The purpose of the tutorials is twofold: First, they help you better understand the lecture material. Secondly, they provide exam preparation material. You are not expected to complete all questions before the tutorial sessions. Start early and do as many as you have time for.

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.]

(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

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

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

(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

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

(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

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

$$\sigma_s^2 = \frac{P_sD_s^2}{P_sC_s^2 + D_s^2} \tag{18}$$

(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

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

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

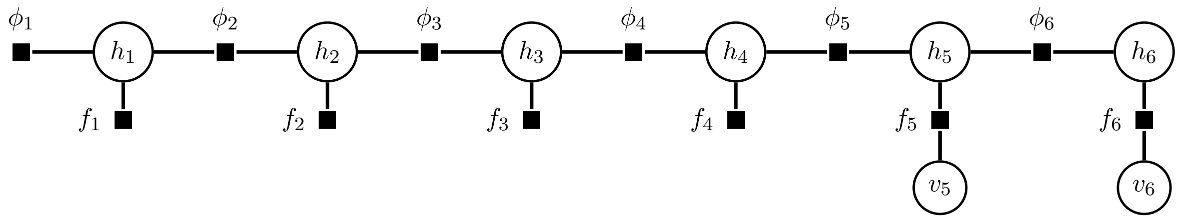
$$K_s = \frac{P_sC_s}{C_s^2P_s + D_s^2} \tag{22}$$

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

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

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

$$f_t(h_t) = p(v_t|h_t) \;\; (t \leq 4) \qquad\qquad f_t(v_t, h_t) = p(v_t|h_t) \;\; (t > 4) \tag{23}$$

$$\phi_1(h_1) = p(h_1) \qquad\qquad \phi_t(h_t, h_{t-1}) = p(h_t|h_{t-1}) \;\; (t > 1) \tag{24}$$

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 $\beta(h_4) = \mu_{\phi_5 \to h_4}(h_4) = 1$.

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

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

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