ME 7247: Advanced Control Systems **Fall 2022–23**

Lecture 14: Stochastic LQR

Tuesday, October 25, 2022

Lecturer: Laurent Lessard *Scribe: Chang Wang*

In this lecture, we derive the stochastic version of the Linear Quadratic Regulator (LQR). Then, we will talk about the controllability and observability Gramians.

1 Stochastic LQR

The model is similar to the one used with LQR, except we now have process noise.

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{1}$$

The noise w_t is normally distributed with mean zero and variance:

$$w_t \sim \mathcal{N}(0, \Sigma_w)$$

With w_t, the state x_t becomes a random variable. Consequently, the standard quadratic cost used for deterministic LQR becomes a random variable. In order for this problem to make sense, the cost must be deterministic. The standard way to achieve this is to consider the *expected* cost. To this end, we define

$$J = \underset{u_0,\dots,u_{N-1}}{\text{minimize}} \quad \mathbf{E}\left[\sum_{t=0}^{N-1} (x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t) + x_N^\mathsf{T} Q_f x_N\right] \tag{2}$$

is also random. In order for this optimization to make sense, we need to make the cost function non-random. The standard way to do this is using expectation. So we will define

We will use dynamic programming like we did with deterministic LQR, but this will require a modified value function and a new version of the recursion we used in the deterministic case. We define the value function as:

$$V_t(z) \quad := \quad \underset{u_t,\dots,u_{N-1}}{\text{minimize}} \quad \mathbf{E}\left[\sum_{k=t}^{N-1} (x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t) + x_N^\mathsf{T} Q_f x_N \;\middle|\; x_t = z\right] \tag{3}$$

The modified recursion is

$$V_t(z) = \min_u \left(z^\mathsf{T} Q z + u^\mathsf{T} R u + \mathbf{E}\, V_{t+1}(Az + Bu + w_t)\right) \tag{4}$$

Note the expectation inside the minimization. This expectation is taken with respect to the random variable w_t. Eq. (4) is called the **principle of optimality**.

We *cannot* prove (4) using the same approach as we did in the deterministic case, because the expectation makes things more complicated. Specifically, we would have to exchange the order of minimization and expectation, which, in general, is not possible. For more details on how to prove the principle of optimality, please see the supplementary notes.

2 Optimal Solution for Stochastic LQR

We are going to use the same approach of last time. The difference this time is that the value function will be quadratic-plus-constant rather than being a pure quadratic. In other words, we will prove that $V_t(z) = z^\mathsf{T} P_t z + r_t$. At the terminal timestep, we have

$$P_N = Q_f \qquad \text{and} \qquad r_N = 0.$$

We use induction to prove that V_t has the required form. It holds for N, so assume it holds for $t+1$ and we will show that it holds for t. Substitute the formula for V_{t+1} into (4) and obtain

$$\begin{aligned}
V_t(z) &= \min_u \left(z^\mathsf{T} Q z + u^\mathsf{T} R u + \mathbf{E}_w \Big[(Az + Bu + w)^\mathsf{T} P_{t+1} (Az + Bu + w) + r_{t+1} \Big] \right) \\
&= \min_u \left(z^\mathsf{T} Q z + u^\mathsf{T} R u + (Az + Bu)^\mathsf{T} P_{t+1} (Az + Bu) + \mathbf{tr}(P_{t+1}\Sigma_w) + r_{t+1} \right) \\
&= \underbrace{\min_u \begin{bmatrix} z \\ u \end{bmatrix}^\mathsf{T} \begin{bmatrix} A^\mathsf{T} P_{t+1} A + Q & A^\mathsf{T} P_{t+1} B \\ B^\mathsf{T} P_{t+1} A & B^\mathsf{T} P_{t+1} B + R \end{bmatrix} \begin{bmatrix} z \\ u \end{bmatrix}}_{\text{same as deterministic LQR}} + \underbrace{\mathbf{tr}(P_{t+1}\Sigma_w) + r_{t+1}}_{\text{constant term}}
\end{aligned}$$

