Finite-Time Error Bounds For Linear Stochastic Approximation and TD Learning

Chang Cao

April 15, 2019

Main Statement

Derive the finite error bounds on the moments of the error of the linear stochastic approximation algorithm:

$$\Theta_{k+1} = \Theta_k + \epsilon(A(X_k)\Theta_k + b(X_k)) \tag{1}$$

1. $\{X_k, k \geq 0\}$ is an underlying Markov chain
2. $A(X_k)$ is a random matrix; $b(X_k)$ is a random vector; Θ_k is a random vector
3. algorithm updates Θ_k using recursion (1)
4. ϵ is a constant step size

Outline

1. Motivation: $TD(0)$
2. Linear Stochastic Approximation
3. Finite-Time Error Bounds

TD Learning: TD(0)

Setup:

1. MDP over a finite space $\mathcal{S} = \{1, \ldots, N\}$
2. Fix a stationary policy μ
3. $\{Z_k\}$: the resulting Markov chain
4. Value function

$$V(i) := \mathbb{E}\left[\sum_{k=0}^{\infty} \alpha^k c(Z_k, \mu(Z_k), Z_{k+1}) \middle| Z_0 = i\right] \tag{2}$$

   where c is one-step reward.

5. Purpose: estimate the value function V associated with μ by observing a trajectory $\{z_0, z_1, z_2, \ldots\}$

TD(0): Linear Approximation

1. V satisfies the Bellman equation: $V = T_\mu V$

$$V(i) = \mathbb{E}_j[c(i,\mu(i),j) + \alpha V(j)] = \mathbb{E}[c(i,\mu(i),j)] + \alpha \sum_j p_{ij} V(j) \tag{3}$$

   denote $\overline{c} := (\mathbb{E}[c(1,\mu(i),j)], \ldots, \mathbb{E}[c(N,\mu(i),j)])^t$

2. If the transition probabilities p_{ij} are known, we can solve (3) to get V.
3. still, when $N = |\mathcal{S}|$ is large, we approximate value function V by a linear function of feature functions $\phi^t(i) = (\phi_1(i), \ldots \phi_d(i))$:

$$V(i) \cong \sum_{k=1}^{d} \theta_k \phi_k(i) \tag{4}$$

   where d is small compared to N. Now: estimate weights θ_k

TD Learning: Algorithm Design

1. Goal: approximate V by a member from $\mathcal{L} = \{\phi^t\theta : \theta \in \mathbb{R}^d\}$
2. Minimizing L^2-error

$$\theta^* = \arg\min_{\theta \in \mathbb{R}^d} \|V - \phi^t\theta\|_\xi^2 \tag{5}$$

   where

$$\|f\|_\xi^2 := \int_{\mathcal{S}} f^2(s)\xi(ds) \tag{6}$$

3. $\Pi_{\mathcal{L}} :=$ projection operator onto $\mathcal{L}$ with respect to $\|\|_\xi^2$; Solve the projected Bellman equation:

$$\Pi_{\mathcal{L}} T_\mu(\phi^t\theta) = \phi^t\theta \tag{7}$$

4. since θ^* should satisfy

$$T_\mu(\phi^t\theta^*) \cong V \tag{8}$$

TD Learning: Algorithm Design

1. one can show $\Pi_{\mathcal{L}} T_\mu$ is a contraction mapping when ξ is chosen to be the stationary distribution of $\{Z_k\}$
2. by solving (7), one can show it is equivalent to solving for θ^* so that

$$\mathbb{E}[\phi(i)(\phi(i)^t\theta^* - \alpha\phi(j)^t\theta^* - c(i, \mu(i), j)] = 0 \tag{9}$$

3. observe that

$$\theta^* - \epsilon\mathbb{E}[\phi(i)(\phi(i)^t\theta^* - \alpha\phi(j)^t\theta^* - c(i, \mu(i), j)] = \theta^* \tag{10}$$

4. for an episode $\{Z_0, Z_1, \dots\}$,

$$\Theta_{k+1} = \Theta_k - \epsilon\phi(Z_k)\left(\phi^t(Z_k)\Theta_k - c(Z_k) - \alpha\phi^t(Z_{k+1})\Theta_k\right) \tag{11}$$

where Θ_k is the estimate of θ^* at time k, $\epsilon \in (0, 1)$ is a constant

TD(0): Convergence

Theorem (Tsitsiklis, Van Roy 1997)

Θ_k converges to θ^* where

$$\Pi_{\mathcal{L}} T_\mu(\phi^t \theta^*) = \phi^t \theta^* \tag{12}$$

Srikant and Ying 2019 provides finite-time error bounds on $\mathbb{E}\|\Theta_k - \theta^*\|^2$. Rewrite (11) as

$$\Theta_{k+1} = \Theta_k + \epsilon(A(X_k)\Theta_k + b(X_k)) \tag{13}$$

where

$$X_k := (Z_k, Z_{k+1}), \quad A(X_k) := -\phi(Z_k)(\phi^t(Z_k) - \alpha\phi^t(Z_{k+1})) \tag{14}$$

and

$$b(X_k) := c(Z_k)\phi(Z_k) - A(X_k)\theta^*, \quad \Theta \leftarrow \Theta - \theta^* \tag{15}$$

