

## Problem 14.4

Let us first state some context regarding Hidden Markov Models by following the exposition given in [1]. Given two measurable spaces  $(X, \mathcal{X})$  and  $(Y, \mathcal{Y})$ ,  $Q$  a Markov kernel on  $(X, \mathcal{X})$  and  $G$  a Markov kernel from  $(X, \mathcal{X})$  to  $(Y, \mathcal{Y})$  we define the following transition kernel on the product space:

$$T((x, y), C) = \int 1_C(x, y) G(x', dy') Q(x, dx')$$

Together with the initial distribution  $\nu \otimes G$ , this defines a Markov chain  $(X_k, Y_k)_{k \geq 0}$ .

For the case considered in the problem,  $(X, \mathcal{X}) = (\{1, \dots, \kappa\}, \mathcal{P}(\{1, \dots, \kappa\}))$ ,  $(Y, \mathcal{Y}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and the model is *fully dominated*, meaning that both kernels have densities. Indeed,  $\forall x \in X, Q(x, \cdot) \ll N_X$  where  $N_X$  denotes the counting measure on  $X$ , with density  $x' \mapsto \mathbb{P}_{xx'}$ . In addition,  $\forall x \in X, G(x, \cdot) \ll \lambda$  with density  $y \mapsto f(y|x)$ .

It follows that the transition kernel  $T((x, y), \cdot)$  has density  $(x', y') \mapsto t((x, y), (x', y')) = p(x, x') f(y'|x')$ .

By a standard result on Markov chains, the joint distribution of  $(X_0, Y_0, \dots, Y_t, Y_t)$  is given by the equality

$$\begin{aligned} E(g(X_0, Y_0, \dots, Y_t, Y_t)) &= \int g(x_0, y_0, \dots, x_t, y_t) \prod_{j=0}^{t-1} T((x_j, y_j), d(x_{j+1}, y_{j+1})) \nu \otimes G(d(x_0, y_0)) \\ &= \int \int \int g(x_0, y_0, \dots, x_t, y_t) \prod_{j=0}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=0}^t f(y_j|x_j) d\lambda^{\otimes t+1}(y_{0:t}) dN_X^{\otimes t}(x_{1:t}) d\nu(x_0) \end{aligned}$$

for every bounded measurable function  $g$ . Consequently, the joint distribution of  $(X_1, Y_1, \dots, Y_t, Y_t)$  is given by

$$E(g(X_1, Y_1, \dots, X_t, Y_t)) = \int \int g(x_1, y_1, \dots, x_t, y_t) \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) d\lambda^{\otimes t}(y_{1:t}) dN_X^{\otimes t}(x_{1:t})$$

where  $p(x_1) = \int \int \mathbb{P}_{x_0 x_1} f(y_0|x_0) d\lambda(y_0) d\nu(x_0) = \int \mathbb{P}_{x_0 x_1} d\nu(x_0)$ .

This implies that the joint density of  $(X_1, Y_1, \dots, Y_t, Y_t)$  with respect to the measure  $\lambda^{\otimes t} \otimes N_X^{\otimes t}$  is

$$(x_1, y_1, \dots, x_t, y_t) \mapsto \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1)$$

We will use this result repeatedly in the problem.

- (a) We are asked to compute the conditional densities of  $X_1, \dots, X_t$  given  $Y_{1:t} = y_{1:t}$  and  $X_1, \dots, X_t$  given  $Y_{1:t-1} = y_{1:t-1}$ . It suffices to compute the ratio of the joint density and the relevant marginal density.

$$\begin{aligned} p(x_{1:t}|y_{1:t}) &= \frac{\prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1)}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\ &= \frac{\prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1)}{\prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\ &\quad \times \frac{1}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\ &\quad \times \frac{1}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\ &= f(y_t|x_t) p(x_{1:t}|y_{1:t-1}) \frac{1}{p(y_t|y_{1:t-1})} \end{aligned}$$

The red term is the joint density of  $(X_{1:t}, Y_{1:t-1})$  obtained by marginalizing with respect to  $Y_t$ , where we used  $\int f(y_t|x_t) d\lambda(y_t) = 1$ . The blue term is the joint density of  $Y_{1:t-1}$  obtained by marginalizing

with respect to  $(X_{1:t}, Y_t)$ .

