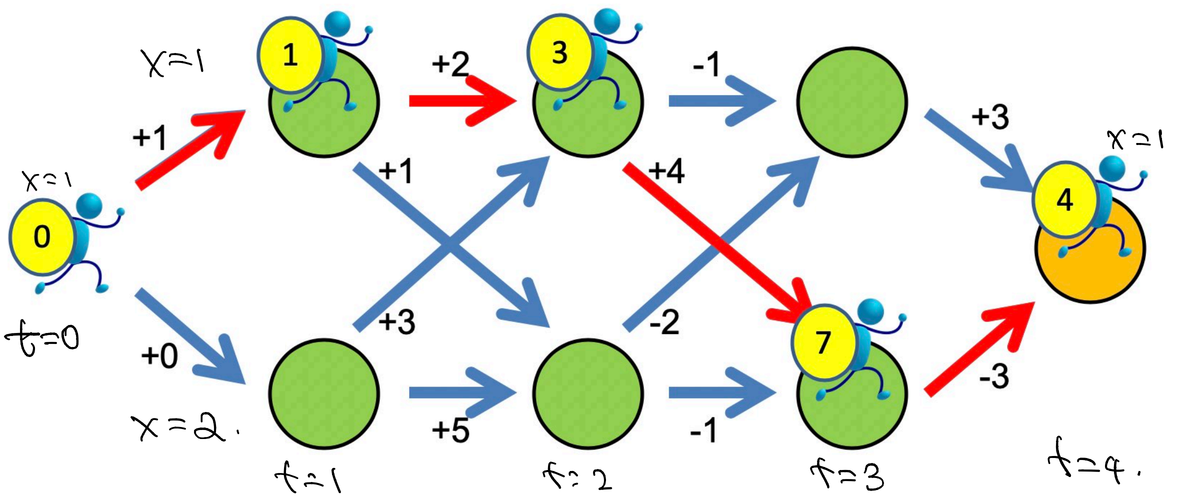
Lecture 4



x=1
1
+2
3
-1
+3
+1
x=1
x=1
+1
+4
0
4
t=0
+3
-2
7
+0
-3
x=2.
+5
-1
t=1
t=2
t=3
t=4.

$x = 1, 2, \cdots, N$.

$t = 0, 1, 2, \cdots, T-1$

\# of paths $= N^T$

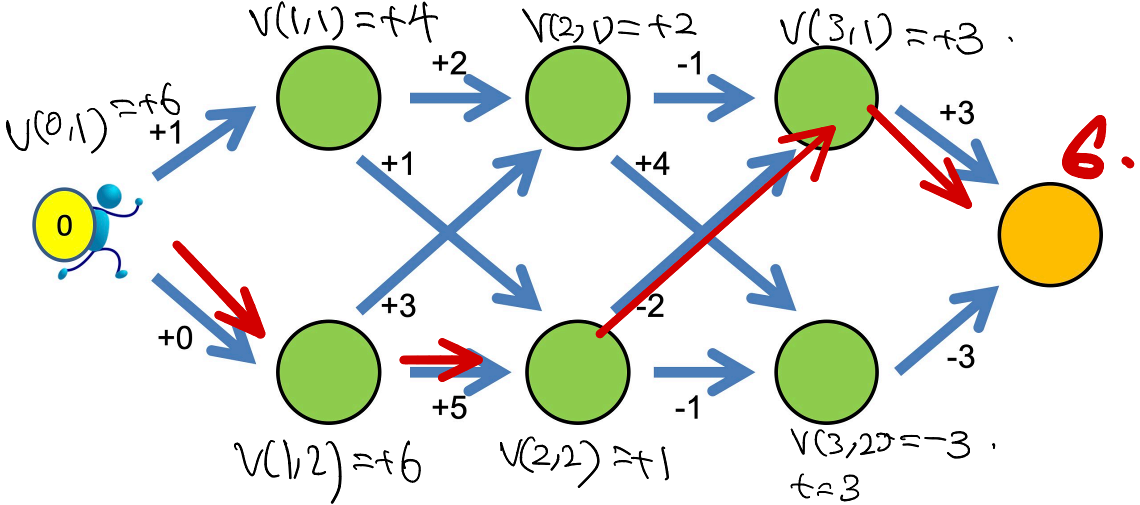
Define

- $S(t)$: state at time t $\quad (S(t) = 1, 2)$
- $V(t, x) := \max \left\{ \sum_{s=t+1}^{T} \underbrace{R(t)}_{\text{score, reward.}} : S(t) = x \right\}$.
- $V(0, 1)$ is the best we can do
- Goal : find $V(0,1)$ and optimal strategy.

