4 Markov processes with imperfect state information

We consider a Markov Decision Process (MDP) with *imperfect information* with finite state and action spaces $\mathcal{X}, \mathcal{U}$. Recall that an MDP with complete information takes the form

$$X_{t+1} = f_t(X_t, U_t, W_t) \tag{1}$$

with primitive random variables $X_0, W_0, W_1, \dots$ which are mutually independent. Under this model, the information available at time t is X_0^t, U_0^{t-1} which can be used by the policy to take an action. We will generalize this model to the situation where the states $X_0, \dots, X_t$ are partially observable.

4.1 Model

In addition to the state dynamics equation in (1), define an observation space $\mathcal{Y}$ which is also finite. There is an underlying observation model

$$Y_t = h_t(X_t, V_t) \tag{2}$$

with observation disturbances $V_0, V_1, \dots$ that will now be added to the primitive random variables. Naturally we assume that $X_0, W_0, W_1, \dots, V_0, V_1, \dots$ are mutually independent. In contrast to the perfect information setting, we assume that we only have knowledge of the history of observations Y_0^t and actions U_0^{t-1} at time t.

As usual, the problem of stochastic optimal control over a finite horizon $T \geq 1$ is to minimize the expected cost of a policy $g = (g_0, \dots, g_{T-1})$:

$$J(g) = \mathbf{E}^g\Big[\sum_{t=0}^{T-1} c_t(X_t, U_t) + c_T(X_T)\Big] = \mathbf{E}^g\Big[\sum_{t=0}^{T-1} c_t(X_t, g_t(Y_0^t, U_0^{t-1})) + c_T(X_t)\Big]. \tag{3}$$

Notice that unlike the perfect information setting, in which the policy was a function of the history of states and actions, the policy g_t in the imperfect observation case is a function of the observation-action history. Here, it is important to note that, while the cost *functions* c_t are known in advance, the realization of the cost $c_t(X_t, U_t)$ incurred at each time t is not observable.

Given a causal policy $\{g_t\}$, where $g_t : \mathcal{Y}_0^t \times \mathcal{U}_0^{t-1} \to \mathcal{U}$, the states, observations, and actions

evolve according to the following recursion:

$$\begin{aligned}
X_0 &\sim \mu \\
Y_0 &= h_0(X_0, V_0) \\
U_0 &= g_0(Y_0) \\
X_1 &= f_0(X_0, U_0, W_0) \\
Y_1 &= h_1(X_1, V_1) \\
U_1 &= g_1(Y_0^1, U_0) \\
&\dots \\
X_t &= f_{t-1}(X_{t-1}, U_{t-1}, W_{t-1}) \\
Y_t &= h_t(X_t, V_t) \\
U_t &= g_t(Y_0^t, U_0^{t-1}) \\
X_{t+1} &= f_{t+1}(X_t, U_t, W_t) \\
&\dots
\end{aligned}$$

As in the case of complete observations, once we fix a policy g, the states, observations, and actions are deterministic functions of the primitive random variables variables.

Similar to the perfect information setting, we are interested in answering the following questions:

1. Given an appropriate notion of *state*, are Markov policies optimal?
2. Is there an analogue of dynamic programming (DP) as in the perfect observation case that can be used to find optimal policies?

Before attempting to answer the above questions, we first consider a simpler model in which there is no underlying control.

