Lecture Note 14

1 Convergence of $TD(\lambda)$

In this lecture, we will continue to analyze the behavior of $TD(\lambda)$ for autonomous systems. We assume that the system has stage costs $g(x)$ and transition matrix P.

Recall that we want to approximate J^* by $J^* \approx \Phi\tilde{r}$. We find successive approximations $\Phi r_0, \Phi r_1, \dots$ by applying $TD(\lambda)$:

$$r_{k+1} = r_k + \gamma_k d_k z_k \tag{1}$$

$$d_k = g(x_k) + \alpha(\Phi r_k)(x_{k+1}) - (\Phi r_k)(x_k) \tag{2}$$

$$z_k = \alpha\lambda z_{k-1} + \phi(x_k) = \sum_{\tau=0}^{k} (\alpha\lambda)^\tau \phi(x_\tau) \tag{3}$$

We make the following assumptions:

Assumption 1 *The Markov chain characterized by P is irreducible and aperiodic with stationary distribution π.*

Assumption 2 *The basis functions are orthonormal with respect to $\|\cdot\|_{2,D}$, where $D = \mathrm{diag}(\pi)$, i.e., $\Phi^T D\Phi = I$.*

In the previous lecture, we introduced and analyzed *approximate value iteration* (AVI). The main idea is that $TD(\lambda)$ may be interpreted as a stochastic approximations version of AVI. Before finishing the analysis of $TD(\lambda)$, we review the main points related to AVI.

Recall the operators T_λ and Π:

$$T_\lambda J = (1-\lambda)\sum_{m=0}^{\infty} \lambda^m T^{m+1} J,$$

$$\Pi J = \Phi < \Phi, J >_D .$$

Then AVI is given by

$$\Phi r_{k+1} = \Pi T \Phi r_k, \tag{4}$$

and we have the following theorem characterizing its limiting behavior:

Theorem 1 *If*

$$D = \begin{bmatrix} \pi_1 & 0 & \dots & 0 \\ 0 & \pi_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \pi_{|\mathcal{S}|} \end{bmatrix}$$

and $\pi^T P = \pi^T$*, then* $r_k \to r^*$*, and*

$$\|J^* - \Phi r^*\|_{2,D} \leq \frac{1}{\sqrt{1-k^2}} \|J^* - \Pi J^*\|_{2,D}$$

where $k = \frac{\alpha(1-\lambda)}{1-\alpha\lambda} \leq \alpha$.

We can think $TD(\lambda)$ as a stochastic approximations version of AVI. Recall that the main idea in stochastic approximation algorithms is as follows. We would like to solve a system of equations $r = Hr$, but only have access to noisy observations $Hr = w$ for any given r. Then we attempt to solve $r = Hr$ iteratively by considering

$$r_{k+1} = r_k + \gamma_k (Hr_k - r_k + w_k).$$

Hence in order to show that $TD(\lambda)$ is a stochastic approximations version of AVI, we would like to show that

$$\Phi r_{k+1} = \Pi T_\lambda \Phi r_k - \Phi r_k + w_k,$$

for some noise w_k.

The following lemma expresses (4) in a format that is more amenable to our analysis.

Lemma 1 *The AVI equations (4) can be rewritten as*

$$\Phi r_{k+1} = \Phi < \Phi, T_\lambda \Phi r_k >_D, \tag{5}$$

or, equivalently,

$$r_{k+1} = A r_k + b, \tag{6}$$

where

$$A = (1-\lambda)\Phi^T D \left(\sum_{t=0}^{\infty} \lambda^t (\alpha P)^{t+1} \right) \Phi \tag{7}$$

and

$$b = \Phi^T D \sum_{t=0}^{\infty} (\alpha\lambda)^t P^t g. \tag{8}$$

Proof: (5) follows immediately from the definition of Π. Now note that

$$\begin{aligned}
r_{k+1} &= < \Phi, T_\lambda \Phi r_k >_D \\
&= \Phi^T D T_\lambda \Phi r_k \\
&= (1-\lambda)\Phi^T D \sum_{m=0}^{\infty} \lambda^m \left[\sum_{t=0}^{m} (\alpha P)^t g + \alpha^{m+1} P^{m+1} \Phi r_k \right] \\
&= (1-\lambda)\Phi^T D \sum_{m=0}^{\infty} \lambda^m (\alpha P)^{m+1} \Phi r_k + (1-\lambda) \sum_{t=0}^{\infty} (\alpha P)^t g \sum_{m=t}^{\infty} \lambda^m \\
&= A r_k + \sum_{t=0}^{\infty} (\lambda \alpha P)^t g \\
&= A r_k + b.
\end{aligned}$$

