**EE 223: Stochastic Estimation and Control** **Spring 2007**

# Lecture 10 — February 15

*Lecturer: Venkat Anantharam* *Scribe: Assane Gueye*

So far we have considered dynamical systems where the controller directly observes the states and chooses the control causally, i.e. the control to use at time $k$ is determined as a function of the sequence of states up to and including time $k$.

In most problems, however, the controller only has access to partial information about the states. In this lecture, we will study how the controller can optimally causally choose its controls in those cases where only partial information about the state is available (we will name it *partially observed stochastic control*). As before we limit ourselves to the finite horizon case.

We assume the same dynamical system and the same objective (written informally) as in the fully observed case:

$$x_{k+1} = f_k(x_k, u_k, w_k) \ , \quad k = 0, 1, \ldots, N-1 \ ; \tag{10.1}$$

$$\text{Aim} : Minimize E\left[\sum_{k=0}^{N-1} g_k(X_k, U_k, w_k) + g_N(X_N)\right] \ . \tag{10.2}$$

However, the set of strategies over which the minimization of the objective is to be done is different from the fully observed case. The controller observes a function of the state at each time:

$$\begin{aligned} y_0 &= h_0(x_0, v_0) \ , \\ y_k &= h_k(x_k, u_{k-1}, v_k) \ , \quad k = 1, 2, \ldots, N, \end{aligned} \tag{10.3}$$

where $v_k$ may be thought of as observation noise. The random variables

$$w_0, \ldots, w_{N-1}, v_0, \ldots, v_N, x_0 \ ,$$

are assumed to be mutually independent. The difference now is that we are allowed to choose the control $u_k$ at each time $k = 0, 1, \ldots, N-1$ based only on the observation $y_0, \ldots, y_k$ up to and including that time.

**Remark 1.** *For notational convenience later, we have assumed that there is an observation also at time $N$. You may think of this observation as being trivial (always equal to some fixed value) if you wish. However, our discussion remains valid whether or not this observation is trivial.*

**Remark 2.** *A more general formulation of the problem would allow the control $u_k$ to be chosen as a* random *function of the observations $y_0, \ldots, y_k$ at each time $k$, but for the sake of simplicity we will consider only deterministic controls.*

Formally we would like to solve the problem

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

where the minimization is over the set of all possible strategies $\Psi$. Here, a strategy $\Psi$ is comprised of functions $\nu_k(y_0, \ldots, y_k)$, $k = 0, 1, \ldots, N-1$.

**Remark 3.** *We could have dropped the $w_k$ terms in the cost functions. In fact we can always reduce the problem to one with new cost functions defined as:*

$$\tilde{g}_k(x_k, u_k) = E_{w_k}[g_k(x_k, u_k, w_k)] .$$

*This is because $w_k$ is independent of the states, observations, and controls up to and including time $k$.*

The main insight is that we can convert the partially observed problem into a fully observed problem for another stochastic system whose state at any time is the conditional law of the current state of the original system given the knowledge of the controller up to an including that time. What does this law look like?

We will use the notation $\eta_k$ for $(y_0, \ldots, y_k, u_0, \ldots, u_{k-1})$, $k = 0, 1 \ldots, N$. If the strategy $\Psi$ is in effect, the information available to the controller at time $k$ is given by $(Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi})$, $k = 0, 1 \ldots, N$. (Note that there is no control action at time $N$.) In fact, under our assumption that the controls are chosen deterministically as a function of the available information, the information available in $(Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi})$ is identical to that available in $(Y_0^{\Psi}, \ldots, Y_k^{\Psi})$, $k = 0, 1 \ldots, N$. However, we find it convenient to use the more extended description. We will use that notation

$$I_k^{\Psi} = (Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi}) , \quad k = 0, 1, \ldots, N .$$

Let us compute the probability distribution of $X_k^{\Psi}$ given $(Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi})$. First notice that this this should be thought of as a *probability distribution valued random variable*. It takes values in probability distribution on $\mathcal{X}_k$ (the state space at time $k$) and is a deterministic function of the random variable $I_k^{\Psi}$. It is also often called the *conditional law* of $X_k^{\Psi}$ given $I_k^{\Psi}$. It can be thought of as the mapping

$$A \mapsto P(X_k^{\Psi} \in A | Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi}), \quad \forall A \subset \mathcal{X}_k$$

because this is indeed how one would describe the distribution of $X_k^{\Psi}$ given the information $I_k^{\Psi} = (Y_0^{\Psi}, \ldots, Y_k^{\Psi}, U_0^{\Psi}, \ldots, U_{k-1}^{\Psi})$. For notational convenience we will sometimes write this conditional law as if it has a density. Thus we might write:

$$P(X_k^{\Psi} \in A | I_k^{\Psi}) = \int_A p^{\Psi}(x_k | I_k^{\Psi}) dx_k$$

Occasionally we will use the notation $\lambda_k^{\Psi}$ for this conditional law. Occasionally we will write it as $P(X_k^{\Psi} \in \cdot \mid I_k^{\Psi})$.

