EE 221A: Special Lecture on the Linear Quadratic Regulator

Somil Bansal

1 Linear Quadratic Regulator (LQR)

Consider a discrete linear time-invariant dynamical system:

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t,\ t \in \{0, 1, \ldots, N\} \\ x_0 &= x^{init} \end{aligned} \tag{1}$$

Our goal is to minimize a quadratic cost functional:

$$J(U, x_0) = \sum_{\tau=0}^{N-1} \left(x_\tau^T Q x_\tau + u_\tau^T R u_\tau \right) + x_N^T Q_f x_N, \tag{2}$$

where $Q, R, Q_f \succeq 0$ are positive semidefinite matrices and $U := (u_0, u_1, \ldots, u_{N-1})$. In equation (2):

- N is called time horizon (we will consider $N = \infty$ later);
- $x_t^T Q x_t$ measures state deviation cost at $\tau = t$;
- $u_t^T R u_t$ measures input authority cost at $\tau = t$;
- $x_N^T Q_f x_N$ measures the terminal cost.

Thus, we would like to solve the following optimization problem:

$$\min_U \{J(U, x_0) \text{ subject to } (1)\} \tag{3}$$

The optimization problem in (3) can be solved in various ways; in this handout, we will use Dynamic Programming (making use of the Principle of Optimality) to solve (3).

2 LQR solution via Dynamic Programming

The basic idea of dynamic programming is that the optimal control sequence over the entire horizon remains optimal at intermediate points in time. To begin this discussion, we will embed the optimization problem which we are solving in a larger class of problems. More specifically, we will consider the original cost function of equation (2) from an initial time $t \in \{0, 1, \ldots, N\}$ by considering the cost function on the discrete interval $\{t : N\}$ (a shorthand for $\{t, t+1, \ldots, N\}$):

$$J(U_t, x_t) = \sum_{\tau=t}^{N-1} \left(x_\tau^T Q x_\tau + u_\tau^T R u_\tau \right) + x_N^T Q_f x_N, \tag{4}$$

where $U_t = (u_t, u_{t+1}, \ldots, u_{N-1})$. Bellman's principle of optimality says that if we have found the optimal trajectory on the interval from $\{0 : N\}$ by solving the optimal control problem on that interval, the resulting trajectory is also optimal on all subintervals of this interval of the form $\{t : N\}$ with $t > 0$, provided that the initial condition x_t at time t was obtained from running the system forward along the optimal trajectory from time 0. The optimal value of $J(U_t, x_t)$ is referred to as the "cost-to-go" starting from state x_t at time t and is denoted as $J_t^*(x_t)$ here on. Our goal is to find $J_0^*(x_0)$.

Now suppose at time t, we know $J_{t+1}^*(z)$ for all states z and are interested in finding the optimal u_t (and hence $J_t^*(x_t)$). Our choice of u_t affects:

- current cost incurred (through $u_t^T R u_t$)
- our next state, x_{t+1} (hence, the min-cost-to-go from x_{t+1}.)

The principle of optimality suggests that the optimal u_t is the solution to the following optimization problem:

$$J_t^*(x_t) = \min_{u_t} \left(x_t^T Q x_t + u_t^T R u_t + J_{t+1}^*(Ax_t + Bu_t) \right). \tag{5}$$

In the above equation:

- $x_t^T Q x_t + u_t^T R u_t$ is the cost incurred at time t;
- $J_{t+1}^*(Ax_t + Bu_t)$ is the minimum cost-to-go from where we land at $t + 1$.

We can thus find optimal control sequence by solving the above equation backwards starting at $t = N$. This is formalized in the theorem below:

Theorem 1. *The optimal cost-to-go and the optimal control at time t are given by:*

$$\begin{aligned} J_t^*(z) &= z^T P_t z \\ u_t^* &= -K_t z, \end{aligned} \tag{6}$$

where $t \in \{0, 1, \ldots, N-1\}$ and

$$\begin{aligned} P_t &= Q + K_t^T R K_t + (A - BK_t)^T P_{t+1} (A - BK_t), \ P_N = Q_f \\ K_t &= (R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A. \end{aligned} \tag{7}$$

Proof. We will prove the claim by induction. Note that for $i = N$, the LQR cost is independent of control (see (4)) and the optimal cost-to-go is trivially given by $z^T Q_f z$. Thus, $P_N = Q_f$.

Now assume that the theorem holds for $i = k$; our goal is to prove the claim for $i = k - 1$.

