ECE276B: Planning & Learning in Robotics
Lecture 14: Linear Quadratic Control

Nikolay Atanasov

natanasov@ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Outline

Finite-Horizon Deterministic Optimal Control

- Recall the finite-horizon deterministic optimal control (DOC) problem:
  - no disturbances, i.e., $\mathbf{w}_t \equiv 0$
  - closed-loop control does not offer any advantage over open-loop control
- Assume $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{U} = \mathbb{R}^m$
- Given $\mathbf{x}_0 \in \mathcal{X}$, construct an optimal control sequence $\mathbf{u}_{0:T-1}$ such that:

$$\begin{aligned}\min_{\mathbf{u}_{0:T-1}} \;& V_0^{\mathbf{u}_{0:T-1}}(\mathbf{x}_0) = \mathfrak{q}(\mathbf{x}_T) + \sum_{t=0}^{T-1} \ell(\mathbf{x}_t, \mathbf{u}_t) \\ \text{s.t.} \;& \mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t), \; t = 0, \ldots, T-1\end{aligned}$$

- The DOC problem can be solved via Dynamic Programming
- The DOC problem can also be viewed as an equality-constrained optimization for which we can obtain first-order necessary conditions for optimality

Finite-Horizon Deterministic Optimal Control

- Introduce Lagrange multipliers $\mathbf{p}_{1:T}$ to relax the constraints $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$
- The Lagrange multiplier $\mathbf{p}_t$ is called the system **costate**
- Lagrangian:

$$L(\mathbf{x}_{1:T}, \mathbf{u}_{0:T-1}, \mathbf{p}_{1:T}) = \mathfrak{q}(\mathbf{x}_T) + \sum_{t=0}^{T-1} \ell(\mathbf{x}_t, \mathbf{u}_t) + (f(\mathbf{x}_t, \mathbf{u}_t) - \mathbf{x}_{t+1})^\top \mathbf{p}_{t+1}$$

- Lagrangian gradients:

$$\begin{aligned}
\nabla_{\mathbf{x}_T} L &= \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}_T) - \mathbf{p}_T \\
\nabla_{\mathbf{x}_t} L &= \nabla_{\mathbf{x}} \ell(\mathbf{x}_t, \mathbf{u}_t) + \nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{u}_t)^\top \mathbf{p}_{t+1} - \mathbf{p}_t, && t = 1, \ldots, T-1 \\
\nabla_{\mathbf{p}_{t+1}} L &= f(\mathbf{x}_t, \mathbf{u}_t) - \mathbf{x}_{t+1}, && t = 0, \ldots, T-1 \\
\nabla_{\mathbf{u}_t} L &= \nabla_{\mathbf{u}} \ell(\mathbf{x}_t, \mathbf{u}_t) + \nabla_{\mathbf{u}} f(\mathbf{x}_t, \mathbf{u}_t)^\top \mathbf{p}_{t+1}, && t = 0, \ldots, T-1
\end{aligned}$$

- An optimal primal-dual sequence $\mathbf{x}^*_{1:T}$, $\mathbf{u}^*_{0:T-1}$, $\mathbf{p}^*_{1:T}$ satisfies:

$$\begin{aligned}
\nabla_{\mathbf{p}_t} L(\mathbf{x}^*_{1:T}, \mathbf{u}^*_{0:T-1}, \mathbf{p}^*_{1:T}) &= 0, && t = 1, \ldots, T \\
\nabla_{\mathbf{x}_t} L(\mathbf{x}^*_{1:T}, \mathbf{u}^*_{0:T-1}, \mathbf{p}^*_{1:T}) &= 0, && t = 1, \ldots, T \\
\nabla_{\mathbf{u}_t} L(\mathbf{x}^*_{1:T}, \mathbf{u}^*_{0:T-1}, \mathbf{p}^*_{1:T}) &= 0, && t = 0, \ldots, T-1
\end{aligned}$$

Pontryagin's Minimum Principle

- Define the **Hamiltonian**:

$$H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \ell(\mathbf{x}, \mathbf{u}) + \mathbf{p}^\top f(\mathbf{x}, \mathbf{u})$$

