Lecture Note 8

1 Lyapunov Function Analysis

In this lecture, we want to study the convergence of

$$r_{t+1} = r_t + \gamma_t S(r_t, w_t)$$

to some γ^* with $\mathrm{E}\left[S(r^*, w_t)\right] = 0$. Recall the Lyapunov function analysis in deterministic case that we pick a function $V(r)$ such that

- $V(r) \geq 0$, $\forall\, r$,
- $\nabla V(r)^T S(r) < 0$, if $r \neq r^*$,
- $\nabla V(r^*) = 0$.

The argument for convergence is that we observe $V(r_t)$ decreasing over time and lower bounded; therefore, $V(r_t)$ converges to some limit. With technical conditions on V and S, we can show that $r_t \to r^*$.

We now proceed to the stochastic case. Let $\mathcal{F}_t$ denote the history of the process up to stage t. Explicitly, we can have $\mathcal{F}_t$ as

$$\mathcal{F}_t = \{r_l, l \leq t,\, w_l, l < t,\, \gamma_t, l \leq t\}\,.$$

Note that the step size γ_t can depend on the history which is stochastic, but not on the disturbance w_t.

We define the Euclidean norm $\|V\|_2 = (V^T V)^{\frac{1}{2}}$.

Theorem 1 *Suppose that* $\exists V$ *such that*

(a) $V(r) \geq 0,\ \forall r$,

(b) $\exists L$ *such that* $\|\nabla V(r) - \nabla V(\bar{r})\|_2 \leq L\|r - \bar{r}\|_2$ *(Lipschitz continuity),*

(c) $\exists K_1, K_2$ *such that* $\mathrm{E}\left[\|S(r_t, w_t)\|_2^2 \,\middle|\, \mathcal{F}_t\right] \leq K_1 + K_2\|\nabla V(r_t)\|_2^2$,

(d) $\exists c$ *such that* $\nabla V(r_t)^T \mathrm{E}\left[S(r_t, w_t) \,\middle|\, \mathcal{F}_t\right] \leq -c\|\nabla V(r_t)\|_2^2$.

Then, if γ_t *satisfies* $\sum_{t=0}^{\infty} \gamma_t = \infty$ *and* $\sum_{t=0}^{\infty} \gamma_t^2 < \infty$, *we have*

- $V(r_t)$ *converges,*
- $\lim_{t\to\infty} \nabla V(r_t) = 0$.
- *every limit point* $\bar{r}$ *of* r_t *satisfies* $\nabla V(\bar{r}) = 0$.

We will prove the convergence for a special case where $V(r) = \frac{1}{2}\|r - r^*\|_2^2$ for some r^*.

Theorem 2 *Suppose* $V(r) = \frac{1}{2}\|r - r^*\|_2^2$ *satisfies*

(a) $\exists K_1, K_2$ *such that* $\mathrm{E}\left[\|S(r_t, w_t)\|_2^2 \,\middle|\, \mathcal{F}_t\right] \leq K_1 + K_2 V(r_t)$,

(b) $\exists c$ *such that* $\nabla V(r_t)^T \mathrm{E}\left[S(r_t, w_t) \,\middle|\, \mathcal{F}_t\right] \leq -cV(r_t)$.

Then, if $\gamma_t > 0$ *with* $\sum_{t=0}^{\infty} \gamma_t = \infty$ *and* $\sum_{t=0}^{\infty} \gamma_t^2 < \infty$,

$$r_t \to r^*, \qquad w.p.\ 1.$$

We use the following Supermartingale convergence theorem to prove Theorem 2.

Theorem 3 (Supermartingale Convergence Theorem) *Suppose that* X_t, Y_t *and* Z_t *are nonnegative random variables and* $\sum_{t=1}^{\infty} Y_t < \infty$ *with probability 1. Suppose also that*

$$\mathrm{E}\left[X_{t+1}\middle|\mathcal{F}_t\right] \leq X_t + Y_t - Z_t, \qquad w.p.\ 1.$$

Then

1. X_t *converges to a limit with probability 1,*
2. $\sum_{t=1}^{\infty} Z_t < \infty$.

The key idea for the proof of Theorem 2 is to show that $V(r_t)$ is a supermartingale, so that $V(r_t)$ converges and then show that it converges to zero w.p. 1.

