Chapter 5

Problems with Partial Information

In most real-world applications, controllers operate under partial information, i.e., they do not have a complete and perfect measurement of the full state of the system at each time. If some unobserved part of the state cannot be simply ignored for reasonable performance, then the standard strategy is to augment the controller with an observer, in charge of constructing a state estimate. Optimal control problems under partial information are typically very difficult to solve and require some kind of approximation, some of which we will cover in the second part of the course. Indeed they couple an already intractable problem, optimal nonlinear filtering, to the control problem. Both aspects (estimation and control) can influence each other in general. For example, the controller might need to drive the state temporarily in a region where the estimator has better performance even if this means a higher cost in the short term. So our discussion of the optimal control problem under partial information will be relatively superficial. Still, we cover one of the tractable cases, the Linear-Quadratic Gaussian problem, and give a introductory discussion of problems with discrete-state spaces (often called Partially Observable Markov Decision Processes or POMDPs).

References: [Ber07, chapter 5], [TBF05].

5.1   Partial Information Model

DP Using the Full Information Vector

Often the basic problem formulation of chapter 1 needs to be modified to consider the situation where the controller does not have perfect knowledge of the state x_k. Instead, it only has access to measurements $y_k, k = 0, \ldots, N-1$, and must design the control policy based on these measurements only (as well as the memory of the past controls). These measurements give only a partial description of the state in general, and moreover they can be subject to noise. Let us assume that the measurement y_k of the state x_k is made before the control u_k is applied (see Fig. 5.1. Another ordering of events, with control preceding observation, leads to straightforward modifications). The
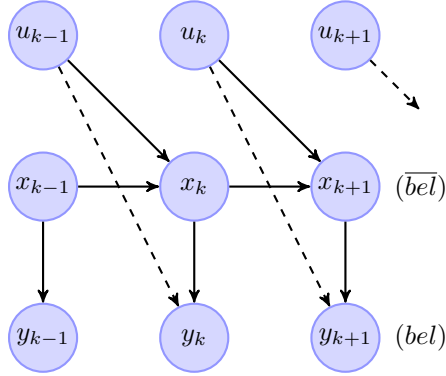
45

Figure 5.1: Dynamic Bayes network characterizing the evolution of controls, states and measurements.

observations are then of the form

$$y_0 = g_0(x_0, v_0), \ \ y_k = g_k(x_k, u_{k-1}, v_k), \ k = 1, \ldots, N - 1,$$

where the measurement y_k belongs to a given observation space Y_k, and the *observation disturbance* v_k to a space V_k. As for the process noise, we assume that these disturbances can be described probabilistically. The variable v_k has a known probability distribution which could a priori depend on the whole history of the system up to the end of period $k - 1$, denoted

$$\mathcal{H}_{k-1} = (x_{k-1}, \ldots, x_0, u_{k-1}, \ldots, u_0, w_{k-1}, \ldots, w_0, v_{k-1}, \ldots, v_0)$$

as well as the current state x_k. In the following, it is useful to introduce the notation $x_{i:j}$ to denote a set of variables $x_i, x_{i+1}, \ldots, x_j$. The realization of v_k follows the conditional distribution $P_{v_k}(\cdot | x_k, \mathcal{H}_{k-1})$. The initial state x_0 itself is random with known distribution P_{x_0}. The rest of the model is as in chapter 1, except that we assume that the control set U_k doe not depend on x_k (otherwise we would not even know with certainty which constraints our control input should respect!).

The *information vector* (or *information state*) $\mathcal{I}_k$ consists of all the information available at time k to the controller to make its decision u_k:

$$\mathcal{I}_0 = y_0,$$
$$\mathcal{I}_k = (y_0, y_1, \ldots, y_k, u_0, u_1, \ldots, u_{k-1}), \ k = 1, 2, \ldots, N.$$

The controller can only implement an *admissible* policy $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, where each function μ_k maps the information vector $\mathcal{I}_k$ (instead of the state in the perfect information case) into the control space U_k. Note the dramatic *increase in complexity* with respect to the full-information problem: the policy at time k should specify a control for every possible state of information at

time k, and this information state depends on past observations and controls. We want to find an admissible policy π minimizing the cost function

$$E\left\{c_N(x_N) + \sum_{k=0}^{N-1} c_k(x_k, \mu_k(\mathcal{I}_k), w_k)\right\}.$$

Note in passing that the controller does not observe directly the incurred stage costs $c_k(x_k)$ since it does not have access to the internal state x_k.

