EE378: Final Examination Solutions

Wednesday, June 9, 8:30-11:30 AM

The total number of points is 100. In addition, there are 15 bonus points.

1. (32 points) Consider a sextuple of discrete random variables $(X_1, X_2, X_3, X_4, X_5, X_6)$ satisfying the following three conditional independence relations:

   - $X_1 - X_2 - (X_3, X_4, X_5, X_6)$
   - $(X_2, X_3) - X_4 - X_5$
   - $(X_2, X_3, X_4) - X_5 - X_6$

   (a) (6 points) Show that the joint PMF of $(X_1, X_2, X_3, X_4, X_5, X_6)$ has a factorization of the form

   $$p(x_1, x_2, x_3, x_4, x_5, x_6) = \phi_a(x_1, x_2)\phi_b(x_2, x_3, x_4)\phi_c(x_4, x_5)\phi_d(x_5, x_6). \tag{1}$$

   (b) (5 points) Draw the graph associated with the factorization in (1).

   (c) (21 points) For each of the following relations, indicate whether or not it necessarily holds. If it does, argue why. If it does not, provide a counter-example (that is, describe a sextuple satisfying the above three conditional independence relations yet not satisfying the relation in question).

   i. (3 points) $(X_1, X_2, X_3) - (X_4, X_5) - X_6$
   ii. (3 points) $X_1 - X_3 - X_5$
   iii. (3 points) $X_4 - (X_2, X_3, X_5) - (X_1, X_6)$
   iv. (3 points) $(X_1, X_2) - X_3 - (X_4, X_5, X_6)$
   v. (3 points) $(X_1, X_2, X_3) - (X_4, X_6) - X_5$
   vi. (3 points) $X_2 - (X_1, X_3) - X_5$
   vii. (3 points) $X_1 - X_4 - X_6$

I

Solution to 1:

(a) $P(x_1, x_2, X_3, X_4, X_5, X_6)$

$= P(X_1, X_2) \cdot P(X_3, X_4, X_5, X_6 \mid X_1, X_2)$

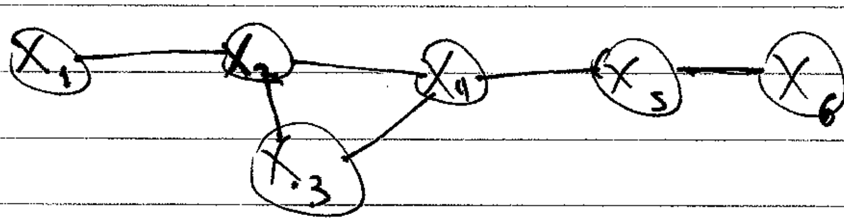
1st cond. indep. relation $= P(X_1, X_2) \cdot P(X_3, X_4, X_5, X_6 \mid X_2)$

$= P(X_1, X_2) \cdot P(X_3, X_4 \mid X_2) \cdot P(X_5 \mid X_2, X_3, X_4)$

2nd relation $\times P(X_6 \mid X_2, X_3, X_4, X_5)$

$= \underbrace{P(X_1, X_2)}_{\phi_a} \cdot \underbrace{P(X_3, X_4 \mid X_2)}_{\phi_b} \cdot \underbrace{P(X_5 \mid X_4)}_{\phi_c} \cdot \underbrace{P(X_6 \mid X_5)}_{\phi_d}$ ← 3rd relation

(b)



(c) (i) holds. (have to pass through (X_5, X_5) to get from (X_1, X_2, X_3) to X_6)

(ii) does not necessarily hold.

Counter-example: Consider $X_1 = X_2 = X_4 = X_5 = X_6 = X$

$X_3 = Y$, X, Y are independent.

~~The three cond. indep.~~

II

(iii) holds. (X_2, X_3, X_5) are all the nearest neighbors of X_4.

(iv) does not nec. hold. Same counter-example as part (ii).

(v) holds. (X_4, X_6) are all the neighbors of X_5

(vi) does not nec. hold. say $X_1 = X$
$X_2 = X_4 = X_5 = X_6 = Y$
$X_3 = Z$
X, Y, Z are independent.

(vii) holds. obviously $(X_1, X_2, X_3) - X_4 - (X_5, X_6)$
which implies $X_1 - X_4 - X_6$.

2. (36 points + 5 bonus points) Let $\{X_n : -\infty < n < \infty\}$, $X_n \in \{1, -1\}$, be a stationary first-order Markov binary symmetric process with transition probability q, that is,

$$P(X_{n+1} = 1 | X_n = -1) = P(X_{n+1} = -1 | X_n = 1) = q,$$

and

$$P(X_0 = 1) = P(X_0 = -1) = \frac{1}{2},$$

where parameter $q \in (0, 1/2)$ is known.