4.2 Hidden Markov Model (HMM)

An HMM is defined by the tuple $(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ is the "hidden" state space and $\mathcal{Y}$ is the observation space. There is an underlying state dynamics

$$X_{t+1} = f_t(X_t, W_t) \tag{4}$$

and observation model

$$Y_t = h_t(X_t, V_t) \tag{5}$$

with mutually independent random variables $X_0, W_0, W_1, \dots, V_0, V_1, \dots$. The goal is to compute the posterior probability of the state given the history of observations. We first prove the following:

Proposition 4.1. *The random process $\{Z_t\}$ with $Z_t = (X_t, Y_t) \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ is a Markov chain.*

Proof: It suffices to show that there exists a sequence of deterministic functions $\tilde{f}_t$ and a sequence of independent random variables $\tilde{W}_t$ that are independent of Z_0, such that $Z_{t+1} = \tilde{f}_t(Z_t, \tilde{W}_t)$. To see this, we write

$$\begin{aligned} Z_{t+1} &= (X_{t+1}, Y_{t+1}) \\ &= [f_t(X_t, W_t), h_{t+1}(X_{t+1}, V_{t+1}] \\ &= [f_t(X_t, W_t), h_{t+1}(f_t(X_t, W_t), V_{t+1})] \\ &= \tilde{f}_t(X_t, Y_t, \tilde{W}_t) \\ &= \tilde{f}_t(Z_t, \tilde{W}_t), \end{aligned}$$

where $\tilde{W}_t := (W_t, V_{t+1})$ and $\tilde{f}_t(x, y, \tilde{w}) := (f_t(x, w), h_{t+1}(f_t(x, w), v))$. ■

Remark 4.1. *Although $Z_t = (X_t, Y_t)$ form a Markov chain, the observation process $\{Y_t\}$ may not be a Markov chain.*

The joint distribution of Z_0^t can now be decomposed as a function of the initial state distribution μ, state transition probability matrix $P^{(t)}(x, x') = \mathbf{P}[f_t(x, W_t) = x']$, and the conditional distribution of observation y given state x: $M^{(t)}(x, y) = \mathbf{P}[h_t(x, V_t) = y]$:

$$\begin{aligned} \mathbf{P}[Z_0^t = z_0^t] &= \mathbf{P}[X_0^t = x_0^t, Y_0^t = y_0^t] \\ &= \mu(x_0)M^{(0)}(x_0, y_0)P^{(0)}(x_0, x_1)M^{(1)}(x_1, y_1) \dots P^{(t-1)}(x_{t-1}, x_t)M^{(t)}(x_t, y_t). \end{aligned}$$

For each $s, t \geq 0$, let $\pi_{t|s}$ denote the conditional distribution of X_t given Y_0^s:

$$\pi_{t|s}(x) := \mathbf{P}[X_t = x | Y_0^s = y_0^s].$$

Although $\pi_{t|s}$ is a function of the observations Y_0^s, we suppress this dependence to keep the notation simple. Depending on the relative values of s and t, we distinguish three problems associated with computing these conditional distributions:

1. If $t = s$, the problem is termed "filtering."
2. If $t > s$, the problem is named "prediction."
3. If $t < s$, the problem is called "smoothing."

We focus on filtering, where, for each $x \in \mathcal{X}$, we compute the posterior probability of being in state x at time t given observations $Y_0^t = y_0^t$, i.e., the probability row vector $\pi_{t|t}$.

From now on, we can assume, without loss of generality, that the state space $\mathcal{X}$ is the set of integers from 1 to n, $\mathcal{X} = \{1, \dots, n\}$. The set

$$\Delta_n := \left\{ \pi = (\pi(1), \dots, \pi(n)) : \pi(i) \geq 0, \sum_{i=1}^{n} \pi(i) = 1 \right\}$$

is the collection of all probability distributions on $\mathcal{X}$, which we identify with n-dimensional row vectors with nonnegative coordinates that sum to one. Note that Δ_n is a closed and bounded subset of $\mathbb{R}^n$.

Denote $\pi_t(\cdot) = \pi_{t|t}(\cdot|Y_0^t)$. We will show that π_{t+1} can be computed from π_t and Y_{t+1} using a *deterministic* update rule $\mathsf{F}_t : \Delta_n \times \mathcal{Y} \to \Delta_n$: $\pi_{t+1} = \mathsf{F}_t(\pi_t, Y_{t+1})$. The function F_t maps the posterior state distribution at time t and a new observation at time $t+1$ onto a the posterior state distribution at time $t+1$.