Example: Slotted Aloha, Bertsekas p.219.

Setting the computational difficulties aside for the moment, we can in theory treat the problem in the standard framework by taking the information vector as the new state. First, we know its dynamics

$$\mathcal{I}_{k+1} = (\mathcal{I}_k, y_{k+1}, u_k).$$

Here y_{k+1} plays the role of the disturbance entering the system dynamics, as in (1.1) (y_{k+1} is independent of the past information $\mathcal{I}_{k-1}$ given $\mathcal{I}_k$ and u_k, since this information is already part of $\mathcal{I}_k$). Using the tower property of the conditional expectation

$$E[c_k(x_k, u_k, w_k)] = E[E[c_k(x_k, u_k, w_k)|\mathcal{I}_k, u_k]]$$

and defining $\tilde{c}_k(\mathcal{I}_k, u_k) = E[c_k(x_k, u_k, w_k)|\mathcal{I}_k, u_k]$, the cost function can be rewritten in our original framework as

$$E\left\{\tilde{c}_N(\mathcal{I}_N) + \sum_{k=0}^{N-1} \tilde{c}_k(\mathcal{I}_k, \mu_k(\mathcal{I}_k))\right\}.$$

The DP algorithm is then simply

$$J_N^*(\mathcal{I}_N) = \tilde{c}_N(\mathcal{I}_N)$$
$$J_k^*(\mathcal{I}_k) = \min_{u_k \in \mathsf{U}_k} \left\{\tilde{c}_k(\mathcal{I}_k, u_k) + E\left[J_{k+1}^*((\mathcal{I}_k, y_{k+1}, u_k))\Big|\mathcal{I}_k, u_k\right]\right\}, \qquad (5.1)$$
$$k = 0, \ldots, N-1.$$

Recall that $\mathcal{I}_0 = y_0$. The optimal expected cost is then $J^* = E_{y_0}[J_0^*(y_0)]$. Now this naive formulation is obviously really difficult to use. Even to initialize the algorithm at time N, one has to consider all possible histories of observations and controls $\mathcal{I}_N$, and for each such information vector, compute the conditional expectation $\tilde{c}_N(\mathcal{I}_n) = E[c_N(x_N)|\mathcal{I}_N]$.

Sufficient Statistics

As mentioned previously, an important drawback of the DP algorithm above is that the information vector $\mathcal{I}_k$ (our new state after transformation to an

equivalent full information problem) grows with k. For estimation and control purposes, it turns out however that one can often find equivalent representations of the information contained in the observations that provide a significant reduction of the data to remember. We now introduce the concept of *sufficient statistic*. Classically, a *statistic* is just a function of the observations $f(Y_1, \ldots, Y_n)$. In the problems where this concept appears, the observations have a distribution which depends on some underlying unknown parameters. Intuitively, a statistic is called *sufficient* when no other statistic which can be calculated from the same observations provides any additional information for the estimation or control purpose, see [Fis22]. For example, given a sample $X_1, \ldots, X_n$ where the X_j are i.i.d., with normal distribution $N(\theta, 1)$ with unknown mean θ, the sample mean $\bar{X} = (X_1 + \ldots X_n)/n$ turns out to be a sufficient statistic for the estimation of the unknown parameter θ. Hence if our purpose is simply the estimation of this parameter, there is no need to remember all the observations. Recording only their sample mean is enough. The degree of data compression achieved by a sufficient statistics T depends on the target space of this function. You can define the amount of compression realized by a statistics as

$$r_c = \frac{\text{Number of bits of memory required to store } T(X)}{\text{Number of bits of memory required to store } X}.$$

Note that there is always the trivial sufficient statistics

$$T(X_1, \ldots, X_n) = (X_1, \ldots, X_n),$$

which achieves no compression ($r_c = 1$)!

