Lecture 13 — February 27

Lecturer: Venkat Anantharam *Scribe: Humberto Gonzalez*

13.1 Example: Dual role of control in partially observed systems.

The controller has to both elicit information and try to get good performance, and there is a tension between these goals.

Suppose $\mathcal{X}$ is a finite set and that $x_k \in \mathcal{X}$, $k = 0, \ldots, N$, evolves as a controlled Markov chain where the transition probabilities depend on a parameter $\theta \in \Theta$, where Θ is a finite set. The controller can observe x_k at time k. However, the true value of the parameter is unknown to the controller.

Let $\mathcal{U}$ be a finite set, representing the set of allowed controls. Let $\mathbb{P}(u, \theta) = [p_{ij}(u, \theta)]$ be the transition probability matrix, where $u \in \mathcal{U}$, and $\theta \in \Theta$. Note that this depends on θ. The aim is to minimize (informally):

$$E\left[\sum_{k=0}^{N-1} g_k(X_k, U_k) + g_N(X_N)\right] \tag{13.1}$$

with initial distribution $p(x_0)$ on the state and $p(\theta)$ on the parameter (drawn independently of x_0).

This is actually a partially observed problem, even though the controller can see x_k at time k. x_k is not the "true" state. The true state is (x_k, θ) and the observation x_k is therefore partial.

If we rewrite the transition probabilities in terms of the "true" state, the transition are from $\mathcal{X} \times \Theta$ to $\mathcal{X} \times \Theta$ and the transition probability matrix for control choice u is:

$$\mathbb{P}(u) = [p_{i\alpha,j\beta}(u)] = [\delta(\alpha, \beta) p_{ij}(u, \alpha)] \tag{13.2}$$

where δ is the Kronecker delta function:

$$\delta(\alpha, \beta) = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases} \tag{13.3}$$

Now, following the general technique for solving partially observed dynamic programming problems, we need to write the evolution equations of $p_{k|k}(i, \alpha|\eta_k)$, where

$$\eta_k = \{x_0, \ldots, x_k, u_0, \ldots, u_{k-1}\} \ .$$

Note that now x_l is the observation at time l. We know that, along the dynamics:

$$p_{k|k}(i, \alpha|\eta_k) = \delta(i, x_k) \pi_{k|k}(\alpha|\eta_k) \tag{13.4}$$

where $\pi_{k|k}(\alpha \mid \eta_k)$ is the conditional distribution of the parameter given the observations.

The evolution of the information state of the "true" state, i.e. the conditional law of the "true" state given I_k^Ψ is described in terms of the functions:

$$p_{k+1|k+1}(j,\beta|\eta_{k+1}) \propto \sum_{(i,\alpha)} p_{k|k}(i,\alpha|\eta_k) p_{i\alpha,j\beta}(u) p(x_{k+1}|j,\beta,u_k) \tag{13.5}$$

Since

$$p_{k+1|k+1}(j,\beta|\eta_{k+1}) = \delta(j,x_{k+1})\pi_{k+1|k+1}(\beta \mid \eta_{k+1}) \ ,$$

this can be written as:

$$\delta(j,x_{k+1})\pi_{k+1,k+1}(\beta|\eta_{k+1}) \propto \sum_{(i,\alpha)} \delta(i,x_k)\pi_{k|k}(\alpha|\eta_k)\delta(\alpha,\beta)p_{ij}(u_k,\alpha)\delta(j,x_{k+1}) \ ,$$

where we used (13.2) and the fact that $p(x_{k+1}|j,\beta,u_k) = \delta(j,x_{k+1})$. Simplifying, we get:

$$\pi_{k+1,k+1}(\beta|\eta_{k+1}) \propto \pi_{k|k}(\beta|\eta_k) p_{x_k,x_{k+1}}(u_k,\beta) \tag{13.6}$$

i.e. the update equation for the information "true" state is just the Bayesian update equation for the parameter estimate.

Here the controller can directly observe the x_k (which is, however, not the "true" state) and the objective depends only on the x_k and the controls. Nevertheless, the DP theory tells us that the controls have to be chosen as functions of the $(\pi_{k|k}(\theta \mid I_k^\Psi), \theta \in \Theta)$, i.e. of the conditional law at each time of the parameter given the observations. Thus we see explicitly in this example that the controller is working to influence the uncertainty in the system model, while at the same time it is trying to achieve optimal performance. The control is implicitly required to consider both how it influences the $\pi_{k|k}(\cdot \mid \eta_k)$ and how it influences the cost (13.1). This is the dual role of control.