$$\begin{aligned}
p(x_{1:t}|y_{1:t-1}) &= \frac{\prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1)}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\
&= \mathbb{P}_{x_{t-1} x_t} \frac{\prod_{j=1}^{t-2} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1)}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\
&= \mathbb{P}_{x_{t-1} x_t} \frac{p(x_{1:t-1}, y_{1:t-1})}{p(y_{1:t-1})} = \mathbb{P}_{x_{t-1} x_t} p(x_{1:t-1}|y_{1:t-1})
\end{aligned}$$

(b) Combining the first actualization equation with the second yields

$$p(x_{1:t}|y_{1:t}) = \frac{f(y_t|x_t) \mathbb{P}_{x_{t-1} x_t}}{p(y_t|y_{1:t-1})} p(x_{1:t-1}|y_{1:t-1})$$

To compute the filtering density, one needs to get a hold of  $p(x_{1:t-1}|y_{1:t-1})$  and  $p(y_t|y_{1:t-1})$ . The second quantity is linked to the first one in the following way:

$$\begin{aligned}
p(y_t|y_{1:t-1}) &= \frac{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} \\
&= \int \mathbb{P}_{x_{t-1} x_t} f(y_t|x_t) \frac{\prod_{j=1}^{t-2} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1)}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t})} dN_X^{\otimes t}(x_{1:t}) \\
&= \int \mathbb{P}_{x_{t-1} x_t} f(y_t|x_t) p(x_{1:t-1}|y_{1:t-1}) dN_X^{\otimes t}(x_{1:t}) \\
&= \sum_{x_1, \dots, x_t \in X^t} \mathbb{P}_{x_{t-1} x_t} f(y_t|x_t) p(x_{1:t-1}|y_{1:t-1})
\end{aligned}$$

As a result, computing  $p(y_t|y_{1:t-1})$  has the same complexity as computing  $p(x_{1:t}|y_{1:t})$ .

(c) From the actualization equations we already know  $p(x_{1:t}|y_{1:t-1})$ .  $p(x_t|y_{1:t-1})$  can then be computed by a simple marginalization:

$$\begin{aligned}
p(x_t|y_{1:t-1}) &= \int p(x_t|y_{1:t-1}) dN_X^{\otimes t-1}(x_{1:t-1}) = \int \mathbb{P}_{x_{t-1} x_t} p(x_{1:t-1}|y_{1:t-1}) dN_X^{\otimes t-1}(x_{1:t-1}) \\
&= \sum_{x_{1:t-1} \in X^{t-1}} \mathbb{P}_{x_{t-1} x_t} p(x_{1:t-1}|y_{1:t-1})
\end{aligned}$$

## Problem 14.5

(a) • Computing  $\alpha_1(i)$  boils down to finding the joint density of  $(X_1, Y_1)$ . Marginalizing like before,

$$\begin{aligned}
p(y_1, x_1) &= p(x_1) f(y_1|x_1) \int \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \underbrace{\prod_{j=2}^t f(y_j|x_j) d\lambda^{\otimes t-1}(y_{2:t})}_{=1} dN_X^{\otimes t-1}(x_{2:t}) \\
&= p(x_1) f(y_1|x_1) \int \dots \underbrace{\int \mathbb{P}_{x_{t-1} x_t} dN_X(x_t) \mathbb{P}_{x_{t-2} x_{t-1}} dN_X(x_{t-1}) \dots dN_X(x_2)}_{=1} \\
&= p(x_1) f(y_1|x_1)
\end{aligned}$$

Hence  $\alpha_1(i) = p(y_1, i) = p(i) f(y_1|i) = \pi_i f(y_1|X_1 = i)$  where the last equality enforces notations used in the statement of the problem.

