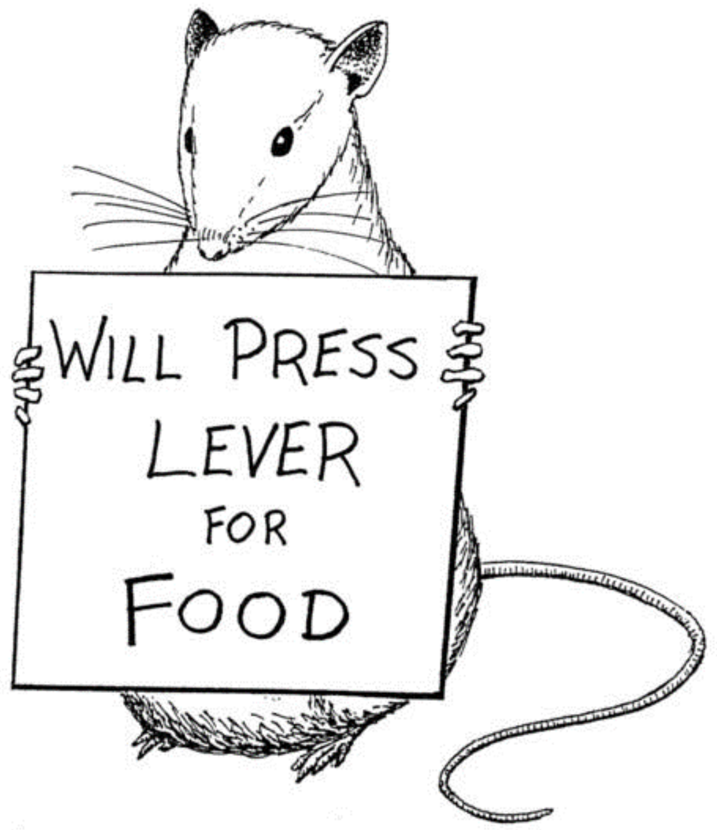
CS 277: Control and Reinforcement Learning

Winter 2024

Lecture 7: Optimal Control

Roy Fox
Department of Computer Science
School of Information and Computer Sciences
University of California, Irvine



Logistics

assignments

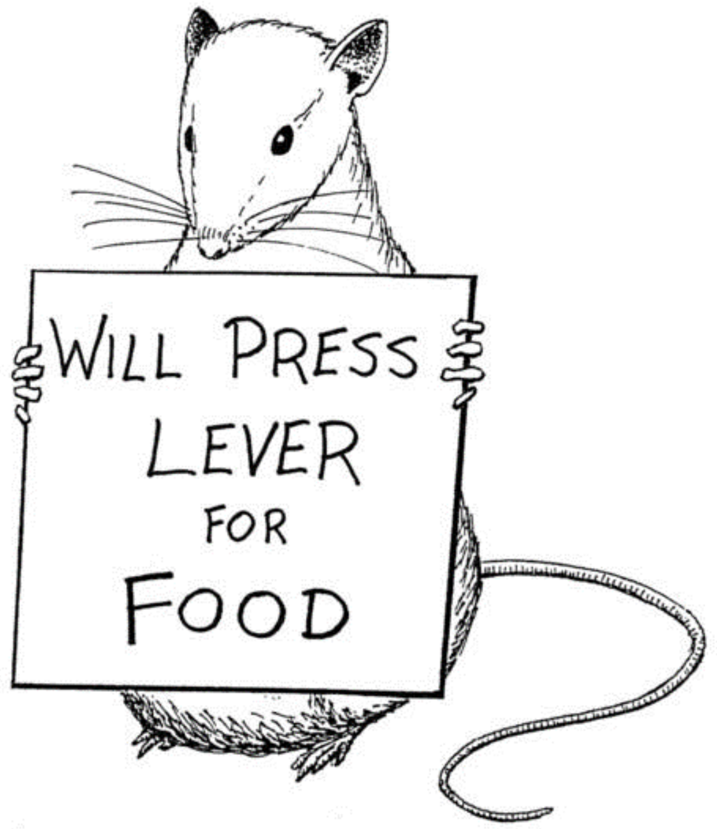
- Exercise 2 and Quiz 4 due next Monday

videos

- Video on trust-region methods on the course website
- Might help with Quiz 4

State of the Course

- Model-Free RL: done!
- Up next:
  - Model-Based RL (related: Optimal Control)
  - Twists and turns! Exploration, Partial observability
  - Advanced settings! RLHF, Inverse RL, Bounded RL, & more



