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

Lecture 23: risk-sensitive control

Tuesday, Dec 02, 2022

Lecturer: Laurent Lessard *Scribe: Chang Wang*

In this lecture, we cover *risk-sensitive control*, which is a similar to stochastic LQR, but the nature of the disturbance will be changed. We start by reformulating the LQR problem slightly differently. Write the system dynamics by including an output z_t:

$$x_{t+1} = Ax_t + Bu_t + w_t,$$
$$z_t = Fx_t + Hu_t,$$

The goal is to

$$\underset{u_0,u_1,\dots}{\text{minimize}} \quad \sum_{t=0}^{\infty} \|z_t\|^2 \tag{1}$$

To recover the LQR case, we can simply pick

$$F = \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix}, \qquad H = \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} \tag{2}$$

1 Risk-sensitive Control

Even when we use the optimal LQR gain, running the system for a finite time will incur different costs every time due to the random disturbances w_t. In other words, the cost J is a random variable. The LQR problem was to find the K that minimizes $\mathbf{E}(J)$, the expected cost. However, there are cases where the cost may have a large variance, and we would rather use a more conservative strategy that may be worse on average, but whose worst-case is not as bad. This is called *risk-sensitive control*. Another way to think about it is that we imagine the noise w_t as not being random, but rather being slightly biased or adversarial, such that we pick a strategy that anticipates some level of unluckiness.

1.1 First approach: soft-constrained minimax

There are several ways to achieve a more robust controller. We will show two of them. The first one is to change the problem into a deterministic problem. In this problem, the w is not going to be a random noise anymore, but will be chosen by an adversary who is trying to *maximize* the cost rather than minimize it. We cannot give our adversary full freedom, however, otherwise they could make the cost $+\infty$. Therefore, we use a *soft-constrained formulation*, which is a way of limiting our adversary's power. Here is the optimization problem.

$$\begin{aligned} \underset{u_{0:N-1}}{\text{minimize}} \quad \underset{w_{0:N-1}}{\text{maximize}} \quad & \sum_{t=0}^{N-1} \left(\|z_t\|^2 - \gamma^2 \|w_t\|^2 \right) \\ \text{s.t.} \quad & x_{t+1} = Ax_t + Bu_t + w_t \\ & z_t = Fx_t + Hu_t \end{aligned} \tag{3}$$

If $\gamma = 0$, we recover the standard LQR cost, which means our adversary could just make the noise w_t infinitely large and the inner maximization would be $+\infty$. However, if γ is large, the adversary's power will be limited and it will become more difficult for the adversary to sabotage us. As $\gamma \to \infty$, the noise is so heavily penalized that the best option for the adversary is to pick $w_t = 0$, which recovers the standard deterministic LQR problem.

1.2 Second approach: linear exponential regulator

Another way to model risk aversion is to change our cost such that larger costs are more greatly penalized. Let C be the standard LQR cost:

$$C = \sum_{t=0}^{N-1} \|z_t\|^2$$

Recall that when w_t is random noise (say, Gaussian), C is a random variable. The standard stochastic LQR setting optimizes the average quadratic cost $\mathbf{E}(C)$. However, if we want to place a greater penalty on large values of C, we can replace the cost by

$$\mathbf{E}(f(C))$$

where $f(x)$ is a function that increases faster than x. This ensures that values of C that are larger than average are penalized more than values of C that are below average. One way to achieve this is to make f an exponential function. This leads to the *linear exponential quadratic regulator* (LEQR)

$$J = \gamma^2 \log \mathbf{E}(e^{\frac{1}{\gamma^2}C}) \tag{4}$$

Note that the $\gamma^2 \log$ does not change the optimal policy (it's just makes the algebra simpler). Also note that we can't interchange the order of the expected value an the exponential. When γ is smaller, the exponential curve has a steeper growth (larger penalty on large C).