- The recursion for  $\alpha_{t+1}(j)$  follows from computing the joint density of  $(X_{t+1}, Y_{1:t+1})$ :

$$\begin{aligned}
p(y_{1:t+1}, x_{t+1}) &= f(y_{t+1}|x_{t+1}) \int \prod_{j=1}^t \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t}(x_{1:t}) \\
&= f(y_{t+1}|x_{t+1}) \int \mathbb{P}_{x_t x_{t+1}} \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t-1}(x_{1:t-1}) dN_X(x_t) \\
&= f(y_{t+1}|x_{t+1}) \int \mathbb{P}_{x_t x_{t+1}} p(y_{1:t}, x_t) dN_X(x_t) \\
&= f(y_{t+1}|x_{t+1}) \sum_{x_t \in X} \mathbb{P}_{x_t x_{t+1}} \alpha_t(x_t)
\end{aligned}$$

Hence

$$\alpha_{t+1}(j) = f(y_{t+1}|j) \sum_{x_t \in X} \mathbb{P}_{x_t j} \alpha_t(x_t) = f(y_{t+1}|X_{t+1} = j) \sum_{i=1}^{\kappa} \mathbb{P}_{ij} \alpha_t(i)$$

where the last equality enforces notations used in the statement of the problem.

- **There is a mistake in the statement of the problem (2nd edition of the book).**

$\beta_t(i)$  should be defined as  $p(y_{t+1:T}|x_t = i)$  (the conditioning sign is missing). Computing the ratio between the joint and marginal densities yields

$$\begin{aligned}
p(y_{t+1:T}|x_t) &= \frac{\int \prod_{j=1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^T f(y_j|x_j) p(x_1) dN_X^{\otimes T-1}(x_{1:T \setminus \{t\}}) d\lambda^{\otimes t}(y_{1:t})}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t-1}(x_{1:t-1}) d\lambda^{\otimes t}(y_{1:t})} \\
&= \frac{\int \prod_{j=1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+1}^T f(y_j|x_j) p(x_1) dN_X^{\otimes T-1}(x_{1:T \setminus \{t\}})}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} p(x_1) dN_X^{\otimes t-1}(x_{1:t-1})} \\
&= \frac{\int \prod_{j=1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+1}^T f(y_j|x_j) p(x_1) dN_X^{\otimes T-1}(x_{1:T \setminus \{t\}})}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} p(x_1) dN_X^{\otimes t-1}(x_{1:t-1})} \\
&= \frac{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} p(x_1) dN_X^{\otimes t-1}(x_{1:t-1}) \cdot \int \prod_{j=t}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+1}^T f(y_j|x_j) dN_X^{\otimes T-t}(x_{t+1:T})}{\int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} p(x_1) dN_X^{\otimes t-1}(x_{1:t-1})} \\
&= \int \prod_{j=t}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+1}^T f(y_j|x_j) dN_X^{\otimes T-t}(x_{t+1:T}) \\
&= \int \mathbb{P}_{x_t x_{t+1}} f(y_{t+1}|x_{t+1}) \int \prod_{j=t+1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+2}^T f(y_j|x_j) dN_X^{\otimes T-t-1}(x_{t+2:T}) dN_X(x_{t+1})
\end{aligned}$$

The red term is identical to the blue one, except that  $t$  is replaced by  $t+1$ . Since the blue term is  $p(y_{t+1:T}|x_t)$ , the red one is  $p(y_{t+2:T}|x_{t+1})$ , hence

$$\begin{aligned}
p(y_{t+1:T}|x_t) &= \int \mathbb{P}_{x_t x_{t+1}} f(y_{t+1}|x_{t+1}) p(y_{t+2:T}|x_{t+1}) dN_X(x_{t+1}) \\
&= \sum_{x_{t+1} \in X} \mathbb{P}_{x_t x_{t+1}} f(y_{t+1}|x_{t+1}) p(y_{t+2:T}|x_{t+1})
\end{aligned}$$

Consequently,

$$\begin{aligned}
\beta_t(i) &= \sum_{x_{t+1} \in X} \mathbb{P}_{i x_{t+1}} f(y_{t+1}|x_{t+1}) \beta_{t+1}(x_{t+1}) = \sum_{j=1}^{\kappa} \mathbb{P}_{ij} f(y_{t+1}|j) \beta_{t+1}(j) \\
&= \sum_{j=1}^{\kappa} \mathbb{P}_{ij} f(y_{t+1}|X_{t+1} = j) \beta_{t+1}(j)
\end{aligned}$$

where the last equality enforces notations used in the statement of the problem.