Remark. In numerous applications, and certainly in the dynamic problems we have in mind, the samples X_j above appear sequentially. We would like to start the estimation process as soon as the first sample appear, and moreover, we want to make sure that if a new sample arrives, we do not need the past samples to update the value of the sufficient statistic! For this purpose, one typically has *recursive* versions of the classical sufficient statistics. For example, denoting the sample mean above M_n, it is easy to see that

$$M_{n+1} = M_n + \frac{1}{n+1}[X_{n+1} - M_n].$$

Recursive algorithms of this type will come up later on in the course.

The topic of sufficient statistics is quite interesting but giving a detailed presentation of it would take us too far away from our subject. You can refer to any classical textbook on statistics, e.g., [BD06, LC98], or [Dud03] for a measure-theoretic presentation. Still, it's good to know a few things about this topic. Consider a random variable X with discrete values and probability distribution P_θ depending on an unknown parameter θ. A statistic T is sufficient for θ if

$$P_\theta(X = x | T = t) = f(x, t)$$

where f is a function that does not depend on θ. Hence once we have computed T, there is no need to keep the initial data X. There is a similar definition if X is a continuous variable with a density, and also a general definition that you can find in the references above. This definition is not easy to use directly. In practice, we use the following result. Suppose the family of distributions $\{P_\theta\}_\theta$ is such that we can write $dP_\theta(x) = f(\theta, x)d\mu(x)$, for some fixed measure μ. The basic cases are X a discrete random variable, where $f(\theta, x)$ is the probability of the event $\{X = x\}$ (and μ is the counting measure), and X a continuous random variable with density $f(\theta, x)$ (and μ the Lebesgue measure). Then T is a sufficient statistic for θ if the we have

$$f(\theta, x) = G(\theta, T(x))h(x),$$

for some functions G and h. This is called the (Fisher) *factorization theorem*, and is usually much easier to verify than the definition. For example, the order statistics $T(X_1, \ldots, X_n) = [X_{(1)}, \ldots X_{(n)}]$, where $X_{(i)}$ denotes the ith smallest value in the sample ($X_{(1)} \leq \ldots X_{(n)}$) achieves a compression ratio $r_c = 1/2$ (the time stamps are removed) and is sufficient if the $X_i's$ are i.i.d. since the joint pdf can be factorized as

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} f(x_{(i)}).$$

Coming back to the DP algorithm, since our observations form the information vector $\mathcal{I}_k$, we are looking for a statistic $T_k(\mathcal{I}_k)$ for each k such that the minimization problem (5.1) can be rewritten as

$$\min_{u_k \in \mathsf{U}_k} H_k(T_k(\mathcal{I}_k), u_k),$$

for some function H_k. Then the control policy can clearly be given as a function of $T_k(\mathcal{I}_k)$ as well

$$\mu_k^*(\mathcal{I}_k) = \bar{\mu}_k(T_k(\mathcal{I}_k)).$$

To achive data compression with respect to the full vector $\mathcal{I}_k$, we would like the range space of T_k to be of smaller dimension than its domain (where the information vector lives). Moreover, since $T_k(\mathcal{I}_k)$ will then serve as our new state in the DP recursion, we will need a recursive way of computing $T_{k+1}(\mathcal{I}_{k+1})$ from $T_k(\mathcal{I}_k)$, u_k and y_{k+1}, to obtain the equivalent of the system dynamics equation.