Theorem: Discrete-time Pontryagin Minimum Principle

If $\mathbf{x}^*_{0:T}$, $\mathbf{u}^*_{0:T-1}$ is an optimal state-control trajectory for the finite-horizon deterministic optimal control problem, then there exists a **costate trajectory $\mathbf{p}^*_{1:T}$** such that:

$$\begin{aligned}
\mathbf{x}^*_{t+1} &= \nabla_{\mathbf{p}} H(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1}) = f(\mathbf{x}^*_t, \mathbf{u}^*_t), && \mathbf{x}^*_0 = \mathbf{x}_0 \\
\mathbf{p}^*_t &= \nabla_{\mathbf{x}} H(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1}) = \nabla_{\mathbf{x}} \ell(\mathbf{x}^*_t, \mathbf{u}^*_t) + \nabla_{\mathbf{x}} f(\mathbf{x}^*_t, \mathbf{u}^*_t)^\top \mathbf{p}^*_{t+1}, && \mathbf{p}^*_T = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}^*_T) \\
\mathbf{0} &= \nabla_{\mathbf{u}} H(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1})
\end{aligned}$$

Pontryagin's Minimum Principle

- Consider a finite-horizon deterministic optimal control problem with
  - time-varying stage cost ℓ_t and motion model f_t
  - terminal constraint $\mathbf{x}_T \in \mathcal{T}$
  - input constraints $\mathcal{U}$

- Hamiltonian with normality variable $\eta \in \{0, 1\}$:

$$H_t(\mathbf{x}, \mathbf{u}, \mathbf{p}, \eta) = \eta\ell_t(\mathbf{x}, \mathbf{u}) + \mathbf{p}^\top f_t(\mathbf{x}, \mathbf{u})$$

Theorem: Discrete-time Pontryagin Minimum Principle

If $\mathbf{x}^*_{0:T}$, $\mathbf{u}^*_{0:T-1}$ is an optimal state-control trajectory, then there exists a costate trajectory $\mathbf{p}^*_{1:T}$ and $\eta \in \{0, 1\}$ such that:

$$\begin{aligned}
(\eta, \mathbf{p}^*_1, \ldots, \mathbf{p}^*_T) &\neq \mathbf{0} \\
\mathbf{x}^*_{t+1} &= \nabla_{\mathbf{p}} H_t(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1}, \eta), \qquad & \mathbf{x}^*_0 &= \mathbf{x}_0 \\
\mathbf{p}^*_t &= \nabla_{\mathbf{x}} H_t(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1}, \eta), \qquad & \mathbf{p}^*_T - \eta\nabla_{\mathbf{x}}\mathfrak{q}(\mathbf{x}^*_T) &\perp_{\mathbf{x}^*_T} \mathcal{T} \\
-\nabla_{\mathbf{u}} H_t(\mathbf{x}^*_t, \mathbf{u}^*_t, \mathbf{p}^*_{t+1}, \eta) &\perp_{\mathbf{u}^*_t} \mathcal{U}
\end{aligned}$$

- $\mathbf{g} \perp_{\mathbf{x}} \mathcal{X}$ means that $\mathbf{g}$ is orthogonal to the tangent cone of $\mathcal{X}$ at $\mathbf{x}$

Pontryagin's Minimum Principle

- PMP provides an efficient way to evaluate the value function gradient with respect to $\mathbf{u}_t$ and thus optimize control trajectories locally and numerically
- Given initial state $\mathbf{x}_0$ and trajectory $\mathbf{u}_{0:T-1}$, let $\mathbf{x}_{1:T}$, $\mathbf{p}_{1:T}$ be such that:

$$\begin{aligned} \mathbf{x}_{t+1} &= f(\mathbf{x}_t, \mathbf{u}_t), & \mathbf{x}_0 \text{ given} \\ \mathbf{p}_t &= \nabla_{\mathbf{x}} \ell(\mathbf{x}_t, \mathbf{u}_t) + [\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{u}_t)]^\top \mathbf{p}_{t+1}, & \mathbf{p}_T = \nabla_{\mathbf{x}} \mathfrak{q}(\mathbf{x}_T) \end{aligned}$$