Therefore, we can write that $V_t(z) = z^\mathsf{T} P_t z + r_t$, with:

$$\begin{aligned}
P_N &= Q_f \\
P_t &= A^\mathsf{T} P_{t+1} A + Q - A^\mathsf{T} P_{t+1} B (B^\mathsf{T} P_{t+1} B + R)^{-1} B^\mathsf{T} P_{t+1} A \\
r_N &= 0 \\
r_t &= \mathbf{tr}(P_{t+1}\Sigma_w) + r_{t+1} \\
K_t &= -(B^\mathsf{T} P_{t+1} B + R)^{-1} B^\mathsf{T} P_{t+1} A
\end{aligned}$$

So the optimal policy for stochastic LQR is $u_t = K_t x_t$, where K_t is the same LQR gain as in the deterministic LQR case.

The total cost using the optimal policy is:

$$\begin{aligned}
V_0(x_0) &= x_0^\mathsf{T} P_0 x_0 + r_0 \\
&= x_0^\mathsf{T} P_0 x_0 + \sum_{t=1}^{N} \mathbf{tr}(P_t \Sigma_w)
\end{aligned}$$

The cost of deterministic LQR is simply $x_0^\mathsf{T} P_0 x_0$. In stochastic LQR, there is an *additional cost* given by the sum of $\mathbf{tr}(P_t\Sigma_w)$ terms, which is due to the process noise.

Note that if x_0 is also random, say $x_0 \sim \mathcal{N}(\mu_x, \Sigma_x)$, then we would instead obtain:

$$\begin{aligned}
V_0(x_0) &= \mathbf{E}(x_0^\mathsf{T} P_0 x_0) + r_0 \\
&= \mu_x^\mathsf{T} P_0 \mu_x + \mathbf{tr}(P_0 \Sigma_w) + \sum_{t=1}^{N} \mathbf{tr}(P_t \Sigma_w)
\end{aligned}$$

In stochastic LQR, we do not have $x_t \to 0$. The process noise added at every timestep causes the state to meander about zero, and it never quite settles down. This is why the cost has this ever-accumulating term that will go to infinity as N grows large. For this reason, it does not make sense to talk about the "steady-state" or "infinite-horizon" cost.

As the horizon goes to infinity, so does the cost. However, we can talk about *average cost*. This is found by taking the average and then the limit:

$$\begin{aligned} J_{\text{avg}} &= \lim_{N\to\infty} \frac{1}{N} \left(x_0^\mathsf{T} P_0 x_0 + \sum_{t=1}^{N} \mathbf{tr}(P_t \Sigma_w) \right) \\ &= \mathbf{tr}(P\Sigma_w) \end{aligned}$$

Where $P := \lim_{t\to\infty} P_t$ is the solution to the Discrete Algebraic Riccati Equation (DARE) we found for the deterministic LQR problem.

Based on what we just derived, we conclude that there is a fundamental equivalence between the deterministic and stochastic versions of the LQR problem. Specifically, the two following quantities are the same:

1. The *expected infinite-horizon cost* of a deterministic LQR problem (no process noise), where the initial state is $x_0 \sim \mathcal{N}(0, \Sigma)$.
2. The *average cost* of a stochastic LQR problem where the process noise is $w_t \sim \mathcal{N}(0, \Sigma)$.

3 Evaluating a suboptimal policy

If instead of using the optimal infinite-horizon LQR policy $u_t = Kx_t$, we used some other policy $u_t = \hat{K}x_t$, we can calculate the cost it would incur by substituting directly into the formula for the standard cost so that we can find the corresponding cost-to-go matrix $\hat{P}$

$$J = \sum_{t=0}^{\infty} (x_t^\mathsf{T} Q x_t + u_t^\mathsf{T} R u_t) = \sum_{t=0}^{\infty} x_t^\mathsf{T} (Q + \hat{K}^\mathsf{T} R \hat{K}) x_t$$

Now use the fact that $x_{t+1} = (A + B\hat{K})x_t$ so $x_t = (A + BK)^t x_0$ and we get:

$$J = \sum_{t=0}^{\infty} x_0^\mathsf{T} (A^\mathsf{T} + \hat{K}^\mathsf{T} B^\mathsf{T})^t (Q + \hat{K}^\mathsf{T} R \hat{K})(A + B\hat{K})^t x_0 = x_0^\mathsf{T} \hat{P} x_0$$

