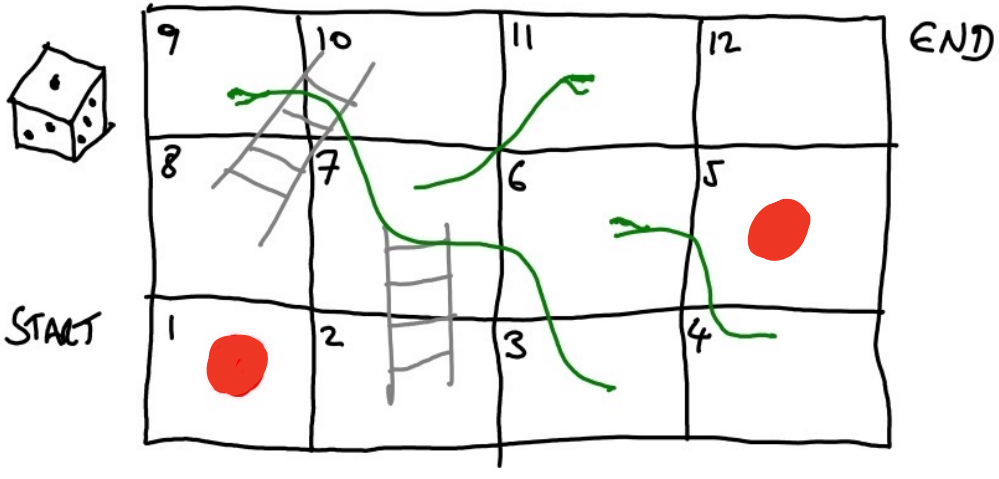
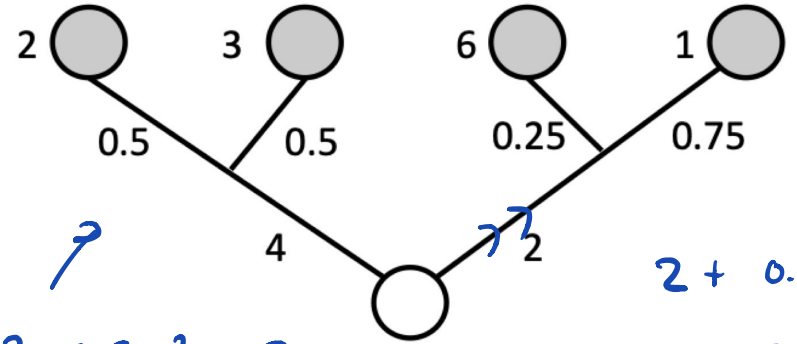
3. Markov Decision Processes

A MARKOV CHAIN WITH DECISIONS:



[CHOOSE WHICH TO SOLVE]

SOLVING AN MDP:



$4 + 2 \times 0.5 + 3 \times 0.5$

$= 4 + 2.5$

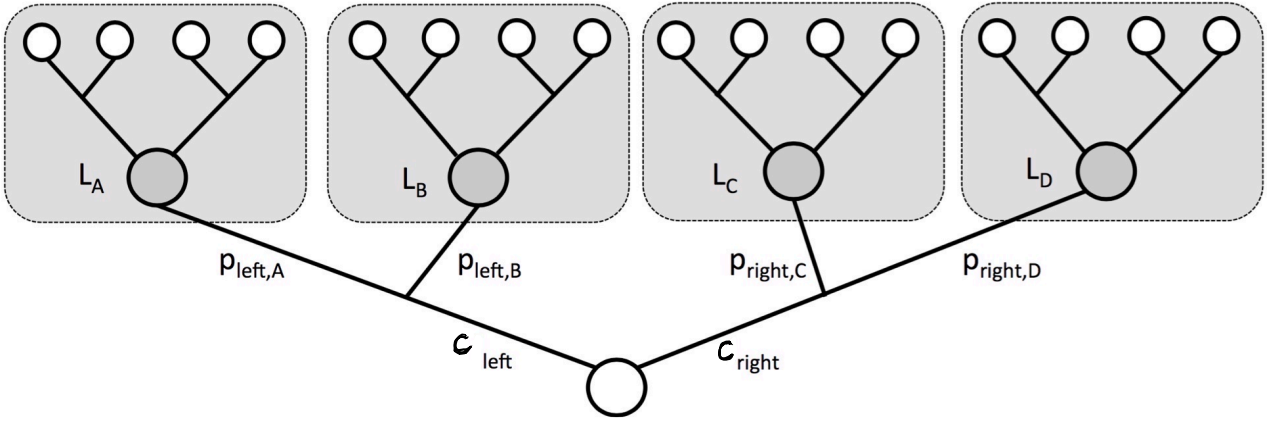
$= 6$

$2 + 0.25 \times 6 + 0.75 \times 1$

$= 2 + 2.25$

$= 4.25$

SOLVING AN MDP:



$$L(x) = \min_{a \in \{\text{LEFT}, \text{RIGHT}\}} \{ c(a) + \mathbb{E}_{\hat{x},a}[L(\hat{x})] \}$$

ABSTRACT DEFINITION:

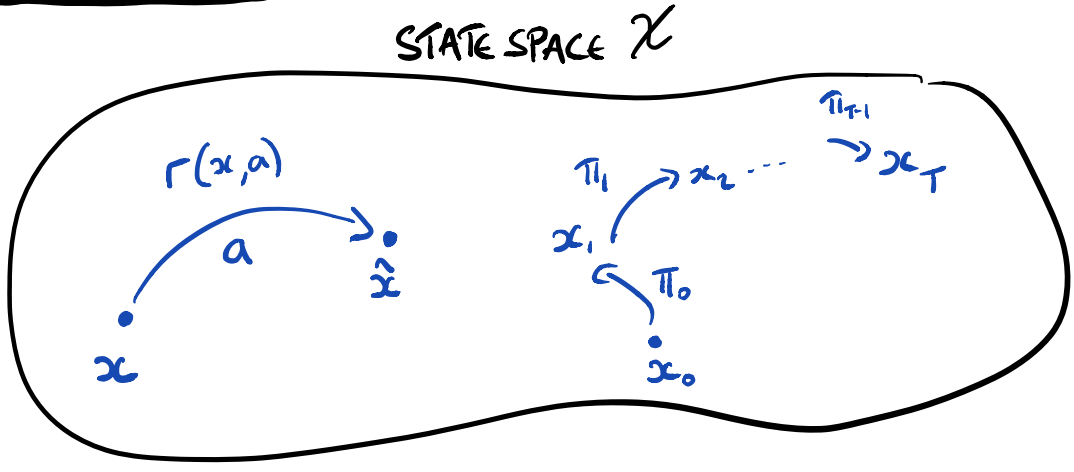
STATE $x \in \mathcal{X}$

ACTION $a \in \mathcal{A}$

REWARD $r(x,a)$

NEXT STATE

$$\hat{x} = f(x,a,u)$$



POLICY $\pi(x)$ OR $\pi_t(x)$

OBJECTIVE: [MAXIMIZE SUM OF REWARDS]

$$V_T(x_0) := \max \underbrace{\mathop{\mathbb{E}}_{x_0}\left[\sum_{t=0}^{T-1} r(x_t,\pi_t) + r(x_T)\right]}_{R_T(x,\pi)} \text{ OVER } \pi \in \underbrace{\mathcal{P}_t}_{\text{POLICY OVER } t \text{ STEPS}}$$

$V_T(x_0)$: VALUE FUNCTION

THE BELLMAN EQUATION

$$\mathbb{E}_{x,a}[V(\hat{x})] = \mathbb{E}[V(X_1) \mid X_0 = x, A_0 = a]$$
$$= \mathbb{E}[V(f(x,a,u))]$$

THEOREM:

$$V_t(x) = \max_{a \in A} \left\{ r(x,a) + \mathbb{E}_{x,a}\left[V_{t-1}(\hat{x})\right] \right\}, \qquad V_0(x) = r(x)$$

