Exercise 1. *Kalman filtering*

We here consider filtering for hidden Markov models with Gaussian transition and emission distributions. For simplicity, we assume one-dimensional hidden variables and observables. We denote the probability density function of a Gaussian random variable x with mean μ and variance σ^2 by $\mathcal{N}(x|\mu, \sigma^2)$,

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{1}$$

The transition and emission distributions are assumed to be

$$p(h_s|h_{s-1}) = \mathcal{N}(h_s|A_s h_{s-1}, B_s^2) \tag{2}$$

$$p(v_s|h_s) = \mathcal{N}(v_s|C_s h_s, D_s^2). \tag{3}$$

The transition and emission distributions correspond to the following update and observation equations

$$h_s = A_s h_{s-1} + B_s \xi_s, \tag{4}$$

$$v_s = C_s h_s + D_s \eta_s, \tag{5}$$

where ξ_s and η_s are independent standard normal random variables, e.g. $\xi_s \sim \mathcal{N}(\xi_s|0,1)$ — independent from each other and from the h_s and v_s. The equations mean that h_s is obtained by scaling h_{s-1} and by adding noise with variance B_s^2. The observed value v_s is obtained by scaling the hidden h_s and by corrupting it with Gaussian observation noise of variance D_s^2.

The distribution $p(h_1)$ is assumed Gaussian with known parameters. The A_s, B_s, C_s, D_s are also assumed known.

(a) *Show that*

$$\int \mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|Ax, B^2)\mathrm{d}x \propto \mathcal{N}(y|A\mu, A^2\sigma^2 + B^2) \tag{6}$$

[While this result can be obtained by direct integration, an approach that avoids this is as follows: First note that $\mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|Ax, B^2)$ is proportional to the joint pdf of x and y. We can thus consider the integral to correspond to the computation of the marginal of y from the joint. Using the equivalence of Equations (2)-(3) and (4)-(5), and the fact that the weighted sum of two Gaussian random variables is a Gaussian random variable then allows one to obtain the result.]

Solution. We follow the procedure outlined above. The two Gaussian densities correspond to the equations

$$x = \mu + \sigma\xi \tag{S.1}$$

$$y = Ax + B\eta \tag{S.2}$$

where ξ and η are independent standard normal random variables. The mean of y is

$$\begin{aligned} \mathbb{E}(y) &= A\mathbb{E}(x) + B\mathbb{E}(\eta) && \text{(S.3)} \\ &= A\mu && \text{(S.4)} \end{aligned}$$

where we have use the linearity of expectation and $\mathbb{E}(\eta) = 0$. The variance of y is

$$\begin{aligned} \mathbb{V}(y) &= \mathbb{V}(Ax) + \mathbb{V}(B\eta) \qquad \text{(since } x \text{ and } \eta \text{ are independent)} && \text{(S.5)} \\ &= A^2\mathbb{V}(x) + B^2\mathbb{V}(\eta) \qquad \text{(by properties of the variance)} && \text{(S.6)} \\ &= A^2\sigma^2 + B^2 && \text{(S.7)} \end{aligned}$$

Since y is the (weighted) sum of two Gaussians, it is Gaussian itself, and hence its distribution is completely defined by its mean and variance, so that

$$y \sim \mathcal{N}(y|A\mu, A^2\sigma^2 + B^2). \tag{S.8}$$

Now, the product $\mathcal{N}(x|\mu, \sigma^2)\mathcal{N}(y|Ax, B^2)$ is proportional to the joint pdf of x and y, so that the integral can be considered to correspond to the marginalisation of x, and hence its result is proportional to the density of y, which is $\mathcal{N}(y|A\mu, A^2\sigma^2 + B^2)$.