The matrix $\hat{P}$ satisfies the Lyapunov equation:

$$(A + B\hat{K})^\mathsf{T} \hat{P} (A + B\hat{K}) - \hat{P} + (Q + \hat{K}^\mathsf{T} R \hat{K}) = 0$$

So to find the cost for this suboptimal $\hat{K}$, we solve the Lyapunov equation above for $\hat{P}$, and then our cost is $x_0^\mathsf{T} \hat{P} x_0$. Note that $(A + B\hat{K})$ must be stable, otherwise the cost will be infinite.

4 Controllability

Controllability means you can reach any state from any other state. But, it doesn't mean every state is easy to reach. In this section, we are going to quantify the notion of how reachable different states are, and how much can you reach with a given input budget. Assume a system:

$$x_{t+1} = Ax_t + Bu_t \tag{5}$$

Suppose $x_0 = 0$, and we want to see what x_t are reachable with a fixed budget of energy

$$\sum_{t=0}^{\infty} \|u_t\|^2 \le \alpha^2 \tag{6}$$

The state x_t can be expressed in terms of the inputs $u_0, \dots, u_{t-1}$ as:

$$\begin{aligned} x_t &= A^t x_0 + A^{t-1}Bu_0 + A^{t-2}Bu_1 + \cdots + Bu_{t-1} \\ &= \underbrace{\begin{bmatrix} A^{t-1}B & \cdots & AB & B \end{bmatrix}}_{\mathcal{C}_t} \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{t-1} \end{bmatrix} \end{aligned} \tag{7}$$

where we defined $\mathcal{C}_t$ to be the controllability matrix. As we can see, a state x_t is reachable if $x_t \in \operatorname{range}(\mathcal{C}_t)$. Moreover, if $\operatorname{rank}(\mathcal{C}_t) = n$, then all states are reachable. In general, we have $\operatorname{range}(\mathcal{C}_0) \subseteq \operatorname{range}(\mathcal{C}_1) \subseteq \cdots \subseteq \operatorname{range}(\mathcal{C}_n) = \operatorname{range}(\mathcal{C}_{n+1}) = \dots$. This follows from the Cayley–Hamilton theorem, since A^n is a linear combination of $\{I, A, \dots, A^{n-1}\}$.

If we have a limited energy budget, say $\|u\| \le \alpha$, then we can ask what x are reachable. This is the set $\{\mathcal{C}u \mid \|u\| \le \alpha\}$. As seen in Lecture 7, this set is an ellipsoid. We can find this ellipsoid by taking the SVD of $\mathcal{C}_t$:

$$\mathcal{C}_t = U\Sigma V^\mathsf{T}$$

The easiest direction to reach is u_1 (first left singular vector), and this requires picking $u = v_1$ (first right singular vector). Likewise, the most difficult direction to reach is u_n, achieved by picking $u = v_n$. As t increases, the ellipsoid of reachable states will grow. But what happens as $t \to \infty$? It turns out that the ellipsoid converges to a limiting ellipsoid, which tells us how far we could reach if we had an infinite amount of time.

We would like to find this limiting ellipsoid, but it requires taking the SVD of a matrix with infinitely many columns! Remember that the singular values of $\mathcal{C}_t$ are the square roots of the eigenvalues of $P = \mathcal{C}_t\mathcal{C}_t^\mathsf{T}$, which is a fixed-size $n \times n$ matrix (for any t). So let's see if we can find this limiting matrix. Observe that:

$$P = \lim_{t\to\infty} \mathcal{C}_t\mathcal{C}_t^\mathsf{T} = BB^\mathsf{T} + ABB^\mathsf{T}A^\mathsf{T} + A^2BB^\mathsf{T}(A^\mathsf{T})^2 + \cdots$$

If A is Schur-stable, this sum will converge, and P satisfies the Lyapunov equation

$$\boxed{APA^\mathsf{T} - P + BB^\mathsf{T} = 0}$$

P is called the **controllability Gramian**. It turns out when A is Schur-Stable, $P \succ 0$ if and only if (A, B) is controllable (see supplementary notes on Lyapunov equations for a proof). The eigenvectors of the controllability Gramian are the u_i vectors (principal directions).