Remark 4.2. *Given an HMM with unobservable state space $\mathcal{X}$ and observation space $\mathcal{Y}$, we can define a fully observable Markov chain with a new state space Δ_n. This allows us to avoid storing the history of observations Y_0^t which may be memory-expensive; alternatively we just have to store the previous 'state' $\pi_t \in \Delta_n$ and update the state when given a new observation y_{t+1} to $\pi_{t+1} = \mathsf{F}(\pi_t, y_{t+1})$.*

Using Bayes' rule, we write

$$\pi_{t+1}(x_{t+1}) = \mathbf{P}[X_{t+1} = x_{t+1}|Y_0^{t+1} = y_0^{t+1}] = \frac{\mathbf{P}[X_{t+1} = x_{t+1}, Y_0^{t+1} = y_0^{t+1}]}{\mathbf{P}[Y_0^{t+1} = y_0^{t+1}]}. \tag{6}$$

The joint distribution of the history of states and observations up to time $t+1$ can be decomposed into marginal and conditional probabilities:

$$\begin{aligned}
\mathbf{P}[X_{t+1} = x_{t+1}, Y_0^{t+1} = y_0^{t+1}] &= \mathbf{P}[X_{t+1} = x_{t+1}, Y_0^t = y_0^t]\mathbf{P}[Y_{t+1} = y_{t+1}|X_{t+1} = x_{t+1}, Y_0^t = y_0^t] \\
&\overset{(a)}{=} \mathbf{P}[X_{t+1} = x_{t+1}, Y_0^t = y_0^t]M^{(t+1)}(x_{t+1}, y_{t+1}) \\
&= \mathbf{P}[X_{t+1} = x_{t+1}|Y_0^t = y_0^t]\mathbf{P}[Y_0^t = y_0^t]M^{(t+1)}(x_{t+1}, y_{t+1}) \\
&= \pi_{t+1|t}(x_{t+1})M^{(t+1)}(x_{t+1}, y_{t+1}) \cdot \mathbf{P}[Y_0^t = y_0^t]
\end{aligned} \tag{7}$$

where (a) is due to the fact that the observation y_{t+1} is conditionally independent of past observations

y_0^t and states x_0^t given the state x_{t+1}.[1] Combining (6) with (7) we obtain

$$\begin{aligned}\pi_{t+1}(x_{t+1}) &= \frac{\pi_{t+1|t}(x_{t+1})M^{(t+1)}(x_{t+1},y_{t+1})\mathbf{P}[Y_0^t=y_0^t]}{\mathbf{P}[Y_{t+1}=y_{t+1}|Y_0^t=y_0^t]\mathbf{P}[Y_0^t=y_0^t]}\\ &= \frac{\pi_{t+1|t}(x_{t+1})M^{(t+1)}(x_{t+1},y_{t+1})}{\mathbf{P}[Y_{t+1}=y_{t+1}|Y_0^t=y_0^t]}\\ &= \frac{\pi_{t+1|t}(x_{t+1})M^{(t+1)}(x_{t+1},y_{t+1})}{\sum_{x'}\mathbf{P}[X_{t+1}=x',Y_{t+1}=y_{t+1}|Y_0^t=y_0^t]}\\ &\overset{(b)}{=} \frac{\pi_{t+1|t}(x_{t+1})M^{(t+1)}(x_{t+1},y_{t+1})}{\sum_{x'}M^{(t+1)}(x',y_{t+1})\pi_{t+1|t}(x',y_{t+1})}\end{aligned}$$

where (b) follows from the identity $\mathbf{P}[X_{t+1}=x',Y_{t+1}=y_{t+1}]=M^{(t+1}(x',y_{t+1})\pi_{t+1|t}(x',y_{t+1})$.