- Then:

$$\nabla_{\mathbf{u}_t} V_0^{\mathbf{u}_{0:T-1}}(\mathbf{x}_0) = \nabla_{\mathbf{u}} H(\mathbf{x}_t, \mathbf{u}_t, \mathbf{p}_{t+1}) = \nabla_{\mathbf{u}} \ell(\mathbf{x}_t, \mathbf{u}_t) + \nabla_{\mathbf{u}} f(\mathbf{x}_t, \mathbf{u}_t)^\top \mathbf{p}_{t+1}$$

- The states $\mathbf{x}_t$ can be found in a forward pass and then the costates $\mathbf{p}_t$ and value function gradients $\nabla_{\mathbf{u}_t} V_0^{\mathbf{u}_{0:T-1}}(\mathbf{x}_0)$ can be found in a backward pass
- **Claim**: $\mathbf{p}_t = \nabla_{\mathbf{x}_t} V_t^{\mathbf{u}_{t:T-1}}(\mathbf{x}_t)$:
  - Base case: $\mathbf{p}_T = \nabla_{\mathbf{x}_T} \mathfrak{q}(\mathbf{x}_T)$
  - Induction: identical with costate difference equation

$$\underbrace{\nabla_{\mathbf{x}_t} V_t^{\mathbf{u}_{t:T-1}}(\mathbf{x}_t)}_{=\mathbf{p}_t} = \nabla_{\mathbf{x}} \ell(\mathbf{x}_t, \mathbf{u}_t) + [\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{u}_t)]^\top \underbrace{\nabla_{\mathbf{x}_{t+1}} V_{t+1}^{\mathbf{u}_{t+1:T-1}}(\mathbf{x}_{t+1})}_{=\mathbf{p}_{t+1}}$$

Outline

Finite-horizon Linear Quadratic Regulator

- **Linear Quadratic Regulator** (LQR) problem:

$$\begin{aligned} \min_{\pi_{0:T-1}} \quad & V_0^\pi(\mathbf{x}) := \frac{1}{2}\mathbf{x}_T^\top \mathbb{Q} \mathbf{x}_T + \sum_{t=0}^{T-1} \left(\frac{1}{2}\mathbf{x}_t^\top Q \mathbf{x}_t + \frac{1}{2}\mathbf{u}_t^\top R \mathbf{u}_t \right) \\ \text{s.t.} \quad & \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t, \;\; \mathbf{x}_0 = \mathbf{x} \\ & \mathbf{x}_t \in \mathbb{R}^n, \; \mathbf{u}_t = \pi_t(\mathbf{x}_t) \in \mathbb{R}^m \end{aligned}$$

  where $Q = Q^\top \succeq 0$, $\mathbb{Q} = \mathbb{Q}^\top \succeq 0$, and $R = R^\top \succ 0$

- This is a special case of the finite-horizon deterministic optimal control problem, which can be solved via Dynamic Programming

Finite-horizon Linear Quadratic Regulator

- At $t = T$, the value function equals the terminal cost which is quadratic in $\mathbf{x}$:

$$V_T^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M_T \mathbf{x} := \frac{1}{2}\mathbf{x}^\top \mathbb{Q} \mathbf{x}$$

- Iterate backwards in time $t = T-1, \ldots, 0$:

$$V_t^*(\mathbf{x}) = \min_{\mathbf{u}} \underbrace{\left\{ \frac{1}{2}\left(\mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u}\right) + V_{t+1}^*(A\mathbf{x} + B\mathbf{u}) \right\}}_{Q_t^*(\mathbf{x},\mathbf{u})}$$

- At $t = T-1$:

$$V_{T-1}^*(\mathbf{x}) = \min_{\mathbf{u}} \frac{1}{2}\left\{\mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} + (A\mathbf{x} + B\mathbf{u})^\top M_T (A\mathbf{x} + B\mathbf{u})\right\}$$

