Note: Problems (or parts of problems) marked with a star (★) are required for graduate students to receive 4 credit hours; undergraduate students who solve these problems will receive extra credit points.

Submission: Write your name, netid, and U for undergrad/G for grad in the upper right-hand corner of the first page of your written solutions. Typewritten solutions will receive 5 extra credit points.

Problems to be handed in

In Problems 1–3, you will fill in the details from the lecture on the Kalman filter.

1 Let X and Y be two random variables with joint pdf f_{XY}. The *conditional pdf* of X given $Y = y$ is defined as

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)},$$

whenever it exists. Here, $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x$ is the marginal pdf of Y. The *conditional expectation* (or *conditional mean*) of X given $Y = y$ is defined as

$$\mathbf{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x,$$

whenever it exists. Note that $\mathbf{E}[X|Y = y]$ is a function of y. In this problem, we will explore several properties of conditional pdfs and conditional means.

(a) Suppose that X and Y are independent random variables. Prove that $\mathbf{E}[X|Y = y] = \mathbf{E}[X]$ for any y.

(b) Let U be a random variable with pdf f_U, which is independent of X. Define $Y = aX + U$, where $a \in \mathbb{R}$ is some constant. Show that $f_{Y|X}(y|x) = f_U(y - ax)$.

Hint: You will first need to compute the joint pdf of X and Y. A good way of doing this is by exploiting the Law of the Unconscious Statistician: the expected value of any function $g(X,Y)$ with respect to X and Y, defined as

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

can also be written down in terms of X and U:

$$\mathbf{E}[g(X,Y)] = \mathbf{E}[g(X, aX + U)].$$

Use this relation, together with the fact that X and U are independent random variables, to show that

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_X(x) f_U(y - ax)\mathrm{d}x\mathrm{d}y$$

for any g.

(c) Consider the setting of part (b), where $X \sim N(m_X, \sigma_X^2)$ and $U \sim N(0, \sigma_U^2)$. Prove that the conditional pdf $f_{Y|X}$ is Gaussian with mean $m_{Y|X=x} = ax$ and variance σ_U^2.

(d) Continuing with the setting of part (c), prove that the conditional pdf $f_{X|Y}$ is Gaussian with mean

$$m_{X|Y=y} = \frac{a\sigma_X^2 y + m_X \sigma_U^2}{a^2\sigma_X^2 + \sigma_U^2}$$

and variance

$$\sigma_{X|Y}^2 = \frac{\sigma_X^2 \sigma_U^2}{a^2\sigma_X^2 + \sigma_U^2}$$

(note that $m_{X|Y=y}$ is a function of y, while $\sigma_{X|Y}^2$ is a constant, which is why we write $\sigma_{X|Y}^2$).

Hint: Using Bayes' rule,

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}.$$

Use the results of parts (b,c) to prove that

$$f_{X|Y}(x|y) \propto \exp\left(-\frac{1}{2\sigma_U^2}(y - ax)^2 - \frac{1}{2\sigma_X^2}(x - m_X)^2\right),$$

where the constant of proportionality is a function of y. Complete the square in the exponent to extract $m_{X|Y=y}$ and $\sigma_{X|Y}^2$ and thus show that $f_{X|Y}$ is a Gaussian pdf.

2 In class, we have derived the Bayesian filtering recursion for a discrete-state hidden Markov model. In this problem, we will derive such a recursion for a certain type of continuous-state hidden Markov models. Let $X_0, U_1, U_2, \dots, V_1, V_2, \dots$ be independent real-valued random variables, where the distribution of X_0 has a pdf f_0, the U_t's are i.i.d. with common pdf g, and the V_t's are i.i.d. with common pdf h. The evolution of the real-valued hidden state signal $(X_t)_{t\in\mathbb{N}}$ and the real-valued observation signal $(Y_t)_{t\in\mathbb{N}}$ is given by

$$\begin{aligned} X_t &= aX_{t-1} + U_t, \\ Y_t &= cX_t + V_t, \end{aligned}$$

where $a, c \in \mathbb{R}$ are known coefficients.