We have established a relationship between the filtering distribution $\pi_{t+1}(x_{t+1})$ and the predictive distribution $\pi_{t+1|t}(x_{t+1})$ as follows:

$$\pi_{t+1}=\mathsf{C}_t(\pi_{t+1|t},y_{t+1}),$$

where $\mathsf{C}_t:\Delta_n\times\mathcal{Y}\to\Delta_n$ is a function defined by

$$[\mathsf{C}_t(\pi,y)](x)=\frac{\pi(x)M^{(t+1)}(x,y)}{\sum_{x'}\pi(x')M^{(t+1)}(x',y)},\qquad \pi\in\Delta_n,\ (x,y)\in\mathcal{X}\times\mathcal{Y}. \tag{8}$$

We call the map $\mathsf{C}_t:(\pi_{t+1|t},y_{t+1})\mapsto\pi_{t+1}$ the *correction map*, and the process of updating $\pi_{t+1|t}$ to π_{t+1} the correction step. We next consider the *prediction step*, i.e., the process of updating π_t to $\pi_{t+1|t}$:

$$\begin{aligned}\pi_{t+1|t}(x_{t+1}) &= \mathbf{P}[X_{t+1}=x_{t+1}|Y_0^t=y_0^t]\\ &= \sum_{x'}\mathbf{P}[X_{t+1}=x_{t+1},X_t=x'|Y_0^t=y_0^t]\\ &= \sum_{x'}\mathbf{P}[X_{t+1}=x_{t+1}|X_t=x',Y_0^t=y_0^t]\mathbf{P}[X_t=x'|Y_0^t=y_0^t]\\ &= \sum_{x'}\mathbf{P}[X_{t+1}=x_{t+1}|X_t=x',Y_0^t=y_0^t]\pi_{t|t}(x')\\ &= \sum_{x'}P^{(t)}(x',x_{t+1})\pi_t(x').\end{aligned}$$

[1] Let (X,Y,Z) be a triple of jointly distributed random variables. We say that X and Z are conditionally independent given Y if the following holds for all measurable subsets A,B,C of the appropriate spaces:

$$\mathbf{P}[X\in A|Y\in B,Z\in C]=\mathbf{P}[X\in A|Y\in B],$$

or, equivalently,

$$\mathbf{P}[X\in A,Z\in C|Y\in B]=\mathbf{P}[X\in A|Y\in B]\mathbf{P}[Z\in C|Y\in B].$$

If we now define the *prediction map* $\mathsf{P}_t : \Delta_n \to \Delta_n$ by $\mathsf{P}_t(\pi) = \pi P^{(t)}$, where the row vector π is multiplied on the right by the state transition probability matrix $P^{(t)}$, we see that This gives rise to the prediction step:

$$\pi_{t+1|t}(x_{t+1}) = \sum_x P^{(t)}(x, x_{t+1})\pi_t(x) \equiv [\mathsf{P}_t(\pi)](x), \qquad x \in \mathcal{X}. \tag{9}$$

The composition $\mathsf{F}_t := \mathsf{C}_t \circ \mathsf{P}_t$ of the predition and the correction steps gives rise to the *nonlinear filter equation*:

$$\pi_{t+1} = \mathsf{F}_t(\pi_t, y_{t+1}) := \mathsf{C}_t[\mathsf{P}_t(\pi_t), y_{t+1}]. \tag{10}$$

We can now summarize the procedure for updating the filtering distribution π_t (also known as the *belief state* at time t):

1. Given the belief state π_t at time t, compute the predictive distribution $\pi_{t+1|t} = \mathsf{P}(\pi_t)$ according to (9).

2. After observing Y_{t+1}, correct the predictive state distribution to $\pi_{t+1} = \mathsf{C}_t(\pi_{t+1|t}, Y_{t+1})$ according to (8).