Cost, $V(t,x)$, need N computations.

\# $V(t,x) = NT$

\# of operations $\Rightarrow N^2 T$



V(0,1)=+6
V(1,1)=+4
V(2,1)=+2
V(3,1)=+3
+1
+0
+2
-1
+3
+1
+4
+3
-2
+5
-1
-3
0
6
V(1,2)=+6
V(2,2)=+1
V(3,2)=-3
t=3

Dynamic Programming Principle

Bolza problem

$$\inf_{\underline{\theta} \in L^\infty} J[\underline{\theta}] = \int_{t_0}^{t_1} L(t, x(t), \theta(t))\, dt + \Phi(x(t_1))$$

$$\text{s.t.} \quad \dot{x}(t) = f(t, x(t), \theta(t)), \quad x(t_0) = x_0 .$$

Define the value function $V : [t_0, t_1] \times \mathbb{R}^d \to \mathbb{R}$.

$$V(s, z) = \inf_{\underline{\theta}} \int_{s}^{t_1} L(t, x(t), \theta(t))\, dt + \Phi(x(t_1))$$

$$\text{s.t.} \quad \dot{x}(t) = f(t, x(t), \theta(t)), \quad x(s) = z$$

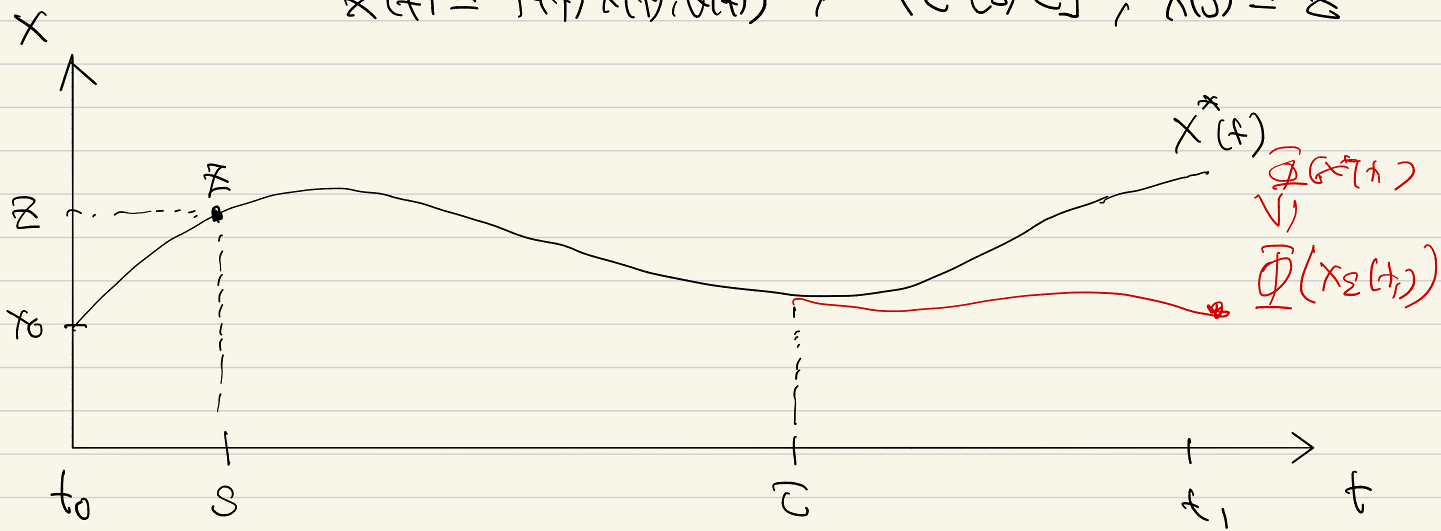
$s = 0,\ z = x_0 \Rightarrow V(0, x_0)$

Theorem (Dynamic Prog. Principle, DPP)

For every $\tau, s \in [t_0, t_1]$, $s \le \tau$, $z \in \mathbb{R}^d$, we have.

$$V(s,z) = \inf_{\theta} \left\{ \int_s^\tau L(t, x(t), \theta(t))\,dt + V(\tau, x(\tau)) \right\}$$

$$\dot{x}(t) = f(t, x(t), \theta(t)), \quad t \in [s, \tau], \quad x(s) = z$$



Proof:

Define $J^\tau := \inf_{\underline{\theta}} \left\{ \int_s^\tau L(t, x(t), \theta(t))\,dt + V(\tau, x(\tau)) \right\}$.

