114-1 Robot Perception and Learning

CartPole Optimal Control: A Case Study

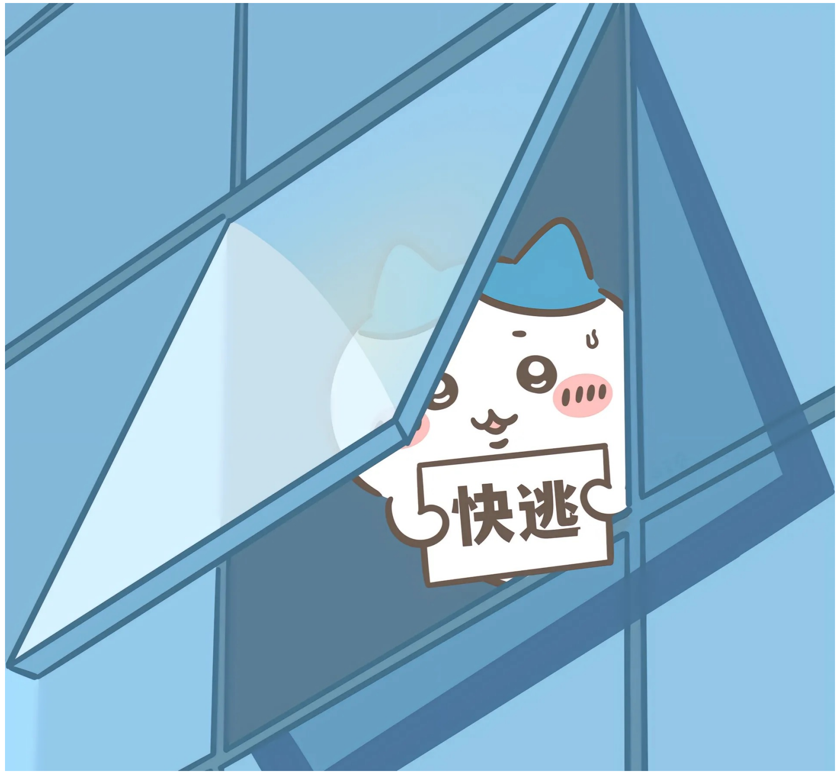
Jia-Wei Liao

Ph.D. Candidate in Computer Science

National Taiwan University

Important Reminder

The withdrawal deadline for this course is **Friday, October 17**. Requests submitted after this date will **not be accepted**.