(b) *Show that*

$$\mathcal{N}(x|m_1, \sigma_1^2)\mathcal{N}(x|m_2, \sigma_2^2) \propto \mathcal{N}(x|m_3, \sigma_3^2) \tag{7}$$

where

$$\sigma_3^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{8}$$

$$m_3 = \sigma_3^2\left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2}\right) = m_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(m_2 - m_1) \tag{9}$$

Solution. We show the result using a classical technique called "completing the square", see e.g. https://en.wikipedia.org/wiki/Completing_the_square.
We work in the (negative) log-domain and use that

$$\begin{aligned}
-\log\left[\mathcal{N}(x|m, \sigma^2)\right] &= \frac{(x-m)^2}{2\sigma^2} + \text{const} &\text{(S.9)}\\
&= \frac{x^2}{2\sigma^2} - x\frac{m}{\sigma^2} + \frac{m^2}{2\sigma^2} + \text{const} &\text{(S.10)}\\
&= \frac{x^2}{2\sigma^2} - x\frac{m}{\sigma^2} + \text{const} &\text{(S.11)}
\end{aligned}$$

where const indicates terms not depending on x. We thus obtain

$$\begin{aligned}
-\log\left[\mathcal{N}(x|m_1, \sigma_1^2)\mathcal{N}(x|m_2, \sigma_2^2)\right] &= -\log\left[\mathcal{N}(x|m_1, \sigma_1^2)\right] - \log\left[\mathcal{N}(x|m_2, \sigma_2^2)\right] &\text{(S.12)}\\
&= \frac{(x-m_1)^2}{2\sigma_1^2} + \frac{(x-m_2)^2}{2\sigma_2^2} + \text{const} &\text{(S.13)}\\
&= \frac{x^2}{2\sigma_1^2} - x\frac{m_1}{\sigma_1^2} + \frac{x^2}{2\sigma_2^2} - x\frac{m_2}{\sigma_2^2} + \text{const} &\text{(S.14)}\\
&= \frac{x^2}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - x\left(\frac{m_1}{\sigma_1^2} + \frac{m_1}{\sigma_1^2}\right) + \text{const} &\text{(S.15)}\\
&= \frac{x^2}{2\sigma_3^2} - \frac{x}{\sigma_3^2}\sigma_3^2\left(\frac{m_1}{\sigma_1^2} + \frac{m_1}{\sigma_1^2}\right) + \text{const}, &\text{(S.16)}
\end{aligned}$$

where

$$\frac{1}{\sigma_3^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}. \tag{S.17}$$

Comparison with (S.11) shows that we can further write

$$\frac{x^2}{2\sigma_3^2} - \frac{x}{\sigma_3^2}\sigma_3\left(\frac{m_1}{\sigma_1^2} + \frac{m_1}{\sigma_1^2}\right) = \frac{(x-m_3)^2}{2\sigma_3^2} + \text{const} \tag{S.18}$$

where

$$m_3 = \sigma_3 \left(\frac{m_1}{\sigma_1^2} + \frac{m_1}{\sigma_1^2} \right) \tag{S.19}$$

so that

$$-\log \left[\mathcal{N}(x|m_1, \sigma_1^2) \mathcal{N}(x|m_2, \sigma_2^2) \right] = \frac{(x - m_3)^2}{2\sigma_3^2} + \text{const} \tag{S.20}$$

and hence

$$\mathcal{N}(x|m_1, \sigma_1^2) \mathcal{N}(x|m_2, \sigma_2^2) \propto \mathcal{N}(x|m_3, \sigma_3^2). \tag{S.21}$$

Note that the identity

$$m_3 = \sigma_3^2 \left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2} \right) = m_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(m_2 - m_1) \tag{S.22}$$

is obtained as follows

$$\sigma_3^2 \left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2} \right) = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2} \right) \tag{S.23}$$

$$= m_1 \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} + m_2 \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{S.24}$$

$$= m_1 \left(1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) + m_2 \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{S.25}$$

$$= m_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(m_2 - m_1) \tag{S.26}$$

(c) *In the lecture, we showed that $p(h_t|v_{1:t}) \propto \alpha(h_t)$ where $\alpha(h_t)$ can be computed recursively via the "alpha-recursion"*

$$\alpha(h_1) = p(h_1) \cdot p(v_1|h_1) \qquad \alpha(h_s) = p(v_s|h_s) \sum_{h_{s-1}} p(h_s|h_{s-1}) \alpha(h_{s-1}). \tag{10}$$

