Optimal Control for the Linear Quadratic Regulator

Lucas Janson and Sham Kakade

CS/Stat 184: Introduction to Reinforcement Learning
Fall 2022

Today

- Feedback from last lecture
- Recap
- Derivation of optimal LQR policy
- Extensions

Feedback from feedback forms

1. Thank you to everyone who filled out the forms!
2.

Today

- Feedback from last lecture
- Recap
- Derivation of optimal LQR policy
- Extensions

Recap: LQR

Problem Statement:

$$\arg \min_{\pi_0,\ldots,\pi_{T-1}:\mathbb{R}^d\to\mathbb{R}^k} \mathbb{E}\left[s_T^\top Q s_T + \sum_{t=0}^{T-1}(s_t^\top Q s_t + a_t^\top R a_t)\right]$$

Used to be Σ, but $\sigma^2 I$ is simpler

such that $s_{t+1} = As_t + Ba_t + w_t, \quad s_0 \sim \mu_0, \quad a_t = \pi_t(s_t), \quad w_t \sim N(0, \sigma^2 I)$

Value function for a policy $\pi = (\pi_0, \ldots, \pi_{T-1})$:

$$V_t^\pi(s) = \mathbb{E}\left[s_T^\top Q s_T + \sum_{i=t}^{T-1}(s_i^\top Q s_i + a_i^\top R a_i) \;\middle|\; a_i = \pi_i(s_i)\ \forall i \geq t,\ s_t = s\right]$$

And corresponding Q function:

$$Q_t^\pi(s,a) = \mathbb{E}\left[s_T^\top Q s_T + \sum_{i=t}^{T-1}(s_i^\top Q s_i + a_i^\top R a_i) \;\middle|\; a_t = a,\ a_i = \pi_i(s_i)\ \forall i > t,\ s_t = s\right]$$

Today

- ✓ Feedback from last lecture
- ✓ Recap
- Derivation of optimal LQR policy
- Extensions

LQR Optimal Control

$$V_t^\star(s) = \min_\pi V_t^\pi(s) = \min_{\pi_t, \pi_{t+1}, \ldots, \pi_{T-1}} \mathbb{E}\left[s_T^\top Q s_T + \sum_{i=t}^{T-1} (s_i^\top Q s_i + a_i^\top R a_i) \;\middle|\; a_i = \pi_i(s_i)\ \forall i \geq t,\ s_t = s \right]$$

Theorem:

1. $V_t^\star$ is a quadratic function, i.e., $V_t^\star(s) = s^\top P_t s + p_t$ for some $P_t \in \mathbb{R}^{d \times d}$ and $p_t \in \mathbb{R}^d$
2. The optimal policy $\pi_t^\star$ is linear, i.e., $\pi_t^\star(s) = -K_t s$ for some $K_t \in \mathbb{R}^{k \times d}$
3. P_t, p_t, and K_t can be computed exactly

Today: prove the above theorem, deriving the optimal policy along the way

Key Steps in the Proof

Dynamic programming (finite-horizon), stepping backwards in time from T to 0

1. **Base case:** Show that $V_T^\star(s)$ is quadratic

2. **Inductive hypothesis:** Assuming $V_{t+1}^\star(s)$ is quadratic,
   a) Show that $Q_t^\star(s, a)$ is quadratic (in both s and a)
   b) Derive the optimal policy $\pi_t^\star(s) = \arg\min_a Q_t^\star(s, a)$, and show that it's linear
   c) Show $V_t^\star(s)$ is quadratic

3. **Conclusion:** $V_t^\star(s)$ is quadratic and $\pi_t^\star(s)$ is linear and we'll have their formulas

Base case at T

Recall the value function at a given t is:

$$V_t^\pi(s) = \mathbb{E}\left[s_T^\top Q s_T + \sum_{i=t}^{T-1} (s_i^\top Q s_i + a_i^\top R a_i) \;\middle|\; a_i = \pi_i(s_i)\ \forall i \geq t,\ s_t = s\right]$$

For V_T^π, everything disappears except first term $s_T^\top Q s_T = s^\top Q s$:

$$V_T^\star(s) = s^\top Q s$$

Denoting $P_T := Q$ and $p_T := 0$, we get

$$V_T^\star(s) = s^\top P_T s + p_T$$

(P_t and p_t didn't do much here, but we're going to define them recursively in the next step)

Induction Step

Assume $V^{\star}_{t+1}(s) = s^{\top} P_{t+1} s + p_{t+1}$, for all s, where $P_{t+1} \in \mathbb{R}^{d \times d}$ and $p_{t+1} \in \mathbb{R}^{d}$

