Markov decision processes (MDP) / stochastic optimal control problems (SOCP)

goal: mathematical model for optimal control with random dynamics / cost

refs: *Neuro-Dynamic Programming*

Dimitri P. Bertsekas and John N. Tsitsiklis

Chapter 2

Reinforcement Learning

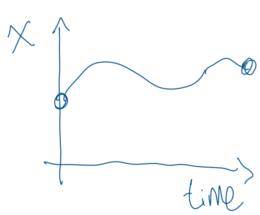
An Introduction

Richard S. Sutton and Andrew G. Barto

Chapter 3

---

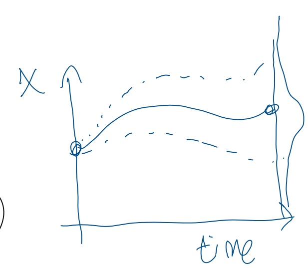
- analogous to how MP/SDE naturally generalizes deterministic dynamics, we can consider optimal control problems with random dynamics:




$$\min_u c(x,u) \quad \text{s.t. } x^+ = F(x,u)$$

$$\longrightarrow$$

$$\min_u \cancel{c(x,u)} \quad \text{s.t. } x^+ \sim P(x,u)$$




$\longrightarrow$ what should we use for the cost function?

i.e. what statistic should we choose? (e.g. $E[\cdot]$, $Var[\cdot]$)

– $E[c(x,u)]$ is the average / "expected" cost

$\rightsquigarrow$ useful when system will run many times and/or unlikely outcomes are ok

- $\mathrm{Var}[c(x,u)]$ is the "spread" in cost
  ⇝ useful when system will run few times and/or unlikely outcomes catastrophic
- also consider order statistics (median, interquartile, confidence intervals)

* we'll focus on minimizing expected cost:
$$\min_{u} E[c(x,u)] \quad \text{s.t. } x^+ \sim P(x,u)$$

where
$$c(x,u) = \ell(t, x_t) + \sum_{s=0}^{t-1} \mathcal{L}(s, x_s, u_s) \quad \text{— finite-horizon}$$
or
$$= \lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \mathcal{L}(s, x_s, u_s) \quad \text{— infinite-horizon, time-averaged}$$
or
$$= \lim_{t\to\infty} \sum_{s=0}^{t} \gamma^s \cdot \mathcal{L}(s, x_s, u_s) \quad \text{— infinite-horizon, exponentially-discounted}$$

$\gamma \in (0,1)$ termed a <u>discount factor</u>

<u>def</u>: <u>Markov decision process</u> (MDP) / <u>stochastic optimal control prob.</u> (SOCP) specified by $(X, \mathcal{U}, P, c)$ where $X, \mathcal{U}$ are sets,

$P : X \times \mathcal{U} \to \Delta(X)$, $c : X^T \times \mathcal{U}^T \to \mathbb{R}$, $T$ is a time interval, e.g. $= [0,t]$ or $= [0,\infty)$

$X^T = \{x : T \to X\}$