*The purpose of this tutorial sheet is to help you better understand the lecture material. Start early and do as many as you have time for. Even if you are unable to make much progress, you should still attend your tutorial.*

## 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 $\mu$ and variance $\sigma^2$ 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 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 $h_s$ and $v_s$ as defined in the update and observation equations

$$h_s = A_s h_{s-1} + B_s \xi_s \tag{4}$$
$$v_s = C_s h_s + D_s \eta_s \tag{5}$$

follow the conditional distributions in (2) and (3). The random variables $\xi_s$ and $\eta_s$ are independent from the other variables in the model and follow a standard normal Gaussian distribution, e.g. $\xi_s \sim \mathcal{N}(\xi_s|0, 1)$.
Hint: For two constants $c_1$ and $c_2$, $y = c_1 + c_2 x$ is Gaussian if $x$ is Gaussian. In other words, an affine transformation of a Gaussian is Gaussian.

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 $D_s^2$.

(b) Show that

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

Hint: 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.

(c) 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

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

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

*Hint: Work in the negative log domain.*

(d) In the lecture, we have seen 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 $\mu_1$ and its variance by $\sigma_1^2$ 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

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

(e) Use Equation (7) to show that

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

where

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

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

(f) 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.

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