Today's lecture

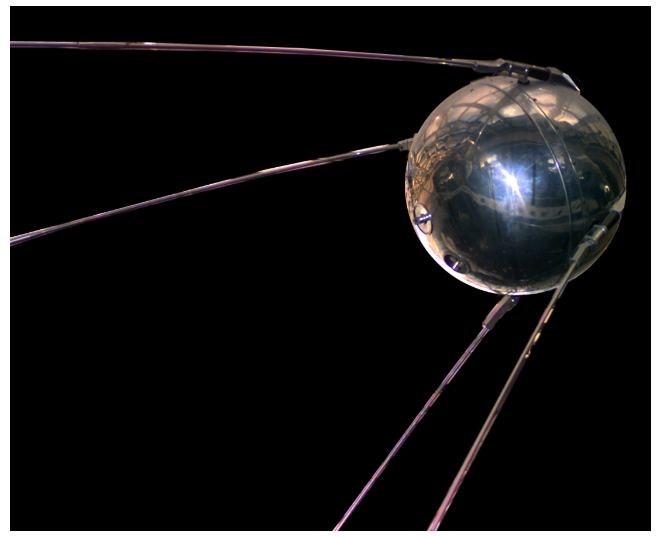
Stability, reachability, stabilizability

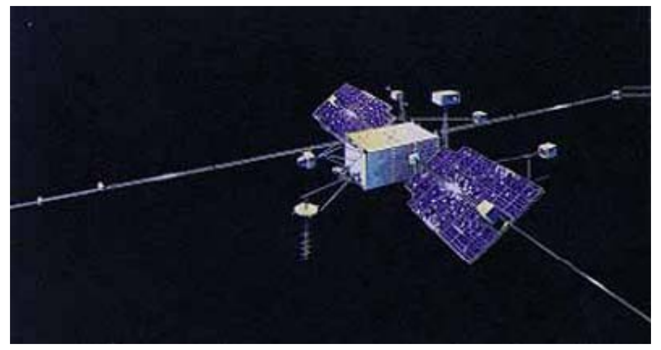
Linear Quadratic Regulator

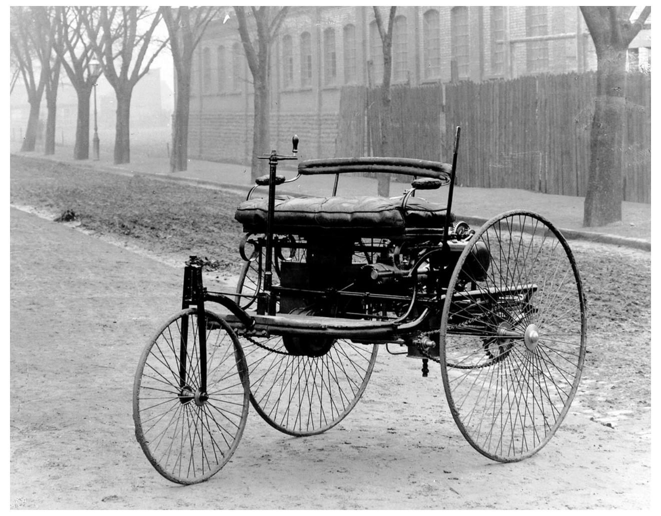
Hamiltonian

Why Optimal Control?