- Marginalizing yields

$$\begin{aligned}
p(x_t, y_{1:T}) &= \int \prod_{j=1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^T f(y_j | x_j) p(x_1) dN_X^{\otimes T-1}(x_{1:T \setminus \{t\}}) \\
&= \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j | x_j) p(x_1) dN_X^{\otimes t-1}(x_{1:t-1}) \\
&\quad \times \int \prod_{j=t}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+1}^T f(y_j | x_j) dN_X^{\otimes T-t}(x_{t+1:T}) \\
&= p(x_t, y_{1:t}) p(y_{t+1:T} | x_t)
\end{aligned}$$

The last equality follows from identification of each of the integrals: the first one is an obvious marginalization that yields  $p(x_t, y_{1:t})$  and the second is the blue term computed a few moments ago, and we proved it is equal to  $p(y_{t+1:T} | x_t)$ . Besides,

$$p(y_{1:T}) = \int p(x_t, y_{1:T}) dN_X(x_t) = \int p(x_t, y_{1:t}) p(y_{t+1:T} | x_t) dN_X(x_t)$$

Hence

$$p(x_t | y_{1:T}) = \frac{p(x_t, y_{1:t}) p(y_{t+1:T} | x_t)}{\int p(x_t, y_{1:t}) p(y_{t+1:T} | x_t) dN_X(x_t)} = \frac{\alpha_t(x_t) \beta_t(x_t)}{\sum_{x_t \in X} \alpha_t(x_t) \beta_t(x_t)}$$

and finally

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^{\kappa} \alpha_t(j) \beta_t(j)}$$

- (b) Let us assume first that all the  $\alpha_t(i)$  and  $\beta_t(i)$  have already been computed and stored in memory. We begin with a naive evaluation: for fixed  $t$  and  $i$ , computing  $\gamma_t(i)$  requires computing  $\sum_{j=1}^{\kappa} \alpha_t(j) \beta_t(j)$ , which takes  $O(\kappa)$ , hence a global time complexity of  $O(T\kappa^2)$ .

However, this is clearly inefficient since  $\sum_{j=1}^{\kappa} \alpha_t(j) \beta_t(j)$  may be only evaluated once and stored in memory. For a given  $t$ , we compute all the  $\alpha_t(i) \beta_t(i)$ , then the sum  $\sum_{j=1}^{\kappa} \alpha_t(j) \beta_t(j)$  and finally the ratios, yielding all the  $\gamma_t(i)$  with a time complexity of  $O(\kappa + \kappa + \kappa) = O(\kappa)$ , hence a global complexity of  $O(T\kappa)$ .

For a more realistic evaluation we assume that the  $\alpha_t(i)$  and  $\beta_t(i)$  are not known and need to be computed beforehand. Since recursive programming results in a lot of repeated computations, it is inefficient and we turn to dynamic programming instead. Computing all the  $\alpha_1(i)$  takes  $O(\kappa)$ . By storing  $\alpha_t$  in memory, evaluating  $\alpha_{t+1}(j)$  for each  $j$  is  $O(\kappa)$ , hence computing  $\alpha_{t+1}$  takes  $O(\kappa^2)$ , yielding a total time complexity of  $O(\kappa) + \sum_{t=2}^T O(\kappa^2) = O(T\kappa^2)$  to compute  $\alpha$ . In a similar fashion, the evaluation of  $\beta$  requires  $O(T\kappa^2)$  in time. Finally, computing  $\gamma$  from scratch has complexity  $O(T\kappa^2) + O(T\kappa^2) + O(T\kappa) = O(T\kappa^2)$ .

- (c) Let us compute the joint density of  $(X_t, X_{t+1}, Y_{1:T})$  by marginalizing:

$$\begin{aligned}
p(x_t, x_{t+1}, y_{1:T}) &= \int \prod_{j=1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^T f(y_j | x_j) p(x_1) dN_X^{\otimes T-2}(x_{1:T \setminus \{t, t+1\}}) \\
&= \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j | x_j) p(x_1) dN_X^{\otimes t-1}(x_{1:t-1}) \\
&\quad \times \mathbb{P}_{x_t x_{t+1}} f(y_{t+1} | x_{t+1}) \\
&\quad \times \int \prod_{j=t+1}^{T-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=t+2}^T f(y_j | x_j) dN_X^{\otimes T-t-1}(x_{t+2:T}) \\
&= p(y_{1:t}, x_t) \mathbb{P}_{x_t x_{t+1}} f(y_{t+1} | x_{t+1}) p(y_{t+2:T} | x_{t+1})
\end{aligned}$$

