Optimal Control Theory and Linear Quadratic Regulators

Recap:
Finite horizon Markov Decision Process

$$\mathscr{M} = \left\{S, A, r, P, H, \mu_0\right\},$$
$$r : S \times A \mapsto [0,1], H \in \mathbb{N}^+, P : S \times A \mapsto \Delta(S), s_0 \sim \mu_0$$

We need to consider time-dependent policies, i.e.,

$$\pi := \left\{\pi_0, \pi_1, \ldots, \pi_{H-1}\right\}, \pi_h : S \mapsto A, \forall h$$

Policy interacts with the MDP as follows:

$$\tau = \{s_0, a_0, s_1, a_1, \ldots, s_{H-1}, a_{H-1}, s_H\}, s_0 \sim \mu_0, a_0 = \pi_0(s_0), s_1 \sim P(\cdot \mid s_0, a_0), a_1 = \pi_1(s_1), \ldots$$

Recap: V/Q functions in Finite horizon MDP

$$V_h^\pi(s) = \mathbb{E}\left[\sum_{\tau=h}^{H-1} r(s_\tau, a_\tau) \,|\, s_h = s, a_\tau = \pi_\tau(s_\tau), s_{\tau+1} \sim P(\cdot \,|\, s_\tau, a_\tau)\right]$$

$$Q_h^\pi(s, a) = \mathbb{E}\left[\sum_{\tau=h}^{H-1} r(s_\tau, a_\tau) \,|\, (s_h, a_h) = (s, a), a_\tau = \pi_\tau(s_\tau), P\right]$$

Bellman Equation:

$$Q_h^\pi(s, a) = r(s, a) + \mathbb{E}_{s' \sim P(\cdot|s,a)}\left[V_{h+1}^\pi(s')\right]$$

Example: Compute Optimal Policy

Example (Continue)

Recap: Formalizing the process

$$\pi^\star = \{\pi_0^\star, \pi_1^\star, \ldots, \pi_{H-1}^\star\}$$

We use Dynamic Programming, and do DP backward in time; start at $H-1$

$$Q_{H-1}^\star(s,a) = r(s,a) \qquad \pi_{H-1}^\star(s) = \arg\max_a Q_{H-1}^\star(s,a)$$

$$V_{H-1}^\star(s) = \max_a Q_{H-1}^\star(s,a) = Q_{H-1}^\star(s, \pi_{H-1}^\star(s))$$

Recap: Formalizing the process

$$\pi^\star = \{\pi_0^\star, \pi_1^\star, \ldots, \pi_{H-1}^\star\}$$

We use Dynamic Programming, and do DP backward in time; start at $H-1$

$$Q_{H-1}^\star(s,a) = r(s,a) \qquad \pi_{H-1}^\star(s) = \arg\max_a Q_{H-1}^\star(s,a)$$

$$V_{H-1}^\star(s) = \max_a Q_{H-1}^\star(s,a) = Q_{H-1}^\star(s, \pi_{H-1}^\star(s))$$

Now assume that we have already computed $V_{h+1}^\star$, $h \leq H-2$

(i.e., we know how to perform optimally at $h+1$)

$$Q_h^\star(s,a) = r(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)} V_{h+1}^\star(s')$$

$$\pi_h^\star(s) = \arg\max_a Q_h^\star(s,a)$$

State-action Distributions

Given $\pi := \{\pi_0, \ldots, \pi_{H-1}\}$

Define $\mathbb{P}_h^\pi(s, a; \mu_0)$: the probability of reaching (s, a) at time step h following π from μ_0

State-action Distributions

Given $\pi := \{\pi_0, \ldots, \pi_{H-1}\}$

Define $\mathbb{P}_h^\pi(s, a; \mu_0)$: the probability of reaching (s, a) at time step h following π from μ_0