**Claim 1.** *Using density notation for convenience, there is a sequence of $p_{k|k}(x_k|\eta_k)$, $k = 0, 1, \ldots, N$, that does not depend on $\Psi$, such that $p^\Psi(x_k|I_k^\Psi)$ equals $p_{k|k}(x_k|I_k^\Psi)$ for all $k = 0, 1, \ldots, N$.*

The mapping $x_k \mapsto p_{k|k}(x_k|I_k^\Psi)$ (in density notation), or $\lambda_k^\Psi$ or $P(X_k^\Psi \in \cdot \mid I_k^\Psi)$ in general notation, is called the information state at time $k$.

To prove the claim, we also need to consider the conditional laws of $X_{k+1}^\Psi$ given $(Y_0^\Psi, \ldots, Y_k^\Psi, U_0^\Psi, \ldots, U_k^\Psi)$ for $k = 0, 1, \ldots, N-1$. Again, pretend that these have a density, this is corresponds to considering the mappings $x_{k+1} \mapsto p^\Psi(x_{k+1}|I_k^\Psi, U_k^\Psi)$, $k = 0, 1, \ldots, N-1$. When we want to use a more general notation we will write these as $P(X_{k+1}^\Psi \in \cdot \mid I_k^\Psi, U_k^\Psi)$, $k = 0, 1, \ldots, N-1$.

**Claim 2.** *Using density notation for convenience, there is a sequence of functions $p_{k+1|k}(x_{k+1}|\eta_k, u_k)$, $k = 0, 1, \ldots, N-1$,* which do not depend upon $\Psi$ *such that $p^\Psi(x_{k+1}|I_k^\Psi, U_k^\Psi)$ equals $p_{k+1|k}(x_{k+1}|I_k^\Psi, U_k^\Psi)$, $k = 0, 1, \ldots, N-1$.*

**Remark 4.** *It is also convenient to write $p^\Psi(x_0)$, which equals the initial distribution $p(x_0)$ and does not depend on $\Psi$, as $p_{0|-1}(x_0)$. This is consistent with our notational system.*

We now proceed to prove these claims. We have, for $k = 0, 1, \ldots, N-1$,

$$\begin{aligned} p^\Psi(x_{k+1}|\eta_{k+1}) &= \frac{p^\Psi(y_{k+1}|x_{k+1}, \eta_k, u_k) p^\Psi(x_{k+1}|\eta_k, u_k) p^\Psi(\eta_k, u_k)}{p^\Psi(\eta_{k+1})} \\ &= \frac{p(y_{k+1}|x_{k+1}, u_k) p^\Psi(x_{k+1}|\eta_k, u_k)}{p^\Psi(y_{k+1}|\eta_k, u_k)} \end{aligned} \tag{10.4}$$

where in equation (10.4) the first term of the numerator is completely determined by the observation function (10.3). Another important observation is that the denominator is just a normalization term: it is equal to the integral of the numerator over $\mathcal{X}_{k+1}$. This may not be obvious when you look at the right hand side of equation (10.4), but it has to be true because the left hand side of this equation integrates to 1 over $\mathcal{X}_{k+1}$ and the denominator on the right hand side of this equation does not depend on $x_{k+1}$. This is just an instance of Bayes's rule. We may thus write equation (10.4) as

$$p^\Psi(x_{k+1}|\eta_{k+1}) = \frac{p(y_{k+1}|x_{k+1}, u_k) p^\Psi(x_{k+1}|\eta_k, u_k)}{\int p(y_{k+1}|x_{k+1}, u_k) p^\Psi(x_{k+1}|\eta_k, u_k) dx_{k+1}} \ . \tag{10.5}$$