$\square$

At the same time, we have the following

Lemma 2 *$TD(\lambda)$'s equations (1) can be rewritten as*

$$r_{k+1} = r_k + \gamma_k(A_k r_k - r_k + b_k), \tag{9}$$

where

$$\begin{aligned} \lim_{k\to\infty} \mathrm{E}A_k &= A, \\ \lim_{k\to\infty} \mathrm{E}b_k &= b, \end{aligned}$$

where A and b are given by (7) and (8), respectively.

Proof: It is easy to verify that (1) is equivalent to (9), where

$$\begin{aligned} A_k &= z_k(\alpha\phi(x_{k+1}) - \phi(x_k)) + I, \\ b_k &= z_k g(x_k). \end{aligned}$$

We will study the limit of $\mathrm{E}A_k$ and $\mathrm{E}b_k$. For all J, we have

$$\begin{aligned} \lim_{k\to\infty} \mathrm{E}\left[z_k J_k\right] &= \lim_{k\to\infty} \mathrm{E}\left[\sum_{\tau=0}^{k}(\alpha\lambda)^{k-\tau}\phi(x_\tau)J(x_k)\right] \\ &= \lim_{k\to\infty} \mathrm{E}\left[\sum_{\tau=0}^{k}(\alpha\lambda)^{\tau}\phi(x_{k-\tau})J(x_k)\right] \qquad (\because P^k(x,y)\to\pi(y)) \\ &= \mathrm{E}\left[\sum_{\tau=0}^{\infty}(\alpha\lambda)^{\tau}\phi(x_0)J(x_\tau)|x_0\sim\pi\right] \qquad (P(x_\tau = x|x_0) = P^\tau(x_0,x)) \\ &= \mathrm{E}\left[\sum_{\tau=0}^{\infty}(\alpha\lambda)^{\tau}\phi(x_0)(P^\tau J)(x_0)|x_0\sim\pi\right] \\ &= \sum_{\tau=0}^{\infty}(\alpha\lambda)^{\tau} < \Phi, P^\tau J >_D \end{aligned}$$

Letting

$$J(x_k, x_{k+1}) = \alpha\phi(x_{k+1}) - \phi(x_k),$$

we conclude that

$$\begin{aligned} \lim_{k\to\infty} \mathrm{E}A_k &= \sum_{\tau=0}^{\infty}(\alpha\lambda)^{\tau} < \Phi, \alpha P^{\tau+1}\Phi - P^\tau\Phi >_D + I \\ &= \Phi^T D\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^{\tau+1}P^{\tau+1}\Phi - \Phi^T D\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^\tau P^\tau\Phi + I \\ &= \Phi^T D\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^{\tau+1}P^{\tau+1}\Phi - \Phi^T D\sum_{\tau=1}^{\infty}\lambda^\tau\alpha^\tau P^\tau\Phi - \Phi^T D\Phi + I \\ &= \Phi^T D\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^{\tau+1}P^{\tau+1}\Phi - \lambda\Phi^T D\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^{\tau+1}P^{\tau+1}\Phi \\ &= (1-\lambda)\sum_{\tau=0}^{\infty}\lambda^\tau\alpha^{\tau+1}P^{\tau+1}\Phi \\ &= A. \end{aligned}$$

In the fourth equality we have used the assumption that $\Phi^T D \Phi = I$.