The Conditional State Distribution as a Sufficient Statistic and the Bayes Filter

In this section, we make the following extra assumption. Conditioned on x_{k+1}, w_k and u_k, the observation v_{k+1} is independent of the past states, controls and disturbances $x_{0:k}, u_{0:k-1}, w_{0:k-1}, v_{0:k}$. Then, besides the identity function, there is another function that always provide a sufficient statistic for our partial

information control problem. This function is the conditional distribution of the system state x_k given the information $\mathcal{I}_k$, i.e., function $T_k(\mathcal{I}_k) = P_{x_k}(\cdot|\mathcal{I}_k)$. It is important to understand what this means, in particular the fact that this does not achieve any data compression in general, and could be even worse than the representation in terms of $\mathcal{I}_k$. Indeed, this function maps the space where $\mathcal{I}_k$ lives, which is typically finite dimensional although growing with time, to a space of probability measures, which is infinite dimensional! For example, let us assume that the measurements $y_0, \ldots, y_k$ and controls $u_0, \ldots, u_{k-1}$ are all real valued. Then $\mathcal{I}_k \in \mathbb{R}^{2k+1}$. Now assume that the state x_k is also real valued (we will encounter such a situation in section (5.2)). Then $T_k(\mathcal{I}_k)$, being a conditional distribution, is a *function* that associates to every (Borel measurable) subset of $\mathbb{R}$ a number between 0 and 1:

$$T_k(\mathcal{I}_k)(A) = P(X_k \in A|\mathcal{I}_k), \forall A \in \mathcal{B}(\mathbb{R}).$$

Here $B(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$, which consists of pretty much all subsets of $\mathbb{R}$ that you can think of (closed sets, open sets, and countable intersections and unions of these; so even the Cantor set is Borel measurable since it is closed). In this case, admittedly I'm exaggerating a bit because for x_k real we can characterize $P_{x_k}(\cdot|\mathcal{I}_k)$ completely by its distribution function $x \mapsto F(x|\mathcal{I}_k) = P(X_k \leq x|\mathcal{I}_k)$, but the range space of $T_k(\mathcal{I}_k)$ is still infinite-dimensional.

In any case, this sufficient statistic still has value in a number of very important cases, because in some instances the conditional probability distribution can be summarized by a finite number of parameters, in fact a much smaller number than the dimension of the space of $\mathcal{I}_k$. That is, there are important cases where we can find another statistic $\hat{T}(\mathcal{I}_k)$ such that $P_{x_k}(\cdot|\mathcal{I}_k) = G_k(\hat{T}_k(\mathcal{I}_k))$ for some function G_k, and this implies then that $T_k(\mathcal{I}_k)$ is also a sufficient statistic. We could have tried to find this sufficient statistic $\hat{T}_k$ directly starting from $\mathcal{I}_k$, but it seems that it is typically easier to do this with the representation in terms of conditional distributions. For example, if we can show that $P_{x_k}(\cdot|\mathcal{I}_k)$ is a Gaussian distribution, then its mean and covariance matrix can serve as a finite dimensional sufficient statistic $\hat{T}_k$, and the dimension of this statistic remains the same as time evolve if the dimension of the state space X_k does not change. This case actually arises in the important Linear-Quadratic-Gaussian control problem of section 5.2.

As mentioned previously, we also need a recursive procedure to update the sufficient statistic when a new measurement is obtained. This procedure is provided by *Bayes' rule*, which gives a dynamics equation of the form

$$P_{x_{k+1}}(\cdot|\mathcal{I}_{k+1}) = \Phi_k(P_{x_k}(\cdot|\mathcal{I}_k), u_k, y_{k+1}). \tag{5.2}$$

Indeed, assuming we know the conditional distribution of x_k, we need to update it to obtain the conditional distribution of x_{k+1} based on the additional knowledge of the new control and observation u_k, y_{k+1}. Writing the recursion in the general case is useful, because it can be taken as the basis for developing practical algorithms, including Kalman filters and particle filters.

Recall that we assumed the time evolution as in fig. 5.1. The procedure to compute the evolution (5.2) is usually carried in two steps for discrete time problems. In the first step, called the *propagation step*, we compute the conditional probability according to our knowledge of the control input u_k and the model of the dynamics, but without incorporating the measurement y_{k+1}, i.e., we propagate the dynamics of the system through the probabilistic model. We have

$$P(X_{k+1}|\mathcal{I}_k, u_k) = \int P(X_{k+1}|\mathcal{I}_k, u_k, x_k) \, dP(x_k|\mathcal{I}_k, u_k)$$

$$= \int P(X_{k+1}|u_k, x_k) \, dP(x_k|\mathcal{I}_k). \tag{5.3}$$

Here $P(X_k|\mathcal{I}_k, u_k) = P(X_k|\mathcal{I}_k)$ since u_k itself is a function of $\mathcal{I}_k$ (recall $u_k = \mu_k(\mathcal{I}_k)$). The distribution $P(X_k|\mathcal{I}_k)$ is assumed known from the previous step, and $P(X_{k+1}|u_k, X_k)$ can be computed directly from the distribution of the system disturbance w_k (or specified directly in the controlled Markov chain model).