Let $\{Z_n : -\infty < n < \infty\}$ be an i.i.d. noise process independent of $\{X_n\}$, where

$$P(Z_n = 1) = P(Z_n = -1) = \frac{1}{2}.$$

Let the observation process $Y_n = X_n + Z_n$.

(a) (6 points) Find the autocorrelation function of X.
(Hint: show that the autocorrelation function is of the form $R_X(k) = c^{|k|}$ and identify c explicitly.)

For the following parts, let $q = 1/5$.

(b) (6 points) Find the causal Wiener filter $H(e^{j\omega})$ for estimating X_n based on $\{Y_k : -\infty < k \leq n\}$

(c) (6 points) Find the MSE of the causal Wiener filter in part (b).

(d) (6 points) Let $\{\tilde{X}_n\}$ and $\{\tilde{Y}_n\}$ be Gaussian random processes such that

$$\tilde{X}_n = a\tilde{X}_{n-1} + W_n,$$
$$\tilde{Y}_n = \tilde{X}_n + N_n,$$

where a is a constant, $\{W_n\}$ and $\{N_n\}$ are i.i.d. processes ~~independent of~~ $\cancel{\{\tilde{X}_t\}}$, $W_n \sim \mathcal{N}(0, \sigma_W^2)$, and $N_n \sim \mathcal{N}(0, \sigma_N^2)$.
Find a, σ_W^2, and σ_N^2 such that

$$R_{\tilde{X}}(k) = R_X(k), \quad R_{\tilde{Y}}(k) = R_Y(k), \quad R_{\tilde{X}\tilde{Y}}(k) = R_{XY}(k).$$

(e) (6 points) For the values of a, σ_W^2, and σ_N^2 you found in part (d), use the Kalman filter to find $\lim_{n\to\infty} \mathrm{Var}(\tilde{X}_n | \tilde{Y}^n)$. Compare it with the MSE in part (c).

(f) (6 points) Find the optimal non-linear causal filter $\mathrm{E}(X_n | Y^n = y^n)$.
(Hint: consider the sets

$$\mathcal{S}_{n,0} = \{y^n : y_i = 0 \text{ for } 1 \leq i \leq n\}$$
$$\mathcal{S}_{n,k}^+ = \{y^n : \ y_k = 2, \ y_i = 0 \text{ for } k+1 \leq i \leq n\}$$
$$\mathcal{S}_{n,k}^- = \{y^n : \ y_k = -2, \ y_i = 0 \text{ for } k+1 \leq i \leq n\}$$

for $k = 1, 2, \ldots, n$ and note that $\mathrm{E}(X_n | Y^n = y^n)$ is the same for all $y^n \in \mathcal{S}_{n,k}^+$ and similarly for $\mathcal{S}_{n,k}^-$ and $\mathcal{S}_{n,0}$.)

(g) (Bonus: 5 points) Find the MSE of the optimal causal filter in part (f) as $n \to \infty$. Compare it with the MSE's in parts (c) and (e).

Solution:

(a) Consider

$$R_X(0) = \mathrm{E}(X_n^2) = \frac{1}{2} \cdot 1^2 + \frac{1}{2} \cdot (-1)^2 = 1.$$

Next, for $k > 0$,

$$\begin{aligned} R_X(k) &= \mathrm{E}(X_{n+k} X_n) \\ &= P(X_n = 1, X_{n+k} = 1) + P(X_n = -1, X_{n+k} = -1) \\ &\quad - P(X_n = 1, X_{n+k} = -1) - P(X_n = -1, X_{n+k} = -1) \\ &= P(X_n = 1) P(X_{n+k} = 1 | X_n = 1) + P(X_n = -1) P(X_{n+k} = -1 | X_n = -1) \\ &\quad - P(X_n = 1) P(X_{n+k} = -1 | X_n = 1) - P(X_n = -1) P(X_{n+k} = -1 | X_n = -1). \end{aligned}$$

The k-step transition probability matrix is

$$\begin{aligned} \begin{bmatrix} 1-q & q \\ q & 1-q \end{bmatrix}^k &= \left(\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1-2q \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \right)^k \\ &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-2q)^k \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1+(1-2q)^k & 1-(1-2q)^k \\ 1-(1-2q)^k & 1+(1-2q)^k \end{bmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} R_X(k) &= \frac{1}{2} \left(\frac{1+(1-2q)^k}{2} + \frac{1+(1-2q)^k}{2} - \frac{1-(1-2q)^k}{2} - \frac{1-(1-2q)^k}{2} \right) \\ &= (1-2q)^k. \end{aligned}$$