From equation (5), we have:

$$J_{k-1}^*(z) = \min_{u_{k-1}} \left(z^T Q z + u_{k-1}^T R u_{k-1} + J_k^*(Az + Bu_{k-1}) \right)$$

By induction hypothesis, $J_k^*(x) = x^T P_k x$. Therefore,

$$J_{k-1}^*(z) = \min_{u_{k-1}} \left(z^T Q z + u_{k-1}^T R u_{k-1} + (Az + Bu_{k-1})^T P_k (Az + Bu_{k-1}) \right) \tag{8}$$

The optimal u_{k-1} can be derived by setting the derivative of RHS equal to zero:

$$2u_{k-1}^T R + 2(Az + Bu_{k-1})^T P_k B = 0$$

Hence the optimal control is

$$\begin{aligned} u_{k-1}^* &= -(R + B^T P_k B)^{-1} B^T P_k A z \\ &= -K_{k-1} z. \end{aligned} \tag{9}$$

To get the optimal cost-to-go, we can substitute optimal control in (9) in equation (8):

$$\begin{aligned} J_{k-1}^*(z) &= z^T Q z + {u_{k-1}^*}^T R u_{k-1}^* + (Az + Bu_{k-1}^*)^T P_k (Az + Bu_{k-1}^*) \\ &= z^T \left(Q + K_{k-1}^T R K_{k-1} + (A - BK_{k-1})^T P_k (A - BK_{k-1}) \right) z \\ &= z^T P_{k-1} z. \end{aligned} \tag{10}$$

□

Remark 1. Theorem 1 implies that the optimal control is a linear function of the state (called *linear state feedback*). The open-loop dynamics should thus be modified as:

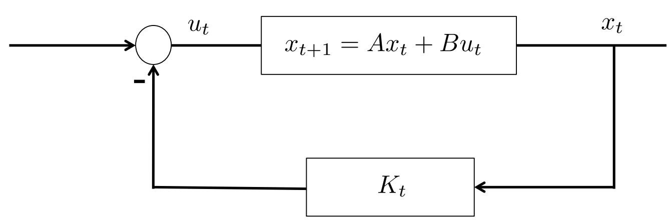

Figure 1: In LQR, the current state x_t is measured and is fed back in the system after passing it through a time-varying gain block, whose gain is given by K_t.

Moroever, the optimal cost-to-go under the optimal control policy is a quadratic function of the state.

Remark 2. Matrices P_t and K_t in the above theorem can be computed recursively backward in time starting from $t = N - 1$.

3 Extensions of LQR

Next, we discuss a few extensions of the basic LQR problem in (3).

3.1 Infinite horizon LQR

If $N = \infty$, we can still apply the dynamic programming principle; however, the recursive equation (7) reaches a steady state solution in this case:

$$\begin{aligned} P_{ss} &= Q + K_{ss}^T R K_{ss} + (A - BK_{ss})^T P_{ss} (A - BK_{ss}), \\ K_{ss} &= (R + B^T P_{ss} B)^{-1} B^T P_{ss} A. \end{aligned} \tag{11}$$

The above equations can be equivalently written as:

$$P_{ss} = Q + A^T P_{ss} A - A^T P_{ss} B (R + B^T P_{ss} B)^{-1} B^T P_{ss} A, \tag{12}$$

which is called the (DT) algebraic Riccati equation (ARE). P_{ss} in ARE can be found by iterating the Riccati recursion, or by direct methods. The optimal input in this case is given by a linear *constant* state feedback:

$$u_t = -K_{ss} x_t, \quad K_{ss} = (R + B^T P_{ss} B)^{-1} B^T P_{ss} A.$$

3.2 State affine systems

Suppose that the system dynamics in (1) are now given by:

$$\begin{aligned} x_{t+1} &= A x_t + B u_t + c, \; t \in \{0, 1, \ldots, N\} \\ x_0 &= x^{init} \end{aligned}$$

The optimal LQR control policy still remains linear and the optimal cost-to-go function still remains quadratic. To derive the optimal control policy, we can re-define the state as $z_t := [x_t, 1]$, then we have:

$$z_{t+1} = \begin{bmatrix} x_{t+1} \\ 1 \end{bmatrix} = \begin{bmatrix} A & c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ 1 \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_t = A' z_t + B' u_t, \tag{13}$$

which is in the standard LQR form discussed earlier. The optimal control policy is thus given by $u_t^* = -K_t z_t$.

LQR can also be readily extended to handle time-varying systems, for trajectory tracking problems, etc. We will see a few examples in homework and discussion session.

Our treatment of LQR in this handout is based on [1, 2, 3, 4].

References

[1] Pieter Abbeel. Advanced robotics lecture notes. *https://people.eecs.berkeley.edu/~pabbeel/cs287-fa15/slides/lecture5-LQR.pdf.*

[2] Stephen Boyd. Ee363 lecture notes. *http://stanford.edu/class/ee363/lectures/dlqr.pdf.*

[3] Peter Lancaster and Leiba Rodman. *Algebraic riccati equations.* Clarendon press, 1995.

[4] Claire Tomlin. Ee291e lecture notes. *https://inst.eecs.berkeley.edu/~ee291e/sp16/handouts/lecture8.pdf.*