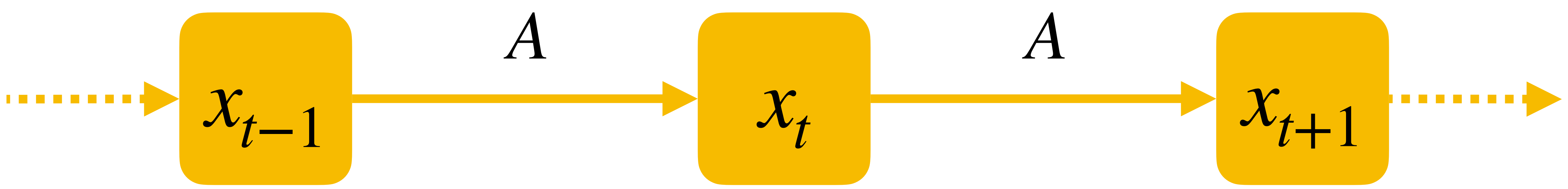
- Optimal Control involves environments simple enough to solve directly
  - Important applications
  - Powerful and profound theory
  - Useful insights / components for harder domains







Linear Time-Invariant (LTI) systems



- Continuous state space: $x_t \in \mathbb{R}^n$

- Simplest system — linear: $x_{t+1} = Ax_t \qquad A \in \mathbb{R}^{n \times n}$
  - Linear Time-Invariant (LTI): A does not depend on t

- How does the system evolve over time?

$$x_t = A^t x_0$$

- Adding drift b doesn't add much insight, won't do it today (well, ok, once)

Stability

- To analyze: use eigenvectors $\lambda e = Ae$

- Consider a basis of eigenvectors $e_1, \ldots, e_n \in \mathbb{C}^n$

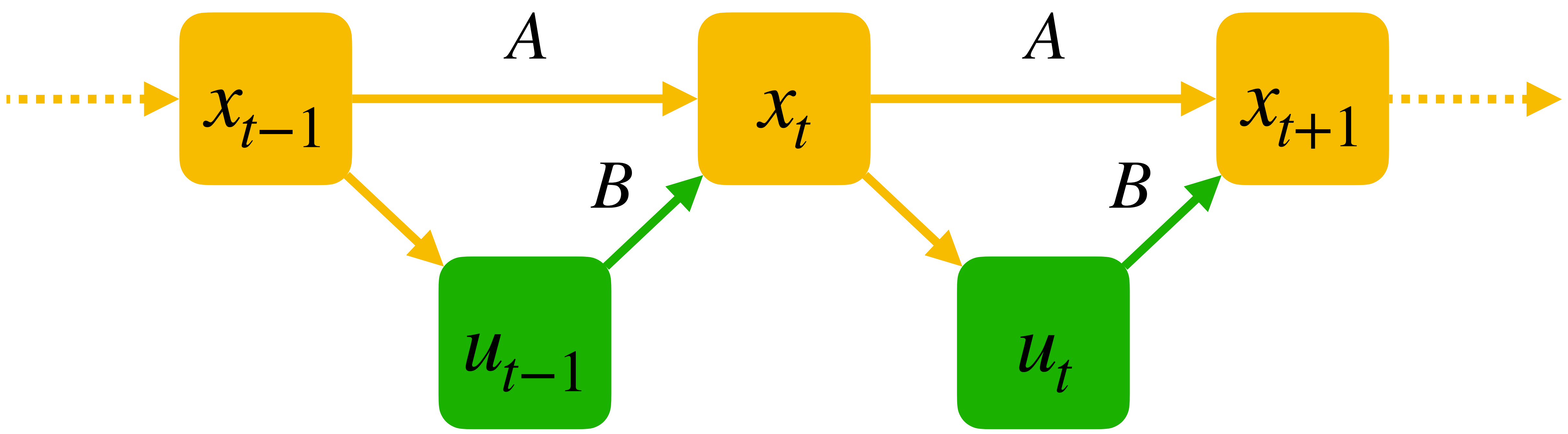
$$x_0 = \sum_i \alpha_i e_i \implies x_1 = Ax_0 = \sum_i \alpha_i \lambda_i e_i \implies x_t = \sum_i \alpha_i \lambda_i^t e_i$$

- Instability: some $\|\lambda_i\| > 1$, so that $\lim_{t\to\infty} \|x_t\| \to \infty$