Proof: [Theorem 2]

$$\begin{aligned}
\mathrm{E}\left[V(r_{t+1})\middle|\mathcal{F}_t\right] &= \mathrm{E}\left[\frac{1}{2}\|r_{t+1} - r^*\|_2^2\middle|\mathcal{F}_t\right] \\
&= \mathrm{E}\left[\frac{1}{2}(r_t + \gamma_t S_t - r^*)^T(r_t + \gamma_t S_t - r^*)\middle|\mathcal{F}_t\right] \qquad (S_t \triangleq S(r_t, w_t)) \\
&= \frac{1}{2}(r_t - r^*)^T(r_t - r^*) + \gamma_t (r_t - r^*)^T \mathrm{E}\left[S_t\middle|\mathcal{F}_t\right] + \frac{\gamma_t^2}{2}\mathrm{E}\left[S_t^T S_t\middle|\mathcal{F}_t\right]
\end{aligned}$$

Since $V(r_t) = \frac{1}{2}\|r_t - r^*\|_2^2$, $\nabla V(r_t) = (r_t - r^*)$. Then

$$\begin{aligned}
\mathrm{E}\left[V(r_{t+1})\middle|\mathcal{F}_t\right] &= V(r_t) + \gamma_t (r_t - r^*)^T \mathrm{E}\left[S_t\middle|\mathcal{F}_t\right] + \frac{\gamma_t^2}{2}\mathrm{E}\left[\|S_t\|_2^2\middle|\mathcal{F}_t\right] \\
&= V(r_t) + \gamma_t \nabla V(r_t)^T \mathrm{E}\left[S_t\middle|\mathcal{F}_t\right] + \frac{\gamma_t^2}{2}\mathrm{E}\left[\|S_t\|_2^2\middle|\mathcal{F}_t\right] \\
&\leq V(r_t) - \gamma_t c V(r_t) + \frac{\gamma_t^2}{2}\left(K_1 + K_2 V(r_t)\right) \\
&\leq \underbrace{V(r_t)}_{X_t} - \underbrace{\left(\gamma_t c - \frac{\gamma_t^2 K_2}{2}\right) V(r_t)}_{Z_t} + \underbrace{\frac{\gamma_t^2}{2} K_1}_{Y_t}
\end{aligned}$$

Since $\gamma_t > 0$ and $\sum_{t=0}^{\infty} \gamma_t^2 < \infty$, γ_t must converge to zero, and $Z_t \geq 0$ for all large enough t. Moreover,

$$\sum_{t=0}^{\infty} Y_t = \frac{K_1}{2} \sum_{t=0}^{\infty} \gamma_t^2 < \infty.$$

Therefore, by Supermartingale convergence theorem,

$$V(r_t) \text{ converges w. p. 1, and}$$
$$\sum_{t=0}^{\infty}\left(\gamma_t c - \frac{\gamma_t^2 K_2}{2}\right)V(r_t) < \infty, \qquad \text{w. p. 1.}$$

Suppose that $V(r_t) \to \epsilon > 0$. Then, by hypothesis that $\sum_{t=0}^{\infty}\gamma_t = \infty$ and $\sum_{t=0}^{\infty}\gamma_t^2 < \infty$, we must have

$$\sum_{t=0}^{\infty}\left(\gamma_t c - \frac{\gamma_t^2 K_2}{2}\right)V(r_t) = \infty$$

which is a contradiction. Therefore

$$\lim_{t\to\infty}\|r_t - r^*\|_2^2 = 0 \quad \text{w.p. 1} \Rightarrow r_t \to r^* \text{ w.p. 1.}$$