- Since $R \succ 0$, the cost above is a positive-definite quadratic function of $\mathbf{u}$
- Taking the gradient and setting it equal to 0:

$$\pi_{T-1}^*(\mathbf{x}) = -\left(B^\top \mathbb{Q} B + R\right)^{-1} B^\top \mathbb{Q} A \mathbf{x}$$

$$V_{T-1}^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M_{T-1} \mathbf{x}$$

$$M_{T-1} = A^\top M_T A + Q - A^\top M_T B \left(B^\top M_T B + R\right)^{-1} B^\top M_T A$$

Finite-horizon Linear Quadratic Regulator

- At $t = T - 2$:

$$V_{T-2}^*(\mathbf{x}) = \min_{\mathbf{u}} \frac{1}{2} \left\{ \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} + (A\mathbf{x} + B\mathbf{u})^\top M_{T-1} (A\mathbf{x} + B\mathbf{u}) \right\}$$

- Since $R \succ 0$, the cost above is a positive-definite quadratic function of $\mathbf{u}$

- Taking the gradient and setting it equal to 0:

$$\pi_{T-2}^*(\mathbf{x}) = -\left(B^\top M_{T-1} B + R\right)^{-1} B^\top M_{T-1} A \mathbf{x}$$

$$V_{T-2}^*(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top M_{T-2} \mathbf{x}$$

$$M_{T-2} = A^\top M_{T-1} A + Q - A^\top M_{T-1} B \left(B^\top M_{T-1} B + R\right)^{-1} B^\top M_{T-1} A$$

- The optimal value and policy are determined by a **Riccati equation** for M_t:

$$\pi_t^*(\mathbf{x}) = -\left(B^\top M_{t+1} B + R\right)^{-1} B^\top M_{t+1} A \mathbf{x}$$

$$V_t^*(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top M_t \mathbf{x}$$

$$M_t = A^\top M_{t+1} A + Q - A^\top M_{t+1} B \left(B^\top M_{t+1} B + R\right)^{-1} B^\top M_{t+1} A, \quad M_T = \mathbb{Q}$$

Finite-horizon Linear Quadratic Regulator

- **Batch formulation**: instead of using the DP algorithm, express the system evolution as a large matrix system

$$\underbrace{\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} I \\ A \\ \vdots \\ A^T \end{bmatrix}}_{\mathcal{A}} \mathbf{x}_0 + \underbrace{\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ B & 0 & \cdots & 0 \\ AB & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ A^{T-1}B & \cdots & \cdots & B \end{bmatrix}}_{\mathcal{B}} \underbrace{\begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{T-1} \end{bmatrix}}_{\mathbf{v}}$$

- The batch formulation of LQR is a quadratic program in **s** and **v**:

$$\min_{\mathbf{s},\mathbf{v}} V_0^{\pi}(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{s}^T \mathcal{Q} \mathbf{s} + \mathbf{v}^T \mathcal{R} \mathbf{v}\right) \qquad \mathcal{Q} := \mathbf{diag}(Q, \ldots, Q, \mathbb{Q}) \succeq 0$$

$$\text{s.t.} \quad \mathbf{s} - \mathcal{B}\mathbf{v} = \mathcal{A}\mathbf{x}_0 \qquad \mathcal{R} := \mathbf{diag}(R, \ldots, R) \succ 0$$

Finite-horizon Linear Quadratic Regulator

- Express $V_0^\pi(\mathbf{x}_0)$ only in terms of the initial condition $\mathbf{x}_0$ and the control sequence $\mathbf{v}$ by using the batch dynamics $\mathbf{s} = \mathcal{A}\mathbf{x}_0 + \mathcal{B}\mathbf{v}$:

$$V_0^\pi(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{v}^\top\left(\mathcal{B}^\top\mathcal{Q}\mathcal{B} + \mathcal{R}\right)\mathbf{v} + 2\mathbf{x}_0^\top\left(\mathcal{A}^\top\mathcal{Q}\mathcal{B}\right)\mathbf{v} + \mathbf{x}_0^\top\mathcal{A}^\top\mathcal{Q}\mathcal{A}\mathbf{x}_0\right)$$

- $V_0^\pi(\mathbf{x}_0)$ is a positive-definite quadratic function of $\mathbf{v}$ since $\mathcal{R} \succ 0$