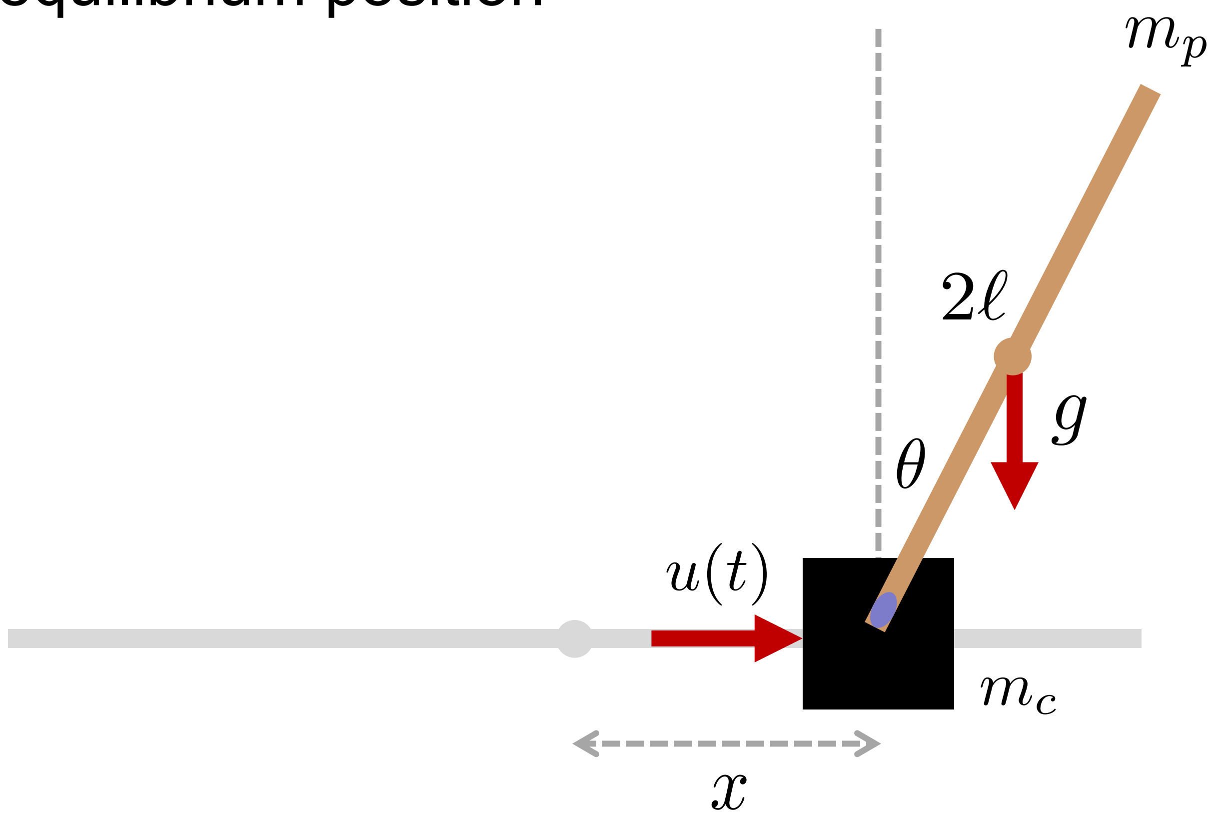


Outline

- CartPole Problem
- Discrete-time LQR
- Steady-state Regulator
- Iterative LQR

CartPole Problem

We aim to control the cart by applying a horizontal force $u(t)$ such that the inverted pendulum mounted on top of the cart remains stabilized near the upright equilibrium position



- Cart mass m_c
- Pole mass m_p
- Pole length 2ℓ
- Total mass $M = m_c + m_p$
- Cart displacement x
- Pole angle θ
- Gravity g

Continuous-time Dynamical System

Nonlinear

$$\ddot{\theta} = \frac{g\sin\theta - \cos\theta \frac{u + m_p \ell \dot{\theta}^2 \sin\theta}{M}}{\ell\left(\frac{4}{3} - \frac{m_p}{M}\cos^2\theta\right)}$$

$$\ddot{x} = \frac{u + m_p \ell \dot{\theta}^2 \sin\theta}{M} - \frac{m_p \ell \cos\theta}{M}\ddot{\theta}$$