(a) Show that, for any $t \in \mathbb{N}$, the random variables $X_0^t = (X_0, \dots, X_t)$, $Y_1^t = (Y_1, \dots, Y_t)$ have the joint pdf

$$f_{X_0^t, Y_1^t}(x_0^t, y_1^t) = f_0(x_0) \prod_{s=1}^{t} g(x_s - ax_{s-1}) h(y_s - cx_s).$$

(b) Just as in the discrete case, the Bayesian filtering problem entails the computation of the conditional pdf's

$$\pi_t(x_t) \triangleq f_{X_t|Y_1^t}(x_t|y_1^t) = \frac{f_{X_t, Y_1^t}(x_t, y_1^t)}{f_{Y_1^t}(y_1^t)}.$$

For any $s, t \in \mathbb{N}$, let $\pi_{s|t}(x_s)$ denote the conditional pdf of X_s given Y_1^t. Show that the computation of $\pi_t \equiv \pi_{t|t}$ can be decomposed into the prediction and the correction steps,

$$\pi_{t|t} \xrightarrow{\text{prediction}} \pi_{t+1|t} \xrightarrow{\text{correction}} \pi_{t+1|t+1},$$

where the prediction step is given by

$$\pi_{t+1|t}(x) = \int_{-\infty}^{\infty} \pi_t(u) g(x - au) \mathrm{d}u$$

and the correction step is given by

$$\pi_{t+1|t+1}(x) = \frac{\pi_{t+1|t}(x) h(y_{t+1} - cx)}{\int_{-\infty}^{\infty} \pi_{t+1|t}(u) h(y_{t+1} - cu), \mathrm{d}u},$$

with the initial condition $\pi_0 = \pi_{0|0} = f_0$.

3 In general, the computation of π_t is intractable, just like in the discrete case. However, when X_0, the U_t's, and the V_t's are independent Gaussian random variables, the filtering update simplifies considerably and amounts to recursive computation of conditional means $m_t = \mathbf{E}[X_t | Y_1^t = y_1^t]$ and variances $\Sigma_t = \mathrm{Var}[X_t | Y_1^t = y_1^t]$. This recursion is known as the *Kalman filter*, after its inventor Rudolf E. Kalman. We assume the following:

- The initial state $X_0 \sim N(\mu_0, \sigma_0^2)$, $U = (U_t)_{t \in \mathbb{N}} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_U^2)$, $V = (V_t)_{t \in \mathbb{N}} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_V^2)$ are mutually independent random variables.
- Just like in Problem 2, the evolution of the hidden state $(X_t)_{t \in \mathbb{N}}$ and the observation $(Y_t)_{t \in \mathbb{N}}$ is given by

$$\begin{aligned} X_t &= aX_{t-1} + U_t, \\ Y_t &= cX_t + V_t \end{aligned}$$

where $a, c \in \mathbb{R}$ are the given coefficients.

(a) Show that, for each t, the random variables X_0^t, Y_1^t are jointly Gaussian, and therefore it suffices to keep track of the conditional mean

$$m_t \triangleq \mathbf{E}[X_t | Y_1^t = y_1^t]$$

and the conditional variance

$$\Sigma_t \triangleq \mathrm{Var}[X_t | Y_1^t = y_1^t].$$

(b) For any s, t, define $m_{s|t} \triangleq \mathbf{E}[X_s | Y_1^t]$ and $\Sigma_{s|t} \triangleq \mathrm{Var}[X_s | Y_1^t = y_1^t]$. Thus, $m_t = m_{t|t}$ and $\Sigma_t = \Sigma_{t|t}$. Show that the prediction step amounts to

$$\begin{aligned} m_{t+1|t} &= a m_{t|t}, \\ \Sigma_{t+1|t} &= a^2 \Sigma_{t|t} + \sigma_U^2 \end{aligned}$$

and the correction step is given by

$$m_{t+1|t+1} = \frac{c y_{t+1} \Sigma_{t+1|t} + m_{t+1|t}\sigma_V^2}{c^2 \Sigma_{t+1|t} + \sigma_V^2},$$
$$\Sigma_{t+1|t+1} = \frac{\Sigma_{t+1|t}\sigma_V^2}{c^2 \Sigma_{t+1|t} + \sigma_V^2},$$

with the initial condition $m_0 = m_{0|0} = \mu_0$ and $\Sigma_0 = \Sigma_{0|0} = \sigma_0^2$.