Since the autocorrelation function is even, $R_X(k) = (1-2q)^{|k|}$, that is, $c = 1 - 2q$. Alternatively,

$$\mathrm{E}(X_{n+1} | X_n) = (1-q) X_n + q(-X_n) = (1-2q) X_n.$$

Then for $k > 0$,

$$\begin{aligned} R_X(k) &= \mathrm{E}(X_{n+k} X_n) = \mathrm{E}(\mathrm{E}(X_{n+k} X_n | X_n^{n+k-1})) \\ &= \mathrm{E}(X_n \, \mathrm{E}(X_{n+k} | X_n^{n+k-1})) = \mathrm{E}(X_n \, \mathrm{E}(X_{n+k} | X_{n+k-1})) \\ &= (1-2q) \, \mathrm{E}(X_{n+k-1} X_n) = (1-2q) R_X(k-1). \end{aligned}$$

Therefore, $R_X(k) = (1-2q)^{|k|}$.

(b) The causal Wiener filter is given by

$$H(e^{j\omega}) = \frac{1}{S_Y^+(e^{j\omega})} \left\{ \frac{S_X(e^{j\omega})}{S_Y^+(e^{-j\omega})} \right\}_+ .$$

The power spectral density of $\{X_n\}$ is

$$\begin{aligned}
S_X(e^{j\omega}) &= \sum_{k=-\infty}^{\infty} R_X(k) e^{-j\omega k} = \sum_{k=-\infty}^{\infty} (1-2q)^{|k|} e^{-j\omega k} \\
&= \sum_{k=-\infty}^{-1} (1-2q)^{-k} e^{-j\omega k} + \sum_{k=0}^{\infty} (1-2q)^{k} e^{-j\omega k} \\
&= \frac{(1-2q)e^{j\omega}}{1-(1-2q)e^{j\omega}} + \frac{1}{1-(1-2q)e^{-j\omega}} \\
&= \frac{4q(1-q)}{(1-(1-2q)e^{j\omega})(1-(1-2q)e^{-j\omega})} \\
&= \frac{16}{25(1-\frac{3}{5}e^{j\omega})(1-\frac{3}{5}e^{-j\omega})}.
\end{aligned}$$

Since $\{Z_n\}$ is independent of $\{X_n\}$ and $S_Z(\omega) = R_Z(0) = 1$,

$$\begin{aligned}
S_Y(e^{j\omega}) &= S_X(e^{j\omega}) + S_Z(e^{j\omega}) \\
&= \frac{16}{25(1-\frac{3}{5}e^{j\omega})(1-\frac{3}{5}e^{-j\omega})} + 1 \\
&= \frac{9(1-\frac{1}{3}e^{j\omega})(1-\frac{1}{3}e^{-j\omega})}{5(1-\frac{3}{5}e^{j\omega})(1-\frac{3}{5}e^{-j\omega})}.
\end{aligned}$$

Thus,

$$S_Y^+(e^{j\omega}) = \frac{3(1-\frac{1}{3}e^{-j\omega})}{\sqrt{5}(1-\frac{3}{5}e^{-j\omega})}.$$

Next, consider

$$\begin{aligned}
\left\{ \frac{S_X(e^{j\omega})}{S_Y^+(e^{-j\omega})} \right\}_+ &= \left\{ \frac{16}{25(1-\frac{3}{5}e^{j\omega})(1-\frac{3}{5}e^{-j\omega})} \frac{\sqrt{5}(1-\frac{3}{5}e^{j\omega})}{3(1-\frac{1}{3}e^{j\omega})} \right\}_+ \\
&= \frac{16\sqrt{5}}{75} \left\{ \frac{1}{(1-\frac{3}{5}e^{-j\omega})(1-\frac{1}{3}e^{j\omega})} \right\}_+ \\
&= \frac{16\sqrt{5}}{75} \left\{ \frac{5}{4(1-\frac{3}{5}e^{-j\omega})} + \frac{5e^{j\omega}}{12(1-\frac{1}{3}e^{j\omega})} \right\}_+ \\
&= \frac{4\sqrt{5}}{15(1-\frac{3}{5}e^{-j\omega})}.
\end{aligned}$$

Therefore, the causal Wiener filter is

$$H(e^{j\omega}) = \frac{\sqrt{5}(1-\frac{3}{5}e^{-j\omega})}{3(1-\frac{1}{3}e^{-j\omega})}\frac{4\sqrt{5}}{15(1-\frac{3}{5}e^{-j\omega})} = \frac{4}{9(1-\frac{1}{3}e^{-j\omega})}.$$

