1 - \alpha\varphi(n_0)},$$

with probability exceeding

$$1 \;-\; 2d \sum_{n \geq n_0} (n - n_0)e^{-D\delta^2/\beta(n)}, \; 0 < \delta \leq C\varphi(n_0),$$

$$1 \;-\; 2d \sum_{n \geq n_0} (n - n_0)e^{-D\delta/\beta(n)}, \quad \delta > C\varphi(n_0).$$

Thank You!

Appendix: A Concentration Inequality

Let $\{M_n\}$ be a real valued martingale difference sequence with respect to an increasing family of σ-fields $\{\mathcal{F}_n\}$. Assume that there exist $\varepsilon, C > 0$ such that

$$E\left[e^{\varepsilon|M_n|}\middle|\mathcal{F}_{n-1}\right] \leq C \quad \forall\ n \geq 1, \text{a.s.}$$

Let $S_n := \sum_{m=1}^{n} \xi_{m,n} M_m$, where $\xi_{m,n},\ m \leq n,$, for each n, are a.s. bounded $\{\mathcal{F}_n\}$-previsible random variables, i.e., $\xi_{m,n}$ is $\mathcal{F}_{m-1}$-measurable $\forall m \geq 1$, and $|\xi_{m,n}| \leq A_{m,n}$ a.s. for some constant $A_{m,n}$, $\forall\ m, n$. Suppose

$$\sum_{m=1}^{n} A_{m,n} \leq \gamma_1, \ \max_{1\leq m\leq n} A_{m,n} \leq \gamma_2 \omega(n),$$

for some $\gamma_i, \omega(n) > 0,\ i = 1, 2; n \geq 1$. Then we have:

Theorem 5.1. *There exists a constant $D > 0$ depending on $\varepsilon, C, \gamma_1, \gamma_2$ such that for $\epsilon > 0$,*

$$P(|S_n| > \epsilon) \leq 2e^{-\frac{D\epsilon^2}{\omega(n)}}, \quad \text{if } \epsilon \in \left(0, \frac{C\gamma_1}{\varepsilon}\right], \tag{45}$$

$$2e^{-\frac{D\epsilon}{\omega(n)}}, \quad \text{otherwise.} \tag{46}$$

This is a variant of Theorem 1.1 of [22]. See [3], Theorem A.1, pp. 21-23, for details.