We define average state-action distribution as:

$$d^\pi(s, a) = \frac{1}{H} \sum_{h=0}^{H-1} \mathbb{P}_h^\pi(s, a; \mu_0)$$

State-action Distributions

Given $\pi := \{\pi_0, \ldots, \pi_{H-1}\}$

Define $\mathbb{P}_h^\pi(s, a; \mu_0)$: the probability of reaching (s, a) at time step h following π from μ_0

We define average state-action distribution as:

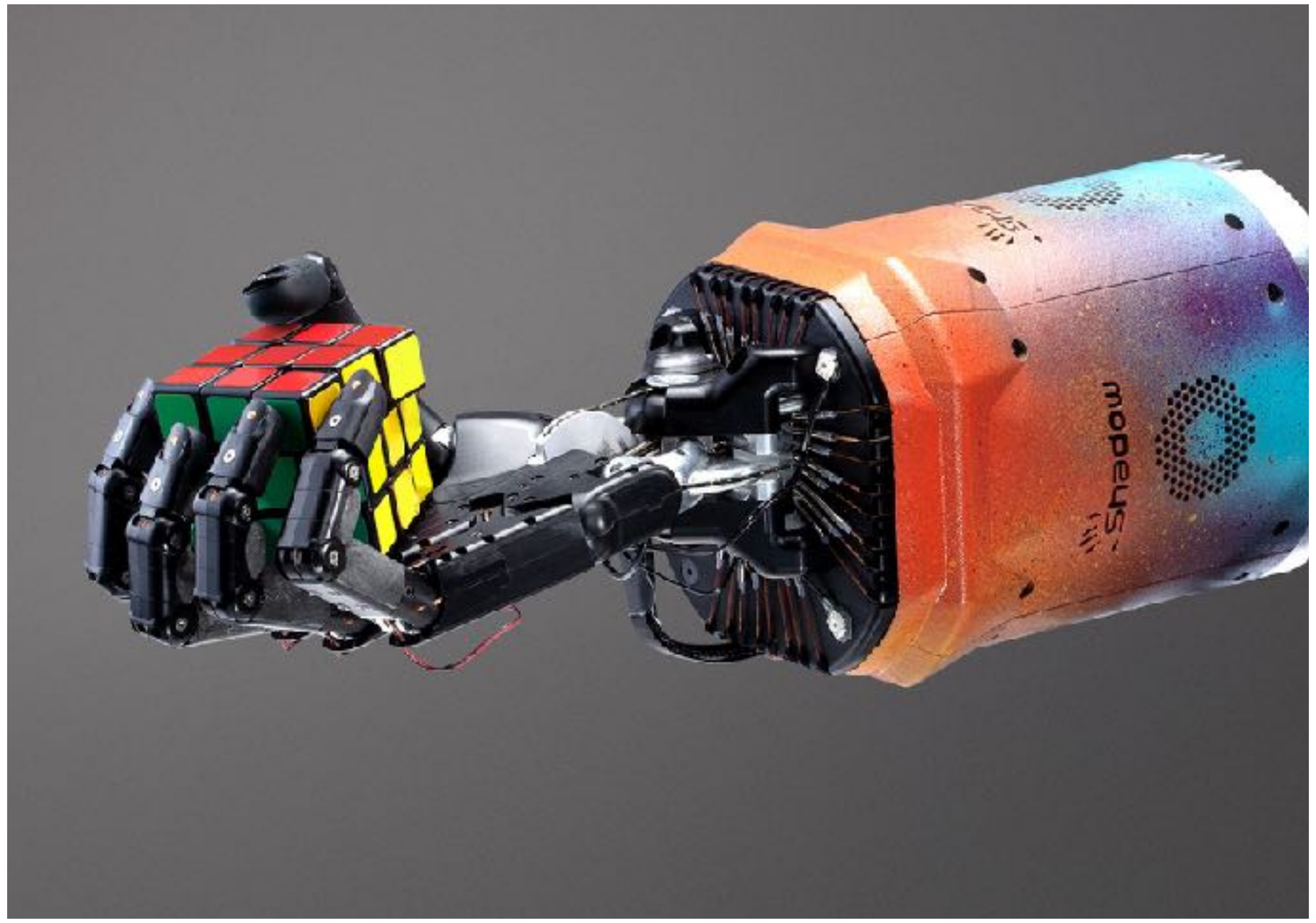
$$d^\pi(s, a) = \frac{1}{H} \sum_{h=0}^{H-1} \mathbb{P}_h^\pi(s, a; \mu_0)$$

