3/10: Dynamic Programming

Announcements:
- HW 3 out, due 3/26
- Mid-semester survey out, due 3/17
- Project milestone due 3/19
  - 2-6 pg.
  - intro/related work, problem def, timeline

Last time:
- intro to optimal control
- LQR (as a QP)

Today:
- LQR (via dynamic programming)

Recall: optimal control

$$\min_{x_{0:T},\, u_{0:T-1}} \quad c_T(x_T) + \Sigma_{t=0}^{T-1} c_t(x_t, u_t)$$

s.t.
$x_{t+1} = f(x_t, u_t)$
$x_0 = x(0)$
$x_t \in \mathcal{X}$ collision avoidance
$u_t \in \mathcal{U}$ input constraints

1. Trajectory optimization
   open-loop $x_{0:T}^*, u_{0:T-1}^*$

2. Dynamic programming
   closed-loop $\pi^*(x_t) = u_t^*$

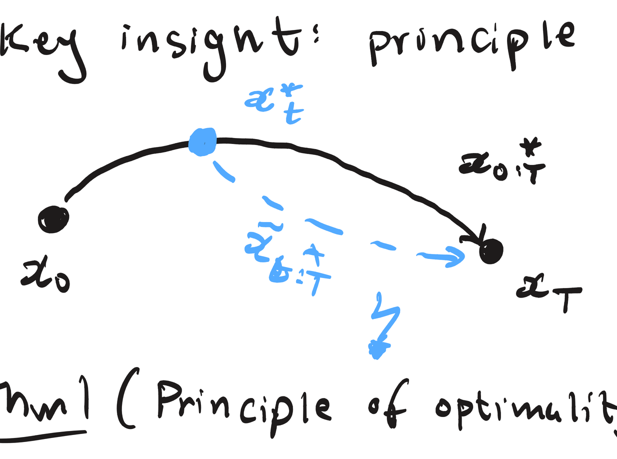
LQR: $$\min_{x_{0:T},\, u_{0:T-1}} \quad \tfrac{1}{2} x_T^T Q_T x_T + \Sigma_i \tfrac{1}{2} x_t^T Q_t x_t + \tfrac{1}{2} u_t^T Q_t u_t$$
$$x_{t+1} = A x_t + B u_t$$

solution:
$$\boxed{\begin{aligned} u_t^* &= -K_t x_t \\ K_t &= (R_t + B^T P_{t+1} B)^{-1} B^T P_{t+1} A \\ P_t &= Q_t + A^+ P_{t+1} (A - BK_t) \end{aligned}}$$

(i) "backward pass" $P_T = Q_T \rightarrow (K_t, P_t)$
(ii) "forward pass" $x_0 = x(0) \rightarrow (x_{0:T}^*, u_{0:T-1}^*)$

Dynamic programming

Key insight: principle of optimality



"subproblems have the same solution as the full problem"

<u>Thm 1</u> (Principle of optimality)
Consider the optimal control subproblem from time τ, optimal x_τ^*

$$\min_{x_{\tau:T},\, u_{\tau:T-1}} \quad c_T(x_T) + \Sigma_{t=\tau}^{T-1} c_t(x_t, u_t)$$

s.t. $x_\tau = x_\tau^*$ $\quad x_t \in \mathcal{X}$
$x_{t+1} = f(x_t, u_t)$ $\quad u_t \in \mathcal{U}$

Then, the original solution $(x_{\tau:T}^*, u_{\tau:T-1}^*)$ of the full problem is a solution to the subproblem.

Key idea: break down long horizons into shorter subproblem

Define $V_t^*(x_t)$ as "cost-to-go" of the optimal trajectory from x_t ("value function")

Bellman's principle implies:

$$V_t^*(x_t) = \min_{u \in \mathcal{U}} \left\{ c_t(x_t, u_t) + V_{t+1}^*(f(x_t, u_t)) \right\}$$

$$V_T^*(x_T) = c_T(x_T)$$

$$\pi^*(x_t) = \arg\min_{u \in \mathcal{U}} \left\{ c_t(x_t, u_t) + V_{t+1}^*(f(x_t, u_t)) \right\}$$

<u>Alg 1</u> (Dynamic programming)
$V_T^*(x_T) \leftarrow c_T(x_T) \quad \forall x_T \in \mathcal{X}$
for $t = T-1, T-2, \dots, 0$
$$V_t^*(x_t) = \min_{u \in \mathcal{U}} \left\{ c_t(x_t, u_t) + V_{t+1}^*(f(x_t, u_t)) \right\} \quad \forall x_t \in \mathcal{X}$$
return $V_0^*, \dots, V_T^*$

Pros: + optimal <u>feedback policy</u>
Cons: − "blows up" with $|\mathcal{X}|$

---

e.g. LQR

$$\min_{x_{0:T},\, u_{0:T-1}} \quad \tfrac{1}{2} x_T^T Q_T x_T + \Sigma_{t=0}^{T-1} \tfrac{1}{2} x_t^T Q_t x_t + \tfrac{1}{2} u_t^T R u_t$$

s.t. $x_{t+1} = A x_t + B u_t$

$$V_T^*(x_T) = \tfrac{1}{2} x_T^T Q_T x_T \equiv \tfrac{1}{2} x_T^T P_T x_T$$

"ansatz" $V_t^*(x_t) = \tfrac{1}{2} x_t^T P_t x_t$

By induction: suppose we have P_{t+1}

$$V_t^*(x_t) = \min_u \Big\{ \underbrace{\tfrac{1}{2} x_t^T Q_t x_t}_{\text{(neglect)}} + \tfrac{1}{2} u^T R u + \tfrac{1}{2} x_{t+1}^T P_{t+1} x_{t+1} \Big\}$$
$$= \min_u \left\{ \tfrac{1}{2} u^T R_t u + \tfrac{1}{2} (A x_t + B u)^T P_{t+1} (A x_t + B u) \right\}$$

$\nabla_u = 0$
$$R_t u^* + B^T P_{t+1} (A x_t + B u^*) = 0$$

$$u^* = -(R_t + B^T P_{t+1} B)^{-1} B^T P_{t+1} A x_t$$
$R_t \succ 0$
$$= -K_t x_t \qquad (A x_t + B u^* = (A - BK) x_t)$$

$$V_t^*(x_t) = \tfrac{1}{2} x_t^T Q_t x_t + \tfrac{1}{2} x_t^T K_t^T R_t K_t x_t + \tfrac{1}{2} x_t (A - BK)^T P_{t+1} (A - BK) x_t$$

$$P_t = Q_t + K_t^T R_t K_t + (A - BK_t) P_{t+1} (A - BK_t)$$

* same as the open-loop case *

upshot: for LQR, equivalence b/w open- and closed-loop

Interesting extension: $T \rightarrow \infty$, $Q_t = Q$, $R_t = R$
if (A, B) is "stabilizable"
running $P_t \rightarrow P_{-\infty}$ $\qquad P_t = P_{t+1}$

$$P = Q + A^T P (A - B(R + B^T Q B)^{-1} B^T P A)$$
"discrete agebraic Ricati equation (DARE)"

Takeaways:
- dynamic programming uses recursion to find $\pi^*(x_t)$
- generally intractable, but clean for LQR