□

Example 1 (Stochastic Gauss-Seidel) *Consider*[1]

$$\begin{aligned} r_{t+1}(i_t) &= r_t(i_t) + \gamma_t\left((Fr_t)(i_t) - r_t(i_t)\right), \\ r_{t+1}(i) &= r_t(i_t), \forall\, i \neq i_t. \end{aligned}$$

Suppose that F is a $\|\cdot\|_2$ contraction. Suppose also that i_t, $t = 1, 2, \ldots,$ are chosen i.i.d. with $P(i_t = i) = \pi_i > 0$. Then

$$r_{t+1}(i) = r_t(i) + \gamma_t\pi_i\left((Fr_t)(i) - r_t(i)\right) + \gamma_t \underbrace{\left[\mathbf{1}(i_t = i) - \pi_i\right]\left[(Fr_t)(i) - r_t(i)\right]}_{w_t(i)}$$

Define

$$\Pi = \begin{bmatrix} \pi_1 & 0 & 0 & \ldots & 0 \\ 0 & \pi_2 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \pi_n \end{bmatrix}$$

then

$$r_{t+1} = r_t + \gamma_t \underbrace{\Pi(Fr_t - r_t)}_{\mathrm{E}[S_t|\mathcal{F}_t]} + \gamma_t w_t.$$

Let $V(r) = \frac{1}{2}(r - r^*)^T\Pi^{-1}(r - r^*) \geq 0$. Then we have

$$\nabla V(r) = \Pi^{-1}(r - r^*) \qquad \textit{(Lipschitz continuity holds).}$$

We also have

$$\begin{aligned} \nabla V(r_t)^T \mathrm{E}\left[S_t \middle| \mathcal{F}_t\right] &= (r_t - r^*)^T\Pi^{-1}\Pi(Fr_t - r_t) = (r_t - r^*)^T(Fr_t - r^* + r^* - r_t) \\ &= -(r_t - r^*)^T(r_t - r^*) + (r_t - r^*)^T(Fr_t - r^*) \\ &\leq -\|r_t - t^*\|_2^2 + \|r_t - r^*\|_2\|Fr_t - r^*\|_2 \\ &\leq -\|r_t - t^*\|_2^2 + \alpha\|r_t - t^*\|_2^2 \\ &\leq -(1-\alpha)\min_i \pi_i^2\|\nabla V(r_t)\|_2^2. \end{aligned}$$

[1]Recall the AVI: $r_{t+1}(i_t) = (Fr_t)(i_t)$

We finally have

$$\begin{aligned} \mathrm{E}\left[\|S_t\|_2^2|\mathcal{F}_t\right] &= \mathrm{E}\left[(Fr_t)(i_t)-r_t(i_t))^2|\mathcal{F}_t\right] \\ &\leq \mathrm{E}\left[\|Fr_t-r_t\|_2^2|\mathcal{F}_t\right] \\ &= \|Fr_t-r_t\|_2^2 \\ &\leq \|Fr_t-r^*\|_2^2+\|r_t-r^*\|_2^2 \\ &\leq (1+\alpha)\|r_t-r^*\|_2^2 \\ &\leq (1+\alpha)\max_i \pi_i^2\|\nabla V(r_t)\|_2^2. \end{aligned}$$

We conclude by Theorem 1 that stochastic Gauss-Seidel converges.

2 Q-learning

Recall that the Q-learning algorithm updates the Q factor according to

$$Q_{t+1}(x_t,a_t)=Q_t(x_t,a_t)+\gamma_t(g_{a_t}(x_t)+\alpha\min_{a'}Q_t(x_{t+1},a')-Q_t(x_t,a_t)).$$