5 Observability

In a manner analogous to our study of controllability, we will now study observability. When using measurements to estimate a state, some measurements can be more informative (our estimation error will be less sensitive to error in the measurements). Here, we use the system:

$$x_{t+1} = Ax_t \qquad \text{and} \qquad y_t = Cx_t$$

We observe the measurements $y_0, \dots, y_{t-1}$ and our task is to estimate the initial state x_0. We assume our measurements have some fixed amount of uncertainty

$$\sum_{t=0}^{\infty} \|y_t\|^2 \le \alpha^2$$

The initial state x_0 is related to the measurements $y_0, \dots, y_{t-1}$ via:

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{t-1} \end{bmatrix}}_{\mathcal{O}_t} x_0$$

where we defined $\mathcal{O}_t$ to be the observability matrix. If $\text{rank}(\mathcal{O}_t) = n$, then x_0 can be perfectly determined from measurements. In general, we have $\text{null}(\mathcal{O}_0) \supseteq \text{range}(\mathcal{O}_1) \supseteq \cdots \supseteq \text{null}(\mathcal{O}_n) = \text{null}(\mathcal{O}_{n+1}) = \dots$. This follows from the Cayley–Hamilton theorem, like the controllability case.

We can ask what the uncertainty set would be for x_0 given the measurements lie in ball $\|y\| \le \alpha$. This is the set $\{x \mid \|\mathcal{O}_t x\| \le \alpha\}$. As seen in Lecture 7, this set is an ellipsoid. We can find it by taking the SVD of $\mathcal{O}_t$:

$$\mathcal{O}_t = U\Sigma V^\mathsf{T}$$

The easiest direction to estimate is v_1 (which corresponds to measurements aligned with u_1), since this is the direction where a fixed change in y causes the smallest possible change in x_0 (the amplification from x_0 to y is as large as possible). So a measurement in the direction v_1 will yield the most certainty in our estimate of x_0.

As $t \to \infty$, we have a limiting ellipsoid, and we can find it by looking at the eigenvalues of $Q = \mathcal{O}_t^\mathsf{T}\mathcal{O}_t$, which is an $n \times n$ matrix. Observe that:

$$Q = \lim_{t\to\infty} \mathcal{O}_t^\mathsf{T}\mathcal{O}_t = C^\mathsf{T}C + A^\mathsf{T}C^\mathsf{T}CA + (A^\mathsf{T})^2C^\mathsf{T}CA^2 + \cdots$$

If A is Schur-stable, this sum will converge, and Q satisfies the Lyapunov equation

$$\boxed{A^\mathsf{T}QA - Q + C^\mathsf{T}C = 0}$$

Q is called the **observability Gramian**. It turns out when A is Schur-Stable, $Q \succ 0$ if and only if (A, C) is observable (see supplementary notes on Lyapunov equations for a proof). The eigenvectors of the observability Gramian are the v_i vectors (principal directions).

6 Spring-mass-damper example

Here, we revisit the spring-mass-damper example from Lectures 11 and 12. We assume a sequence of forces is applied to the third mass, and we ask which states are easier or harder to reach. Note that the *state* here is a vector in $\mathbb{R}^6$ (position and velocity for each of the three masses).

Solving the Lyapunov equation to find the controllability Gramian in Matlab, we obtain

```
>> P = dlyap(A,B*B')
P =
    0.3916    0.6781    0.8210    0.0000   -0.0068   -0.0074
    0.6781    1.2136    1.4992    0.0068   -0.0000   -0.0141
    0.8210    1.4992    1.8935    0.0074    0.0141    0.0000
    0.0000    0.0068    0.0074    0.2110    0.2863    0.2859
   -0.0068   -0.0000    0.0141    0.2863    0.5009    0.5696
   -0.0074   -0.0141    0.0000    0.2859    0.5696    0.7873
```