To interpret what the LEQR model does, consider a Taylor approximation. Using the fact that

$$\begin{aligned} \exp(x) &= 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots \\ \log(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots \end{aligned}$$

we can write Eq. (4) as

$$\begin{aligned}
\gamma^2 \log \mathbf{E}(e^{\frac{1}{\gamma^2}C}) &= \gamma^2 \log \mathbf{E}\left[1 + \frac{1}{\gamma^2}C + \frac{1}{2\gamma^2}C^2 + \cdots\right] \\
&= \gamma^2 \log \left[1 + \frac{1}{\gamma^2}\mathbf{E}(C) + \frac{1}{2\gamma^2}\mathbf{E}(C^2) + \cdots\right] \\
&= \gamma^2 \left[\left(\frac{1}{\gamma^2}\mathbf{E}(C) + \frac{1}{2\gamma^2}\mathbf{E}(C^2) + \cdots\right) - \frac{1}{2}\left(\frac{1}{\gamma^2}\mathbf{E}(C) + \frac{1}{2\gamma^2}\mathbf{E}(C^2) + \cdots\right)^2 + \cdots\right] \\
&= \underbrace{\mathbf{E}(C)}_{\text{LQR cost}} + \frac{1}{2\gamma^2}\underbrace{\left(\mathbf{E}(C^2) - \mathbf{E}(C)^2\right)}_{\mathbf{Var}(C)} + \frac{1}{\gamma^4}(\dots)
\end{aligned}$$

So roughly, the LEQR objective function is a trade-off between $\mathbf{E}(C)$ (the mean cost) and $\mathbf{Var}(C)$ (the variance of the cost). This Taylor approximation is valid when γ is large (so $1/\gamma$ is small). As we make γ smaller, we place more weight on the variance of the cost, which causes the optimal strategy to be more conservative and risk-averse.

Comparison of both approaches. Amazingly, the soft-constrained minimax formulation and the LEQR formulation yield the *exact same optimal policy*. This is not obvious! In these notes, we will only derive the solution to the soft-constrained minimax problem using a dynamic programming approach. The LEQR problem can also be solved using dynamic programming.

To solve the minimax problem, we would like to use dynamic programming and employ the same technique as we did with deterministic LQR, but this would require replacing the minimization and maximization in Eq. (3) by an alternating version that looks like:

$$\min_{u_0} \max_{w_0} \min_{u_1} \max_{w_1} \cdots \min_{u_{N-1}} \max_{w_{N-1}}$$

In general, this is not allowed. One cannot simply swap the order of minimization and maximization unless certain specific conditions are met. We will take a detour to study what these conditions are and then return to solving this minimax problem.

2 von Neumann's minimax theorem

Theorem. If $f(x, y)$ is convex in x (for all y) and concave in y (for all x), then

$$\min_x \max_y f(x, y) = \max_y \min_x f(x, y)$$

An example of a function that satisfies this requirement is $f(x, y) = x^2 - y^2$. When graphed, this function has a *saddle point* at $(0, 0)$.

Proof. Firstly, we start by proving that $\min_x \max_y f(x, y) \geq \max_y \min_x f(x, y)$. Start with

$$\max_y f(x, y) \geq f(x, y) \geq \min_x f(x, y) \qquad \text{for all } x, y.$$

Therefore we have

$$\underbrace{\max_y \quad f(x,y)}_{\text{function of } x \text{ only}} \quad \geq \quad \underbrace{\min_x \quad f(x,y)}_{\text{function of } y \text{ only}} \qquad \text{for all } x, y.$$

Since each side is a function of a different variable, we can minimize the left-hand side with respect to x and maximize the right-hand side with respect to y and the inequality will still be true. Therefore,

$$\min_x \max_y \; f(x,y) \geq \max_y \min_x \; f(x,y) \tag{5}$$

as required. Now, we prove the more difficult direction of the inequality. We start by converting the problem into epigraph form.

$$\begin{aligned} \min_x \max_y \; f(x,y) \quad &= \quad \min_{x,t} \quad t \\ &\qquad\quad \text{s.t.} \quad \max_y \; f(x,y) \leq t \\ &= \quad \min_{x,t} \quad t \\ &\qquad\quad \text{s.t.} \quad f(x,y) \leq t \quad \text{for all } y \end{aligned}$$

This doesn't seems like progress because we started with an unconstrained optimization problem and now we have a problem with an extra variable and infinitely many constraints... But stay with it, it will get better!