- Stability: all $\|\lambda_i\| < 1$, so that $\lim_{t\to\infty} x_t = 0$

  - When $\|\lambda_i\| = 1$, component never vanishes or explodes; still called unstable

Linear control systems



- Continuous action (control) space: $u_t \in \mathbb{R}^m$

- Controlled LTI system: $x_{t+1} = Ax_t + Bu_t \qquad B \in \mathbb{R}^{n \times m}$

$$x_t = A^t x_0 + A^{t-1}Bu_0 + \cdots + ABu_{t-2} + Bu_{t-1}$$

$$x_t = A^t x_0 + \begin{bmatrix} B & AB & \cdots & A^{t-1}B \end{bmatrix} \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{bmatrix}$$

Reachability

$$x_t = A^t x_0 + \begin{bmatrix} B & AB & \cdots & A^{t-1}B \end{bmatrix} \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{bmatrix}$$

- Can we reach a given state x_t at time t?
  - If and only if $x_t - A^t x_0 \in \operatorname{span} \begin{bmatrix} B & AB & \cdots & A^{t-1}B \end{bmatrix}$
- Cayley-Hamilton: A satisfies $p_A(\lambda) = |\lambda I - A|$
  - Sufficient to take $t = n$, controllability matrix: $\mathscr{C}_{n \times nm} = \begin{bmatrix} B & AB & \cdots & A^{n-1}B \end{bmatrix}$

p_A **has degree** n

$\Rightarrow A^n$ **spanned by** $I, A, \ldots, A^{n-1}$

- Reachability: can we reach all states eventually?
  - If and only if $\operatorname{span}\mathscr{C} = \mathbb{R}^n \iff \operatorname{rank}\mathscr{C} = n \implies \mathscr{C}\mathscr{C}^+ = I$ ($\mathscr{C}^+$= pseudo-inverse)
- To reach x: control $\vec{u} = \mathscr{C}^+(x - A^n x_0)$

Stabilizability

$$x_t = A^t x_0 + \begin{bmatrix} B & AB & \cdots & A^{t-1}B \end{bmatrix} \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{bmatrix}$$

- Can we reach $x = 0$ eventually?
- For each mode e_i (eigenvector of A):
  - Is $\|\lambda_i\| < 1$? $\Rightarrow$ stable, otherwise unstable
    - Stable modes reach 0 on their own
  - If unstable, is $e_i \in \operatorname{span}\mathscr{C}$? $\Rightarrow$ stabilizable, otherwise unstabilizable
    - Stabilizable modes = unstable, but controllable
- The system (A, B) is stabilizable if all modes are stable or stabilizable

Today's lecture

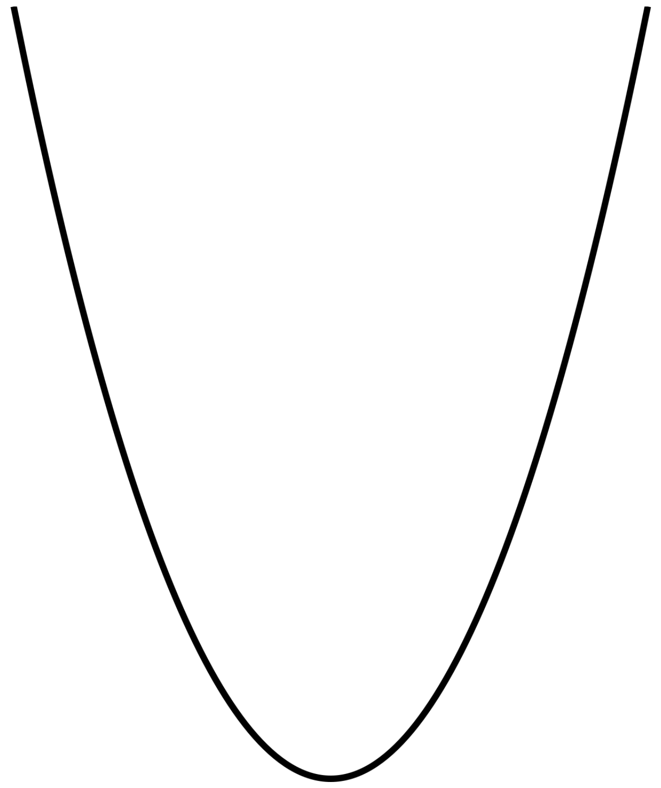
Stability, reachability, stabilizability