① " $J^\tau \leq V(s, z)$ "

Fix $\varepsilon > 0$, pick $\underline{\theta} : [s, t_1] \to$ (int) s.t.

$$J[\underline{\theta}] \leq V(s, z) + \varepsilon . \qquad (\text{by defn of inf in } V)$$

Under this control, we have.

$$V(\tau, x(\tau)) \leq \int_\tau^{t_1} L(t, x(t), \theta(t))\,dt + \Phi(x(t_1))$$

$$\Rightarrow \quad J^\tau \leq \int_s^\tau L(t, x(t), \theta(t))\,dt + V(\tau, x(\tau))$$

$$\leq \left(\int_s^\tau + \int_\tau^{t_1} \right) L(t, x(t), \theta(t))\,dt + \Phi(x(t_1)) = J[\underline{\theta}]$$

$$\leq V(s, z) + \varepsilon .$$

$$\Rightarrow \quad J^\tau \leq V(s, z) .$$

Step 2: " $J^\tau \geq V(s,z)$ "

Fix $\varepsilon > 0$, then $\exists$ $\underline{\theta}_1 : [s,\tau] \to \Theta$ s.t.

$$\int_s^\tau L(t, x(t), \theta_1(t))\,dt + V(\tau, x(\tau)) \leq J^\tau + \varepsilon .$$

$\exists$ $\underline{\theta}_2 : [\tau, t_1] \to \Theta$ s.t.

$$\int_\tau^{t_1} L(t, x(t), \theta_2(t))\,dt + \Phi(x(t_1)) \leq V(\tau, x(\tau)) + \varepsilon .$$

$+$

Define

$$\theta(t) = \begin{cases} \theta_1(t) & t \in [s,\tau] \\ \theta_2(t) & t \in [\tau, t_1] \end{cases}$$

$$\underbrace{\int_s^{t_1} L(t, x(t), \theta(t))\,dt + \Phi(x(t_1))}_{V(s,z) \leq J[\underline{\theta}]} \leq J^\tau + 2\varepsilon .$$

$\Rightarrow$ $V(s,z) \leq J^\tau \cancel{+ 2\varepsilon}$. $\square$.

Hamilton-Jacobi-Bellman Equation (HJB)

$\dot{x} = f \cdots \quad x(s) = z.$

DPP : $V(s,z) = \inf_{\theta} \left\{ \int_s^{\tau} L(t, x(t), \theta(t))\,dt + V(\tau, x(\tau)) \right\}$

$\downarrow$ infinitesimal $\tau = s + \Delta s.$

$$V(s,z) = \inf_{\theta} \left\{ \int_s^{s+\Delta s} L(\cdots)\,dt + V(s+\Delta s, x(s+\Delta s)) \right\}$$

$$\begin{aligned} x(s+\Delta s) &= x(s) + \int_s^{s+\Delta s} f(t, x(t), \theta(t))\,dt \\ &= \underbrace{x(s)}_{z} + \Delta s\, f(s, x(s), \theta(s)) + o(\Delta s) \end{aligned}$$

$$\begin{aligned} V(s+\Delta s, x(s+\Delta s)) &= V(s,z) + \partial_s V(s,z) \cdot \Delta s \\ &\quad + [\nabla_z V(s,z)]^T f(s, z, \theta(s))\,\Delta s + o(\Delta s) \end{aligned}$$

$$\int_S^{S+\Delta S} L(t, x(t), \theta(t))\, dt = L(S, z, \theta(S))\,\Delta S + o(\Delta S)$$

$$\cancel{V(S,z)} = \inf_{\theta} \left\{ \cancel{V(S,z)} + \Delta S \left[\partial_S V(S,z) + [\nabla_z V(S,z)]^T f(S,z,\theta(S)) + L(S,z,\theta(S)) \right] \cancel{+ o(\Delta S)} \right\}$$

$$\Rightarrow \begin{cases} \partial_S V(S,z) + \inf_{\theta \in \Theta} \left\{ [\nabla_z V(S,z)]^T f(S,z,\theta) + L(S,z,\theta) \right\} = 0 \\ V(t_1, z) = \Phi(z). \end{cases}$$

(terminal cost)

Hamilton-Jacobi-Bellman Equation.