Next, for $k = 0, 1, \ldots, N-1$, we may write

$$\begin{aligned} p^\Psi(x_{k+1}|\eta_k, u_k) &= \int p^\Psi(x_{k+1}, x_k|\eta_k, u_k) dx_k \\ &= \int p^\Psi(x_{k+1}|x_k, \eta_k, u_k) p^\Psi(x_k|\eta_k, u_k) dx_k \\ &= \int p(x_{k+1}|x_k, u_k) p^\Psi(x_k|\eta_k, u_k) dx_k \end{aligned} \tag{10.6}$$

where in equation (10.6), $p(x_{k+1} \mid x_k, u_k)$, which appears in the last step, is given by the dynamical equation (10.1). We now notice that

$$p^{\Psi}(x_k|\eta_k, u_k) = p^{\Psi}(x_k|\eta_k) \ ,$$

because $u_k = \nu_k(y_0, \ldots y_k)$ is a deterministic function of $\eta_k$. Thus we may rewrite equation (10.6) as

$$p^{\Psi}(x_{k+1}|\eta_k, u_k) = \int p(x_{k+1}|x_k, u_k) p^{\Psi}(x_k|\eta_k) dx_k$$

To start with, we have

$$p^{\Psi}(x_0|y_0) = \frac{p(y_0|x_0)p(x_0)}{\int p(y_0 \mid x_0)p(x_0)dx_0} \ .$$

The right hand side of this equation can be thought of as defining a function $p_{0|0}(x_0 \mid \eta_0)$ that does not depend on the strategy $\Psi$. Indeed, this can itself be thought of as defined in terms of the initial distribution, denoted $p_{0|-1}(x_0)$ in our notational system, in a manner consistent with equation (10.5). By induction over the k, starting from this observation at $k = 0$, we conclude that $p^{\Psi}(x_k|\eta_k)$ can be written as $p_{k|k}(x_k|\eta_k)$ for a function $p_{k|k}(x_k|\eta_k)$ which does not depend on $\Psi$, for $k = 0, 1 \ldots, N$, and also that we conclude that $p^{\Psi}(x_{k+1}|\eta_k, u_k)$ can be written as $p_{k+1|k}(x_{k+1}|\eta_k, u_k)$ for a function $p_{k+1|k}(x_{k+1}|\eta_k, u_k)$ which does not depend on $\Psi$, for $k = 0, 1 \ldots, N-1$. These functions satisfy the update equations:

$$p_{k+1|k}(x_{k+1}|\eta_k, u_k) = \int p(x_{k+1}|x_k, u_k) p_{k|k}(x_k|\eta_k) dx_k \tag{10.7}$$

$$p_{k+1|k+1}(x_{k+1}|\eta_{k+1}) \propto p(y_{k+1}|x_{k+1}, u_k) p_{k+1|k}(x_{k+1}|\eta_k, u_k) \tag{10.8}$$

The equations above (10.7 and 10.8) define the state and the dynamic model of the new dynamical system in terms of which the original partially observed control problem will be converted into a fully observed control problem. Next time we will see that the original cost function can be expressed as a cost function for the new fully observed dynamical system, in terms of its own states, i.e the information states. We will also need to argue that the evolution of the dynamical system above can be thought of, at each time step, as being driven by a random variable, such that these random variables and the (new) observation noise random variables are mutually independent and independent of the (new) initial condition.

To make the discussion in this lecture more concrete, consider a finite state time-invariant Markov Chain with $d$ states and probability transition matrix $[p_{i,j}(a)] = \mathbb{P}(a)$. We observe the chain through a time-invariant channel with input-output relation $q(y|i)$ for $y \in \mathcal{Y}$. Here $\mathcal{Y}$ denotes the set of channel outputs, and $1 \leq i \leq d$.
The information state can be organized as a row matrix

$$\pi_{k|k}(\eta_k) = \left[p_{k|k}(1|\eta_k), \ldots, p_{k|k}(d|\eta_k)\right]$$

The updates are as follows:

$$\pi_{k+1|k+1}(\eta_{k+1}) \propto \pi_{k|k}(\eta_k) \mathbb{P}(u_k) D(u_k, y_{k+1})$$

where $D(u_k, y_{k+1})$ is a diagonal matrix with entries $p(y_{k+1}|i, u_k), \quad 1 \leq i \leq d$.
Later we will see that the optimal strategy at time $k$ depends only on $\pi_{k|k}(I_k^\Psi)$ (where $I_k^\Psi$ itself comes from the application of the optimal strategy at preceding times).