⊛ when the dynamics are unknow, but have data:

first look at value prediction / Policy evaluation

$$V^\pi(s) = \mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i r(s_t, a_t) \mid s_0 = s\right]$$

$a_t \sim \pi(\cdot \mid s_t)$

$s_{t+1} \sim P(\cdot \mid s_t, a_t)$

$(s_t, a_t, r_t, s_{t+1})_{t=0}^{T}$ $\Rightarrow$ Q: how to estimate $V^\pi(s)$?

1) $$V^\pi(s) = r^\pi(s) + \gamma \sum_{s' \in S} P^\pi(s' \mid s) V^\pi(s')$$

If $v \in \mathbb{R}^{|S| \times 1}$, then

$$V^\pi = r + \gamma P V^\pi \Rightarrow r + (\gamma P - I) V^\pi = 0$$

$$\frac{d}{dt} V(t) = r + \underbrace{(\gamma P - I)}_{A} V(t) \overset{t \to \infty}{\Rightarrow} V(t) \to V^\pi$$

$\text{eig}(A) < 0$

(think: why $\frac{d}{dt} V(t) = (I - \gamma P) V(t) - r \overset{t \to \infty}{\not\Rightarrow} V^\pi$).

2). $$V_{n+1} = V_n + \alpha_n (r + (\gamma P - I) V_n)$$

If the current state is x_n

$$V_{n+1}(x_n) = V_n(x_n) + \alpha_n \underbrace{(r(x_n) + \gamma V(x_{n+1}) - V(x_n))}_{TD}$$

Thm: $\sum_n \alpha_n = \infty$ $\sum \alpha_n^2 < \infty$ $\Rightarrow V_n \to V^\pi$

TD is a stochastic Approximation method to approximate the steady state of the ODE $\dot{x} = h(x)$.

$$x_{n+1} = x_n + \varepsilon_n \left(h(x_n) + \underbrace{M_{n+1}}_{\mathbb{E}[M_{n+1}] = 0} \right)$$

$$\Leftrightarrow x_{n+1} = x_n + \varepsilon_n \hat{h}(x_n) \text{ , where } \mathbb{E}[\hat{h}(x_n) | x_n] = h(x_n)$$

with assumption on h, ε_n, M_n, $x \to x_*$ ← stable equilibrium.

Robbins-Monro 1951

Assumption on h, ε_n, M_n :

① $h \to$ Lipschitz : $\exists L$ s.t $\|h(x) - h(y)\| \le L\|x-y\|$ $\forall x, y$

② $\sum_{n \ge 0} \varepsilon_n = \infty$, $\sum_{n \ge 0} \varepsilon_n^2 < \infty$

③ $\mathbb{E}[M_{n+1} | \mathcal{F}_n] = 0$ & $\mathbb{E}[\|M_{n+1}\|^2 | \mathcal{F}_n] \le k(1 + \|x_n\|^2)$

$(\Leftrightarrow \mathbb{E}[V(s_{t+1})^2 | s_t] - (\mathbb{E}[V(s_{t+1}) | s_t])^2 \le \|V\|^2)$

④ $\sup_{n \ge n} \|x_n\| < \infty$ (← By "Bolkar-Meyn, 2000")

⑤ $\exists f(x)$ s.t $\lim_{\|x\| \to \infty} f(x) \to \infty$ & $f(x) \ge 0$, $f(x) \in C^1$ ← Lyapunov fcn.

s.t. $\langle \nabla f(x), h(x) \rangle \le 0$ w. strict ineq if $f(x) \ne 0$ ←

↓

x^* is a stable equil

For linear ODE $\mathrm{Re}(\mathrm{Eig}(P)) < 0$ is enough (see the proof in the next page.

A linear dynamical system is asymptotically stable if for $\forall$ initial condition $x(0)$, $x(t)$ converges to the origin 0 as $t \to \infty$

① $x(t+1) = Ax(t)$ is asymp stable $\Leftrightarrow |eig(A)| < 1$

1) If $\exists \underset{n}{Q}$ & $\underset{n}{\Lambda}$ s.t $A = Q \Lambda Q^{-1}$

$c(1, \cdots, c_n)$ $diag(\lambda_1, \cdots, \lambda_n)$.

then $x(t) = Q \Lambda^n Q^{-1} x(0)$

$= (c_1, \cdots c_n) \lambda_1^t y_1(0)$

$\lambda_2^t y_2(0)$

$\vdots$

$\lambda_n^t y_n(0)$.

$|x(t)| = \left| \sum_{i=1}^{n} c_i \lambda_i^t y_i(0) \right| \leq \sum_{i=1}^{n} |c_i| |\lambda_i|^t |y_i(0)| \to 0.$

2). Jordan Normal form Thm:

$\forall$ matrix A can be conjugated to $\Lambda + N$ with $N^n = 0$

$(B+N)^t = B^t + C_t' B^{t-1} N + \cdots + C_t^n B^{t-n+1} N^{n-1}$

$B^t \downarrow 0$, $B^{t-1} \downarrow 0$, $B^{t-n+1} \downarrow 0$

② $\dot{x}(t) = Ax(t)$ is asymp stable $\Leftrightarrow$ $Re(eig(A)) < 0$

$x(t) = e^{At} x(0) = U^t x(0)$ with $U = e^A$

$|eig(e^A)| < 1 \quad \Leftrightarrow \quad |e^{eig(A)}| < 1 \Leftrightarrow Re(eig(A)) < 0$

"The ODE method for conv of SA & RL"

"Borkar-Meyn Thm (2000)"

Lip + stepsize + Var + $\lim_{r\to\infty} \frac{h(rx)}{r} = h_\infty(x)$ for $\forall x$

+ θ is a globally exponentially stable equil of $\dot{\theta} = f_\infty(\theta)$

For linear $f_\infty(\theta) = A\theta$, $\mathrm{Re}(\mathrm{eig}(A)) < 0$

$\Rightarrow \quad x_n \to x^*$.

(*) The convergence rate for standard linear SA method: $O(\frac{1}{\sqrt{t}})$