(c)★ Show that Σ_t is the variance of the state estimation error $X_t - m_t$ at time t. We say that Σ_∞ is the steady-state error variance of the Kalman filter if $\Sigma_t \to \Sigma_\infty$ as $t \to \infty$. It can be shown that, if this limit exists, then it is given by solving the fixed-point equation $\Sigma_{t+1} = \Sigma_t$. What are the conditions on $a, c, \sigma_U^2, \sigma_V^2$ to guarantee this?

4 Hidden Markov models can have very strange behavior. Consider a discrete-state hidden Markov model, where the hidden state signal $X = (X_t)_{t \in \mathbb{Z}_+}$ is a Markov chain with state space $\mathsf{X} = \{1, 2, 3, 4\}$, arbitrary initial state distribution p_0, and one-step transition probability matrix

$$M = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix},$$

and where the observation signal $Y = (Y_t)_{t \in \mathbb{N}}$ is binary-valued, with

$$Y_t = \begin{cases} 1, & \text{if } X_t = 1 \text{ or } X_t = 3 \\ 0, & \text{otherwise} \end{cases}.$$

Consider the following two situations:

(i) We observe the entire signal Y and a single state X_t for some $t \ge 1$.

(ii) We observe the entire signal Y, but not the state process X.

Explain why in the first situation we can recover the entire hidden state process X, while in the second situation there is still uncertainty about the initial state X_0 (and therefore the subsequent states $X_1, X_2, \ldots$).

5(★)

(a) Suppose that X and Y are two jointly Gaussian random variables with means m_X, m_Y, variances σ_X^2, σ_Y^2, and covariance $c_{XY} = \mathbf{E}[(X - m_X)(Y - m_Y)]$. Prove that the conditional pdf $f_{X|Y}$ is Gaussian with mean

$$m_{X|Y=y} = m_X + \frac{c_{XY}}{\sigma_Y^2}(y - m_Y)$$

and variance

$$\sigma_{X|Y}^2 = \sigma_X^2 - \frac{c_{XY}^2}{\sigma_Y^2},$$

and show that the result of Problem 1(d) is a special case of this.

Hint: Consider the random variables

$$U \triangleq m_X + \frac{c_{XY}}{\sigma_Y^2}(Y - m_Y)$$

and $V \triangleq X - U$. Show that $\mathbf{E}[V] = 0$, that Y and V are independent, and that U and V are independent. Use this to prove that $\mathbf{E}[X|Y = y] = \mathbf{E}[U + V|Y = y] = \mathbf{E}[U|Y = y] = m_{X|Y=y}$ and that $\sigma_X^2 = \sigma_U^2 + \sigma_V^2$, and therefore that $\sigma_{X|Y=y}^2 = \sigma_V^2 = \sigma_X^2 - \sigma_U^2$.

(b) Let $X, Y_1, \ldots, Y_n$ be jointly Gaussian random variables, where $\mathbf{E}[Y_i] = 0$ and $\mathbf{E}[Y_i Y_j] = 0$ for all i and all $j \neq i$. Show that the conditional mean $\mathbf{E}[X|Y_1^n = y_1^n]$ is given by

$$\mathbf{E}[X|Y_1^n = y_1^n] = \mathbf{E}[X] + \sum_{i=1}^{n} \frac{\mathbf{E}[XY_i]}{\mathrm{Var}[Y_i]} y_i.$$

Hint: Use the same strategy as in part (a).