Linear Quadratic Regulator

Hamiltonian

Quadratic costs

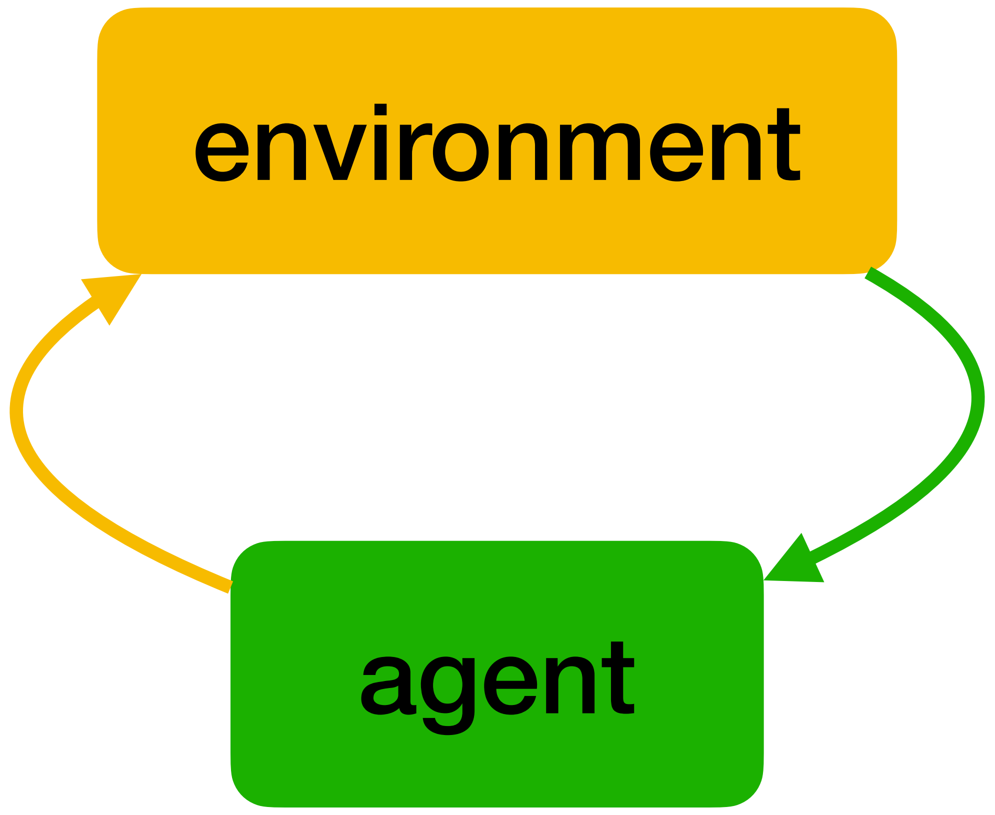
- Linear reward has no maximum ⇒ simplest of interest: concave quadratic
  - Consider negative reward = cost: $c(x_t, u_t) = \frac{1}{2}x_t^\intercal Q x_t + \frac{1}{2}u_t^\intercal R u_t$