In the second equality, the first integral is a straight marginalization that yields  $p(y_{1:t}, x_t)$  and the last integral is the red term computed in (a), which was shown to be  $p(y_{t+2:T} | x_{t+1})$ . Therefore,

$$p(x_t, x_{t+1} | y_{1:T}) = \frac{p(y_{1:t}, x_t) \mathbb{P}_{x_t x_{t+1}} f(y_{t+1} | x_{t+1}) p(y_{t+2:T} | x_{t+1})}{\int p(y_{1:t}, x_t) \mathbb{P}_{x_t x_{t+1}} f(y_{t+1} | x_{t+1}) p(y_{t+2:T} | x_{t+1}) dN_X^{\otimes 2}(x_{t:t+1})}$$

Hence

$$\begin{aligned}\xi_t(i, j) &= p(i, j|y_{1:T}) = \frac{p(y_{1:t}, i) \mathbb{P}_{ij} f(y_{t+1}|j) p(y_{t+2:T}|j)}{\sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} p(y_{1:t}, i) \mathbb{P}_{ij} f(y_{t+1}|j) p(y_{t+2:T}|j)} \\ &= \frac{\alpha_t(i) \mathbb{P}_{ij} f(y_{t+1}|j) \beta_{t+1}(j)}{\sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} \alpha_t(i) \mathbb{P}_{ij} f(y_{t+1}|j) \beta_{t+1}(j)}\end{aligned}$$

Computing  $\alpha$  and  $\beta$  has complexity  $O(T\kappa^2)$  as shown before. For a fixed  $t$ , we compute all the numerators in  $O(\kappa^2)$ , then evaluate the sum  $\sum_{i=1}^{\kappa} \sum_{j=1}^{\kappa} \alpha_t(i) \mathbb{P}_{ij} f(y_{t+1}|j) \beta_{t+1}(j)$  in  $O(\kappa^2)$  and finally compute all the ratios also in  $O(\kappa^2)$ , yielding a total time complexity of  $O(T(\kappa^2 + \kappa^2 + \kappa^2)) = O(T\kappa^2)$ .

(d) Let  $\alpha'_t(i) = \frac{\alpha_t(i)}{c_t}$  and note that

$$\frac{\alpha'_t(i) \beta_t(i)}{\sum_{j=1}^{\kappa} \alpha'_t(j) \beta_t(j)} = \frac{\frac{\alpha_t(i)}{c_t} \beta_t(i)}{\sum_{j=1}^{\kappa} \frac{\alpha_t(j)}{c_t} \beta_t(j)} = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^{\kappa} \alpha_t(j) \beta_t(j)} = \gamma_t(i)$$

Hence the equalities for  $\gamma$  are preserved.

## Problem 14.6

(a) For  $s \geq 3$ ,

$$\begin{aligned}p(x_s, x_{s-1}, y_{1:t}) &= \int \prod_{j=1}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t-2}(x_{1:t} \setminus \{s-1, s\}) \\ &= \int \prod_{j=1}^{s-2} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{s-1} f(y_j|x_j) p(x_1) dN_X^{\otimes s-2}(x_{1:s-2}) \\ &\quad \times \mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \\ &\quad \times \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t}) \\ &= p(x_{s-1}, y_{1:s-1}) \\ &\quad \times \mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \\ &\quad \times \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t}) \quad (*)\end{aligned}$$

Since  $p(x_{s-1}, y_{1:s-1})$  does not depend on  $x_s$ , it vanishes when computing the conditional density:

$$p(x_s|x_{s-1}, y_{1:t}) = \frac{\mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t})}{\int \mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t}) dN_X(x_s)}$$

A similar computation yields

$$\begin{aligned}p(x_s, x_{s-1}, y_{s:t}) &= p(x_{s-1}) \\ &\quad \times \mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \\ &\quad \times \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t})\end{aligned}$$

$p(x_{s-1})$  vanishes and we get

$$p(x_s|x_{s-1}, y_{s:t}) = \frac{\mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t})}{\int \mathbb{P}_{x_{s-1} x_s} f(y_s|x_s) \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t}) dN_X(x_s)} = p(x_s|x_{s-1}, y_{1:t})$$

For  $s = 2$  the equality is proved in a similar fashion.