Writing Expected total reward using $d^\pi(s, a)$:

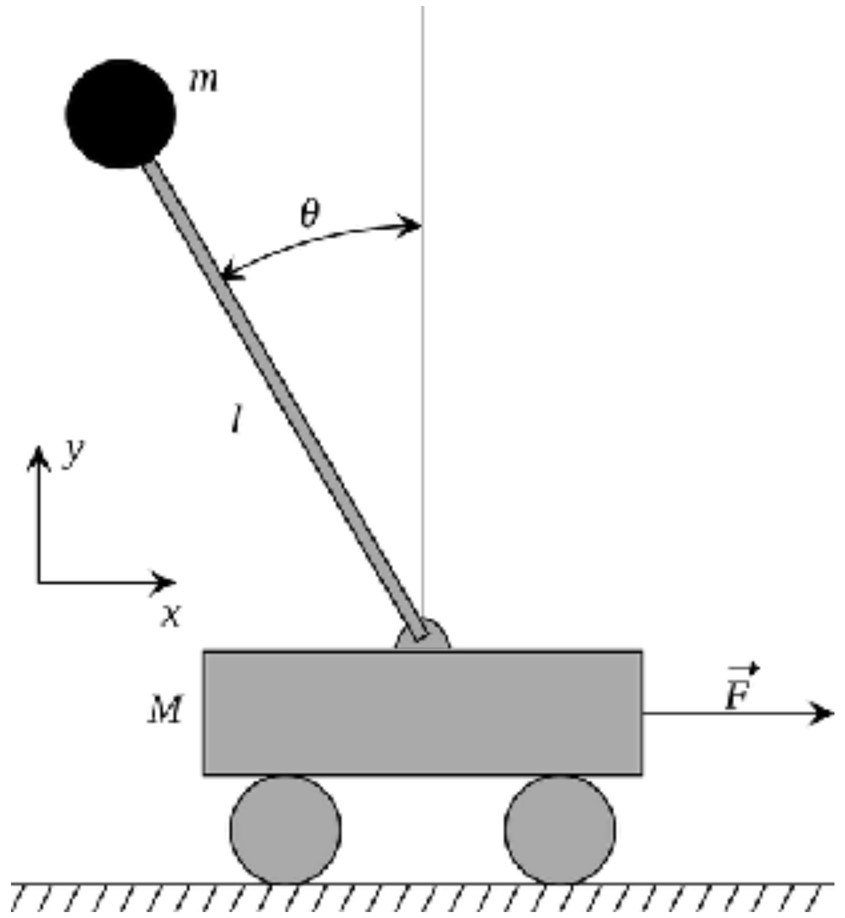
$$\mathbb{E}_{s_0 \sim \mu_0} \left[V_h^\pi(s_0) \right] = H \mathbb{E}_{s,a \sim d^\pi} \left[r(s, a) \right]$$