Assumptions

From now on, we focus on linear stochastic recursion (1). We use 2-norm for all vectors and induced 2-norm for all matrices.
Assumptions:

1. $\{X_k\}$ is a Markov chain with state space $\mathcal{S}$.

$$\lim_{k\to\infty} \mathbb{E}[A(X_k)] = \overline{A}, \quad \lim_{k\to\infty} \mathbb{E}[b(X_k)] = 0 \tag{16}$$

   For mixing time τ_ϵ of $\{X_k\}$ so that for all i and $k \geq \tau_\epsilon$

$$\|\mathbb{E}[b(X_k)\big|X_0 = i]\| \leq \epsilon, \|\mathbb{E}[A(X_k)\big|X_0 = i] - \overline{A}\| \leq \epsilon, \tag{17}$$

   there exists $K \geq 1$ so that $\tau_\epsilon \leq K \log \frac{1}{\epsilon}$.

2. Assumption 2:

$$b_{max} := \sup_{i\in\mathcal{S}} \|b(i)\| < \infty, \quad A_{max} := \sup_{i\in\mathcal{S}} \|A(i)\| \leq 1 \tag{18}$$

3. Assumption 3: A is Hurwitz: all eigenvalues have strictly negative parts

One can check that *TD* algorithms satisfy assumptions 1-3.

Relevant Quantities

1. Fact: there exists a symmetric matrix $P > 0$ so that

$$\overline{A}^t P + P\overline{A}^t = -I \tag{19}$$

   γ_{max} := largest eigenvalue of P; γ_{min} := smallest eigenvalue of P

2. some universal constants

$$\kappa_1 = 62\gamma_{max}(1 + b_{max}), \quad \kappa_2 = 55\gamma_{max}(1 + b_{max})^3, \quad \tilde{\kappa_2} = 2(\kappa_2 + \gamma_{max} b^2_{max}) \tag{20}$$

Theorem Statement

Theorem

For ϵ so that $\kappa_1 \epsilon \tau_\epsilon + \epsilon \gamma_{max} \leq 0.05$ and all $k \geq \tau_\epsilon$,

$$\mathbb{E}[\|\Theta_k\|^2] \leq \frac{\gamma_{max}}{\gamma_{min}} \left(1 - \frac{0.9\epsilon}{\gamma_{max}}\right)^{k-\gamma} (1.5\|\Theta_0\| + 0.5 b_{max})^2 + \frac{\tilde{\kappa_2}\gamma_{max}}{0.9\gamma_{min}} \epsilon \tau \tag{21}$$

1. this is a finite error bound compared to the convergence result from Tsitsiklis and Van Roy 1997
2. if $k \geq \tau_\epsilon + O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$, then $\mathbb{E}\|\Theta_k\|^2 = O(\epsilon \tau_\epsilon)$.
3. step size ϵ is fixed. Not difficult to extend analysis to algorithms with diminishing step sizes.

Theorem: Motivation

A standard way to study (1) is to consider

$$\mathbb{E}[W(\Theta_{k+1}) - W(\Theta_k) \big| H_k] \tag{22}$$

where H_k is some appropriate history.
Two questions:

1. what is a suitable Lyapunov function W?
2. how to decide H_k?

To answer the first question, we rely on intuitions from

1. Stein's Method
2. Stability (equilibrium) of the associated ODE

Stein's Method: Taylor Expansion of Operator

Stochastic Recursion:

$$\Theta_{k+1} = \Theta_k + \epsilon(A(X_k)\Theta_k + b(X_k)) \tag{23}$$

1. think about the problem in steady state + i.i.d. samples
2. for any proper function H,

$$\mathbb{E}[H(\Theta_{k+1}) - H(\Theta_k)] = 0 \tag{24}$$

3. Taylor expansion:

$$\mathbb{E}\left[\nabla^t H(\Theta_k)(\Theta_{k+1} - \Theta_k) + \frac{1}{2}(\Theta_{k+1} - \Theta_k)^t \nabla^2 H(\tilde{\Theta})(\Theta_{k+1} - \Theta_k)\right] = 0 \tag{25}$$

for appropriate $\tilde{\Theta}$

Stein's Method: Poisson Equation

1. set up the Poisson equation:

$$\nabla^t W(\Theta_k)\mathbb{E}[\Theta_{k+1} - \Theta_k \big| \Theta_k] = -\|\Theta_k\|^2, \text{ for each } \Theta_k \tag{26}$$