The non-linearity comes from the normalization in the correction step C_t. We will next state a claim that formally justifies the statement made earlier: storing just the previous belief state π_t, rather than the history of observations Y_0^t, is sufficient to determine the conditional probability distribution of the next belief state π_{t+1}:

Proposition 4.2. *Define the augmented state space $\mathcal{S} := \Delta_n \times \mathcal{Y}$. Given the HMM $\{(X_t, Y_t)\}_{t\geq 0}$, let $S_t := (\pi_t, Y_t)$. Then $\{S_t\}_{t\geq 0}$ is a Markov chain with state space $\mathcal{S}$.*

Proof: Fix any Borel set $A \subseteq \Delta_n$ and any $y \in \mathsf{Y}$. We need to show that, for any t,

$$\mathbf{P}[\pi_{t+1} \in A, Y_{t+1} = y | \pi_0^t, y_0^t] = \tau_t(\pi_t, y_t; y, A),$$

for some function τ_t. Consider the LHS in the above equation:

$$\begin{aligned} \mathbf{P}[\pi_{t+1} \in A, Y_{t+1} = y | \pi_0^t, y_0^t] &= \mathbf{P}[\mathsf{F}_t(\pi_t, Y_{t+1}) \in A, Y_{t+1} = y | \pi_0^t, y_0^t] \\ &= \mathbb{1}\{\mathsf{F}_t(\pi_t, y) \in A\}\mathbf{P}[Y_{t+1} = y | \pi_0^t, y_0^t] \end{aligned} \tag{11}$$

The first term in the product on the RHS, i.e., $\mathbb{1}\{\mathsf{F}_t(\pi_t, y) \in A\}$, is clearly a function of only π_t, y and A. We need to show that the second term in the product, i.e, $\mathbf{P}[Y_{t+1} = y | \pi_0^t, y_0^t]$, is a function of only π_t and y. Note that since $\pi_s = \mathsf{F}_{s-1}(\pi_{s-1}, y_s)$ for each s, π_0^t is a deterministic function of y_0^t.

Therefore, we have:

$$\begin{aligned}\mathbf{P}[Y_{t+1} = y|\pi_0^t, y_0^t] &= \mathbf{P}[Y_{t+1} = y|y_0^t] \\ &= \sum_x \mathbf{P}[X_{t+1} = x, Y_{t+1} = y|y_0^t] \\ &= \sum_x \mathbf{P}[Y_{t+1} = y|X_{t+1} = x, y_0^t]\mathbf{P}[X_{t+1} = x|y_0^t] \\ &= \sum_x M^{(t+1)}(x, y)\pi_{t+1|t}(x) \\ &= \sum_x M^{(t+1)}(x, y)\pi_t P^{(t)}(x) \\ &= \pi_t P^{(t)} M^{(t+1)}(y)\end{aligned}$$

Combining the above equation with (11), we have:

$$\mathbf{P}[\pi_{t+1} \in A, Y_{t+1} = y|\pi_0^t, y_0^t] = \mathbb{1}\{\mathsf{F}_t(\pi_t, y) \in A\}\pi_t P^{(t)} M^{(t+1)}(y) \tag{12}$$

Since $\mathbf{P}[\pi_{t+1} \in A, Y_{t+1} = y|\pi_0^t, y_0^t]$ is simply a function of π_t, y_t, y and A, we see that $\{S_t\}_{t\geq 0} = \{(\pi_t, y_t)\}_{t\geq 0}$ is a Markov chain. ■

Using the above proposition, we can show that the belief state process $\{\pi_t\}_{t\geq 0}$ is also a Markov chain. From the proof it follows that π_{t+1} and Y_0^t are conditionally independent given π_t. Therefore, using (12):

$$\begin{aligned}\mathbf{P}[\pi_{t+1} \in A|\pi_0^t, y_0^t] &= \mathbf{P}[\pi_{t+1} \in A|\pi_0^t] \\ &= \sum_y \mathbf{P}[\pi_{t+1} \in A, Y_{t+1} = y|\pi_0^t, y_0^t] \\ &= \sum_y \mathbb{1}\{\mathsf{F}_t(\pi_t, y) \in A\}\pi_t P^{(t)} M^{(t+1)}(y)\end{aligned}$$

Note that the RHS of the above equation depends only on π_t and A. Therefore, $\{\pi_t\}_{t\geq 0}$ is also a Markov chain. Alternately, we can also compute $\mathbf{P}[\pi_{t+1} \in A|\pi_0^t, y_0^t]$ and observe that it is just a function of π_t.