We have also seen that the alpha-recursion corresponds to sum-product message passing with

$$\mu_{h_s \to \phi_{s+1}}(h_s) = \alpha(h_s) \qquad \mu_{\phi_s \to h_s}(h_s) = \sum_{h_{s-1}} p(h_s|h_{s-1}) \alpha(h_{s-1}) \tag{11}$$

and that $\mu_{\phi_s \to h_s}(h_s) \propto p(h_s|v_{1:s-1})$. For continuous random variables, the sum above becomes an integral so that

$$\alpha(h_s) = p(v_s|h_s) \mu_{\phi_s \to h_s}(h_s) \qquad \mu_{\phi_s \to h_s}(h_s) = \int p(h_s|h_{s-1}) \alpha(h_{s-1}) \mathrm{d}h_{s-1}. \tag{12}$$

For a Gaussian prior distribution for h_1 and Gaussian emission probability $p(v_1|h_1)$, $\alpha(h_1) = p(h_1) \cdot p(v_1|h_1) \propto p(h_1|v_1)$ is proportional to a Gaussian. We denote its mean by μ_1 and its variance by σ_1^2 so that

$$\alpha(h_1) \propto \mathcal{N}(h_1|\mu_1, \sigma_1^2). \tag{13}$$

Assuming $\alpha(h_{s-1}) \propto \mathcal{N}(h_{s-1}|\mu_{s-1}, \sigma_{s-1}^2)$ (which holds for $s = 2$), use Equation (6) to show that

$$\mu_{\phi_s \to h_s}(h_s) \propto \mathcal{N}(h_s|A_s \mu_{s-1}, P_s) \tag{14}$$

where

$$P_s = A_s^2 \sigma_{s-1}^2 + B_s^2. \tag{15}$$

Solution. We can set $\alpha(h_{s-1}) \propto \mathcal{N}(h_{s-1}|\mu_{s-1}, \sigma^2_{s-1})$. Since $p(h_s|h_{s-1})$ is Gaussian, see Equation (2), Equation (12) becomes

$$\mu_{\phi_s \to h_s}(h_s) = \int \mathcal{N}(h_s|A_s h_{s-1}, B_s^2)\mathcal{N}(h_{s-1}|\mu_{s-1}, \sigma^2_{s-1})\mathrm{d}h_{s-1}. \tag{S.27}$$

Equation (6) with $x \equiv h_{s-1}$ and $y \equiv h_s$ yields the desired result,

$$\mu_{\phi_s \to h_s}(h_s) = \mathcal{N}(h_s|A_s\mu_{s-1}, A_s^2\sigma^2_{s-1} + B_s^2). \tag{S.28}$$

With $\alpha(h_{s-1}) \propto p(h_{s-1}|v_{1:s-1})$ and $\mu_{\phi_s \to h_s}(h_s) \propto p(h_s|v_{1:s-1})$, we can understand the equation as follows: To compute the predictive mean of h_s given $v_{1:s-1}$, we forward propagate the mean of $h_{s-1}|v_{1:s-1}$ using the update equation (4). This gives the mean term $A_s\mu_{s-1}$. Since $h_{s-1}|v_{1:s-1}$ has variance σ^2_{s-1}, the variance of $h_s|v_{1:s-1}$ is given by $A_s^2\sigma^2_{s-1}$ plus an additional term, B_s^2, due to the noise in the forward propagation. This gives the variance term $A_s^2\sigma^2_{s-1} + B_s^2$. In the lecture, it was pointed out that $\mu_{\phi_s \to h_s}(h_s)$ is called the "prediction" step in the alpha-recursion. Indeed, we here compute the predictive distribution of h_s given $v_{1:s-1}$, which is the Gaussian in Equation (S.28).

(d) Use Equation (7) to show that

$$\alpha(h_s) \propto \mathcal{N}\left(h_s|\mu_s, \sigma_s^2\right) \tag{16}$$

where

$$\mu_s = A_s\mu_{s-1} + \frac{P_s C_s}{C_s^2 P_s + D_s^2}\left(v_s - C_s A_s \mu_{s-1}\right) \tag{17}$$

$$\sigma_s^2 = \frac{P_s D_s^2}{P_s C_s^2 + D_s^2} \tag{18}$$

Solution. Having computed $\mu_{\phi_s \to h_s}(h_s)$, the final step in the alpha-recursion is

$$\alpha(h_s) = p(v_s|h_s)\mu_{\phi_s \to h_s}(h_s) \tag{S.29}$$

With Equation (3) we obtain

$$\alpha(h_s) \propto \mathcal{N}(v_s|C_s h_s, D_s^2)\mathcal{N}(h_s|A_s\mu_{s-1}, P_s). \tag{S.30}$$