In the second step, called the *update step*, we take the new measurement y_{k+1} into account. This decreases the uncertainty on the state x_{k+1} in general. The equation for this step, using Bayes' rule and our conditional independence assumptions on v_{k+1}, is

$$
\begin{aligned}
P(X_{k+1}|\mathcal{I}_{k+1}) &= P(X_{k+1}|\mathcal{I}_k, u_k, y_{k+1}) \\
&= \mathcal{Z} \; P(y_{k+1}|X_{k+1}, \mathcal{I}_k, u_k) P(X_{k+1}|\mathcal{I}_k, u_k), \\
&= \mathcal{Z} \; P(y_{k+1}|X_{k+1}, u_k) P(X_{k+1}|\mathcal{I}_k, u_k) \tag{5.4}
\end{aligned}
$$

where $\mathcal{Z}$ is a normalization factor

$$\mathcal{Z} = \left(\int P(y_{k+1}|x_{k+1}, u_k) \, dP(x_{k+1}|\mathcal{I}_k, u_k) \right)^{-1}.$$

Note that the distribution $P(X_{k+1}|\mathcal{I}_k, u_k)$ appearing in these equations was obtained from the propagation step. The distribution $P(Y_{k+1}|X_{k+1}, u_k)$ can be computed directly from the distribution of the measurement noise v_{k+1}, or again specified directly in a controlled (hidden) Markov chain model.

These equations can be specialized to the case of discrete distributions or continuous distributions with densities for example. In some fields, in particular artificial intelligence and robotics [TBF05], the conditional distribution $P_{x_k}(\cdot|\mathcal{I}_k)$ is often called the "belief state", reflecting the fact that this function serves as the actual state for planning purposes. The recursion expressing the new conditional distribution in terms of the previous one, the new control and new observation, is sometimes called the *Bayes filter*. To help fix ideas, let us assume that the distributions all have a density. Also, to simplify the notation, let us rename these densities $p(x_{t+1}|\mathcal{I}_k, u_k) =: \overline{bel}(x_{k+1})$ for the density before incorporating the new measurement, and $p(x_{k+1}|\mathcal{I}_{k+1}) =: bel(x_{k+1})$ for the

density of the belief state. Then the recursion in terms of densities is

$$\overline{bel}(x_{k+1}) = \int p(x_{k+1}|u_k, x_k)\, bel(x_k)dx_k,$$

$$bel(x_{k+1}) = \mathcal{Z}\, p(y_{k+1}|x_{k+1}, u_k)\overline{bel}(x_{k+1}),$$

$$\mathcal{Z} = \left(\int p(y_{k+1}|x_{k+1}, u_k)\, \overline{bel}(x_{k+1})dx_{k+1}\right)^{-1}.$$

To link the densities $p(x_{k+1}|u_k, x_k)$ and $p(y_{k+1}|x_{k+1}, u_k)$ to the dynamics and measurement noise for which the stochastic properties are usually given, note that we can write, assuming again that the distributions have densities:

$$p(x_{k+1}|u_k, x_k) = \int p(x_{k+1}|u_k, x_k, w_k)p(w_k|u_k, x_k)dw_k$$

$$= \int \delta\{x_{k+1} = f_k(x_k, u_k, w_k)\}\, p(w_k|u_k, x_k)dw_k, \qquad (5.5)$$

and similarly for the sensor noise.

