ORF 418 : OPTIMAL LEARNING

LECTURE 6 : September 21 . 2025

1) Problem Formulation

2) Dynamic Programming

I. Problem Formulation

Stochastic dynamics is given by

$$x_{k+1} = Ax_k + Bu_k + w_{k+1} \qquad k=0,1,\dots$$

where A, B are as before, and $w_1, w_2, \dots$ is a sequence of independently, identically distributed (i.i.d.) sequence of $\mathbb{R}^d$ valued Gaussian random variables with mean zero and $(d \times d)$ covariance matrix Q.

Infinite horizon cost functional is given by

$$J_\infty(x,u) := \mathbb{E}\left[\sum_{k=0}^{\infty} \rho^k \left(x_k^T M x_k + u_k^T N u_k\right)\right].$$

Important restriction the control is that they must be adapted to the information flow. Namely, u_k can use information upto time k or equivalently $w_1, \dots, w_k$.

We always need $\rho < 1$ to have a finite value function:

$$V_\infty(x) := \inf_{u \in \mathcal{U}} J(x,u) .$$

Example. Consider the problem with $d=b=1$,

$$x_{k+1} = 4x_k + u_k + w_{k+1}$$

where $w_1, w_2, \ldots$ have zero mean and variance σ^2. If we use a linear control $u_k = -fx_k$, then

$$x_{k+1} = (4-f)x_k + w_{k+1}$$

$$\Rightarrow \quad x_n = (4-f)^n x_0 + y_n$$

and

$$y_n = \sum_{k=1}^{n} (4-f)^{n-k} w_k \quad , n=1,2,\ldots$$

Then,

$$\begin{aligned} \operatorname{var}(x_n) &= \operatorname{var}(y_n) \\ &= \sum_{k=1}^{n} (4-f)^{2(n-k)} \operatorname{var}(w_k) \\ &= \sigma^2 \sum_{k=1}^{n} (4-f)^{2(n-k)} \geq \sigma^2 . \end{aligned}$$

So no matter how we choose the gain constant f, x_k will never converge to zero.

II. DYNAMIC PROGRAMMING.

Proceeding exactly as in the deterministic case, we obtain the following equation for the value function v_∞:

$$v_\infty(x) = x^T M x + \inf_u \{ u^T N u + \rho \mathbb{E}[v_\infty(Ax+Bu+\omega_1)]$$

We postulate that

$$v_\infty(x) = x^T V x + C$$

for some positive definite matrix, symmetric matrix V and a constant C. By substituting this form into the dynamic programming equation we obtain the following result:

THEOREM 2.4.1. Let V be the Riccati equation with matrices (A,B,M,N) and discount factor ρ. Then,

$$v_\infty(x) = x^T V x + \frac{\rho}{1-\rho} \text{trace}(VQ)$$

and $u^*(x) = -Fx$ (with the same gain matrix F) is the optimal feedback control.

Proof given in the notes is a straightforward but tedious calculation. Important to note that the form of the optimal control does not change!

Example. Going back to the previous example, we have $d=\ell=1$, $A=4$, $M=N=1$, and $\rho=\frac{1}{2}$. Then, the dynamic programming equation is,

$$v_\infty(x) = x^2 + \min_u \left\{ u^2 + \frac{1}{2}\, \mathbb{E}[v_\infty(4x+u+w_1)] \right\}.$$

We know that $v_\infty(x) = Vx^2 + C$. Hence

$$\mathbb{E}[v_\infty(4x+u+w_1)] = \mathbb{E}[V(4x+u+w_1)^2]$$

$$= V\left[(4x+u)^2 + 2(4x+u)\underbrace{\mathbb{E}[w_1]}_{=0} + \underbrace{\mathbb{E}(w_1^2)}_{=\sigma^2}\right]$$

$$= V\left[(4x+u)^2 + \sigma^2\right].$$

Then,

$$Vx^2 + C = x^2 + \min_u \left\{ u^2 + \frac{1}{2} V(4x+u)^2 \right\} + \sigma^2 \frac{1}{2}$$

By calculus $u^*(x) = -\frac{4V}{2+V}$, exactly as in the deterministic case. Hence,

$$C = \sigma^2/2$$

and
$$V = 1 + \left(\frac{4V}{2+V}\right)^2 + \frac{1}{2}\left(4 - \frac{4V}{2+V}\right)^2 V$$

$$= 1 + \frac{16V^2 + 32V}{(2+V)^2} = 1 + \frac{16V}{2+V}.$$

This is the same Riccati equation as in the deterministic case.