First, observe that the optimization problem is convex, since f is convex in x and therefore each constraint is a convex constraint in (x,t). Moreover, the problem is strictly feasible, since we can always pick a larger t in order to make each inequality strict. Therefore, by Slater's constraint qualification, the problem must exhibit strong duality.

Now compute the dual. Since there is a constraint for every y, there must be a dual variable for every y. In other words, the set of dual variables can be viewed as a function $\lambda(y)$. Let $\lambda(y)$ be the dual variable for the constraint $f(x,y) \leq t$. Now compute the dual function, optimizing over t first.

$$\begin{aligned} F(\lambda) &= \min_{x,t} \quad \left(t + \int \lambda(y)(f(x,y) - t) \mathrm{d}y \right) \\ &= \min_{x,t} \quad t\left[1 - \int \lambda(y) \mathrm{d}y\right] + \int \lambda(y) f(x,y) \, \mathrm{d}y \\ &= \begin{cases} \min_x \quad \int \lambda(y) f(x,y) \mathrm{d}y & \text{if } \int \lambda(y) \, \mathrm{d}y = 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned} \tag{6}$$

Therefore, strong duality tells us that

$$\begin{aligned}
\min_x \max_y f(x,y) \quad = \quad & \max_{\lambda} \quad F(\lambda) \\
& \text{s.t.} \quad \lambda(y) \geq 0 \quad \text{for all } y \\
\max_{\lambda} \min_x \quad & \int \lambda(y) f(x,y)\,\mathrm{d}y \\
& \text{s.t.} \quad \int \lambda(y)\,\mathrm{d}y = 1 \\
& \qquad \lambda(y) \geq 0 \quad \text{for all } y \\
\max_{\lambda} \min_x \quad & \mathbf{E}_{y\sim\lambda} f(x,y) \\
& \text{s.t.} \quad \lambda \text{ is a probability density}
\end{aligned} \tag{7}$$

In the last step, we used the fact that λ is everywhere nonnegative and it integrates to 1. Therefore λ is a *probability distribution*! We therefore interpret y as being a random vector with probability density function λ. The notation $\mathbf{E}_{y\sim\lambda}(\dots)$ means that we are taking the expected value with respect to y, which is distributed according to the density function λ.

We will now use the fact that $f(x,y)$ is a concave function of y. Recall that a concave function $h : \mathbb{R}^n \to \mathbb{R}$ satisfies the following inequality by definition:

$$\alpha h(x) + (1-\alpha)h(y) \leq h(\alpha x + (1-\alpha)y) \qquad \text{for all } x, y \in \mathbb{R}^n \text{ and } \alpha \in [0,1].$$

A similar inequality holds when we have m points. If $h : \mathbb{R}^n \to \mathbb{R}$ is a concave function, then

$$\sum_{i=1}^m \alpha_i h(x_i) \leq h\left(\sum_{i=1}^m \alpha_i x_i\right) \qquad \text{for all } x_i \in \mathbb{R}^n \text{ and } \alpha_i \geq 0 \text{ satisfying } \sum_{i=1}^m \alpha_i = 1.$$

It even holds when we have infinitely many points, which yields *Jensen's inequality.* If $h : \mathbb{R}^n \to \mathbb{R}$ is a concave function, then

$$\int p(x)h(x)\,\mathrm{d}x \leq h\left(\int xp(x)\,\mathrm{d}x\right) \qquad \text{for any probability distribution } p.$$