THERE $\mathbb{E}_{x,a}[V(\hat{x})] = \mathbb{E}[V(X_1) \mid X_0 = x, A_0 = a] = \mathbb{E}[V(f(x,a,u))]$

PROOF:

$$V_t(x) = \max_{\pi_{T-t}, \dots, \pi_{T-1}} \mathbb{E}_{x, \pi_{T-t}} \left[\sum_{s=T-t}^{T-1} r(X_s, \pi_s) + r(X_T) \right]$$

$$\max_{\pi_{T-1}} \max_{\pi_{T-2}, \dots, \pi_{T-1}} \mathbb{E}_{x_{T-t}, \pi_{T-t}} \mathbb{E}_{X_{T-t+1}, \pi_{T-t+1}} \quad \text{MOVE AS FAR TO RIGHT AS POSSIBLE} \rightarrow$$

$$r(x_{T-t}, \pi_{T-t}) + \sum_{s=T-t+1}^{T-1} r(X_s, \pi_s) + r(X_T)$$

$$= \max_{\pi_{T-1}} \Big\{ r(x_{T-t}, \pi_{T-1}) + \mathbb{E}_{\underbrace{x_{T-t}, \pi_{T-t}}_{x}} \Big[\underbrace{\max_{\pi_{T-2}, \dots, \pi_{T-1}} \mathbb{E}_{X_{T-t+1}, \pi_{T-t+1}} \Big[\sum_{s=T-t+1}^{T-1} r(X_s, \pi_s) + r(X_T) \Big]}_{V_{t-1}(\hat{X})} \Big] \Big\}$$

$$= \max_{a} \Big\{ r(x, a) + \mathbb{E}_{x,a} \big[V_{t-1}(\hat{X}) \big] \Big\}$$

□

Ex 20. *You need to sell a car. At every time $t = 0, \ldots, T-1$, you set a price p_t and a customer then views the car. The probability that the customer buys a car at price p is $D(p)$. If the car isn't sold be time T then it is sold for fixed price V , $V < 1$. Maximize the reward from selling the car and find the recursion for the optimal reward when $D(p) = (1-p)_+$.*

SOLD OR NOT

$r(x, a)$

$V_t(x) = \begin{cases} V_t & \text{IF } x = 0 \text{ NOT SOLD} \\ 0 & \text{IF } x = 1 \text{ SOLD} \end{cases}$

ANSWER:

$$V_t = \max_p \{ p\,D(p) + (1-D(p))\,V_{t-1} \} \qquad V_0 = V$$

$$= \max_p \{ p(1-p) + (1-(1-p))V_{t-1} \}$$

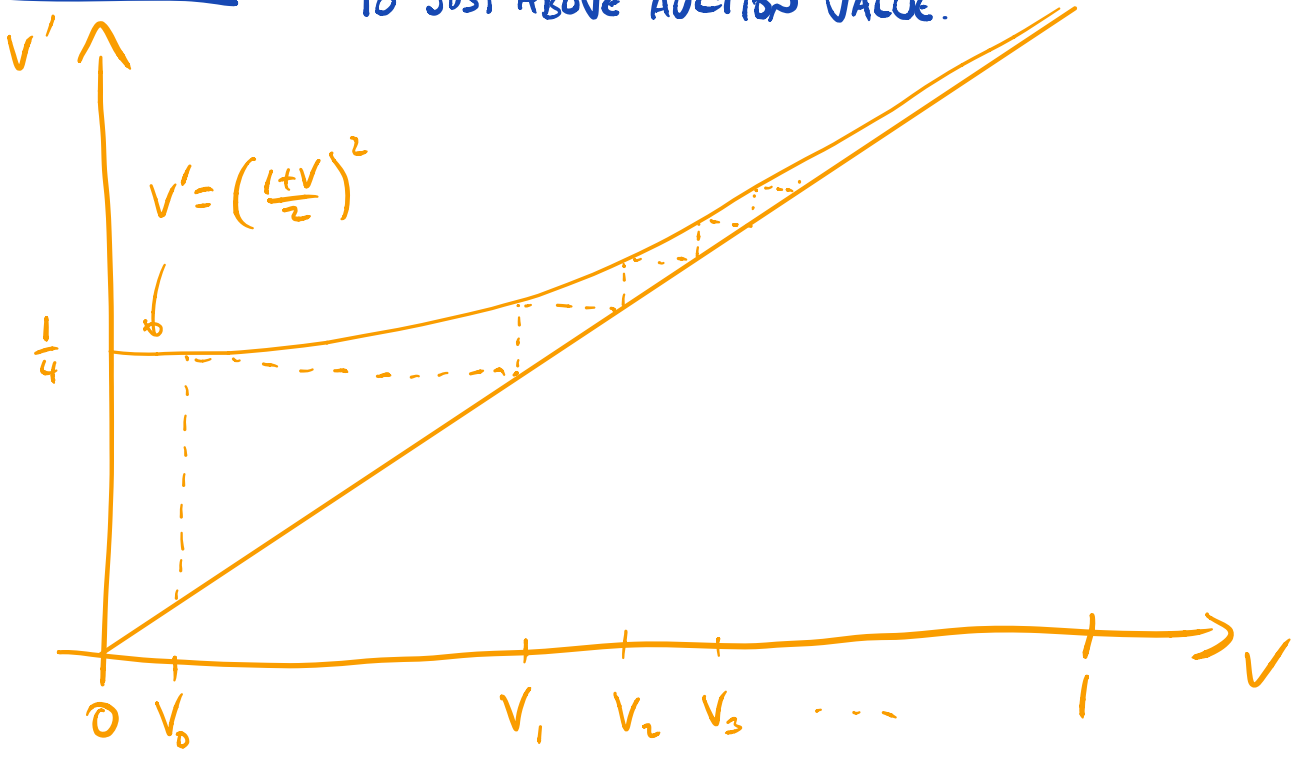
DIFFERENTIATE & SOLVE FOR p GIVES

$$V_t = \left(\frac{1 + V_{t-1}}{2} \right)^2$$

NOTE $V_t > V_{t-1}$ & ONLY FIXED POINT IS $V^* = 1$

SO $V_t \to 1$ AS $t \to \infty$

A PICTURE: START HIGH & SLOWLY DECREASE PRICE TO JUST ABOVE AUCTION VALUE.



Ex 21 (Call Option). *You own a call option with strike price p. Here you can buy a share at price p making profit $X_t - p$ where x_t is the price of the share at time t. The share must be exercised by time T. The price of stock X_t satisfies*

$$X_{t+1} = X_t + \epsilon_t$$

for ϵ_t IIDRV with finite expectation. Show that there exists a decresing sequence $\{a_t\}_{0 \le t \le T}$ such that it is optimal to exercise whenever $X_s \ge a_s$ occurs.

ANSWER: BELLMAN EQN IS

$$V_t(x) = \text{MAX}\{x - p, \mathbb{E}[V_{t-1}(x+\varepsilon)]$$

1). POLICIES WITH t STEPS LEFT CONTAIN ALL POLICIES WITH $t-1$ STEPS

SO $V_t(x) \ge V_{t-1}(x)$.

2) $V_t(x) - x = \text{MAX}\{-p, \mathbb{E}[V_{t-1}(x-\varepsilon) - x]\}$ IS DECREASING CONTINUOUS IN x.

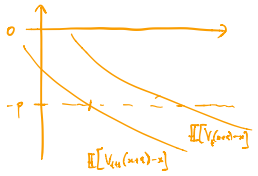
(DECREASING)