4.3 Markov Decision Processes (MDPs) with imperfect observations

We now move on to handling MDPs with imperfect observations, i.e., introduce an action in the HMM model that we analyzed in the previous subsection. We have the following system dynamics:

$$\begin{aligned}X_{t+1} &= f_t(X_t, U_t, W_t) \\ Y_t &= h_t(X_t, V_t),\end{aligned}$$

where $X_0, W_0, W_1, ..., V_0, V_1, ...$ are independent primitive random variables as before. We also have the stagewise cost functions: $c_t : \mathcal{X} \times \mathcal{U} \to \mathbb{R}, t = 0, 1, ..., T-1$, and the terminal cost $c_T : \mathcal{X} \to \mathbb{R}$. A policy uses all the information available until the current time and takes an action. It is defined as $g = (g_t)_{t=0}^{T-1}$, where g_t maps $Z_t = (Y_0^t, U_0^{t-1})$ to the action U_t. We aim to show the following:

1. The *belief state* $\pi_t(\cdot) := \mathbf{P}[X_t = \cdot | Z_t]$ can be updated recursively in a policy-independent manner, i.e., there exist deterministic functions F_t, such that

$$\pi_{t+1} = \mathsf{F}_t(\pi_t, U_t, Y_{t+1})$$

for any admissible policy g.

2. The belief state process $\{\pi_t\}_{t\geq 0}$ is a controlled Markov chain, i.e., for any admissible policy g, the following holds: for any Borel set $A \subseteq \Delta_n$,

$$\mathbf{P}^g[\pi_{t+1} \in A | \pi_0^t, u_0^t] = \mathbf{P}[\pi_{t+1} \in A | \pi_t, u_t],$$

where the right-hand side is policy-independent.

Once we prove the above two statements, we will define new cost functions $\tilde{c}_t : \Delta_n \times \mathcal{U} \to \mathbb{R}, t = 0, 1, ..., T-1$, and $\tilde{c}_T : \Delta_n \to \mathbb{R}$ such that, for any admissible policy g,

$$\mathbf{E}^g \left[\sum_{t=0}^{T-1} c_t(x_t, u_t) + c_T(x_T) \right] = \mathbf{E}^g \left[\sum_{t=0}^{T-1} \tilde{c}_t(\pi_t, u_t) + \tilde{c}_T(\pi_T) \right] \tag{13}$$

The above equation would then imply that Markovian policies, i.e., the ones under which $U_t = g_t(\pi_t)$ for some functions $g_t : \Delta_n \to \mathcal{U}$, are optimal (from our analysis of the fully observed case). Consider the following cost function definitions:

$$\begin{aligned} \tilde{c}_t(\pi, u) &:= \sum_{x \in \mathcal{X}} \pi(x) c_t(x, u), \\ \tilde{c}_T(\pi) &:= \sum_{x \in \mathcal{X}} \pi(x) c_T(x). \end{aligned}$$