Proof that the conditional distribution forms a sufficient statistic for DP. Once we have a recursive equation for the the conditional distribution $P(X_k|\mathcal{I}_k)$, this distribution can play the role of state for control purposes and we can write the corresponding DP algorithm. Hence our goal is to show that we can write

$$J_k^*(\mathcal{I}_k) = \min_{u_k \in \mathsf{U}_k} H_k(P_{x_k}(\cdot|\mathcal{I}_k), u_k) = \overline{J}_k^*(P_{x_k}(\cdot|\mathcal{I}_k)), \qquad (5.6)$$

for some appropriate functions H_k and $\overline{J}_k^*$. For the last period, we have

$$J_N^*(\mathcal{I}_N) = E[c_N(x_N)|\mathcal{I}_N]$$

$$= \int c_N(x_N)\, dP(x_N|\mathcal{I}_N)$$

$$= \overline{J}_k^*(P_{x_N}(\cdot|\mathcal{I}_N)). \qquad (5.7)$$

Let us repeat here the DP recursion in terms of the full information vector

$$J_k^*(\mathcal{I}_k) = \min_{u_k \in \mathsf{U}_k} E\left[c_k(x_k, u_k, w_k) + J_{k+1}^*((\mathcal{I}_k, u_k, y_{k+1}))\,\Big|\,\mathcal{I}_k, u_k\right] \qquad (5.8)$$

Recall that we know the distributions $P(W_k|X_k, u_k)$ and $P(V_{k+1}|X_{k+1}, W_k, u_k)$ for all k. From the observation noise distribution, as we saw above for the process noise, we can deduce the distribution of the observations $P(Y_{k+1}|X_{k+1}, W_k, u_k)$. Then the first term on the right hand side of of the DP recursion (5.8) can be written

$$E[c_k(x_k, u_k, w_k)|\mathcal{I}_k, u_k] = E[E_{w_k}[c_k(x_k, u_k, w_k)|x_k, u_k]\,|\,\mathcal{I}_k, u_k]$$

$$= \int \hat{c}_k(x_k, u_k)\, dP(x_k|\mathcal{I}_k, u_k) \qquad (5.9)$$

$$= \int \hat{c}_k(x_k, u_k)\, dP(x_k|\mathcal{I}_k) = \tilde{c}_k(P_{x_k}(\cdot|\mathcal{I}_k), u_k).$$

with $\hat{c}_k(x_k, u_k) = \int c_k(x_k, u_k, w_k)\, dP(w_k|x_k, u_k)$. We have already seen in the derivation of the Bayes filter that the last equality holds because u_k is constrained to be itself a function of $\mathcal{I}_k$. Similarly for the second term

$$E\left[J_{k+1}^*(\mathcal{I}_k, u_k, y_{k+1})\Big|\mathcal{I}_k, u_k\right] = E\left[E_{y_{k+1}}[J_{k+1}^*(\mathcal{I}_k, u_k, y_{k+1})|x_{k+1}, u_k, w_k]\Big|\mathcal{I}_k, u_k\right].$$

The inner conditional expectation

$$F(x_{k+1}, u_k, w_k) = E_{y_{k+1}}[J_{k+1}^*(\mathcal{I}_k, u_k, y_{k+1})|x_{k+1}, u_k, w_k],$$

is simply an integration with respect to the distribution $P(Y_{k+1}|X_{k+1}, W_k, u_k)$, which is known. Then the computation of

$$E[F(x_{k+1}, u_k, w_k)|\mathcal{I}_k, u_k]$$

requires the distribution $P(X_{k+1}, W_k|\mathcal{I}_k, u_k)$. This distribution is obtained as

$$P(X_{k+1}, W_k|\mathcal{I}_k, u_k) = P(X_{k+1}|W_k, \mathcal{I}_k, u_k)P(W_k|\mathcal{I}_k, u_k).$$