Similarly, letting

$$J(x_k) = g(x_k)$$

yields

$$\begin{aligned}\lim_{k\to\infty} \mathrm{E} b_k &= \sum_{\tau=0}^{\infty} (\alpha\lambda)^{\tau} < \Phi, P^{\tau} g >_D \\ &= b.\end{aligned}$$

If $\lambda = 1$, we have

$$\sum_{\tau=0}^{\infty} (\alpha\lambda)^{\tau} (g + \alpha P^{\tau} \Phi r_k - \Phi r_k) = J^* + (I - \alpha P)^{-1} (\alpha P - I \Phi r) = J^* - \Phi r.$$

If $\lambda < 1$, then

$$\begin{aligned}\sum_{\tau=0}^{\infty} (\lambda\alpha)^{\tau} P^{\tau} &= \sum_{\tau=0}^{\infty} \alpha^{\tau} P^{\tau} (1-\lambda) \sum_{t=\tau}^{\infty} \lambda^t \\ &= (1-\lambda) \sum_{\tau=0}^{t} \lambda^{\tau} \sum_{t=0}^{\tau} \alpha^t P^t\end{aligned}$$

Thus

$$\begin{aligned}\sum_{\tau=0}^{\infty} (\lambda\alpha)^{\tau} P^{\tau} (g + \alpha P \Phi r_k - \Phi r_k) &= (1-\lambda) \sum_{\tau=0}^{\infty} \lambda^{\tau} \sum_{t=0}^{\tau} \alpha^t P^t (g + \alpha P \Phi r - \Phi r) \\ &= (1-\lambda) \sum_{\tau=0}^{\infty} \lambda^{\tau} \sum_{t=0}^{\tau} \Big(\underbrace{\alpha^t P^t g + \alpha^{t+1} P^{t+1} \Phi r}_{T^t \Phi r_k} - \Phi r \Big) \\ &= T_{\lambda} \Phi r_k - \Phi r_k\end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \mathrm{E}\left[z_k d_k\right] = < \Phi, T_{\lambda} \Phi r_k - \Phi r_k >_D$$

□

Comparing Lemmas 1 and 2, it is clear that $TD(\lambda)$ can be seen as a stochastic approximations version of AVI; in particular, TD's equations can be written as

$$r_{k+1} = r_k + \gamma_k (A r_k + b - r_k + w_k),$$

where $w_k = (A_k - A) r_k + (b_k - b)$. If r_k remains bounded, we should have $\lim_{k\to\infty} \mathrm{E} w_k = 0$, so that the noise is zero-mean asymptotically. Note however that the noise is not independent of the past history, and in fact follows a Markov chain, since matrices A_k and b_k are functions of x_k and x_{k+1}. This makes application of the Lyapunov analysis for convergence of $TD(\lambda)$ difficult, and we must resort to the ODE analysis instead. The next theorem provides the convergence result.

Theorem 2 *Suppose that P is irreducible and aperiodic and that $\sum_{k=1}^{\infty} \gamma_k = \infty$ and $\sum_{k=1}^{\infty} \gamma_k^2 < \infty$. Then $r_k \to r^*$ w.p.1, where $\Phi r^* = \Pi T_{\lambda} \Phi r^*$.*

To prove Theorem 2, we first state without proof the following theorem.

Theorem 3 *Let $r_{k+1} = r_k + \gamma_k(A(x_k)r_k + b(x_k))$. Suppose that*

(a) $\sum_{k=1}^{\infty} \gamma_k = \infty$, $\sum_{k=1}^{\infty} \gamma_k^2 < \infty$

(b) x_k *follows a Markov chain and has stationary distribution* π

(c) $A = \mathrm{E}[A(x)|x \sim \pi]$ *is negative definite, and* $b = \mathrm{E}[b(x)|x \sim \pi]$

(d) $\|\mathrm{E}[A(x_k)|x_0] - A\| \leq C\rho^k, \forall\, x_0, \forall k$, *and*
$\|\mathrm{E}[b(x_k)|x_0] - b\| \leq C\rho^k, \forall\, x_0, \forall k$

Then $r_k \to r^*$ *w.p.1, i.e.,* $Ar^* + b = 0$.