Robotics and Controls

Dexterous Robotic Hand Manipulation
OpenAI, 2019



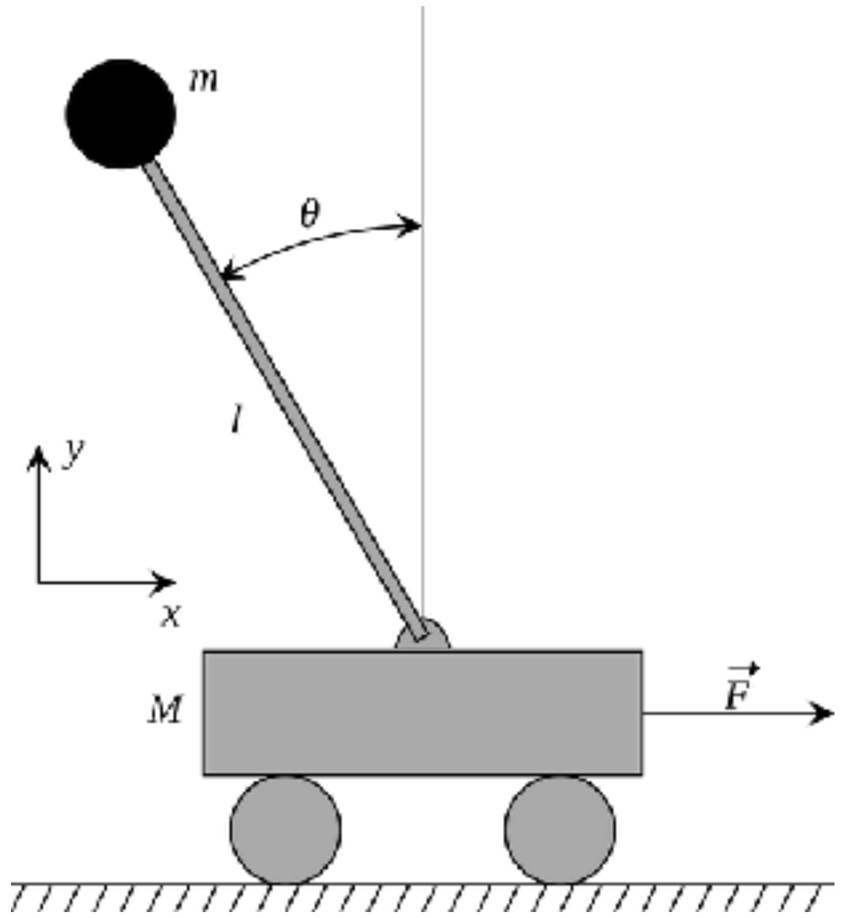
Example: CartPole



State: position and velocity of the cart, angle and angular velocity of the pole

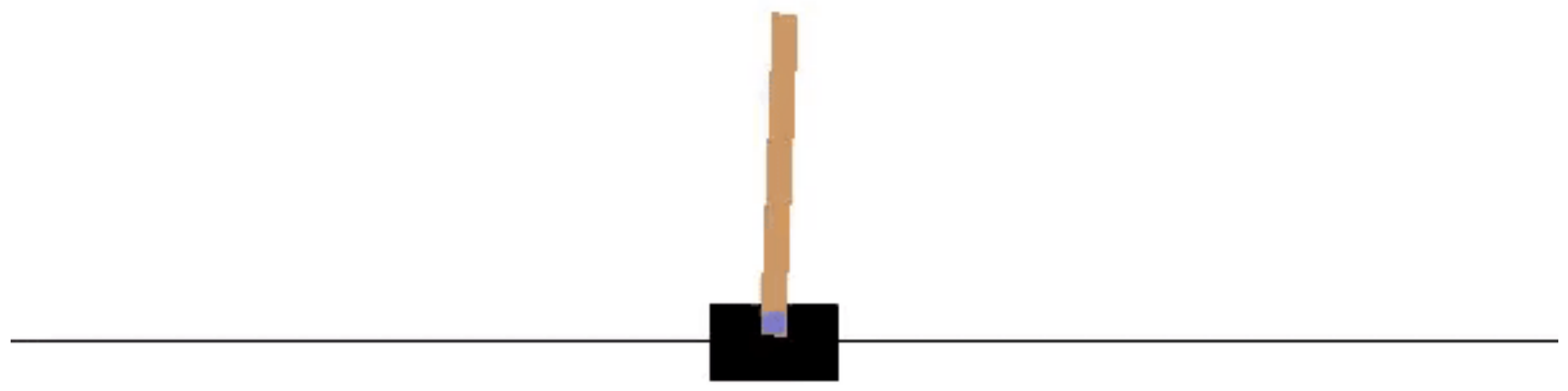
Control: force on the cart

Example: CartPole

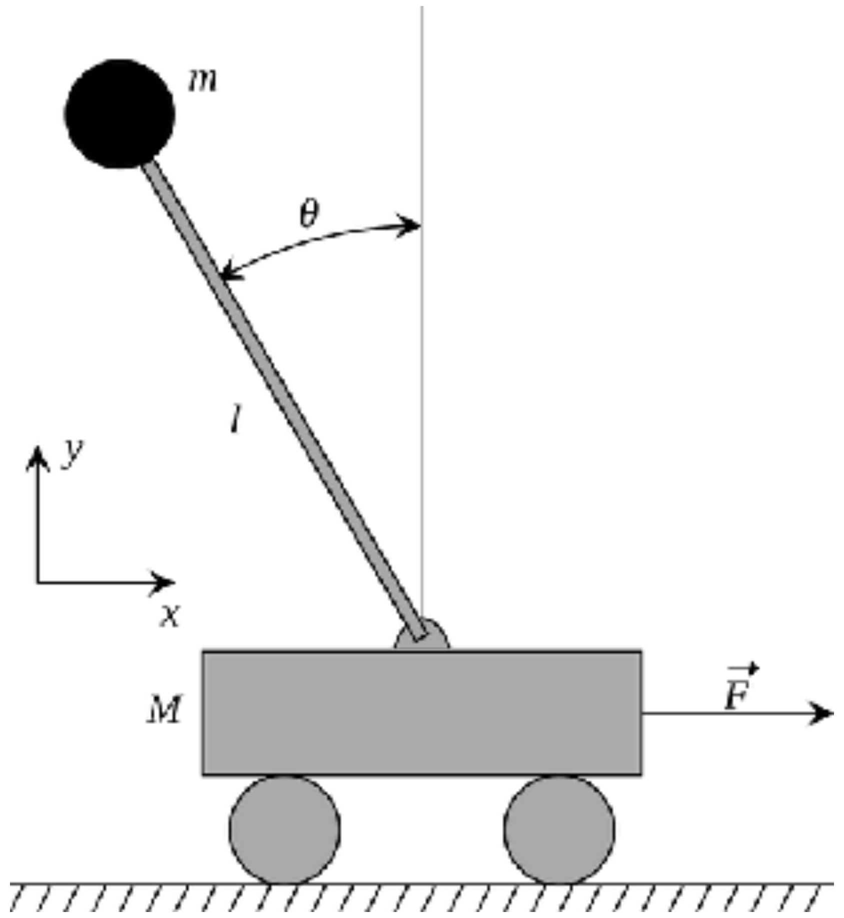


State: position and velocity of the cart, angle and angular velocity of the pole

Control: force on the cart

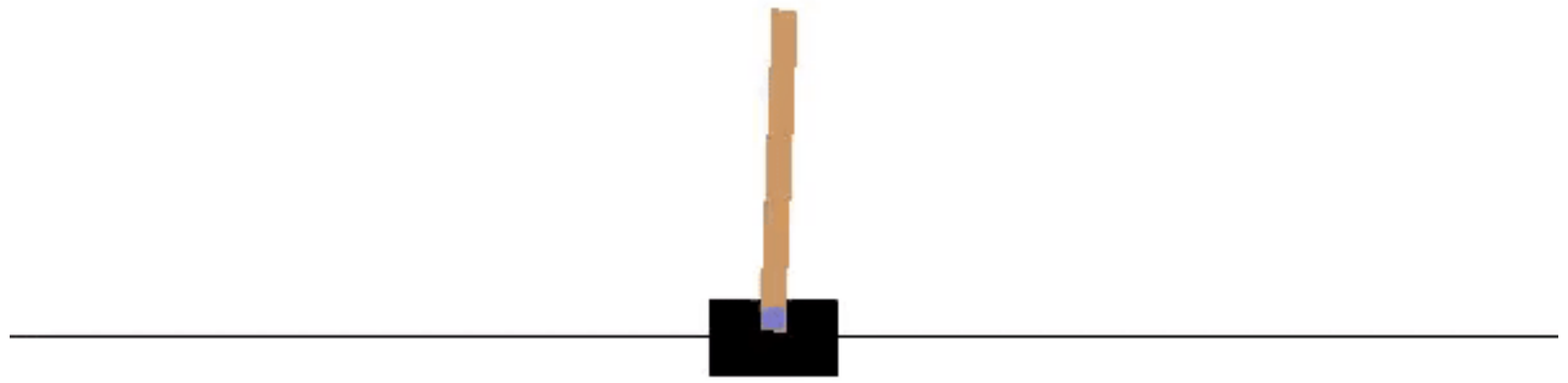


Example: CartPole



State: position and velocity of the cart, angle and angular velocity of the pole

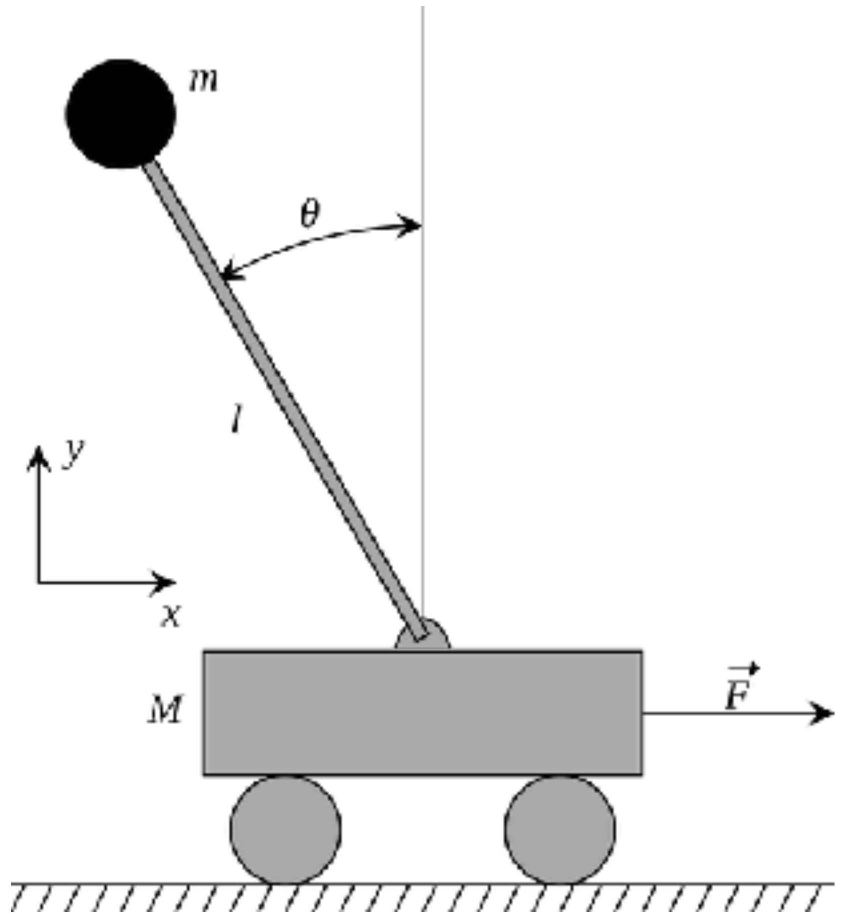
Control: force on the cart



Goal: stabilizing around the point $(x = x^\star, u = 0)$