$\theta \approx 0,\ \dot{\theta} \approx 0$

$\sin\theta \approx \theta$

$\cos\theta \approx 1$

Linear

$$\ddot{\theta} \approx \frac{g}{D}\theta - \frac{1}{DM}u$$

$$\ddot{x} \approx -\alpha g \theta + \frac{1+\alpha}{M}u$$

$$D = \ell\left(\frac{4}{3} - \frac{m_p}{M}\right)$$

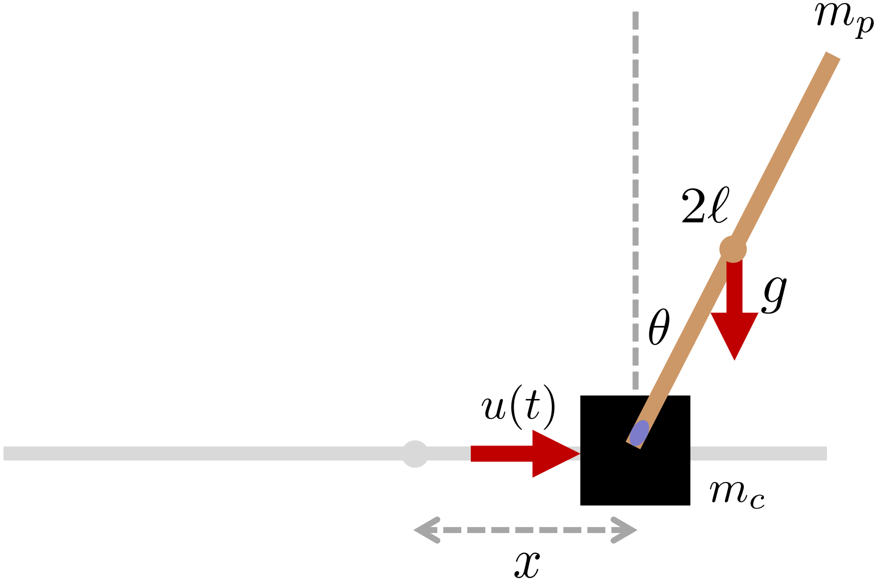
$$\alpha = \frac{3m_p}{4M - 3m_p}$$

Continuous-time Dynamical System

$$\ddot{\theta} \approx \frac{g}{D}\theta - \frac{1}{DM}u$$

$$\ddot{x} \approx -\alpha g\theta + \frac{1+\alpha}{M}u$$

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -\alpha g & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{g}{D} & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1+\alpha}{M} \\ 0 \\ -\frac{1}{DM} \end{bmatrix} u(t)$$



$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + Bu(t)$$

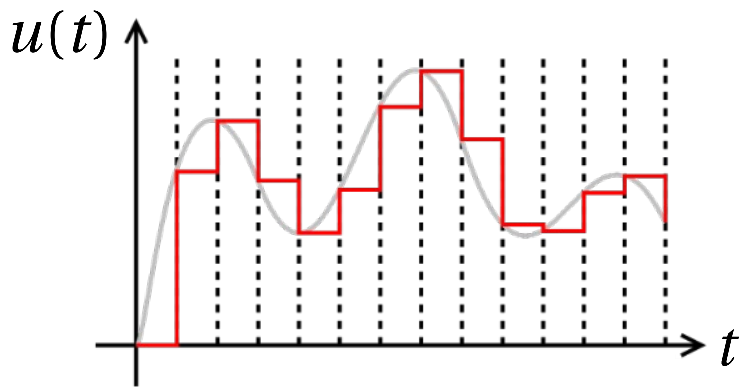
$$\begin{tikzcd} \boldsymbol{u} \arrow[r, "B"] & \dot{\boldsymbol{x}} \arrow[r, "\int \mathrm{d}t"] & \boldsymbol{x} \arrow[l, bend left, "A"] \end{tikzcd}$$

Discretization

Consider the continuous-time linear dynamical system $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + Bu(t)$. By integrating the dynamics from $i\Delta t$ to $(i+1)\Delta t$, we obtain

$$
\begin{aligned}
\boldsymbol{x}_{i+1} &= e^{A\Delta t}\boldsymbol{x}_i + \int_{i\Delta t}^{(i+1)\Delta t} e^{A((i+1)\Delta t-\tau)} B u_i \, d\tau \\
&= e^{A\Delta t}\boldsymbol{x}_i + \left(\int_0^{\Delta t} e^{A\sigma}\, d\sigma\right) B u_i \\
&= A_d \boldsymbol{x}_i + B_d u_i
\end{aligned}
$$

Zero-Order Hold (ZOH)



$u(t) = u_i, \quad t \in [i\Delta t, (i+1)\Delta t)$

Discrete-time Linear Quadratic Regulator

Consider the discrete-time linear dynamical system $\boldsymbol{x}_{i+1} = A_d \boldsymbol{x}_i + B_d u_i$ and the finite-horizon quadratic cost

$$J = \sum_{i=0}^{T-1} \Bigg(\underbrace{\frac{1}{2} \boldsymbol{x}_i^\top Q \boldsymbol{x}_i}_{\text{Error term}} + \overbrace{\frac{1}{2} R u_i^2}^{\text{Control term}} \Bigg) + \underbrace{\frac{1}{2} \boldsymbol{x}_T^\top Q_f \boldsymbol{x}_T}_{\text{Terminal error term}}$$

where Q, Q_f are positive definite and $R > 0$.

The objective is to find the sequence $\{(\boldsymbol{x}_i, u_i)\}_{i=0}^{T}$ to minimize J subject to the system dynamics.