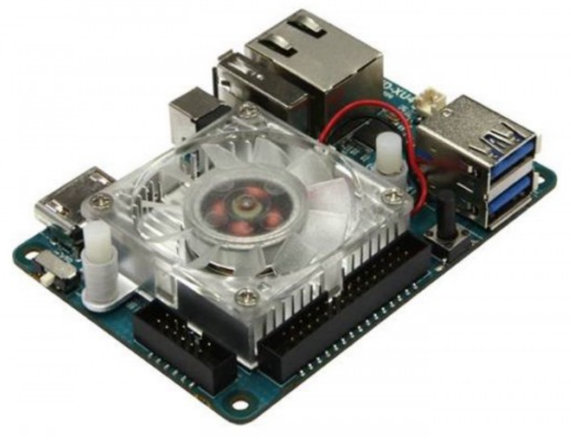


- $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite $Q \succeq 0$: $\frac{1}{2}x^\intercal Q x \geq 0$ for all x
  - No incentive to go to infinity in any direction
- $R \in \mathbb{R}^{m \times m}$ is positive definite $R \succ 0$: $\frac{1}{2}u^\intercal R u > 0$ for all u
  - Incentive for finite control in all directions
- Usually, finite or infinite horizon, no discounting

Linear Quadratic Regulator (LQR)

- Linear Quadratic Regulation (LQR) optimization problem:
  - Given LTI dynamics + quadratic cost (A, B, Q, R)
  - Find the control function $u_t = \pi(x_t)$
  - That minimizes $J^\pi = \sum_{t=0}^{T-1} c(x_t, u_t) = \frac{1}{2} \sum_{t=0}^{T-1} \left(x_t^\intercal Q x_t + u_t^\intercal R u_t \right)$
  - Such that $x_{t+1} = Ax_t + Bu_t$ for all t





Solving the LQR

- Bellman recursion: $V_t(x_t) = \min_{u_t} c(x_t, u_t) + V_{t+1}(x_{t+1})$

$x_{t+1} = Ax_t + Bu_t$

- Let's solve while also proving by induction that V_t is quadratic
  - Base case: $V_T \equiv 0$
  - Assume: $V_{t+1}(x_{t+1}) = \frac{1}{2} x_{t+1}^\intercal S_{t+1} x_{t+1} \qquad S_{t+1} \succeq 0$
  - Solve: $\nabla_{u_t}(c(x_t, u_t) + V_{t+1}(x_{t+1})) = 0$

Bellman optimality

$$V_{t+1}(x_{t+1}) = \tfrac{1}{2}x_{t+1}^\intercal S_{t+1}x_{t+1}$$
$$x_{t+1} = Ax_t + Bu_t$$

$$\begin{aligned} 0 &= \nabla_{u_t}(c(x_t, u_t) + V_{t+1}(x_{t+1})) \\ &= \frac{1}{2}\nabla_{u_t}(x_t^\intercal Qx_t + u_t^\intercal Ru_t + (Ax_t + Bu_t)^\intercal S_{t+1}(Ax_t + Bu_t)) \\ &= Ru_t + B^\intercal S_{t+1}(Ax_t + Bu_t) \end{aligned}$$

$$u_t^* = -(R + B^\intercal S_{t+1}B)^{-1}B^\intercal S_{t+1}Ax_t$$

- Plugging u_t^* into the Bellman recursion and rearranging terms:

$$V_t(x_t) = \frac{1}{2}x_t^\intercal(Q + A^\intercal(S_{t+1} - S_{t+1}B(R + B^\intercal S_{t+1}B)^{-1}B^\intercal S_{t+1})A)x_t$$

- Ricatti equation: $S_t = Q + A^\intercal(S_{t+1} - S_{t+1}B(R + B^\intercal S_{t+1}B)^{-1}B^\intercal S_{t+1})A$

Optimal control: properties

- Linear control policy: $u_t = L_t x_t$
  - Feedback gain: $L_t = -(R + B^\intercal S_{t+1} B)^{-1} B^\intercal S_{t+1} A$
- Quadratic value (cost-to-go) function $V_t(x_t) = \frac{1}{2} x_t^\intercal S_t x_t$
  - Cost Hessian $S_t = \nabla^2_{x_t} V_t$ is the same for all x_t
- Ricatti equation for S_t can be solved recursively backward

$$S_t = Q + A^\intercal (S_{t+1} - S_{t+1} B (R + B^\intercal S_{t+1} B)^{-1} B^\intercal S_{t+1}) A$$

  - Without knowing any actual states or controls (!) = at system design time

Infinite horizon

- Average cost: $J = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} c(x_t, u_t)$