We further note that

$$\mathcal{N}(v_s|C_s h_s, D_s^2) \propto \mathcal{N}\left(h_s|C_s^{-1}v_s, \frac{D_s^2}{C_s^2}\right) \tag{S.31}$$

so that we can apply Equation (7) (with $m_1 = A\mu_{s-1}$, $\sigma_1^2 = P_s$)

$$\alpha(h_s) \propto \mathcal{N}\left(h_s|C_s^{-1}v_s, \frac{D_s^2}{C_s^2}\right)\mathcal{N}(h_s|A_s\mu_{s-1}, P_s) \tag{S.32}$$

$$\propto \mathcal{N}\left(h_s, \mu_s, \sigma_s^2\right) \tag{S.33}$$

with

$$\mu_s = A_s\mu_{s-1} + \frac{P_s}{P_s + \frac{D_s^2}{C_s^2}}\left(C_s^{-1}v_s - A_s\mu_{s-1}\right) \tag{S.34}$$

$$= A_s\mu_{s-1} + \frac{P_sC_s^2}{C_s^2P_s + D_s^2}\left(C_s^{-1}v_s - A_s\mu_{s-1}\right) \tag{S.35}$$

$$= A_s\mu_{s-1} + \frac{P_sC_s}{C_s^2P_s + D_s^2}\left(v_s - C_sA_s\mu_{s-1}\right) \tag{S.36}$$

$$\sigma_s^2 = \frac{P_s\frac{D_s^2}{C_s^2}}{P_s + \frac{D_s^2}{C_s^2}} \tag{S.37}$$

$$= \frac{P_sD_s^2}{P_sC_s^2 + D_s^2} \tag{S.38}$$

$$\tag{S.39}$$

(e) Show that $\alpha(h_s)$ can be re-written as

$$\alpha(h_s) \propto \mathcal{N}\left(h_s|\mu_s, \sigma_s^2\right) \tag{19}$$

where

$$\mu_s = A_s\mu_{s-1} + K_s\left(v_s - C_sA_s\mu_{s-1}\right) \tag{20}$$

$$\sigma_s^2 = (1 - K_sC_s)P_s \tag{21}$$

$$K_s = \frac{P_sC_s}{C_s^2P_s + D_s^2} \tag{22}$$

These are the Kalman filter equations and K_s is called the Kalman filter gain.

Solution. We start from

$$\mu_s = A_s\mu_{s-1} + \frac{P_sC_s}{C_s^2P_s + D_s^2}\left(v_s - C_sA_s\mu_{s-1}\right), \tag{S.40}$$

and see that

$$\frac{P_sC_s}{C_s^2P_s + D_s^2} = K_s \tag{S.41}$$

so that

$$\mu_s = A_s\mu_{s-1} + K_s\left(v_s - C_sA_s\mu_{s-1}\right). \tag{S.42}$$

For the variance σ_s^2, we have

$$\sigma_s^2 = \frac{P_sD_s^2}{P_sC_s^2 + D_s^2} \tag{S.43}$$

$$= \frac{D_s^2}{P_sC_s^2 + D_s^2}P_s \tag{S.44}$$

$$= \left(1 - \frac{P_sC_s^2}{P_sC_s^2 + D_s^2}\right)P_s \tag{S.45}$$

$$= (1 - K_sC_s)P_s, \tag{S.46}$$

which is the desired result.