Solution of Discrete-time LQR

Consider the discrete-time linear dynamical system $\boldsymbol{x}_{i+1} = A_d \boldsymbol{x}_i + B_d u_i$. The problem can be solved using *dynamic programming (DP)* with the terminal condition $P_T = Q_f$:

$$\begin{cases} P_i = Q + A_d^\top P_{i+1} A_d - A_d^\top P_{i+1} B_d (R + B_d^\top P_{i+1} B_d)^{-1} B_d^\top P_{i+1} A_d \\ K_i = (R + B_d^\top P_{i+1} B_d)^{-1} B_d^\top P_{i+1} A_d \\ u_i = -K_i \boldsymbol{x}_i \end{cases}$$

Steady-state Regulator

If the system satisfies the following conditions:

(i) A_d, B_d are fixed (ii) (A_d, B_d) is controllable (iii) $(A_d, Q^{1/2})$ is observable

(iv) Q is positive definite and $R > 0$.

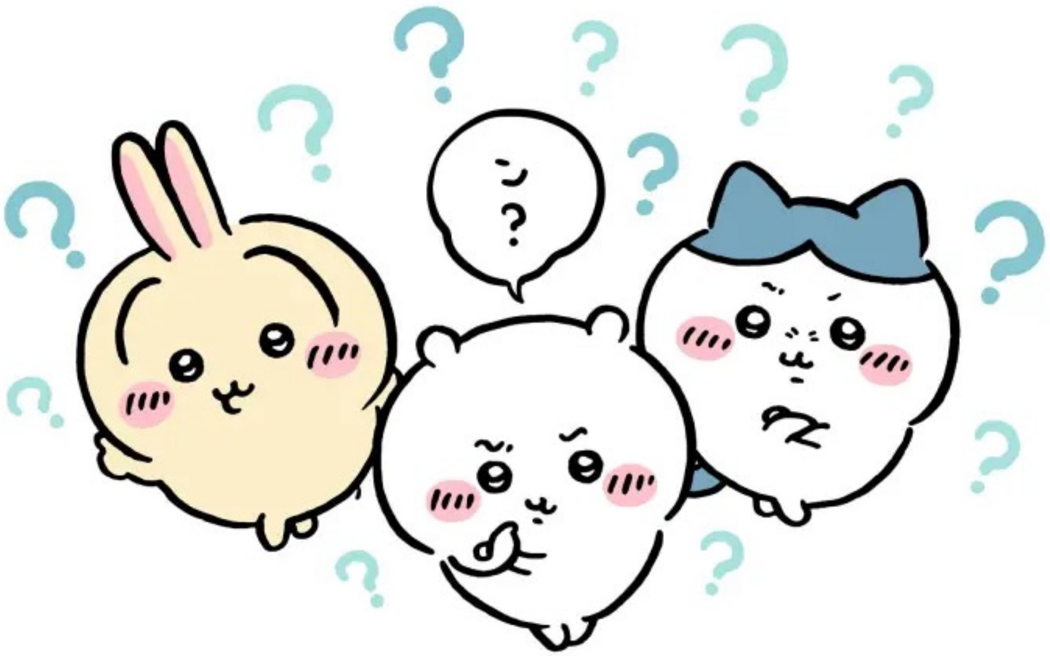
Then the sequence $\{P_i\}$ converges to a steady-state matrix P_∞, which satisfies the *discrete algebraic Riccati equation (DARE)*:

$$P_\infty = Q + A_d^\top P_\infty A_d - A_d^\top P_\infty B_d (R + B_d^\top P_\infty B_d)^{-1} B_d^\top P_\infty A_d$$

The steady-state control gain is $K_\infty = (R + B_d^\top P_\infty B_d)^{-1} B_d^\top P_\infty A_d$ and the corresponding feedback law is

$$u_i = -K_\infty \boldsymbol{x}_i$$

LQR makes life easy, but robots don't always follow linear scripts.
So what's next?