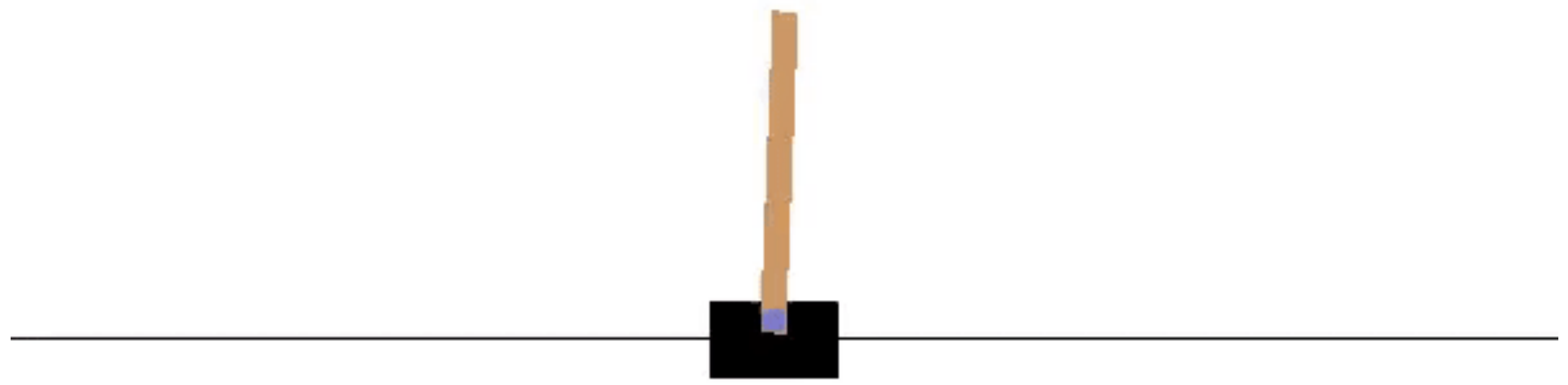
$$c(x_h, u_h) = u_h^\top R u_h + (x_h - x^\star)^\top Q (x_h - x^\star)$$

Example: CartPole



State: position and velocity of the cart, angle and angular velocity of the pole

Control: force on the cart



Goal: stabilizing around the point $(x = x^\star, u = 0)$

$$c(x_h, u_h) = u_h^\top R u_h + (x_h - x^\star)^\top Q (x_h - x^\star)$$

Optimal control:

$$\min_{\pi: X \to U} \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, u_h)\right], \text{s.t., } x_{h+1} = f(x_h, u_h), x_0 \sim \mu_0$$

More Generally: Optimal Control

- a dynamical system is described as
  $x_{h+1} = f_h(x_h, u_h, w_h)$