The filtering result generalises to vector valued latents and visibles where the transition and emission distributions in (2) and (3) become

$$p(\mathbf{h}_s|\mathbf{h}_{s-1}) = \mathcal{N}(\mathbf{h}_s|\mathbf{A}\mathbf{h}_{s-1}, \mathbf{\Sigma}^h), \tag{S.47}$$
$$p(\mathbf{v}_s|\mathbf{h}_s) = \mathcal{N}(\mathbf{v}_s|\mathbf{C}_s\mathbf{h}_s, \mathbf{\Sigma}^v), \tag{S.48}$$

where $\mathcal{N}()$ denotes multivariate Gaussian pdfs, e.g.

$$\mathcal{N}(\mathbf{v}_s|\mathbf{C}_s\mathbf{h}_s, \mathbf{\Sigma}^v) = \frac{1}{|\det(2\pi\mathbf{\Sigma}^v)|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{v}_s - \mathbf{C}_s\mathbf{h}_s)^\top (\mathbf{\Sigma}^v)^{-1}(\mathbf{v}_s - \mathbf{C}_s\mathbf{h}_s)\right). \tag{S.49}$$

We then have

$$p(\mathbf{h}_t|\mathbf{v}_{1:t}) = \mathcal{N}(\mathbf{h}_t|\boldsymbol{\mu}_t, \mathbf{\Sigma}_t) \tag{S.50}$$

where the posterior mean and variance are recursively computed as

$$\boldsymbol{\mu}_s = \mathbf{A}_s\boldsymbol{\mu}_{s-1} + \mathbf{K}_s(\mathbf{v}_s - \mathbf{C}_s\mathbf{A}_s\boldsymbol{\mu}_{s-1}) \tag{S.51}$$
$$\mathbf{\Sigma}_s = (\mathbf{I} - \mathbf{K}_s\mathbf{C}_s)\mathbf{P}_s \tag{S.52}$$
$$\mathbf{P}_s = \mathbf{A}_s\mathbf{\Sigma}_{s-1}\mathbf{A}_s^\top + \mathbf{\Sigma}^h \tag{S.53}$$
$$\mathbf{K}_s = \mathbf{P}_s\mathbf{C}_s^\top \left(\mathbf{C}_s\mathbf{P}_s\mathbf{C}_s^\top + \mathbf{\Sigma}^v\right)^{-1} \tag{S.54}$$

and initialised with $\boldsymbol{\mu}_1$ and $\mathbf{\Sigma}_1$ equal to the mean and variance of $p(\mathbf{h}_1|\mathbf{v}_1)$. The matrix $\mathbf{K}_s$ is then called the Kalman gain matrix.

The Kalman filter is widely applicable, see e.g. https://en.wikipedia.org/wiki/Kalman_filter, and has played a role in now historic events such as the moon landing, see e.g. http://ieeexplore.ieee.org/document/5466132/

An example of the application of the Kalman filter to tracking is shown in Figure 1.

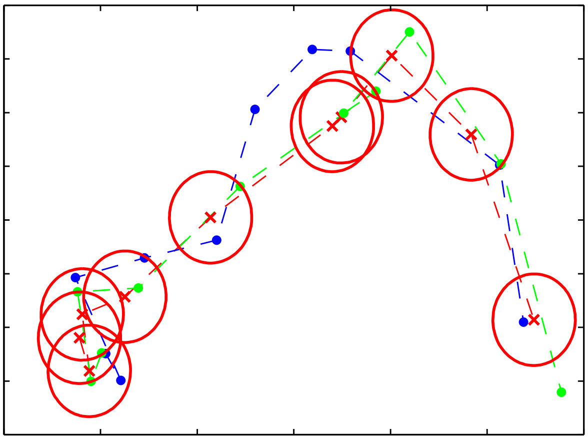


Figure 1: Kalman filtering for tracking of a moving object. The blue points indicate the true positions of the object in a two-dimensional space at successive time steps, the green points denote noisy measurements of the positions, and the red crosses indicate the means of the inferred posterior distributions of the positions obtained by running the Kalman filtering equations. The covariances of the inferred positions are indicated by the red ellipses, which correspond to contours having one standard deviation. (Bishop, Figure 13.22)

(f) *Explain Equation (20) in non-technical terms. What happens if the variance D_s^2 of the observation noise goes to zero?*

Solution. We have already seen that $A_s\mu_{s-1}$ is the predictive mean of h_s given $v_{1:s-1}$. The term $C_sA_s\mu_{s-1}$ is thus the predictive mean of v_s given the observations so far, $v_{1:s-1}$. The difference $v_s - C_sA_s\mu_{s-1}$ is thus the prediction error of observable. Since $\alpha(h_s)$ is proportional to $p(h_s|v_{1:s})$ and μ_s its mean, we thus see that the posterior mean of $h_s|v_{1:s}$ equals the posterior mean of $h_s|v_{1:s-1}$, $A_s\mu_{s-1}$, updated by the prediction error of the observable weighted by the Kalman gain.

