Stochastic Approximation

Shangtong Zhang

University of Virginia

Convergence with pesudo gradient

$$v_{t+1} = v_t + \alpha_t g_t$$

Assumption: There exists a $f : \mathbb{R}^n \to \mathbb{R}$ such that

1. $f(v) \geq 0$
2. ∇f is Lipschitz continuous
3. $\exists c > 0$ such that

$$c\|\nabla f(v_t)\|^2 \leq -\nabla f(v_t)^\top \mathbb{E}\left[g_t | \mathcal{F}_t\right]$$

4. $\exists K_1 > 0, K_2 > 0$ such that

$$\mathbb{E}\left[\|g_t\|^2 | \mathcal{F}_t\right] \leq K_1 + K_2\|\nabla f(v_t)\|^2$$

5. $\sum \alpha_t = \infty, \sum \alpha_t^2 < \infty$

Convergence with pesudo gradient

Results:

1. $\{f(v_t)\}$ converges
2. $\lim_{t\to\infty} \nabla f(v_t) = 0$
3. Every limit point of $\{v_t\}$ is a stationary point of f

Convergence with pesudo contraction

$$v_{t+1}(i) = (1-\alpha_t(i))v_t(i) + \alpha_t(i)\left((\mathcal{T}_t v_t)(i) + \epsilon_t(i) + \xi_t(i)\right)$$

Assumption:

1. $\sum_t \alpha_t(i) = \infty, \sum_t \alpha_t^2(i) < \infty$
2. $\mathbb{E}\left[\epsilon_t(i)|\mathcal{F}_t\right] = 0$, $\mathbb{E}\left[\epsilon_t^2(i)|\mathcal{F}_t\right] \le A + B\|v_t\|^2$
3. $\exists v_*, \gamma > 0$ such that

$$\|\mathcal{T}_t v_t - v_*\|_\infty \le \gamma \|v_t - v_*\|_\infty$$

4. $\exists\{\theta_t\}$ satisfying $\lim_{t\to\infty}\theta_t = 0$ such that w.p. 1

$$|\xi(i)| \le \theta_t\left(\|v_t\| + 1\right)$$

Results:

$$\lim_{t\to\infty} v_t = v_* \quad \text{w.p. } 1$$

Convergence of TD(0)

$$V_{t+1}(s) = \begin{cases} V_t(s) + \alpha_t(s)\left(R_{t+1} + \gamma V_t(S_{t+1}) - V_t(s)\right), & s = S_t, \\ 0, & s \neq S_t \end{cases}$$

$$\begin{aligned} V_{t+1}(s) = & \left(1 - \alpha_t(s)\right) V_t(s) \\ & + \alpha_t(s)\left(\left(\mathcal{T}_\pi V_t\right)(s) + R_{t+1} + \gamma V_t(S_{t+1}) - \left(\mathcal{T}_\pi V_t\right)(s)\right) \end{aligned}$$

Convergence of Q-learning

$$\delta_t \doteq R_{t+1} + \gamma \max_{a'} Q_t(S_{t+1}, a') - Q_t(S_t, A_t)$$

$$Q_{t+1}(s,a) = \begin{cases} Q_t(s,a) + \alpha_t(s,a)\delta_t, & (s,a) = (S_t, A_t), \\ Q_t(s,a), & (s,a) \neq (S_t, A_t) \end{cases}$$

References

- Neuro-Dynamic Programming by Dimitri Bertsekas and John Tsitsiklis