2. Combining Poisson equation and Taylor expansion

$$\mathbb{E}[\|\Theta_k\|^2] = \mathbb{E}\left[\frac{1}{2}(\Theta_{k+1} - \Theta_k)^t \nabla^2 W(\tilde{\Theta})(\Theta_{k+1} - \Theta_k)\right] \tag{27}$$

3. one can use Hession bound to obtain bounds on $\mathbb{E}[\|\Theta_k\|^2]$
4. We focus on Poisson equation (26). By i.i.d. assumption,

$$\nabla^t W(\Theta_k)\overline{A}\Theta_k = -\|\Theta_k\|^2 \tag{28}$$

Stein's Method: Intuition

1. Candidate solution to (28):

$$W(\Theta_k) = \Theta_k^t P \Theta_k \tag{29}$$

   for P a symmetric positive definite matrix

2. Solve P so that

$$\overline{A}^t P + P\overline{A}^t = -I \tag{30}$$

   The solution is unique due to the assumption that $\overline{A}$ is Hurwitz

3. Stein's method (Poisson equation) removes the guesswork for a good Lyapunov function W

ODE

Stochastic Recursion:

$$\Theta_{k+1} = \Theta_k + \epsilon(A(X_k)\Theta_k + b(X_k)) \tag{31}$$

1. the corresponding ODE:

$$\dot{\theta} = \overline{A}\theta \tag{32}$$

2. Fact: Θ_k converges to the equilibrium point of ODE (32)
3. how one could derive bounds on $\|\theta_t\|^2$?

ODE: Same Lyapunov function

Consider

$$W(\theta) = \theta^t P\theta \tag{33}$$

1. consider the time derivative of $W(\theta)$

$$\frac{dW}{dt} = \theta^t \left(\overline{A}^t P + P\overline{A}^t\right)\theta = -\|\theta\|^2 \tag{34}$$

2. $W(\theta) \leq \gamma_{max}\|\theta\|^2 \Rightarrow \frac{dW}{dt} \leq -\frac{1}{\gamma_{max}} W$
3. Thus,

$$\|\theta_t\|^2 \leq \frac{1}{\gamma_{min}} W(\theta_t) \leq \frac{\gamma_{max}}{\gamma_{min}} e^{-t/\gamma_{max}} \|\theta_0\|^2 \tag{35}$$

4. indicates that W is a correct choice of Lyapunov function

Two Methods, One Lyapunov Function and Similar Bounds

1. both Stein's method and analysis of ODE point to the same Lyapunov function W
2. analysis of stochastic system is similar to ODE:
   drift of W versus time derivative of W along the trajectory of ODE

$$\mathbb{E}[\|\Theta_k\|^2] \leq \frac{\gamma_{max}}{\gamma_{min}}\left(1-\frac{0.9\epsilon}{\gamma_{max}}\right)^{k-\gamma}(1.5\|\Theta_0\|+0.5b_{max})^2+\frac{\tilde{\kappa_2}\gamma_{max}}{0.9\gamma_{min}}\epsilon\tau \tag{36}$$

$$\sim \frac{\gamma_{max}}{\gamma_{min}}\left(1-\frac{0.9\epsilon}{\gamma_{max}}\right)^{k-\gamma}\|\Theta_0\|^2 \tag{37}$$

similar to $\frac{\gamma_{max}}{\gamma_{min}}e^{-t/\gamma_{max}}\|\theta_0\|^2$ for small ϵ.

How to Decide H_k?

1. Lyapunov function W as a solution to Poisson equation: applying Stein's method to steady state approximation
2. ODE is determined by the steady states of $A(X_k)$ and $b(X_k)$
3. given history H_k, for drift analysis of W to be effective, we need to wait an initial transient period τ_ϵ for $A(X_k), b(X_k)$ close enough to steady states
4. $H_k := \Theta_{k-\tau}$

Proof of the Theorem

1. Use W as Lyapunov function and obtain bound on the drift

$$\mathbb{E}[W(\Theta_{k+1}) - W(\Theta_k)|\Theta_{k-\tau}] \leq -\frac{0.9\epsilon}{\gamma_{max}}\mathbb{E}[W(\Theta_k)|\Theta_{k-\tau}] + \tilde{k}_2\epsilon^2\tau_\epsilon \tag{38}$$

2. Combine drift bound with

$$\mathbb{E}\|\Theta_k\|^2 \leq \frac{1}{\gamma_{min}}\mathbb{E}[W(\Theta_k)] \tag{39}$$

and various vector inequalities

References

R. Srikant, Lei Ying

Finite-Time Error Bounds For Linear Stochastic Approximation and TD Learning.

Statistical Science, 1997, Vol 12, No.4, 278-300.

Mark Gluzman

Multi-step learning and Value-based approximation methods .

https://rlseminar.github.io/static/files/RL_tutorials2019-0128mark.pdf.