- Take gradient wrt $\mathbf{v}$ and set to 0:

$$\mathbf{v}^* = -\left(\mathcal{B}^\top\mathcal{Q}\mathcal{B} + \mathcal{R}\right)^{-1}\mathcal{B}^\top\mathcal{Q}\mathcal{A}\mathbf{x}_0$$

$$V_0^*(\mathbf{x}_0) = \frac{1}{2}\mathbf{x}_0^\top\left(\mathcal{A}^\top\mathcal{Q}\mathcal{A} - \mathcal{A}^\top\mathcal{Q}\mathcal{B}\left(\mathcal{B}^\top\mathcal{Q}\mathcal{B} + \mathcal{R}\right)^{-1}\mathcal{B}^\top\mathcal{Q}\mathcal{A}\right)\mathbf{x}_0$$

- The optimal sequence of control inputs $\mathbf{u}^*_{0:T-1}$ is a linear function of $\mathbf{x}_0$

- The optimal value function $V_0^*(\mathbf{x}_0)$ is a quadratic function of $\mathbf{x}_0$

Outline

Finite-horizon Linear Quadratic Gaussian

- **Linear Quadratic Gaussian** (LQG) regulation problem:

$$\min_{\pi_{0:T-1}} \quad V_0^{\pi}(\mathbf{x}) = \mathbb{E}\left\{\gamma^T \frac{1}{2}\mathbf{x}_T^\top \mathbb{Q}\mathbf{x}_T + \sum_{t=0}^{T-1} \gamma^t \frac{1}{2}\left(\mathbf{x}_t^\top Q\mathbf{x}_t + 2\mathbf{u}_t^\top P\mathbf{x}_t + \mathbf{u}_t^\top R\mathbf{u}_t\right)\right\}$$

$$\text{s.t.} \quad \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + C\mathbf{w}_t, \;\; \mathbf{x}_0 = \mathbf{x}, \; \mathbf{w}_t \sim \mathcal{N}(0, I)$$

$$\mathbf{x}_t \in \mathbb{R}^n, \; \mathbf{u}_t = \pi_t(\mathbf{x}_t) \in \mathbb{R}^m$$

- **Discount factor**: $\gamma \in [0, 1]$
- **Optimal value**: $V_t^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M_t \mathbf{x} + m_t$
- **Optimal policy**: $\pi_t^*(\mathbf{x}) = -(R + \gamma B^\top M_{t+1} B)^{-1}(P + \gamma B^\top M_{t+1} A)\mathbf{x}$
- **Riccati equation**:

$$M_t = Q + \gamma A^\top M_{t+1} A - (P + \gamma B^\top M_{t+1} A)^\top (R + \gamma B^\top M_{t+1} B)^{-1}(P + \gamma B^\top M_{t+1} A), \quad M_T = \mathbb{Q}$$

$$m_t = \gamma m_{t+1} + \gamma \frac{1}{2}\operatorname{tr}(CC^\top M_{t+1}), \qquad m_T = 0$$

- M_t is independent of the noise amplitude C, which implies that the optimal policy $\pi_t^*(\mathbf{x})$ is **the same for LQG and LQR**!

Infinite-horizon Linear Quadratic Gaussian

- **Linear Quadratic Gaussian** (LQG) regulation problem:

$$
\begin{aligned}
\min_{\pi} \quad & V^{\pi}(\mathbf{x}) := \mathbb{E}\left\{\sum_{t=0}^{\infty} \gamma^t \frac{1}{2}\left(\mathbf{x}_t^\top Q \mathbf{x}_t + 2\mathbf{u}_t^\top P \mathbf{x}_t + \mathbf{u}_t^\top R \mathbf{u}_t\right)\right\} \\
\text{s.t.} \quad & \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + C\mathbf{w}_t, \;\; \mathbf{x}_0 = \mathbf{x}, \; \mathbf{w}_t \sim \mathcal{N}(0, I) \\
& \mathbf{x}_t \in \mathbb{R}^n, \; \mathbf{u}_t = \pi(\mathbf{x}_t) \in \mathbb{R}^m
\end{aligned}
$$

- **Discount factor**: $\gamma \in [0, 1)$

- **Optimal value**: $V^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M \mathbf{x} + m$