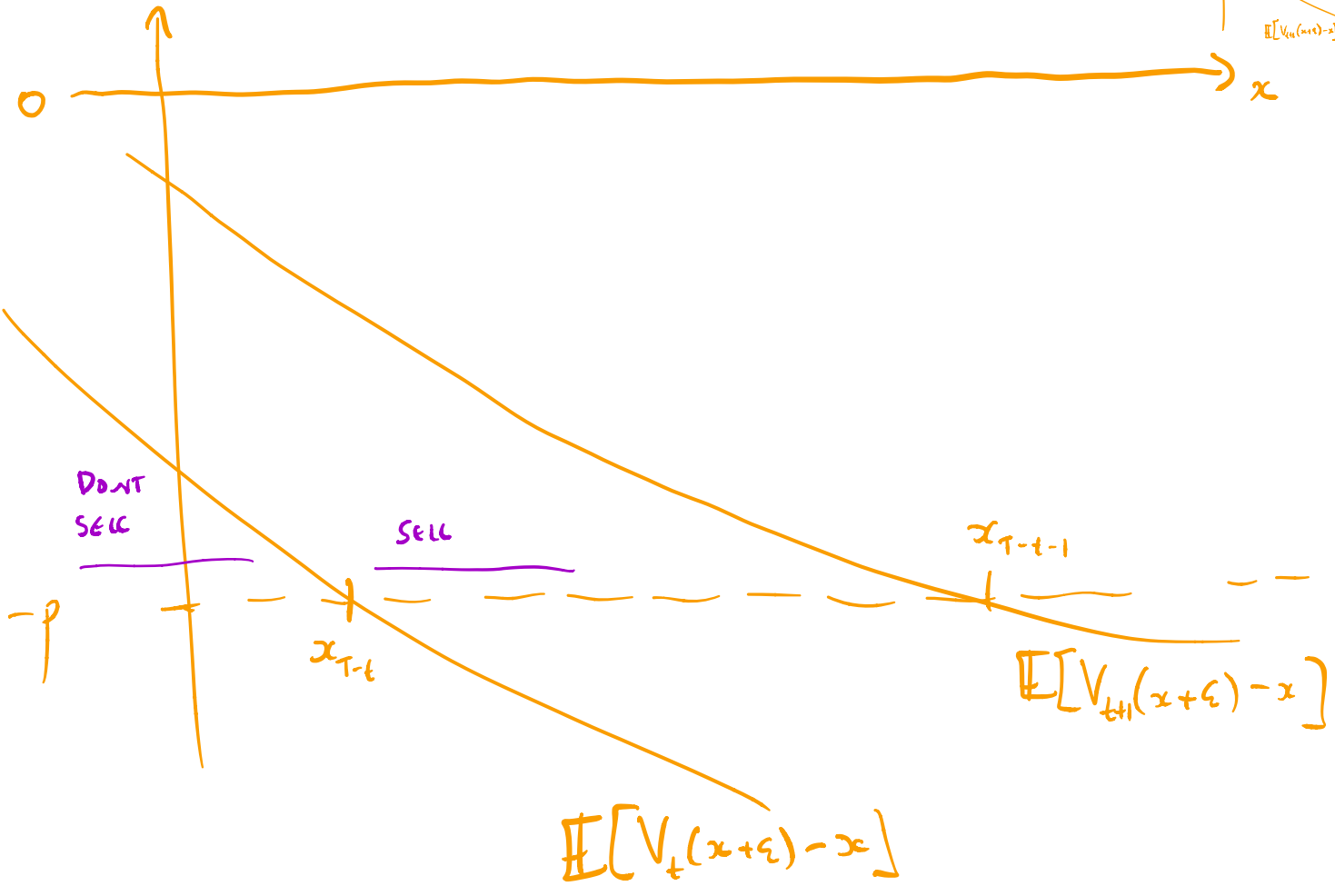
BECAUSE $V_0(x) - x = \text{MAX}\{-p, -x\}$ HAS THESE PROPERTIES. & MAX & EXPECTATION PRESERVES THESE PROPERTIES.

$\therefore$ DECISION SWITCHES AT x_{T-t} s.t. $-p = \mathbb{E}[V_{t-1}(x_{T-t} - \varepsilon) - x_{T-t}]$

x_t DECREASES BECAUSE V_t INCREASES

A PICTURE:


0
-p
$\mathbb{E}[V_{t+1}(x+\xi)-x]$
$\mathbb{E}[V_t(x+\xi)-x]$


0
x
DONT SELL
SELL
x_{T-t-1}
-p
x_{T-t}
$\mathbb{E}[V_{t+1}(x+\xi)-x]$
$\mathbb{E}[V_t(x+\xi)-x]$