(b) • Note that

$$\begin{aligned} p(x_t, x_{t-1}, y_{1:t}) &= \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) \int \prod_{j=1}^{t-2} \mathbb{P}_{x_j x_{j+1}} \prod_{j=1}^{t-1} f(y_j|x_j) p(x_1) dN_X^{\otimes t-2}(x_{1:t-2}) \\ &= \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) p(x_{t-1}, y_{1:t-1}) \end{aligned}$$

Hence  $p(x_t|x_{t-1}, y_{1:t}) = \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) \frac{p(x_{t-1}, y_{1:t-1})}{p(x_{t-1}, y_{1:t})}$  and we define an unnormalized version of this conditional density as  $p_t^*(x_t|x_{t-1}, y_{1:t}) := \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t)$ .

• Let  $2 \leq s \leq t-1$ . From equation (\*) in (a) we have

$$\begin{aligned} p(x_s, x_{s-1}, y_{1:t}) &= p(x_{s-1}, y_{1:s-1}) \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \int \prod_{j=s}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=s+1}^t f(y_j|x_j) dN_X^{\otimes t-s}(x_{s+1:t}) \\ &= p(x_{s-1}, y_{1:s-1}) \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \int \frac{p(x_{s+1}, x_s, y_{1:t})}{p(x_s, y_{1:s})} dN_X(x_{s+1}) \\ &= \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \int p(x_{s+1}|x_s, y_{1:t}) \frac{p(x_s, y_{1:t})}{p(x_s, y_{1:s})} dN_X(x_{s+1}) p(x_{s-1}, y_{1:s-1}) \end{aligned}$$

Hence

$$p(x_s|x_{s-1}, y_{1:t}) = \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \sum_{x_{s+1} \in X} \left[ p(x_{s+1}|x_s, y_{1:t}) \frac{p(x_s, y_{1:t})}{p(x_s, y_{1:s})} \right] \frac{p(x_{s-1}, y_{1:s-1})}{p(x_{s-1}, y_{1:t})}$$

and we let

$$p_s^*(x_s|x_{s-1}, y_{1:t}) := \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \sum_{x_{s+1} \in X} \left[ p(x_{s+1}|x_s, y_{1:t}) \frac{p(x_s, y_{1:t})}{p(x_s, y_{1:s})} \right]$$

• Note that

$$\begin{aligned} p(x_1, y_{1:t}) &= p(x_1) f(y_1|x_1) \int \mathbb{P}_{x_1x_2} f(y_2|x_2) \prod_{j=2}^{t-1} \mathbb{P}_{x_j x_{j+1}} \prod_{j=3}^t f(y_j|x_j) p(x_1) dN_X^{\otimes t-2}(x_{1:t-2}) \\ &= p(x_1) f(y_1|x_1) \int \frac{p(x_2, x_1, y_{1:t})}{p(x_1, y_1)} dN_X(x_2) \end{aligned}$$

where we used (\*) in the second equality.

Hence  $p(x_1|y_{1:t}) = p(x_1) f(y_1|x_1) \sum_{x_2 \in X} \left[ p(x_2|x_1, y_{1:t}) \frac{p(x_1, y_{1:t})}{p(x_1, y_1)} \right] \frac{1}{p(y_{1:t})}$  (\*\*) and we let

$$p_1^*(x_1|y_{1:t}) := p(x_1) f(y_1|x_1) \sum_{x_2 \in X} \left[ p(x_2|x_1, y_{1:t}) \frac{p(x_1, y_{1:t})}{p(x_1, y_1)} \right]$$

• Putting all these relations together we have the following recursion

$$\begin{cases} p_t^*(x_t|x_{t-1}, y_{1:t}) = \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) \\ p_s^*(x_s|x_{s-1}, y_{1:t}) = \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \sum_{x_{s+1} \in X} p_{s+1}^*(x_{s+1}|x_s, y_{1:t}) & 2 \leq s \leq t-1 \\ p_1^*(x_1|y_{1:t}) = p(x_1) f(y_1|x_1) \sum_{x_2 \in X} p_2^*(x_2|x_1, y_{1:t}) \end{cases}$$

We resort to dynamic programming again. For each  $1 \leq s \leq t-1$  and each  $x_s \in \{1, \dots, \kappa\}$ , we must compute a sum of  $\kappa$  terms, hence a global time complexity of  $O(t\kappa^2)$ .