This update can be rewritten as

$$\begin{aligned} Q_{t+1}(x,a)=Q_t(x,a) &+ \gamma_t(x,a)\left[\underbrace{g_a(x)+\alpha\sum_y P_a(x,y)\min_{a'}Q_t(y,a')}_{(HQ)(x,a)}-Q_t(x,a)\right] \\ &+ \alpha\gamma_t(x,a)\left[\underbrace{\min_{a'}Q_t(x_{t+1},a')-\sum_y P_a(x,y)\min_{a'}Q_t(y,a')}_{w_t}\right] \end{aligned}$$

where

$$\begin{aligned} &\gamma_t(x,a)=0, \quad \text{if } (x,a)\neq(x_t,a_t) \\ &\gamma_t(x_t,a_t)=\gamma_t \\ &\mathrm{E}\left[\gamma_t w_t\middle|\mathcal{F}_t\right]=0 \\ &|w_t|\leq\|Q_t\|_\infty. \end{aligned}$$

Then, we have

$$Q_{t+1}=Q_t+\gamma_t(HQ_t-Q_t)+\alpha\gamma_t w_t.$$

We can use the following theorem to show that Q-learning converges, as long as every state and action pair are visited infinitely many times.

Theorem 4 *Let* $r_{t+1}(i)=r_t(i)+\gamma_t(i)\Big((Hr_t)(i)-r_t(i)+w_t(i)\Big)$. *Then, if*

- $\mathrm{E}\left[w_t\middle|\mathcal{F}_t\right]=0$

- $\mathrm{E}\left[w_t^2(i)\middle|\mathcal{F}_t\right] \leq A + B\|r_t\|^2$ *for some norm* $\|\cdot\|$
- $\sum_{t=0}^{\infty} \gamma_t(i) = \infty$, $\sum_{t=0}^{\infty} \gamma_t(i)^2 < \infty$, $\forall\, i$
- H *is a maximum-norm contraction,*

then $r_t \to r^*$ *w.p. 1* $(Hr^* = r^*)$.

Comparing Theorems 2 and 4, note that, if H is a maximum-norm contraction, convergence occurs under weaker conditions than if it is an Euclidean norm contraction.

Corollary 1 *If* $\sum_{t=0}^{\infty} \gamma_t(x,a) = \infty$ *with probability 1 for all* (x,a), *we have*

$$Q_t \to Q^* \quad \text{w.p. 1.}$$

3 ODE Approach

Often times, the behavior of $r_{t+1} = r_t + \gamma_t S(r_t, w_t)$ may be understood by analyzing the following ODE instead:

$$\dot{r_t} = \mathrm{E}\left[S(r_t, w_t)\right].$$

The main idea for the ODE approach is as follows. Look at intervals $[t_m, t_{m+1})$ such that

$$\sum_{t=t_m}^{t_{m+1}-1} \gamma_t = \gamma, \qquad \text{where } \gamma \text{ is small.}$$

Set $r_m \equiv r_{t_m}$. Then

$$r_t \approx r_{t_m} + O(\gamma), \quad \forall\, t \in [t_m, t_{m+1}). \tag{1}$$

Then

$$\begin{aligned} r_{m+1} &= r_{t_{m+1}} = r_m + \sum_{t=t_m}^{t_{m+1}-1} \gamma_t S(r_t, w_t) \\ &\approx r_{t_m} + \sum_{t=t_m}^{t_{m+1}-1} \gamma_t \Big(S(r_t, w_t) + O(\gamma)\Big) && (2) \\ &= r_{t_m} + \gamma \sum_{t=t_m}^{t_{m+1}-1} \frac{\gamma_t}{\gamma} S(r_t, w_t) + O(\gamma^2) \\ &\cong r_m + \gamma \mathrm{E}\left[S(r_m, w)\right] + O(\gamma^2) && (3) \end{aligned}$$

Therefore we can think of the stochastic scheme as a discrete version of the ODE

$$r_{m+1} = r_m + \gamma \mathrm{E}\left[S(r_m, w)\right] \Rightarrow \boxed{\dot{r} = \mathrm{E}\left[S(r, w)\right].}$$

To make the argument rigorous, steps (1), (2) and (3) have to be justified.