  where f_h maps a state $x_h \in R^d$, a control (the action) $u_h \in R^k$, and a disturbance w_h, to the next state $x_{h+1} \in R^d$, starting from an initial state $x_0 \sim \mu_0$.

More Generally: Optimal Control

- a dynamical system is described as
  $x_{h+1} = f_h(x_h, u_h, w_h)$
  where f_h maps a state $x_h \in R^d$, a control (the action) $u_h \in R^k$, and a disturbance w_h, to the next state $x_{h+1} \in R^d$, starting from an initial state $x_0 \sim \mu_0$.

- The objective is to find the control policy π which minimizes the long term cost,
  minimize $\mathbb{E}_\pi \left[c_H(x_H) + \sum_{h=0}^{H-1} c_h(x_h, u_h) \right]$
  such that $x_{h+1} = f_h(x_h, u_h, w_h), u_h = \pi(x_h), x_0 \sim \mu_0$
  where H is the time horizon and where w_h is either statistical (e.g., Gaussian noise) or deterministic (e.g., constant deviation)

Reduce Continuous control to Discrete MDP?

$$x \in \mathbb{R}^d, u \in \mathbb{R}^k$$

Reduce Continuous control to Discrete MDP?

$$x \in \mathbb{R}^d, u \in \mathbb{R}^k$$

Curse of dimensionality:

the number of discretized points are approximately $(1/\epsilon)^d + (1/\epsilon)^k$

Today:
The LQR Model

The Linear Quadratic Regulator (LQR)

$$\min_{\pi_0,\ldots,\pi_{H-1}} \quad E\left[x_H^\top Q x_H + \sum_{h=0}^{H-1} (x_h^\top Q x_h + u_h^\top R u_h)\right]$$

such that $\quad x_{h+1} = Ax_h + Bu_h + w_h, \;\; u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \; w_h \sim N(0, \sigma^2 I)\,,$