- **Optimal policy**: $\pi^*(\mathbf{x}) = -(R + \gamma B^\top M B)^{-1}(P + \gamma B^\top M A)\mathbf{x}$

- **Riccati equation** ('dare' in Matlab):

$$
\begin{aligned}
M &= Q + \gamma A^\top M A - (P + \gamma B^\top M A)^\top (R + \gamma B^\top M B)^{-1}(P + \gamma B^\top M A) \\
m &= \frac{\gamma}{2(1-\gamma)} \operatorname{tr}(CC^\top M)
\end{aligned}
$$

- M is independent of the noise amplitude C, which implies that the optimal policy $\pi^*(\mathbf{x})$ is **the same for LQG and LQR**!

Outline

Pontryagin's Minimum Principle

Linear Quadratic Regulator

Linear Quadratic Gaussian

LQR Methods for Deterministic Optimal Control

Deterministic Optimal Control

- Deterministic optimal control with initial state $\mathbf{x}_0$:

$$\min_{\mathbf{u}_{0:T-1}} \; V_0^{\mathbf{u}_{0:T-1}}(\mathbf{x}_0) = \mathfrak{q}(\mathbf{x}_T) + \sum_{t=0}^{T-1} \ell(\mathbf{x}_t, \mathbf{u}_t)$$
$$\text{s.t. } \; \mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t), \; t = 0, \ldots, T-1$$

- The problem has a closed-form solution when the costs are **quadratic** and the dynamics are **linear**:

$$\begin{aligned}
\mathfrak{q}(\mathbf{x}) &= \frac{1}{2}\mathbf{x}^\top \mathbb{Q}\mathbf{x} + \mathbf{a}^\top \mathbf{x} + a, \qquad \mathbb{Q} \succeq 0, \; Q \succeq 0, \; R \succ 0, \\
\ell(\mathbf{x}, \mathbf{u}) &= \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \frac{1}{2}\mathbf{u}^\top R\mathbf{u} + \mathbf{x}^\top P\mathbf{u} + \mathbf{q}^\top \mathbf{x} + \mathbf{r}^\top \mathbf{u} + q \\
f(\mathbf{x}, \mathbf{u}) &= A\mathbf{x} + B\mathbf{u} + \mathbf{c}
\end{aligned}$$

Deterministic Optimal Control

- **Cost and dynamics**:

$$\mathfrak{q}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbb{Q}\mathbf{x} + \mathbf{a}^\top \mathbf{x} + a, \qquad \mathbb{Q} \succeq 0,\ Q_t \succeq 0,\ R_t \succ 0,$$
$$\ell_t(\mathbf{x}, \mathbf{u}) = \frac{1}{2}\mathbf{x}^\top Q_t \mathbf{x} + \frac{1}{2}\mathbf{u}^\top R_t \mathbf{u} + \mathbf{x}^\top P_t \mathbf{u} + \mathbf{q}_t^\top \mathbf{x} + \mathbf{r}_t^\top \mathbf{u} + q_t$$
$$f_t(\mathbf{x}, \mathbf{u}) = A_t \mathbf{x} + B_t \mathbf{u} + \mathbf{c}_t$$

- **Optimal value**: $V_t^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M_t \mathbf{x} + \mathbf{m}_t^\top \mathbf{x} + m_t$

- **Optimal policy**: $\pi_t^*(\mathbf{x}) = -H_{\mathbf{uu},t}^{-1}(H_{\mathbf{xu},t}^\top \mathbf{x} + \mathbf{h}_{\mathbf{u},t})$