Iterative LQR (iLQR)

$$\min_{u_{T-1}} Q(\boldsymbol{x}_{T-1}, u_{T-1})$$

	0						T
Initialization	$\bar{u}_0$	$\bar{u}_1$	$\bar{u}_2$	$\bar{u}_3$	$\cdots$	$\bar{u}_{T-1}$	$\bar{u}_T$
Forward pass	$\bar{\boldsymbol{x}}_0$	$\bar{\boldsymbol{x}}_1$	$\bar{\boldsymbol{x}}_2$	$\bar{\boldsymbol{x}}_3$	$\cdots$	$\bar{\boldsymbol{x}}_{T-1}$	$\bar{\boldsymbol{x}}_T$
Backward pass	(k_0, K_0)	(k_1, K_1)	(k_2, K_2)	(k_3, K_3)	$\cdots$	(k_{T-1}, K_{T-1})	
Forward pass	$(\bar{\boldsymbol{x}}_0, \bar{u}_0)$	$(\bar{\boldsymbol{x}}_1, \bar{u}_1)$	$(\bar{\boldsymbol{x}}_2, \bar{u}_2)$	$(\bar{\boldsymbol{x}}_3, \bar{u}_3)$	$\cdots$	$(\bar{\boldsymbol{x}}_{T-1}, \bar{u}_{T-1})$	$(\bar{\boldsymbol{x}}_T, \bar{u}_T)$

Repeat until convergence

Iterative LQR (iLQR)

$$\boldsymbol{x}_{i+1} = f(\boldsymbol{x}_i, u_i) \xrightarrow{\text{Linearization}} \boldsymbol{x}_{i+1} \approx \frac{\partial f}{\partial \boldsymbol{x}}(\boldsymbol{x}_i, u_i)\boldsymbol{x}_i + \frac{\partial f}{\partial u}(\boldsymbol{x}_i, u_i)u_i$$

Let $\{(\bar{\boldsymbol{x}}_i, \bar{u}_i)\}_{i=1}^{T}$ be the nominal trajectory. We denote deviations by

$$\delta \boldsymbol{x}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}_i, \quad \delta u_i = u_i - \bar{u}_i.$$

Then we obtain

$$\delta \boldsymbol{x}_{i+1} = A_i \delta \boldsymbol{x}_i + B_i \delta u_i$$

where $A_i = \dfrac{\partial f}{\partial \boldsymbol{x}}(\boldsymbol{x}_i, u_i)$ and $B_i = \dfrac{\partial f}{\partial \boldsymbol{u}}(\boldsymbol{x}_i, u_i)$.

Iterative LQR (iLQR)

Consider the discrete-time linear dynamical system $\boldsymbol{x}_{i+1} = A_i \boldsymbol{x}_i + B_i u_i$ and the finite-horizon quadratic cost

$$J = \sum_{i=0}^{T-1} \underbrace{c(\boldsymbol{x}_i, u_i)}_{\text{Running cost}} + \overbrace{\phi(\boldsymbol{x}_T)}^{\text{Terminal cost}}$$

The objective is to find the sequence $\{(\boldsymbol{x}_i, u_i)\}_{i=0}^{T}$ to minimize J subject to the system dynamics.

Iterative LQR (iLQR)

Let the *value function* be

$$V(\boldsymbol{x}_t, t) := \min_{\{u_k\}_{k=1}^{T-1}} \sum_{k=t}^{T-1} c(\boldsymbol{x}_k, u_k) + \phi(\boldsymbol{x}_T)$$

The recursion has the form (Bellman equation):

$$\begin{aligned} V(\boldsymbol{x}_t, t) &= \min_{\{u_k\}_{k=t}^{T-1}} \left[c(\boldsymbol{x}_t, u_t) + \sum_{k=t+1}^{T-1} c(\boldsymbol{x}_k, u_k) + \phi(\boldsymbol{x}_T) \right] \\ &= \min_{u_t} \left[c(\boldsymbol{x}_t, u_t) + V(\boldsymbol{x}_{t+1}, t+1) \right] \end{aligned}$$

Iterative LQR (iLQR)

Define the *Q-function* as

$$Q(\boldsymbol{x}_t, u_t) = c(\boldsymbol{x}_t, u_t) + V(\boldsymbol{x}_{t+1}, t+1)$$

