Lecture 15: Convergence

Course: Reinforcement Learning Theory

Instructor: Lei Ying
Department of EECS
University of Michigan, Ann Arbor

Supermartingale convergence theorem

Supermartingale convergence theorem

Let Y_t, X_t and Z_t, $t = 0, 1, 2, \ldots$ be three sequences of random variables and let $\mathcal{F}_t$, $t = 0, 1, 2, \ldots$ be sets of random variables such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all t.

Suppose that

- The random variables Y_t, X_t and Z_t are non-negative, and are functions of the random variables in $\mathcal{F}_t$.
- For each t, we have $E[Y_{t+1}|\mathcal{F}_t] \leq Y_t - X_t + Z_t$
- $\sum_{t=0}^{\infty} Z_t < \infty$

Then, we have $\sum_{t=0}^{\infty} X_t < \infty$, and the sequence Y_t converges to a non-negative random variable Y with probability one.

Martingale convergence theorem

Martingale convergence theorem

Let X_t, $t = 0, 1, 2, \ldots$ be a sequence of random variables and let $\mathcal{F}_t$, $t = 0, 1, 2, \ldots$ be sets of random variables such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all t.

Suppose that

(a) The random variable X_t is a function of the random variables in $\mathcal{F}_t$.

(b) For each t, we have $E[X_{t+1}|\mathcal{F}_t] = X_t$

(c) There exists a constant M such that $E[|X_t|] \leq M$ for all t.

Then, the sequence X_t converges to a random variable X with probability one.

Remark: A sequence X_t that satisfies (a) and (b) above, together with $E[|X_t|] < \infty$, is called a martingale.

Martingale convergence theorem

If $E[X_t^2] < M$, then

$$E[|X_t|] \leq E[1 + X_t^2] \leq 1 + M$$

So, if the second moment of a martingale X_t is bounded, the martingale convergence theorem applies.

Proof of the Convergence Theorem

According to the assumption on the gradient of V, Taylor expansion and mean-value theorem,

$$V(\bar{y}) \leq V(y) + \nabla V(y)^T(\bar{y} - y) + \frac{c}{2}\|\bar{y} - y\|^2 \quad \forall \bar{y}, y$$

$$V(Y_{t+1}) \leq V(Y_t) + \beta_t \nabla^T V(Y_t) S_t + \frac{c}{2}\beta_t^2 \|S_t\|^2$$

Taking the Taylor expansion on both sides, conditioned on $\mathcal{F}_t$, and using the assumption on $E[\|S_t\|^2|\mathcal{F}_t]$ and $E[S_t|\mathcal{F}_t]$,

$$\begin{aligned}
E[V(Y_{t+1})|\mathcal{F}_t] &\leq V(Y_t) + \beta_t \nabla V^T(Y_t) E[S_t|\mathcal{F}_t] + \frac{c}{2}\beta_t^2(k_1 + k_2\|\nabla V(Y_t)\|^2) \\
&\leq V(Y_t) - \beta_t \left(c' - \frac{ck_2\beta_t}{2}\right)\|\nabla V(Y_t)\|^2 + \frac{ck_1}{2}\beta_t^2 \\
&= V(Y_t) - X_t + Z_t
\end{aligned}$$

Convergence proofs

$$X_t = \begin{cases} \beta_t(c' - \frac{ck_2\beta_t}{2})\|\nabla V(Y_t)\|^2, & \text{if } ck_2\beta_t \leq 2c' \\ 0, & \text{otherwise} \end{cases}$$

$$Z_t = \begin{cases} \frac{ck_1}{2}\beta_t^2, & \text{if } ck_2\beta_t \leq 2c' \\ \frac{ck_1}{2}\beta_t^2 - \beta_t(c' - \frac{ck_2\beta_t}{2})\|\nabla V(Y_t)\|^2, & \text{otherwise} \end{cases}$$

Note that X_t and Z_t are functions of $\mathcal{F}_t$. Since $\sum_{t=0}^{\infty} \beta_t^2 < \infty$, $\beta_t \to 0$ and so $ck_2\beta_2 < 2c'$ for sufficiently large t.

Thus, there exists T such that $Z_t = \frac{ck_1}{2}\beta_t^2$ for $t \geq T$, and so $\sum_{t=0}^{\infty} Z_t < \infty$.

We can therefore conclude that $V(Y_t)$ converges by the supermartingale convergence theorem.