- **Riccati equations**:

$$M_T = \mathbb{Q}, \qquad \mathbf{m}_T = \mathbf{a}, \qquad m_T = a$$
$$M_t = H_{\mathbf{xx},t} - H_{\mathbf{xu},t} H_{\mathbf{uu},t}^{-1} H_{\mathbf{xu},t}^\top$$
$$\mathbf{m}_t = A_t^\top(\mathbf{m}_{t+1} + M_{t+1}\mathbf{c}_t) + \mathbf{q}_t - H_{\mathbf{xu},t} H_{\mathbf{uu},t}^{-1} \mathbf{h}_{\mathbf{u},t}$$
$$m_t = -\frac{1}{2}\mathbf{h}_{\mathbf{u},t}^\top H_{\mathbf{uu},t}^{-1} \mathbf{h}_{\mathbf{u},t} + \frac{1}{2}\mathbf{c}_t^\top M_{t+1}\mathbf{c}_t + \mathbf{m}_{t+1}^\top \mathbf{c}_t + m_{t+1} + q_t$$
$$\mathbf{h}_{\mathbf{u},t} = B_t^\top(\mathbf{m}_{t+1} + M_{t+1}\mathbf{c}_t) + \mathbf{r}_t$$
$$H_{\mathbf{xx},t} = Q_t + A_t^\top M_{t+1} A_t, \quad H_{\mathbf{uu},t} = R_t + B_t^\top M_{t+1} B_t, \quad H_{\mathbf{xu},t} = P_t + A_t^\top M_{t+1} B_t$$

Iterative LQR (iLQR)

- **iLQR** repeatedly approximates the cost and dynamics as quadratic and affine respectively and solves the resulting LQR problem
- Initialize control and state sequence $\bar{\mathbf{u}}_{0:T-1}$ and $\bar{\mathbf{x}}_{t+1} = f(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)$
- Define errors: $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \bar{\mathbf{x}}_t$ and $\tilde{\mathbf{u}}_t = \mathbf{u}_t - \bar{\mathbf{u}}_t$
- Approximate the cost and dynamics:

$$\mathfrak{q}(\tilde{\mathbf{x}}_T) \approx \underbrace{\mathfrak{q}(\bar{\mathbf{x}}_T)}_{=:a} + \underbrace{\nabla_{\mathbf{x}} \mathfrak{q}(\bar{\mathbf{x}}_T)}_{=:\mathbf{a}}^\top \tilde{\mathbf{x}}_T + \frac{1}{2} \tilde{\mathbf{x}}_T^\top \underbrace{\nabla^2_{\mathbf{x}} \mathfrak{q}(\bar{\mathbf{x}}_T)}_{=:\mathbb{Q}} \tilde{\mathbf{x}}_T$$

$$\begin{aligned}
\ell(\tilde{\mathbf{x}}_t, \tilde{\mathbf{u}}_t) \approx{} & \underbrace{\ell(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:q_t} + \underbrace{\nabla_{\mathbf{x}} \ell(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:\mathbf{q}_t}^\top \tilde{\mathbf{x}}_t + \underbrace{\nabla_{\mathbf{u}} \ell(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:\mathbf{r}_t}^\top \tilde{\mathbf{u}}_t \\
& + \frac{1}{2} \tilde{\mathbf{x}}_t^\top \underbrace{\nabla^2_{\mathbf{xx}} \ell_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:Q_t} \tilde{\mathbf{x}}_t + \frac{1}{2} \tilde{\mathbf{u}}_t^\top \underbrace{\nabla^2_{\mathbf{uu}} \ell_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:R_t} \tilde{\mathbf{u}}_t + \tilde{\mathbf{x}}_t^\top \underbrace{\nabla^2_{\mathbf{xu}} \ell_t(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:P_t} \tilde{\mathbf{u}}_t
\end{aligned}$$

$$\tilde{\mathbf{x}}_{t+1} \approx \underbrace{\frac{\partial f}{\partial \mathbf{x}}(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:A_t} \tilde{\mathbf{x}}_t + \underbrace{\frac{\partial f}{\partial \mathbf{u}}(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)}_{=:B_t} \tilde{\mathbf{u}}_t$$

Iterative LQR (iLQR)