The value function can be written as

$$V(\boldsymbol{x}_t, t) = \min_{u_t} Q(\boldsymbol{x}_t, u_t).$$

Backward Pass

Comparing coefficients of $Q(\boldsymbol{x}_t, u_t) = c(\boldsymbol{x}_t, u_t) + V(\boldsymbol{x}_{t+1}, t+1)$

$$Q(\boldsymbol{x}_t, u_t) \approx Q(\bar{\boldsymbol{x}}_t, \bar{u}_t) + Q_{\boldsymbol{x}_t}^\top \delta \boldsymbol{x}_t + Q_{u_t}^\top \delta u_t + \frac{1}{2}\delta \boldsymbol{x}_t^\top Q_{\boldsymbol{x}_t\boldsymbol{x}_t}\delta \boldsymbol{x}_t + \frac{1}{2}\delta u_t^\top Q_{u_t u_t}\delta u_t + \delta u_t^\top Q_{\boldsymbol{x}_t u_t}\delta \boldsymbol{x}_t$$

$$c(\boldsymbol{x}_t, u_t) \approx c(\bar{\boldsymbol{x}}_t, \bar{u}_t) + c_{\boldsymbol{x}_t}^\top \delta \boldsymbol{x}_t + c_{u_t}^\top \delta u_t + \frac{1}{2}\delta \boldsymbol{x}_t^\top c_{\boldsymbol{x}_t\boldsymbol{x}_t}\delta \boldsymbol{x}_t + \frac{1}{2}\delta u_t^\top c_{u_t u_t}\delta u_t + \delta u_t^\top c_{\boldsymbol{x}_t u_t}\delta \boldsymbol{x}_t$$

$$V(\boldsymbol{x}_{t+1}, t+1) = V(\bar{\boldsymbol{x}}_{t+1}, t+1) + V_{\boldsymbol{x}_{t+1}}^\top A_t \delta \boldsymbol{x}_t + V_{\boldsymbol{x}_{t+1}}^\top B_t \delta u_t + \frac{1}{2}\delta \boldsymbol{x}_t^\top A_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} A_t \delta \boldsymbol{x}_t + \frac{1}{2}\delta u_t^\top B_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} B_t \delta u_t + \delta u_t^\top B_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} A_t \delta \boldsymbol{x}_t$$

$$\boxed{\begin{aligned} Q_{\boldsymbol{x}_t} &= c_{\boldsymbol{x}_t} + A_t^\top V_{\boldsymbol{x}_{t+1}} \\ Q_{u_t} &= c_{u_t} + B_t^\top V_{\boldsymbol{x}_{t+1}} \\ Q_{\boldsymbol{x}_t\boldsymbol{x}_t} &= c_{\boldsymbol{x}_t\boldsymbol{x}_t} + A_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} A_t \\ Q_{u_t u_t} &= c_{u_t u_t} + B_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} B_t \\ Q_{u_t\boldsymbol{x}_t} &= c_{\boldsymbol{x}_t u_t} + B_t^\top V_{\boldsymbol{x}_{t+1}\boldsymbol{x}_{t+1}} A_t \end{aligned}}$$

Backward Pass

$$V(\boldsymbol{x}_t, t) = \min_{u_t} Q(\boldsymbol{x}_t, u_t)$$

$$Q(\boldsymbol{x}_t, u_t) \approx Q(\bar{\boldsymbol{x}}_t, \bar{u}_t) + Q_{\boldsymbol{x}_t}^\top \delta \boldsymbol{x}_t + Q_{u_t}^\top \delta u_t$$
$$+ \tfrac{1}{2} \delta \boldsymbol{x}_t^\top Q_{\boldsymbol{x}_t \boldsymbol{x}_t} \delta \boldsymbol{x}_t + \tfrac{1}{2} \delta u_t^\top Q_{u_t u_t} \delta u_t + \delta u_t^\top Q_{\boldsymbol{x}_t u_t} \delta \boldsymbol{x}_t$$

To minimize Q, we set its first-order condition to zero:

$$0 = \frac{\partial Q}{\partial u_t} = Q_{u_t} + Q_{u_t u_t} \delta u_t + Q_{u_t \boldsymbol{x}_t} \delta \boldsymbol{x}_t$$

Then we have

$$u_t = \bar{u}_t + k_t + K_t \delta \boldsymbol{x}_t$$

where $k_t = -Q_{u_t u_t}^{-1} Q_{u_t}$ and $K_t = -Q_{u_t u_t}^{-1} Q_{u_t \boldsymbol{x}_t}$.