Sketch of Proof of Theorem 2 We verify that conditions (a)-(d) of Theorem 3 are satisfied.

Conditions (a) and (b) are satisfied by assumption.

(c) For all r, we have

$$\begin{aligned} r^T A r &= r^T < \Phi, (1-\lambda) \sum_{\tau=0}^{\infty} \lambda^\tau \alpha^{\tau+1} P^{\tau+1} \Phi r - \Phi r >_D \\ &= \underbrace{< \Phi r, (1-\lambda) \sum_{\tau=0}^{\infty} \lambda^\tau \alpha^{\tau+1} P^{\tau+1} \Phi r >_D}_{\bar{T}_\lambda \phi r, \text{ a contraction w.r.t. } \|\cdot\|_{2,D}} - \|\Phi r\|_{2,D}^2 \\ &\leq \|\Phi r\|_{2,D} \|\bar{T}_\lambda \Phi r\|_{2,D} - \|\Phi r\|_{2,D}^2 \\ &\leq \beta \|\Phi r\|_{2,D}^2 - \|\Phi r\|_{2,D}^2 \qquad (\beta \leq \alpha) \\ &< 0 \end{aligned}$$

Hence, A is negative definite.

(d) We must consider the quantities

$$\begin{aligned} \mathrm{E}[A_k - A] &= \mathrm{E}[z_k(\alpha\phi(x_{k+1}) - \phi(x_k) - A], \\ \mathrm{E}[b_k - b] &= \mathrm{E}[z_k g(x_k) - b]. \end{aligned}$$

This involves a comparison of $\mathrm{E}[\alpha z_k \phi(x_{k+1}]$, $\mathrm{E}[z_k\phi(x_k)]$ and $\mathrm{E}[z_k g(x_k)]$ with their limiting values as k goes to infinity. Let us focus on term $z_k\phi(x_k)$; the other terms involve similar analysis. We have

$$
\begin{aligned}
\mathrm{E}[z_k\phi(x_k)|x_0] &= \mathrm{E}\left[\underbrace{\sum_{t=0}^{k}(\alpha\lambda)^{k-t}\phi(x_t)}_{z_k}\,\phi(x_k)\,|\,x_0=x\right] \\
&= \mathrm{E}\left[\sum_{t=-\infty}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_t\sim\pi, t\le 0\right] \\
&+ \mathrm{E}\left[\sum_{t=0}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0=x\right] \\
&- \mathrm{E}\left[\sum_{t=0}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0\sim\pi\right] \\
&- \mathrm{E}\left[\sum_{t=-\infty}^{-1}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_t\sim\pi\right]
\end{aligned}
$$

It follows from basic matrix theory that $|P(x_t = x|x_0) - \pi(x_t)| \le C\rho^t$, where ρ corresponds to the second highest eigenvalue of P, which is strictly less than one since P is irreducible and aperiodic. Therefore we have

$$
\begin{aligned}
&\Big|\mathrm{E}\left[\sum_{t=0}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0=x\right] - \mathrm{E}\left[\sum_{t=0}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)\Big|x_0\sim\pi\right]\Big| \\
&\le \Big|\mathrm{E}\left[\sum_{t=0}^{\lfloor\frac{k}{2}\rfloor}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0=x\right] - \mathrm{E}\left[\sum_{t=0}^{\lfloor\frac{k}{2}\rfloor}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)\Big|x_0\sim\pi\right]\Big| + \\
&+\Big|\mathrm{E}\left[\sum_{t=\lfloor\frac{k}{2}\rfloor+1}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0=x\right] - \mathrm{E}\left[\sum_{t=\lfloor\frac{k}{2}\rfloor+1}^{k}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)\Big|x_0\sim\pi\right]\Big| \\
&\le M\left((\alpha\lambda)^{k/2}+\rho^{k/2}\right),
\end{aligned}
$$

for some $M < \infty$. Moreover,

$$
\mathrm{E}\left[\sum_{t=-\infty}^{-1}\phi(x_t)(\alpha\lambda)^{k-t}\phi(x_k)|x_0\sim\pi\right] \le M(\alpha\lambda)^k
$$

$\square$