$Q^{\star}_{t}(s, a) =$

Induction Step (continued)

$$Q_t^\star(s,a) = c(s,a) + \mathbb{E}_{s' \sim f(s,a,w_{t+1})}\left[V_{t+1}^\star(s')\right]$$
$$= s^\top \left(Q + A^\top P_{t+1} A\right) s + a^\top \left(R + B^\top P_{t+1} B\right) a + 2 s^\top A^\top P_{t+1} B a + \mathrm{tr}\left(\sigma^2 P_{t+1}\right) + p_{t+1}$$

$$\pi_t^\star(s) = \arg\min_a Q_t^\star(s,a)$$

Set $\nabla_a Q_t^\star(s,a) = 0$ and solve for a:

$$\nabla_a Q_t^\star(s,a) =$$

Concluding the Induction step:

$$Q_t^\star(s,a) = s^\top \left(Q + A^\top P_{t+1} A\right) s + a^\top \left(R + B^\top P_{t+1} B\right) a + 2 s^\top A^\top P_{t+1} B a + \mathrm{tr}\left(\sigma^2 P_{t+1}\right) + p_{t+1}$$

$$\pi_t^\star(s) = -\underbrace{(R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A}_{:=K_t} s$$

$$\begin{aligned} V_t^\star(s) &= Q_t^\star(s, \pi_t^\star(s)) \\ &= s^\top \left(Q + A^\top P_{t+1} A\right) s + s^\top K_t^\top \left(R + B^\top P_{t+1} B\right) K_t s - 2 s^\top A^\top P_{t+1} B K_t s + \mathrm{tr}\left(\sigma^2 P_{t+1}\right) + p_{t+1} \end{aligned}$$

Collecting the quadratic and constant terms together, $V_t^\star(s) = s^\top P_t s + p_t$, where:

$$P_t = Q + A^\top P_{t+1} A - A^\top P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$$ ← **Ricatti Equation**

$$p_t = \mathrm{tr}\left(\sigma^2 P_{t+1}\right) + p_{t+1}$$

Summary:

$$V_T^\star(s) = s^\top Q s, \quad \text{define } P_T = Q, p_T = 0,$$

We have shown that $V_t^\star(s) = s^\top P_t s + p_t$, where:

$$P_t = Q + A^\top P_{t+1} A - A^\top P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$$

$$p_t = \text{tr}\left(\sigma^2 P_{t+1}\right) + p_{t+1}$$

Along the way, we also have shown that $\pi_t^\star(s) = -K_t s$, where:

$$K_t = (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$$

Optimal policy has nothing to do with initial distribution μ_0 or the noise σ^2!

Today

- ✓ Feedback from last lecture
- ✓ Recap
- ✓ Derivation of optimal LQR policy
- Extensions

Time-Dependent Costs and Dynamics

$$\arg \min_{\pi_0,\ldots,\pi_{T-1}:\mathbb{R}^d \to \mathbb{R}^k} \mathbb{E}\left[s_T^\top Q_T s_T + \sum_{t=0}^{T-1} (s_t^\top Q_t s_t + a_t^\top R_t a_t) \right]$$

$$\text{such that} \quad s_{t+1} = A_t s_t + B_t a_t + w_t\,, \quad s_0 \sim \mu_0\,, \quad a_t = \pi_t(s_t)\,, \quad w_t \sim N(0,\sigma^2 I)$$

Exact same derivation, only thing that changes is the Ricatti equation:

$$P_t = Q_t + A_t^\top P_{t+1} A_t - A_t^\top P_{t+1} B_t (R_t + B_t^\top P_{t+1} B_t)^{-1} B_t^\top P_{t+1} A_t$$

More General Quadratic Cost Function

$$\arg \min_{\pi_0,\ldots,\pi_{T-1}:\mathbb{R}^d \to \mathbb{R}^k} \mathbb{E}\left[s_T^\top Q_T s_T + s_T^\top q_T + c_T + \sum_{t=0}^{T-1} (s_t^\top Q_t s_t + a_t^\top R_t a_t + a_t^\top M_t s_t + s_t^\top q_t + a_t^\top r_t + c_t) \right]$$

such that $s_{t+1} = A_t s_t + B_t a_t + v_t + w_t, \quad s_0 \sim \mu_0, \quad a_t = \pi_t(s_t), \quad w_t \sim N(0, \sigma^2 I)$

Derivation is similar—you will work it out on HW3

Tracking a Predefined Trajectory

$$\arg \min_{\pi_0,\ldots,\pi_{T-1}:\mathbb{R}^d \to \mathbb{R}^k} \mathbb{E}\left[(s_T - s_T^\star)^\top Q_T (s_T - s_T^\star) + \sum_{t=0}^{T-1} \left((s_t - s_t^\star)^\top Q_t (s_t - s_t^\star) + (a_t - a_t^\star)^\top R_t (a_t - a_t^\star)\right)\right]$$

such that $s_{t+1} = A_t s_t + B_t a_t + w_t\,, \quad s_0 \sim \mu_0\,, \quad a_t = \pi_t(s_t)\,, \quad w_t \sim N(0, \sigma^2 I)$

Expanding all the quadratic terms produces a special case of the previous slide!

Today

- ✓ Feedback from last lecture
- ✓ Recap
- ✓ Derivation of optimal LQR policy
- ✓ Extensions

Today's summary:

LQR optimal policy/controller

- Used dynamic programming / inductive argument to derive $\pi_t^\star$
- Same argument applies to extensions to some more complicated situations

Next time:

- Applying LQR approximation separately at each time point to get a locally-optimal solution to a nonlinear control problem

1-minute feedback form: https://bit.ly/3RHtlxy