- Solve the general LQR problem with cost and dynamics:

$$\begin{aligned}
\mathfrak{q}(\tilde{\mathbf{x}}) &= \frac{1}{2}\tilde{\mathbf{x}}^\top \mathbb{Q}\tilde{\mathbf{x}} + \mathbf{a}^\top \tilde{\mathbf{x}} + a, \\
\ell_t(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) &= \frac{1}{2}\tilde{\mathbf{x}}^\top Q_t \tilde{\mathbf{x}} + \frac{1}{2}\tilde{\mathbf{u}}^\top R_t \tilde{\mathbf{u}} + \tilde{\mathbf{x}}^\top P_t \tilde{\mathbf{u}} + \mathbf{q}_t^\top \tilde{\mathbf{x}} + \mathbf{r}_t^\top \tilde{\mathbf{u}} + q_t \\
f_t(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) &= A_t \tilde{\mathbf{x}} + B_t \tilde{\mathbf{u}}
\end{aligned}$$

  to obtain an optimal policy $\tilde{\mathbf{u}}_t = \pi_t^*(\tilde{\mathbf{x}}_t)$

- Update the nominal control trajectory: $\bar{\mathbf{u}}_t \leftarrow \bar{\mathbf{u}}_t + \tilde{\mathbf{u}}_t$ for $t = 0, \ldots, T-1$

- Repeat the whole process until convergence (e.g., the change in $\tilde{\mathbf{u}}_{0:T-1}$ or $V_0^{\bar{\mathbf{u}}_{0:T-1}}(\bar{\mathbf{x}}_0)$ between iterations is small)

Differential Dynamic Programming

- The Bellman equation for a deterministic optimal control is:

$$V_t^*(\mathbf{x}_t) = \min_{\mathbf{u}_t} \left\{ \ell(\mathbf{x}_t, \mathbf{u}_t) + V_{t+1}^*(f(\mathbf{x}_t, \mathbf{u}_t)) \right\}$$

- iLQR approximates the cost and dynamics as quadratic and affine. Then, the right-hand side of the Bellman equation is quadratic and can be minimized to find $\tilde{\mathbf{u}}_t^*$

- **DDP** assumes $V_t^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top M_t \mathbf{x} + \mathbf{m}_t^\top \mathbf{x} + m_t$ and directly approximates the right-hand side of the Bellman equation as a quadratic

- DDP is equivalent to iLQR, except the second-roder terms are now:

$$\begin{aligned}
H_{\mathbf{xx},t} &= Q_t + A_t^\top M_{t+1} A_t + \textstyle\sum_{i=1}^d \mathbf{m}_{t+1,i} \nabla^2_{\mathbf{xx}} f_i(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) \\
H_{\mathbf{uu},t} &= R_t + B_t^\top M_{t+1} B_t + \textstyle\sum_{i=1}^d \mathbf{m}_{t+1,i} \nabla^2_{\mathbf{uu}} f_i(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) \\
H_{\mathbf{xu},t} &= P_t + A_t^\top M_{t+1} B_t + \textstyle\sum_{i=1}^d \mathbf{m}_{t+1,i} \nabla^2_{\mathbf{xu}} f_i(\bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t)
\end{aligned}$$

- DDP estimates the Bellman equation more accurately than iLQR but requires computing second-order derivatives

Comments about iLQR and DDP

- Both iLQR and DDP produce an open-loop trajectory $\bar{\mathbf{x}}_{0:T}, \bar{\mathbf{u}}_{0:T-1}$ that is locally optimal and a policy $\pi_t(\mathbf{x}; \bar{\mathbf{x}}_t, \bar{\mathbf{u}}_t) = \bar{\mathbf{u}}_t - H_{\mathbf{uu},t}^{-1}(H_{\mathbf{xu},t}^\top(\mathbf{x} - \bar{\mathbf{x}}_t) + \mathbf{h}_{\mathbf{u},t})$ that is locally optimal for closed-loop tracking

- Since these methods are local optimization techniques, they can get stuck in local minima and require good initialization

- If the second-order terms $H_{\mathbf{xx},t}$ and $H_{\mathbf{uu},t}$ are not positive-semidefinite and positive-definite, respectively, we can try regularizing them (i.e., $H_{\mathbf{xx},t} + \mu I$ and $H_{\mathbf{uu},t} + \mu I$) or projecting them

- The termination criterion is a design choice, e.g., we can stop when either the change in the control trajectory is small, or when the cost improvement is small

- A collection of tips with mathematical details can be found in Ch. 2 of Yuval Tassa's PhD thesis, *"Theory and Implementation of Biomimetic Motor Controllers,"* The Hebrew University of Jerusalem, 2011