- For each finite T we solve with Bellman recursion, affected by end $V_T \equiv 0$
  - In the limit, end effects go away ⇒ converge to time-independent

- Discrete-time algebraic Ricatti equation (DARE):

$$S = Q + A^\intercal(S - SB(R + B^\intercal SB)^{-1}B^\intercal S)A$$

- Optimal cost-to-go function: $V(x) = \frac{1}{2}x^\intercal Sx$; optimal cost: $J = \frac{1}{2}x_0^\intercal Sx_0$

Non-homogeneous case

- More generally, LQR can have lower-order terms

$$x_{t+1} = f_t(x_t, u_t) = A_t x_t + B_t u_t + b_t$$

$$c_t(x_t, u_t) = \frac{1}{2} x_t^\intercal Q_t x_t + \frac{1}{2} u_t^\intercal R_t u_t + u_t^\intercal N_t x_t + q_t^\intercal x_t + r_t^\intercal u_t + s_t$$

- More flexible modeling, e.g. tracking a target trajectory $\frac{1}{2}(x_t - \tilde{x}_t)^\intercal Q (x_t - \tilde{x}_t)$

- Solved essentially the same way
  - Cost-to-go $V_t(x_t)$ will also have lower-order terms

Today's lecture

Stability, reachability, stabilizability

Linear Quadratic Regulator

Hamiltonian

Co-state

$c_t \in \mathbb{R}$ $f_t \in \mathbb{R}^n$

- Consider the cost-to-go $V_t^\pi(x_t) = c(x_t, u_t) + V_{t+1}^\pi(f(x_t, u_t))$

- To study its landscape over state space, consider its spatial gradient

$$\nu_t = \nabla_{x_t} V_t^\pi = \nabla_{x_t} c_t + \nabla_{x_{t+1}} V_{t+1}^\pi \cdot \nabla_{x_t} f_t = \nabla_{x_t} c_t + \nu_{t+1} \cdot \nabla_{x_t} f_t$$

  - Jacobian of the dynamics: $\nabla_{x_t} f_t \in \mathbb{R}^{n \times n}$

- Co-state $\nu_t \in \mathbb{R}^n$ = direction of steepest increase in cost-to-go

  - Linear backward recursion $\nu_t = \nabla_{x_t} c_t + \nu_{t+1} \cdot \nabla_{x_t} f_t$; initialization: $\nu_T = 0$

Hamiltonian

- Cost-to-go recursion: (first-order approximation)

$$V_t^\pi(x_t) = c(x_t, u_t) + V_{t+1}^\pi(x_{t+1}) \approx c(x_t, u_t) + \overset{\text{state } x_{t+1}}{f(x_t, u_t)} \cdot \underset{\text{co-state } \nu_{t+1}}{\nabla_{x_{t+1}} V_{t+1}^\pi}$$

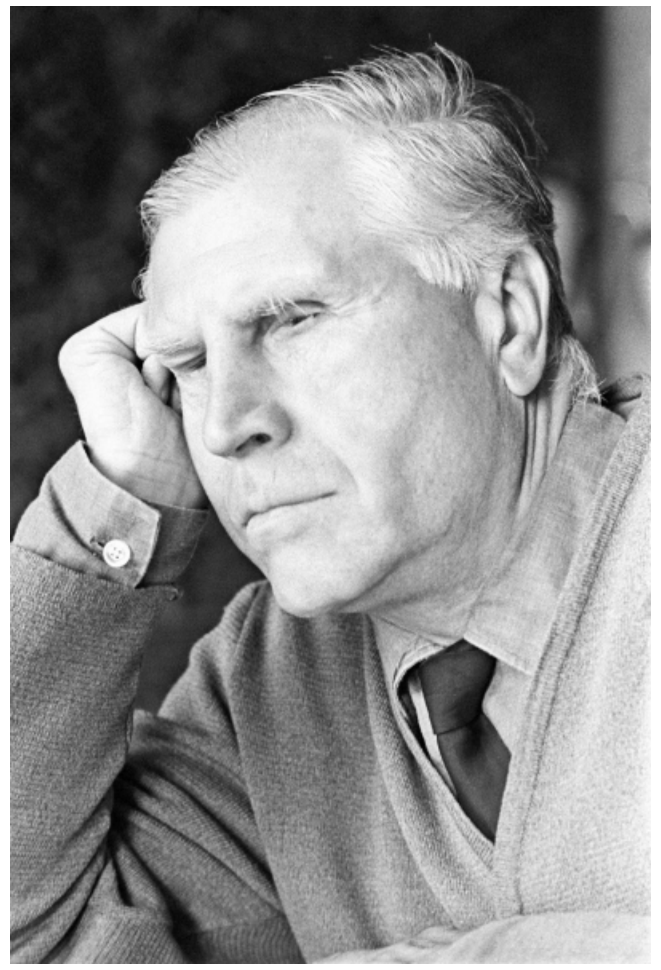
- Hamiltonian = first-order approximation of the cost-to-go