We can write this even more compactly as follows. If $h : \mathbb{R}^n \to \mathbb{R}$ is a concave function, and $x \in \mathbb{R}^n$ is a random vector, we have:

$$\mathbf{E}h(x) \leq h(\mathbf{E}x)$$

By applying Jensen's Inequality, we obtain

$$\begin{aligned}
\min_x \max_y f(x,y) \quad = \quad & \max_{\lambda} \min_x \quad \mathbf{E}_{y\sim\lambda} f(x,y) \\
& \qquad \text{s.t.} \quad \lambda \text{ is a probability density} \\
\leq \quad & \max_{\lambda} \min_x \quad f(x, \mathbf{E}_{y\sim\lambda} y) \\
& \qquad \text{s.t.} \quad \lambda \text{ is a probability density} \\
= \quad & \max_{\bar{y}} \min_x \quad f(x,\bar{y})
\end{aligned} \tag{8}$$

In the last step, we used the fact that the optimization problem only depends on λ through the expected value of y, so it is equivalent to simply optimize over $\bar{y} := \mathbf{E}_{y\sim\lambda} y$.

Now combining (5) and (8), we obtain the desired result. ■

3 Dynamic programming for risk-sensitive control

Now that we have proven the minimax theorem, we can solve the soft-constrained problem via dynamic programming by alternating minimization and maximization. The only catch is that our cost function must be convex in the u_t (the variables we are minimizing over) and concave in the w_t (the variables we are maximizing over). Since everything is quadratic and linear, we can use a quadratic value function $V_t(x) = x^\mathsf{T} P_t x_t$ as in the LQR case. Our dynamic programming recursion becomes

$$V_t(x) = \min_u \max_w \left(x^\mathsf{T} Q x + u^\mathsf{T} R u - \gamma^2 \|w\|^2 + (Ax + Bu + w)^\mathsf{T} P_{t+1} (Ax + Bu + w) \right) \tag{9}$$

Collecting the $\|w\|^2$ terms, we obtain $w^\mathsf{T}(P_{t+1} - \gamma^2 I)w$. This is a concave function whenever $P_{t+1} \prec \gamma^2 I$. So we must check that this remains true at every step of the recursion. After maximizing over w and doing some simplifications, we obtain

$$V_t(x) = \min_u \quad \left(x^\mathsf{T} Q x + u^\mathsf{T} R u + (Ax + Bu)^\mathsf{T} \left(P_{t+1}^{-1} - \gamma^{-2} I \right)^{-1} (Ax + Bu) \right) \tag{10}$$

Make the change of variables: $\tilde{P}_{t+1} := \left(P_{t+1}^{-1} - \gamma^{-2} I \right)^{-1}$. Note that $\tilde{P}_{t+1} \succ 0$ by our previous concavity assumption. Then, (9) becomes exactly the same as the LQR dynamic recursion, except with $\tilde{P}$ rather that P on the right-hand side. Solving it, we obtain:

$$\begin{aligned}
P_N &= Q_f \\
\tilde{P}_{t+1} &= \left(P_{t+1}^{-1} - \gamma^{-2} I \right)^{-1} && \text{for } t = 0, \ldots, N-1 \\
P_t &= A^\mathsf{T} \tilde{P}_{t+1} A + Q - A^\mathsf{T} \tilde{P}_{t+1} B (B^\mathsf{T} \tilde{P}_{t+1} B + R)^{-1} B^\mathsf{T} \tilde{P}_{t+1} A && \text{for } t = 0, \ldots, N-1 \\
K_t &= -(B^\mathsf{T} \tilde{P}_{t+1} B + R)^{-1} B^\mathsf{T} \tilde{P}_{t+1} A && \text{for } t = 0, \ldots, N-1
\end{aligned} \tag{11}$$

And just as with LQR, the optimal policy is $u_t = K_t x_t$.

Note. If $\gamma \to \infty$, we have $\tilde{P} \to P$ so we recover the standard LQR solution. But if γ is too small, $P_{t+1} \prec \gamma^2 I$ will no longer hold. We can view γ as our level of risk aversion. The smaller we make γ, the more paranoid we become, until eventually we are unable to make *any decision at all* because we think our adversary is too powerful.