Then

$$P(W_k|\mathcal{I}_k, u_k) = \int P(W_k|x_k, u_k)\, dP(x_k|\mathcal{I}_k),$$

with $P(W_k|x_k, u_k)$ known. Finally for the distribution $P(X_{k+1}|W_k, \mathcal{I}_k, u_k)$:

$$P(X_{k+1} \in A|W_k, \mathcal{I}_k, u_k) = \int 1\{f_k(x_k, u_k, W_k) \in A\}\, dP(x_k|W_k, \mathcal{I}_k, u_k)$$

and by Bayes' rule

$$P(X_k|W_k, \mathcal{I}_k, u_k) = \mathcal{Z}_1 P(W_k|X_k, u_k)P(X_k|\mathcal{I}_k),$$

where $\mathcal{Z}_1$ is a normalization factor. Hence all the terms appearing on the right-hand side of (5.8) have been written as functions of $P_{x_k}(\cdot|\mathcal{I}_k)$ and u_k, and (5.6) holds. The DP recursion step can be written

$$\overline{J}_k^*(P_{x_k}(\cdot|\mathcal{I}_k)) = \min_{u_k \in \mathsf{U}_k}\left\{\tilde{c}_k(P_{x_k}(\cdot|\mathcal{I}_k), u_k) + E\left[\overline{J}_{k+1}^*(P_{x_{k+1}}(\cdot|\mathcal{I}_{k+1}))\Big|P_{x_k}(\cdot|\mathcal{I}_k), u_k\right]\right\},$$

(5.10)

with the recursion

$$P_{x_{k+1}}(\cdot|\mathcal{I}_{k+1}) = \Phi_k(P_{x_k}(\cdot|\mathcal{I}_k), u_k, y_{k+1}).$$

Finally, note the conceptual importance of the representation of the optimal policy as a sequence of functions of the conditional distribution

$$\mu_k^*(I_k) = \overline{\mu}_k(P_{x_k}(\cdot|\mathcal{I}_k)), k = 0, \ldots, N-1.$$

This representation provides a decomposition of the optimal controller into two parts:

- An *estimator*, which uses at time k the measurement y_k and the control u_{k-1} to generate the probability distribution $P_{x_k}(\cdot|\mathcal{I}_k)$.

- An *actuator*, which generates a control input to the system as a function of the probability distribution $P_{x_k}(\cdot|\mathcal{I}_k)$.

Note that the optimum control problem remains very hard in the general, because of two main factors:

- The practical computation of the Bayes filter in hard in the general case. For example, the propagation and update steps involve the computation of multidimensional integrals, which is difficult in high-dimensions. Simulation methods, such as particle filters, can be used but are computationally intensive.

- Perhaps more importantly, the control u_k has now two roles. Firstly, it should drive the state in a good state from the cost point of view, as in the perfect information case. Secondly, it influences the next observation y_{k+1}, sometimes directly through h_{k+1} but always indirectly through x_{k+1}. Since the next observations are critical for keeping a good state estimate and designing good control inputs in the future, we see that the influence of u_k on the cost of the trajectory becomes quite complicated to characterize. This dual role of the control can be seen as a form of the famous exploitation vs. exploration principle. Inputs that minimize the stage cost in the short term are not necessarily good if they are too detrimental to the state estimation problem.

Hence although the decomposition of the controller in terms of estimation and control holds in terms of implementation it does not mean that the solutions of the optimal estimation and control problems are decoupled. Nonetheless, this decomposition is useful for developing heuristics for hard control problems under imperfect information, see e.g. chapter 11.

5.2   Linear Quadratic Problems

We will now study a particular importance case where the difficulties mentioned at the end of the previous section do not apply, namely where the estimation and control problems actually decouple and can be solved independently. A reference for our approach is [Ber07, section 5.2]. The dynamics of the system are as in chapter 4

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \ k = 0, 1, \ldots, N - 1,$$

assuming that the initial state x_0 is random, with known finite mean $\bar{x}_0$ and covariance matrix Σ_0. In addition, the measurements are linear in the state, of the form

$$y_k = C_k x_k + v_k, \ k = 0, \ldots, N - 1.$$

Here $x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^m, y^k \in \mathbb{R}^p$ and the matrices A_k, B_k, C_k are of appropriate dimensions. As in chapter 4, the variables w_k are assumed to be independent, and independent of x_0, zero mean, and have a finite covariance matrix W_k. Now