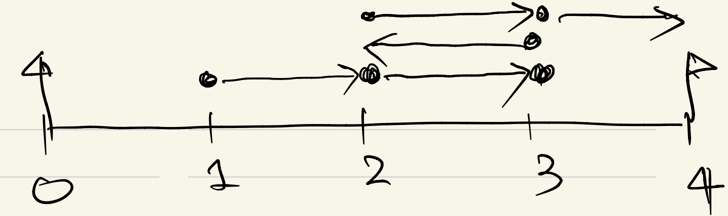
PDE $\Rightarrow$ MRP

ex random walk, ($p=\frac{1}{2}$)



episode →

S_0	R_1	S_1	R_2	S_2	R_3	S_3	R_4	S_4	R_5	S_5
1	1	2	1	3	1	2	1	3	1	4

Gain with discount $\gamma \in (0,1)$

$$G_2 = R_3 + \gamma R_4 + \gamma^2 R_5$$

$$G_0 =$$

Value

$$v(1) = \mathbb{E}[G_t \mid S_t = 1] = ?$$

MRP (Markov Reward Process)

An MRP is a random sequence (episode)

$$\{S_0, R_1, S_1, R_2, S_2 \sim R_T, S_T\}$$

s.t.

- $\mathcal{S}$ is a state space, i.e. $S_t \in \mathcal{S}$
- $\mathcal{P}$ is transition probability matrix. i.e.

$$\mathcal{P}(s, s') = \mathbb{P}(S_{t+1} = s' \mid S_t = s)$$

- $\mathcal{R}$ is reward function

$$\mathcal{R}(s) = \mathbb{E}[R_{t+1} \mid S_t = s]$$

- γ is discount factor
- T is stopping time. (maybe $T = \infty$)

RL on MRP

* Terminal time $T = \inf\{t \geq 0 : S_t \in \underbrace{\partial S}\}$

absorbing state space

* Gain (random)

$$G_t = \sum_{k=0}^{T-t-1} \gamma^k R_{t+k+1}$$

$$= R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T$$

* value (deterministic)

$$v(s) = \mathbb{E}[G_t \mid S_t = s]$$

Bellman eqn ① $\forall s \in \mathcal{S} \setminus \partial \mathcal{S}$

$$v(s) = \mathbb{E}[R_{t+1} + \gamma v(S_{t+1}) \mid S_t = s]$$

$$= R(s) + \gamma \cdot \sum_{s' \in \mathcal{S}} P(s, s') v(s')$$

② $v(s) = 0$, if $s \in \partial \mathcal{S}$.

Pf

$$v(s) = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T \mid S_t = s]$$

Tower property

$$= \mathbb{E}\left[R_{t+1} + \gamma \cdot \mathbb{E}[R_{t+2} + \gamma R_{t+3} + \cdots + \gamma^{T-t-2} R_T \mid S_{t+1}] \,\middle|\, S_t = s\right]$$

$$= \mathbb{E}[R_{t+1} + \gamma \cdot v(S_{t+1}) \mid S_t = s].$$

$\boxed{S}$

⊛ 2-d PDE

$$\begin{cases} \frac{1}{2}\Delta v - v + x_1^2 + x_2^2 - x_1 - x_2 - \frac{3}{2} = 0, & x \in (0,1)^2 = O \\ v(x) = (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 & x \notin (0,1)^2 \end{cases}$$

⊛ GFD with h.

Bellman

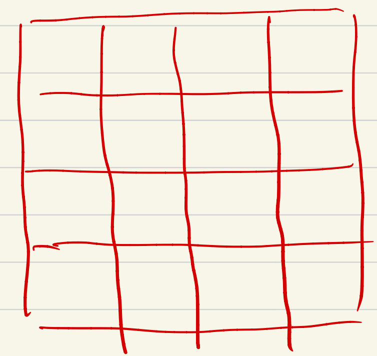
$$\begin{cases} v(x) = \gamma \ell^h(x) + \gamma \sum_{i=1}^{2} \Big(p^h(x+he_i \mid x)\, v(x+he_i) + p^h(x-he_i \mid x_i)\, v(x-he_i) \Big) & x \in O^h \\ v(x) = (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 & x \notin O^h \end{cases}$$

⊛ para

$$\gamma = \frac{2}{2+h^2}, \quad p^h(x \pm h e_i \mid x) = \frac{1}{4}, \quad \ell^h(x) = \frac{h^2}{2}\left(x_1^2 + x_2^2 - x_1 - x_2 - \frac{3}{2}\right)$$

MRP for 2-d

$h = \frac{1}{4}$



① $S = \mathbb{R}^{2,h}$,

$\partial S = S \setminus (0,1)^2$

② If $s \notin \partial S$, then

$$P(s, s') = \begin{cases} \frac{1}{4} & \text{if } \|s' - s\|_1 = h \\ 0 & \text{otherwise} \end{cases}$$

③ Reward $R(s) = \gamma \ell^h(s)$

$$= \frac{h^2}{2+h^2}\left(x_1^2 + x_2^2 - x_1 - x_2 - \frac{3}{2}\right)$$

④ $\gamma = \dfrac{2}{2+h^2}$

$$\|x\|_1 \triangleq \sum_{i=1}^{d} |x_i|$$

A few ways to solve Bellman

① Lin Alg.

② value iteration/Dynamic Programing

③ MC

④ TD (Temporal Difference)