Then, for any policy $g = (g_0, g_1, ..., g_{T-1})$, such that $u_t = g_t(y_0^t, u_0^{t-1})$, we have:

$$\begin{aligned} \mathbf{E}^g[c_t(X_t, U_t)] &= \mathbf{E}^g[\mathbf{E}^g[c_t(X_t, U_t) | Z_t]] \\ &= \mathbf{E}^g[\mathbf{E}^g[c_t(X_t, g_t(Y_0^t, U_0^{t-1})) | Y_0^t, U_0^{t-1}]] \\ &= \mathbf{E}^g \left[\sum_{x \in \mathcal{X}} \mathbf{P}^g\{X_t = x | Y_0^t, U_0^{t-1}\} c_t(x, g_t(Y_0^t, U_0^t)) \right] \\ &= \mathbf{E}^g[\tilde{c}_t(\pi_t, U_t)]. \end{aligned}$$

From the above equation, it can be readily seen that the new cost functions that we have defined, satisfy (13). In order to prove the first point, we need to show that $\pi_{t+1} = \mathsf{F}_t(\pi_t, y_{t+1}, u_t)$. Consider:

$$\begin{aligned} \pi_{t+1|t}(x) &= \mathbf{P}[X_{t+1} = x | Y_0^t, U_0^t] \\ &= \pi_t P_{U_t}^{(t)}(x) \\ &= \sum_{x'} \pi_t(x') P_{U_t}^{(t)}(x', x) \\ \Rightarrow \mathbf{P}^g[X_{t+1} = x | Y_0^t, U_0^t] &= \sum_{x'} \mathbf{P}^g[X_{t+1} = x | X_t = x', Y_0^t, U_0^t] \mathbf{P}^g[X_t = x' | Y_0^t, U_0^t] \end{aligned}$$

Observe that the first term of the product inside the sum on the RHS is $P^{(t)}_{U_t}(x', x)$. Also, since $U_t = g_t(Y_0^t, U_0^{t-1})$, we have $\mathbf{P}^g[\cdot|Z_t, U_t] = \mathbf{P}^g[\cdot|Z_t]$. This implies that the second term of the product inside the sum on the RHS is $\pi_t(x')$. Hence, we have derived the prediction step $(\pi_t, U_t) \mapsto \pi_{t+1|t}$:

$$\pi_{t+1|t} = \mathsf{P}_t(\pi_t, U_t) := \pi_t P^{(t)}_{U_t}.$$

Next, we have

$$\begin{aligned}\mathbf{P}^g[X_{t+1} = x, Y_0^{t+1} = y_0^{t+1}, U_0^t = u_0^t] &= \mathbf{P}^g[X_{t+1} = x, Y^t = y_0^t, Y_{t+1} = y_{t+1}, U_0^t = u_0^t] \\ &= M^{(t+1)}(x, y_{t+1})\mathbf{P}[X_{t+1} = x, y_0^t, u_0^t] \\ &= M^{(t+1)}(x, y_{t+1})\pi_{t+1|t}(x)\mathbf{P}[y_0^t, u_0^t].\end{aligned}$$

Accounting for the normalization, we have that $\pi_{t+1} = \mathsf{C}_t(\pi_{t+1|t}, Y_{t+1})$, where the correction map $\mathsf{C}_t : \Delta_n \times \mathcal{Y} \to \Delta_n$ is defined by

$$[\mathsf{C}_t(\pi, y)](x) := \frac{\pi(x)M^{(t+1)}(x, y)}{\pi M^{(t+1)}(y)}.$$

The belief state update is the composition of the prediction and the correction steps:

$$\mathsf{F}_t = \mathsf{C}_t \circ \mathsf{P}_t, \qquad \pi_{t+1} = \mathsf{C}_t(\mathsf{P}_t(\pi_t, U_t), Y_{t+1}),$$

where the prediction and correction steps in this case are as follows:

$$\begin{aligned}\mathsf{P}_t(\pi, u) &= \pi P^{(t)}_u \\ [\mathsf{C}_t(\pi, y)](\cdot) &= \frac{\pi(\cdot)M^{(t+1)}(\cdot, y)}{\pi M^{(t+1)}(y)}\end{aligned}$$

In the next lecture, we will prove the second important point, i.e., π_t is a controlled Markov chain.