Convergence proofs

Recall that

$$E[V(Y_{t+1})|\mathcal{F}_t] \leq V(Y_t) - \beta_t \left(c' - \frac{ck_2\beta_t}{2}\right) \|\nabla V(Y_t)\|^2 + \frac{ck_1\beta_t^2}{2}.$$

So,

$$\frac{1}{\beta_t}(E[V(Y_{t+1})] - E[V(Y_t)]) + \left(c' - \frac{ck_2\beta_t}{2}\right) E\left[\|\nabla V(Y_t)\|^2\right] \leq \frac{ck_1}{2}\beta_t.$$

Note that $\lim_{t\to\infty} E[V(Y_{t+1})] = \lim_{t\to\infty} E[V(Y_t)]$,

$$\implies \text{(not rigorous)} \lim_{t\to\infty} c' E[\|\nabla V(Y_t)\|^2] \leq 0$$

$$\lim_{t\to\infty} E[\|\nabla V(Y_t)\|^2] = 0$$

Convergence proofs

Therefore, for any y^* such that $y^* = \lim_{t\to\infty} Y_t$, we have

$$\nabla V(y^*) = 0,$$

i.e. y^* is a stationary point of V.

Convergence under contraction

- Martingale approach in general requires a smooth Lyapunov function.
- For value iteration algorithms $J = TJ$ (Q-learning), it is not clear whether we can find a smooth Lyapunov function.

Convergence under contraction

Consider

$$Y_{t+1}(i) = (1-\beta_t)Y_t(i) + \beta_t((HY_t)(i) + W_t(i))$$

Example: data-driven Q-learning

Assumptions: $E[W_t(i)|\mathcal{F}_t] = 0$ and $E[W_t^2(i)|\mathcal{F}_t] \leq A + B\|Y_t\|^2$

H is a weighted maximum norm pseudo-contraction.

- Weighted maximum norm:

$$\|y\|_\xi = \max_i \frac{|y(i)|}{\xi(i)}, \quad \xi > 0$$

- Pseudo-contraction: $\exists y^*, \xi > 0$ and $\alpha \in [0,1)$,

$$\|Hy - y^*\|_\xi \leq \alpha\|y - y^*\|_\xi \quad \forall y$$

Convergence under contraction

Consider

$$Y_{t+1}(i) = (1-\beta_t)Y_t(i) + \beta_t((HY_t)(i) + W_t(i))$$

Example: data-driven Q-learning

Assumptions: $E[W_t(i)|\mathcal{F}_t] = 0$ and $E[W_t^2(i)|\mathcal{F}_t] \leq A + B\|Y_t\|^2$

H **is a weighted maximum norm pseudo-contraction.**

Then, $\sum \beta_t = \infty$ and $\sum \beta_t^2 < \infty$

Y_t converges to y^* with probability one

Convergence under contraction

Intuition: consider a much simpler special case such that $y^* = 0$, H is a pseudo-contraction mapping under $\|\cdot\|_\infty$ norm, i.e., there exists $\alpha \in [0, 1)$,

$$|Hy(i)| \leq \alpha \max_j |y(j)| \quad \forall i.$$

Furthermore, $W_t(i) = 0$, $\forall t, i$ (no noise)

Convergence under contraction

Intuition: consider a much simpler special case such that $y^* = 0$, H is a pseudo-contraction mapping under $\|\cdot\|_\infty$ norm, i.e., there exists $\alpha \in [0, 1)$,

$$|Hy(i)| \le \alpha \max_j |y(j)| \quad \forall i.$$

Furthermore, $W_t(i) = 0$, $\forall t, i$ (no noise)

Now if $\|y(0)\|_\infty \le C$, i.e. $|y_0(j)| \le C \, \forall j$,

then $|y_t(j)| \le C \, \forall j$ under the pseudo-contraction.

Convergence under contraction

Furthermore, when $y(i)$ is updated at time t,

$$|y_{t+1}(i)| \leq \alpha C$$

When all components have been updated at least once by time T,

$$\|y_T\|_\infty \leq \alpha C.$$

Assume all components are updated once every T time slots, we have

$$\lim_{n\to\infty} \|y_{nT}\|_\infty \leq \lim_{n\to\infty} \alpha^n C = 0.$$

Reference

- Chapter 4.3 of Dimitri P. Bertsekas and John Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

Acknowledgements: I would like to thank Alex Zhao for helping prepare the slides, and Honghao Wei and Zixian Yang for correcting typos/mistakes.