Here, $x_h \in \mathbb{R}^d, u_h \in \mathbb{R}^k,$

The Linear Quadratic Regulator (LQR)

$$\min_{\pi_0,\ldots,\pi_{H-1}} \quad E\left[x_H^\top Q x_H + \sum_{h=0}^{H-1} (x_h^\top Q x_h + u_h^\top R u_h)\right]$$

such that $\quad x_{h+1} = Ax_h + Bu_h + w_h, \;\; u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \; w_h \sim N(0,\sigma^2 I)\,,$

Here, $x_h \in \mathbb{R}^d, u_h \in \mathbb{R}^k,$

Studied often in theory, but less relevant in practice (?)
(largely due to that time homogenous, globally linear models are rarely good approximations)

Example: 1-d Vehicle

State $x = (p, v)$, i.e., 1-d position and its velocity

[Example credit: Stanford EE 103]

Example: 1-d Vehicle

State $x = (p, v)$, i.e., 1-d position and its velocity

Control u, 1-d force, Friction force $-\eta v$, Vehicle mass m,

[Example credit: Stanford EE 103]

Example: 1-d Vehicle

State $x = (p, v)$, i.e., 1-d position and its velocity

Control u, 1-d force, Friction force $-\eta v$, Vehicle mass m,

Consider discrete time $t = 0,\ 2\delta,\ 3\delta,\ \ldots,$, for small $\delta,$ we have:

[Example credit: Stanford EE 103]

Example: 1-d Vehicle

State $x = (p, v)$, i.e., 1-d position and its velocity

Control u, 1-d force, Friction force $-\eta v$, Vehicle mass m,

Consider discrete time $t = 0,\ 2\delta,\ 3\delta,\ \ldots,$, for small δ, we have:

$$m\frac{v_{h+1} - v_h}{\delta} \approx u - \eta v_h, \qquad \frac{p_{h+1} - p_h}{\delta} \approx v_h$$

[Example credit: Stanford EE 103]

V/Q functions:

- Define the value function $V_h^\pi : \mathbb{R}^d \to \mathbb{R}$ as

$$V_h^\pi(x) = \mathbb{E}\left[x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \,\Big|\, \pi, x_h = x\right],$$

- and the state-action value $Q_h^\pi : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ as:

$$Q_h^\pi(x, u) = \mathbb{E}\left[x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \,\Big|\, \pi, x_h = x, u_h = u\right],$$

Optimal Value functions:

$$V_h^\star(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E}\left[x_H^T Q x_H + \sum_{t=h}^{H-1} x_t^\top Q x_t + u_t^\top R u_t \,\middle|\, u_t = \pi_t(x_t), x_h = x \right]$$

Optimal Value functions:

$$V_h^\star(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E}\left[x_H^T Q x_H + \sum_{t=h}^{H-1} x_t^\top Q x_t + u_t^\top R u_t \middle| u_t = \pi_t(x_t), x_h = x \right]$$

Theorem:

$V_h^\star$ is a quadratic function, i.e., $V_h^\star(x) = x^\top P_h x + p_h$,

and optimal policy is linear:

$\pi_h^\star(x) = -K_h^\star x$,

and P_h & $K_h^\star$ can be computed exactly

Optimal Value functions:

$$V_h^\star(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E}\left[x_H^T Q x_H + \sum_{t=h}^{H-1} x_t^\top Q x_t + u_t^\top R u_t \,\middle|\, u_t = \pi_t(x_t), x_h = x \right]$$

Theorem:

$V_h^\star$ is a quadratic function, i.e., $V_h^\star(x) = x^\top P_h x + p_h$,

and optimal policy is linear:

$\pi_h^\star(x) = -K_h^\star x$,

and P_h & $K_h^\star$ can be computed exactly

(Derivation? We will do it together next Tuesday)

Next Lecture:

How to compute the optimal policy in closed-form solution