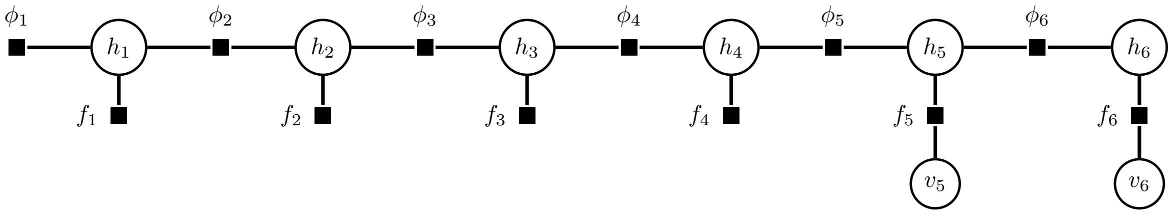
For $D_s^2 \to 0$, $K_s \to C_s^{-1}$ and

$$\begin{aligned}
\mu_s &= A_s\mu_{s-1} + K_s\left(v_s - C_sA_s\mu_{s-1}\right) && \text{(S.55)}\\
&= A_s\mu_{s-1} + C_s^{-1}\left(v_s - C_sA_s\mu_{s-1}\right) && \text{(S.56)}\\
&= A_s\mu_{s-1} + C_s^{-1}v_s - A_s\mu_{s-1} && \text{(S.57)}\\
&= C_s^{-1}v_s, && \text{(S.58)}
\end{aligned}$$

so that the posterior mean of $p(h_s|v_{1:s}$ is obtained by inverting the observation equation. Moreover, $\sigma_s^2 \to 0$, so that with zero observation noise, the value of h_s is known precisely.

Exercise 2. ***Hidden Markov model – beta-recursion***

We consider the following factor graph from the lecture on hidden Markov models.



The factor graph corresponds to the conditional pmf

$$p(h_1, \dots, h_6, v_5, v_6 \mid v_{1:4})$$

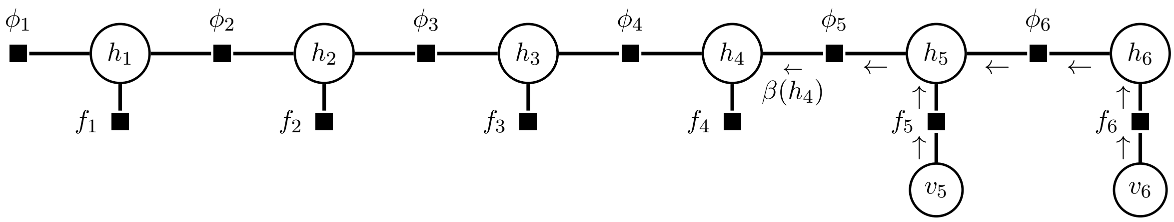
and the factors are defined as

$$\begin{aligned}
f_t(h_t) &= p(v_t|h_t) \;\; (t \le 4) & \qquad f_t(v_t, h_t) &= p(v_t|h_t) \;\; (t > 4) && (23)\\
\phi_1(h_1) &= p(h_1) & \qquad \phi_t(h_t, h_{t-1}) &= p(h_t|h_{t-1}) \;\; (t > 1) && (24)
\end{aligned}$$

We define $\beta(h_s) = \mu_{\phi_{s+1}\to h_s}(h_s)$, which is the message from a factor node "back" to a variable node.

(a) *Show that $\beta(h_4) = \mu_{\phi_5\to h_4}(h_4) = 1$.*

Solution. The arrows in the factor graph below show the messages that need to be computed for the computation of $\beta(h_4)$.