$V(t_1, z) = \Phi(z)$ $\quad \underline{\{V(t)\}}$ flow. $\quad \partial_t V = H(t, \nabla V)$

$$\partial_s V + \inf_{\theta \in \mathbb{R}} \left\{ (\partial_z V)\,\theta + \frac{1}{2}\theta^2 \right\} = 0$$

$$\Rightarrow \quad \underline{\partial_s V - \frac{1}{2}|\partial_z V|^2 = 0} \,.$$

Implications of HJB

$$V(t_0, x_0) = \inf_{\underline{\theta}} J[\underline{\theta}]$$ ← original control problem.

(assume)

Fix s, z $\exists$ $\underline{\theta}^*_{s,z} [s, t_1] \to \Theta$ is optimal.

DPP $\Rightarrow$

$$V(s, z) = \inf_{\theta} \left\{ \int_s^\tau L(t, x(t), \theta(t))\,dt + V(\tau, x(\tau)) \right\}$$

$$= \int_s^\tau L(t, x^*_{s,z}(t), \theta^*_{s,z}(t))\,dt + V(\tau, x^*_{s,z}(\tau))$$

$\{\theta^* := \theta^*_{t_0, x_0}\}$

$\downarrow$ Taylor expansion.

$$-\partial_s V(s, x^*(s)) = \boxed{L(s, x^*(s), \theta^*(s)) + (\nabla_z V(s, x^*(s)))^T f(s, x^*(s), \theta^*(s))}$$

$$= \boxed{\min_{\theta} \left\{ L(\cdots) + (\nabla_z V \quad)^T f(\cdots) \right\}}$$

$$-L(s, x^*(s), \theta^*(s)) - \nabla_z V(s, x^*(s))^T f(s, x^*(s), \theta^*(s))$$
$$\geq -L(s, x^*(s), \theta) - \nabla_z V(s, x^*(s))^T f(s, x^*(s), \theta) \quad \forall\, \theta \in \Theta$$

define $p^*(s) = -\nabla_z V(s, x^*(s))$

$$\underbrace{p^*(s)^T f(s, x^*(s), \theta^*(s)) - L(s, x^*(s), \theta^*(s))}_{H} \geq p^*(s)^T f(s, x^*(s), \theta) - L(s, x^*(s), \theta) \quad \forall\, \theta \in \Theta \qquad (*)$$

$\Rightarrow$ Still a necessary condition.

Sufficient Condition

Assume $\hat{\underline{\theta}} : [t_0, t_1] \to \Theta$ satisfies (*)

$\hat{x}$ be its controlled trajectory.

$$\Rightarrow \underbrace{\partial_t V(t, \hat{x}(t)) + [\nabla_x V(t, \hat{x}(t))]^T f(t, \hat{x}(t), \hat{\theta}(t))}_{\frac{d}{dt} V(t, \hat{x}(t))} + L(t, \hat{x}(t), \hat{\theta}(t)) = 0$$

$\Rightarrow$ integrate from t_0 to t_1

$$\underbrace{V(t_1, \hat{x}(t_1))}_{\Phi(\hat{x}(t_1))} - \underbrace{V(t_0, \hat{x}(t_0))}_{\inf_{\underline{\theta}} J[\theta]} + \underbrace{\int_{t_0}^{t_1} L(\cdots)\, dt}_{\text{running cost}} = 0.$$

$+$

$J[\hat{\theta}]$

$$\Rightarrow \quad J[\hat{\theta}] = \inf_{\theta} J[\theta]$$

Generate optimal control

① Solve HJB. $\rightarrow V$.

② $$\theta^*(t) = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\{ \nabla_x V(t, x^*(t))^T f(t, x^*(t), \theta) + L(t, x^*(t), \theta) \right\}$$

closed loop.