$$\mathscr{H}_t(x_t, \nu_{t+1}, u_t) = c(x_t, u_t) + \nu_{t+1} \cdot f(x_t, u_t)$$

  - Related to, but not the same as the Hamiltonian in physics

- The Hamiltonian is useful to get first-order conditions for optimal control
  - Equivalent to Bellman optimality
  - Even more useful in continuous time (equivalent to Hamilton–Jacobi–Bellman)

Pontryagin's maximum principle



Lev Pontryagin

- Hamiltonian: $\mathscr{H}_t(x_t, \nu_{t+1}, u_t) = c(x_t, u_t) + \nu_{t+1} \cdot f(x_t, u_t)$
- Necessary optimality conditions:

$$\nabla_{\nu_{t+1}} \mathscr{H}_t = x_{t+1} \qquad \nabla_{x_t} \mathscr{H}_t = \nu_t \qquad \nabla_{u_t} \mathscr{H}_t = 0$$

- $\nabla_{\nu_{t+1}} \mathscr{H}_t = f(x_t, u_t) = x_{t+1}$ necessary for x_t to be the state for dynamics f
- $\nabla_{x_t} \mathscr{H}_t = \nabla_{x_t} c_t + \nu_{t+1} \cdot \nabla_{x_t} f_t = \nu_t$ necessary for $\nu_t = \nabla_{x_t} V_t^\pi$ to be a co-state
- Objective: $\min_\pi J$ s.t. $x_{t+1} = f(x_t, u_t)$; Lagrangian: $\underset{\text{optimal when } \nabla_{u_t}\mathscr{H}_t = 0}{\mathscr{L}} = \sum_{t=0}^{T-1} \mathscr{H}_t - \underset{\text{independent of } u_t}{\nu_{t+1} \cdot x_{t+1}}$

Hamiltonian in LQR

- The Hamiltonian is generally high-degree, many local optima, hard to solve

- In LQR, the Hamiltonian is quadratic

$$\mathscr{H}_t = \frac{1}{2}x_t^\intercal Q x_t + \frac{1}{2}u_t^\intercal R u_t + \nu_{t+1}(Ax_t + Bu_t)$$

- This suggests forward–backward recursions for x, ν, and u:

$$x_{t+1} = \nabla_{\nu_{t+1}}\mathscr{H}_t = Ax_t + Bu_t$$

$$\nu_t = \nabla_{x_t}\mathscr{H}_t = \nu_{t+1}A + x_t^\intercal Q$$

$$\nabla_{u_t}\mathscr{H}_t = Ru_t + B^\intercal \nu_{t+1}^\intercal = 0$$

- The solution coincides with the Ricatti equations with $\nu_t^\intercal = S_t x_t \quad u_t = L_t x_t$

Recap

- LQR = simplest dynamics: linear; simplest cost: quadratic
- Can characterize stability, reachability, stabilizability, more.. in terms of (A, B)
- Can use Ricatti equation to find cost-to-go Hessian
- Equivalently: Hamiltonian gives state forward / co-state backward recursions