We start with the leaf variable v_6:

$$\mu_{v_6 \to f_6}(v_6) = 1 \tag{S.59}$$

$$\mu_{f_6 \to h_6}(h_6) = \sum_{v_6} f_6(v_6, h_6)\mu_{v_6 \to f_6}(v_6) \tag{S.60}$$

$$= \sum_{v_6} p(v_6|h_6) \cdot 1 \tag{S.61}$$

$$= 1 \quad \text{since (conditional) pmfs and pdfs are normalised} \tag{S.62}$$

The variable node h_6, having a single incoming message only, copies the message so that

$$\mu_{h_6 \to \phi_6}(h_6) = \beta(h_6) = 1. \tag{S.63}$$

For the next message, which corresponds to the elimination of h_6, we have:

$$\mu_{\phi_6 \to h_5}(h_5) = \sum_{h_6} \phi_6(h_6, h_5)\mu_{h_6 \to \phi_6}(h_6) \tag{S.64}$$

$$= \sum_{h_6} p(h_6|h_5) \cdot 1 \tag{S.65}$$

$$= 1 \quad \text{since (conditional) pmfs and pdfs are normalised.} \tag{S.66}$$

The same kind of calculations show that $\mu_{f_5 \to h_5} = 1$. It follows that

$$\mu_{h_5 \to \phi_5}(x_5) = \mu_{\phi_6 \to h_5}(h_5)\mu_{f_5 \to h_5} \tag{S.67}$$

$$= 1. \tag{S.68}$$

We thus obtain the desired result for $\beta(h_4) = \mu_{\phi_5 \to h_4}(h_4)$:

$$\mu_{\phi_5 \to h_4}(h_4) = \sum_{h_5} \phi_5(h_5, h_4)\mu_{h_5 \to \phi_5}(x_5) \tag{S.69}$$

$$= \sum_{x_5} p(h_5|h_4) \cdot 1 \tag{S.70}$$

$$= 1 \quad \text{since (conditional) pmfs and pdfs are normalised.} \tag{S.71}$$

(b) *Use sum-product message passing to show that the beta-recursion holds*

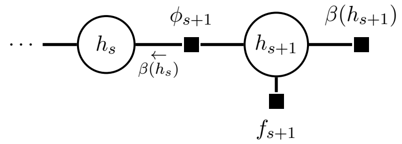
$$\beta(h_4) = 1 \tag{25}$$

$$\beta(h_s) = \sum_{h_{s+1}} p(h_{s+1}|h_s)p(v_{s+1}|h_{s+1})\beta(h_{s+1}) \quad (s < 4) \tag{26}$$

Solution. We defined $\beta(h_s)$ as the message $\mu_{\phi_{s+1} \to h_s}(h_s)$. We thus also have

$$\beta(h_{s+1}) = \mu_{\phi_{s+2} \to h_{s+1}}(h_{s+1}), \tag{S.72}$$

which is the effective factor for h_{s+1} if all variables in all sub-trees attached to ϕ_{s+2}, with exception of the sub-trees attached to h_{s+1}, are eliminated. This gives us the following fragment of a factor graph



Message passing tell us that

$$\beta(h_s) = \mu_{\phi_{s+1} \to h_s}(h_s) = \sum_{h_{s+1}} \phi_{s+1}(h_{s+1}, h_s) \mu_{h_{s+1} \to \phi_{s+1}}(h_{s+1}) \tag{S.73}$$

and that

$$\begin{aligned} \mu_{h_{s+1} \to \phi_{s+1}}(h_{s+1}) &= \mu_{f_{s+1} \to h_{s+1}}(h_{s+1}) \mu_{\beta(h_{s+1}) \to h_{s+1}}(h_{s+1}) && \text{(S.74)} \\ &= f_{s+1}(h_{s+1}) \beta(h_{s+1}), && \text{(S.75)} \\ & && \text{(S.76)} \end{aligned}$$

where for the last equation, we have used that f_{s+1} and $\beta(h_{s+1})$ are leaf factor nodes. We thus obtain

$$\beta(h_s) = \mu_{\phi_{s+1} \to h_s}(h_s) = \sum_{h_{s+1}} \phi_{s+1}(h_{s+1}, h_s) f_{s+1}(h_{s+1}) \beta(h_{s+1}). \tag{S.77}$$

Plugging in the definition of the factors gives

$$\beta(h_s) = \sum_{h_{s+1}} p(h_{s+1}|h_s) p(v_{s+1}|h_{s+1}) \beta(h_{s+1}), \tag{S.78}$$

which is the desired recursion. In our factor graph, the recursion is initialised with $\beta(h_4) = 1$.