We can then find the left singular vectors, which tell us the directions that are easiest (first column) or hardest (last column) to reach:

```
>> [U,S,V] = svd(P)
U =
   -0.3278    0.0071    0.0311   -0.7022   -0.1796    0.6051
   -0.5909    0.0061   -0.0319   -0.3336   -0.0804   -0.7294
   -0.7372   -0.0080    0.0183    0.5793    0.1456    0.3153
   -0.0028   -0.3276   -0.7167   -0.1547    0.5948    0.0361
   -0.0024   -0.5907   -0.3519    0.1706   -0.7057    0.0122
    0.0032   -0.7373    0.6002   -0.0837    0.2971   -0.0294
```

The order are the states is $(x_1, x_2, x_3, v_1, v_2, v_3)$. So the easiest state to reach is when all three masses move in the same direction and have roughly zero velocity, with each mass moving a bit more than the next. This makes sense; it corresponds to all springs being equally compressed.

We can solve the associated minimum-norm control problem to find the sequences of inputs that reach each of these states. See the result in Fig. 1. The final states are normalized so that they have norm-1, but the minimum input norm required to reach these states varies dramatically.

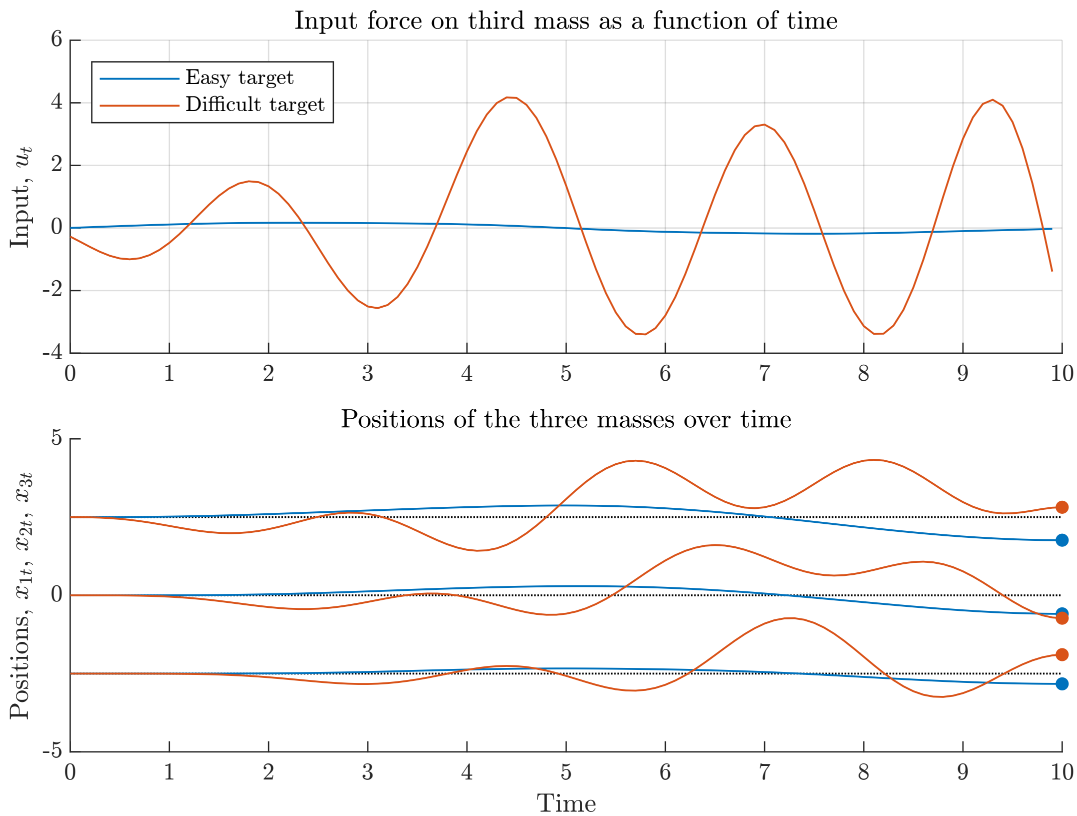

Figure 1: Minimum-norm input signal required to reach either the easiest or most difficult state to reach for the 3-mass spring-mass-damper system.