Forward Pass

Given the feedback gains $\{(k_t, K_t)\}_{t=1}^{T}$, define the updated control around the nominal as

$$u'_t = \bar{u}_t + k_t + K_t \delta \boldsymbol{x}_t$$

$$\boldsymbol{x}'_{t+1} = f(\boldsymbol{x}'_t, u'_t)$$

iLQR Algorithm

Algorithm 1 Iterative Linear Quadratic Regulator (iLQR)

Input: dynamics function f, cost function l, terminal cost ϕ, initial state $\bar{x}_0$, initial control sequence $\{\bar{u}_t\}_{t=0}^{T-1}$, horizon T

Compute initial nominal trajectory $\{\bar{x}_t\}_{t=0}^{T}$ by forward rollout: $\bar{x}_{t+1} = f(\bar{x}_t, \bar{u}_t)$

Compute cost: $J \leftarrow \phi(\bar{x}_T) + \sum_{t=0}^{T-1} l(\bar{x}_t, \bar{u}_t)$

while J does not converge **do**

Linearize dynamics around nominal trajectory:

$$\delta x_t = x_t - \bar{x}_t, \quad \delta u_t = u_t - \bar{u}_t$$
$$A_t = \frac{\partial f}{\partial x}\Big|_{(\bar{x}_t, \bar{u}_t)}, \quad B_t = \frac{\partial f}{\partial u}\Big|_{(\bar{x}_t, \bar{u}_t)}$$

Backward Pass

Initialize terminal value function derivatives: $V_{x_T} = \phi_{x_T}(\bar{x}_T)$, $V_{x_T x_T} = \phi_{x_T x_T}(\bar{x}_T)$

for $t = T-1$ to 0 **do**

Compute Q-function coefficients:

$$\begin{aligned} Q_{x_t} &\leftarrow c_{x_t} + A_t^\top V_{x_{t+1}} \\ Q_{u_t} &\leftarrow c_{u_t} + B_t^\top V_{x_{t+1}} \\ Q_{x_t x_t} &\leftarrow c_{x_t x_t} + A_t^\top V_{x_{t+1} x_{t+1}} A_t \\ Q_{u_t u_t} &\leftarrow c_{u_t u_t} + B_t^\top V_{x_{t+1} x_{t+1}} B_t \\ Q_{u_t x_t} &\leftarrow c_{x_t u_t} + B_t^\top V_{x_{t+1} x_{t+1}} A_t \end{aligned}$$

Compute local optimal feedback gains:

$$\begin{aligned} k_t &\leftarrow -Q_{u_t u_t}^{-1} Q_{u_t} \\ K_t &\leftarrow -Q_{u_t u_t}^{-1} Q_{u_t x_t} \end{aligned}$$

Update value function derivatives:

$$\begin{aligned} V_{x_t} &\leftarrow Q_{x_t} - Q_{u_t x_t}^\top Q_{u_t u_t}^{-1} Q_{u_t} \\ V_{x_t x_t} &\leftarrow Q_{x_t x_t} - Q_{u_t x_t}^\top Q_{u_t u_t}^{-1} Q_{u_t x_t} \end{aligned}$$

end for

Forward Pass

$x'_0 \leftarrow \bar{x}_0$

for $t = 0$ to $T-1$ **do**

Update control action and state:

$$\begin{aligned} u'_t &\leftarrow \bar{u}_t + k_t + K_t(x'_t - \bar{x}_t) \\ x'_{t+1} &\leftarrow f(x'_t, u'_t) \end{aligned}$$

end for

Compute cost: $J \leftarrow \phi(x'_T) + \sum_{t=0}^{T-1} c(x'_t, u'_t)$

Update nominal sequence: $\{(\bar{x}_t, \bar{u}_t)\}_{t=1}^{T} \leftarrow \{(x'_t, u'_t)\}_{t=1}^{T}$

end while

return $\{(x'_t, u'_t)\}_{t=1}^{T}$

Takeaway

- LQR offers a mathematically elegant solution to stabilize linearized CartPole dynamics.
- Steady-state LQR simplifies control with constant gain in infinite-horizon settings.
- iLQR generalizes this idea to nonlinear systems, updating nominal trajectories iteratively for locally optimal performance and forming a fundamental concept in modern trajectory optimization in robotics.

Thank you