13.2 Filling the gap left in the proof of the partially observed problem.

In our general development of the translation of the partially observed DP problem into a fully observed problem for the information state there was a gap. We need to argue that (in dummy density notation), for every control choice u:

$$x_{k+1} \mapsto p_{k+1|k+1}\left(x_{k+1}|I_{k+1}^\Psi\right) \tag{13.7}$$

results from:

$$x_k \mapsto p_{k|k}\left(x_k|I_k^\Psi\right) \tag{13.8}$$

through a stochastic transformation driven by a variable independent of the past variables at time k. Since $I_{k+1}^\Psi = (I_k^\Psi, Y_{k+1}^\Psi, u)$ when the control is u, what we know is that the evolution is driven by Y_{k+1}^Ψ (but this is not yet good enough). We know that if the control is u then:

$$Y_{k+1}^\Psi = h_k(X_{k+1}^\Psi, u, v_{k+1}) \tag{13.9}$$

$$= h_k(f_k(X_k^\Psi, u, \omega_k), u, v_{k+1}) \tag{13.10}$$

where ω_k and v_{k+1} are independent of all past variables. The problem is that this appears to depend on X_k^Ψ.

It turns out that:

$$P(Y_{k+1}^\Psi \in B | I_k^\Psi) = P(Y_{k+1}^\Psi \in B | \lambda_k^\Psi) \quad \forall B \tag{13.11}$$

which comes directly from equation (13.10).

In fact,

$$E\left[Y_{k+1}^\Psi | I_k^\Psi\right] = E\left[B_k(X_k^\Psi, u) | I_k^\Psi\right] \tag{13.12}$$

where

$$B_k(x,u) = E\left[h_k(f_k(x,u,\omega_k),u,v_{k+1})\right] \tag{13.13}$$

and so

$$E\left[Y_{k+1}^\Psi | I_k^\Psi\right] = \int B_k(x,u) P(X_k^\Psi \in dx | I_k^\Psi) \tag{13.14}$$

$$= \int B_k(x,u) \lambda_k^\Psi(dx) \tag{13.15}$$

13.3 The separation principle in the Linear Quadratic partially observed control problem.

Recall the fully observed problem:

$$x_{k+1} = A_k x_k + B_k u_k + \omega_k \quad k = 0, \ldots, N-1 \tag{13.16}$$

where the aim is to minimize (informally):

$$E\left[\sum_{k=0}^{N-1} \left(X_k^T Q_k X_k + U_k^T R_k U_k\right) + X_N^T Q_N X_N\right] \tag{13.17}$$

where $\{Q_k\}_{k=0}^N$ are positive semidefinite matrices and $\{R_k\}_{k=0}^{N-1}$ are positive definite matrices.

An optimal strategy is to put $u_k = L_k x_k$ where:

$$L_k = -\left(R_k + B_k^T K_{k+1} B_k\right)^{-1} B_k^T K_{k+1} A_k \tag{13.18}$$

and

$$K_N = Q_N \tag{13.19}$$

$$K_k = A_k^T K_{k+1} A_k - \Gamma_k + Q_k \tag{13.20}$$

$$\Gamma_k = A_k^T K_{k+1} B_k \left(R_k + B_k^T K_{k+1} B_k\right)^{-1} B_k^T K_{k+1} A_k \tag{13.21}$$

Now, the partially observed problem has the same state dynamics plus observations:

$$y_k = C_k + v_k \tag{13.22}$$

with the same objective but over strategies Ψ that depend causally only on the past observations $\{Y_l^\Psi : l \leq k\}$.

Our general theory already tells that there is an optimal strategy Ψ where the optimal control at time k is a function only of λ_k^Ψ, where:

$$\lambda_k^\Psi = P(X_k^\Psi \in \cdot \,|I_k^\Psi) \tag{13.23}$$

Let $m_k^\Psi = E[X_k^\Psi | I_k^\Psi]$, it turns out that the optimal control strategy uses controls:

$$u_k = L_k m_k^\Psi \ . \tag{13.24}$$

This is known as the separation principle, because it shows that the overall optimal control strategy splits into two parts: (i) the estimation of the conditional mean of the state given the observations, i.e. m_k^Ψ, and (ii) applying linear controls based on m_k^Ψ as if m_k^Ψ were the true state in a fully observed control problem.