(c) It is straightforward to prove that  $p(x_{1:t}|y_{1:t}) = p(x_1|y_{1:t}) \prod_{s=1}^t p(x_s|x_{s-1}, y_{1:t})$ .

Once the  $p_s^*$  have been computed, the normalized versions  $p(x_s|x_{s-1}, y_{1:t})$  can be retrieved without increasing the computational complexity. Given the factorization of the joint distribution, we can sample from it by sequentially generating  $x_1$  from  $p(x_1|y_{1:t})$ ,  $x_2$  from  $p(x_2|x_1, y_{1:t})$  and so on. This process takes  $O(t\kappa^2)$  in time.

(d) Using the formula for  $p(x_{1:t}|y_{1:t})$  proved a bit later in (e), we have for  $s \geq 3$

$$p(x_{1:s}|y_{1:s}) = \mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) p(x_{1:s-1}|y_{1:s-1}) \frac{p(y_{1:s-1})}{p(y_{1:s})}$$

Hence

$$\begin{aligned} \arg \max_{x_{1:t}} p(x_{1:t}|y_{1:t}) &= \arg \max_{x_t} \left[ \arg \max_{x_{1:t-1}} p(x_{1:t}|y_{1:t}) \right] \\ &= \arg \max_{x_t} \left[ \arg \max_{x_{1:t-1}} \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) p(x_{1:t-1}|y_{1:t-1}) \right] \\ &= \arg \max_{x_t} \left[ \arg \max_{x_{t-1}} \left( \mathbb{P}_{x_{t-1}x_t} f(y_t|x_t) \arg \max_{x_{1:t-2}} p(x_{1:t-1}|y_{1:t-1}) \right) \right] \end{aligned}$$

The blue terms show that computing the maximizer is amenable to dynamic programming. One starts by computing

$$\arg \max_{x_1} p(x_{1:2}|y_{1:2}) = \arg \max_{x_1} \mathbb{P}_{x_1x_2} f(y_2|x_2) p(x_1)$$

and then recurses all the way up to  $\arg \max_{x_{1:t}} p(x_{1:t}|y_{1:t})$ .

(e) Equation (\*\*) in (b) rewrites further as

$$p(x_1|y_{1:t}) = \frac{p_1^*(x_1|y_{1:t})}{p(y_{1:t})}$$

Hence  $p(y_{1:t}) = \sum_{x_1 \in X} p_1^*(x_1|y_{1:t})$ , which yields a representation of the observed likelihood.

As an aside,

$$\begin{aligned} p(x_{1:t}|y_{1:t}) &= \frac{p(x_1) f(y_1|x_1) \sum_{x_2 \in X} p_2^*(x_2|x_1, y_{1:t})}{\sum_{x_1 \in X} p_1^*(x_1|y_{1:t})} \prod_{s=1}^{t-1} \frac{\mathbb{P}_{x_{s-1}x_s} f(y_s|x_s) \sum_{x_{s+1} \in X} p_{s+1}^*(x_{s+1}|x_s, y_{1:t})}{\sum_{x_s \in X} p_s^*(x_s|x_{s-1}, y_{1:t})} \\ &\quad \times \frac{\mathbb{P}_{x_{t-1}x_t} f(y_t|x_t)}{\sum_{x_t \in X} p_t^*(x_t|x_{t-1}, y_{1:t})} \\ &= \frac{\prod_{j=1}^{t-1} \mathbb{P}_{x_jx_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1)}{\sum_{x_1 \in X} p_1^*(x_1|y_{1:t})} \\ &= \frac{\prod_{j=1}^{t-1} \mathbb{P}_{x_jx_{j+1}} \prod_{j=1}^t f(y_j|x_j) p(x_1)}{p(y_{1:t})} \end{aligned}$$

(f) As stated in (c), computing the conditional likelihood  $p(x_{1:T}|y_{1:T})$  has complexity  $O(T\kappa^2)$  in time.

## References

- [1] Tobias Ryden Olivier Cappé, Eric Moulines. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer, 2005.