(c) Let $U_n = X_n - \hat{X}_n$ be the error process, where $\hat{X}_n = h_n * Y_n$ and h_n is the causal Wiener filter. Then

$$U_n = X_n - h_n * (X_n + Z_n) = (\delta_n - h_n) * X_n - h_n * Z_n.$$

Thus,

$$\begin{aligned}
S_U(e^{j\omega}) &= |1 - H(e^{j\omega})|^2 S_X(e^{j\omega}) + |H(e^{j\omega})|^2 \\
&= \left|\frac{5-3e^{-j\omega}}{9(1-\frac{1}{3}e^{-j\omega})}\right|^2 \frac{16}{25(1-\frac{3}{5}e^{j\omega})(1-\frac{3}{5}e^{-j\omega})} + \left|\frac{4}{9(1-\frac{1}{3}e^{j\omega})}\right|^2 \\
&= \frac{32}{81(1-\frac{1}{3}e^{j\omega})(1-\frac{1}{3}e^{-j\omega})},
\end{aligned}$$

which is the Fourier transform of

$$R_U(k) = \frac{4}{9}\left(\frac{1}{3}\right)^{|k|}.$$

Therefore, the MSE is $\mathrm{E}(U_n^2) = R_U(0) = 4/9$.

(d) Since $\{\tilde{X}_n\}$ is an AR(1) process,

$$S_{\tilde{X}}(e^{j\omega}) = \frac{\sigma_W^2}{(1-ae^{j\omega})(1-ae^{-j\omega})}.$$

Thus, we have $a = 1-2q = 3/5$ and $\sigma_W^2 = 4q(1-q) = 16/25$. Next,

$$S_{\tilde{Y}}(e^{j\omega}) = S_{\tilde{X}}(e^{j\omega}) + \sigma_N^2.$$

We know that $S_Y(e^{j\omega}) = S_X(e^{j\omega}) + 1$. Therefore, $\sigma_N^2 = \mathrm{Var}(Z) = 1$.

Remark: There is an incorrect condition in this question, so everyone gets 6 points.

(e) Let $\hat{X}_{n|n-1} = \mathrm{E}(X_n|Y^{n-1})$, $\hat{X}_{n|n} = \mathrm{E}(X_n|Y^n)$, $\Lambda_{n|n-1} = \mathrm{Var}(X_n|Y^{n-1})$, and $\Lambda_{n|n} = \mathrm{Var}(X_n|Y^n)$. Define

$$g_t = \frac{h\Lambda_{t|t-1}}{h^2\Lambda_{t|t-1} + \sigma_N^2} = \frac{\Lambda_{t|t-1}}{\Lambda_{t|t-1} + 1}.$$

Then the measurement update of the Kalman filter is

$$\begin{aligned}
\hat{X}_{t|t} &= \hat{X}_{t|t-1} + g_t(y_t - h\hat{X}_{t|t=1}), \\
\Lambda_{t|t} &= (1 - hg_t)\Lambda_{t|t-1},
\end{aligned}$$

and the time update is

$$\hat{X}_{t+1|t} = a\hat{X}_{t|t},$$
$$\Lambda_{t+1|t} = a^2\Lambda_{t|t} + \sigma_W^2.$$

Thus,

$$\Lambda_{t|t} = \frac{\Lambda_{t|t-1}}{\Lambda_{t|t-1}+1} = \frac{\frac{9}{25}\Lambda_{t-1|t-1} + \frac{16}{25}}{\frac{9}{25}\Lambda_{t-1|t-1} + \frac{16}{25} + 1}.$$

Let $\lim_{t\to\infty} \Lambda_{t|t} = \Lambda$. Then we have

$$\Lambda = \frac{9\Lambda + 16}{9\Lambda + 41}.$$

Thus, $9\Lambda^2 - 32\Lambda - 16 = 0$. Since $\Lambda \geq 0$, $\Lambda = 4/9$. Therefore $\lim_{n\to\infty} \mathrm{Var}(X_n|Y^n) = 4/9$, which is the same as the MSE of the causal Wiener filter.

(f) Suppose that $y^n \in \mathcal{S}_{n,0}$. Then

$$\begin{aligned}
P(X_n = 1, Y^n = y^n) &= \sum_{x^{n-1}} P(X^{n-1} = x_{n-1}, X_n = 1, Y^n = y^n) \\
&= \sum_{x^{n-1}} P(X_n = 1) P(X^{n-1} = x_{n-1}|X_n = 1) \\
&\qquad \times P(Y_n = 0|X_n = 1) \prod_{i=1}^{n-1} P(Y_i = 0|X_i = x_i) \\
&= \sum_{x^{n-1}} P(X^{n-1} = x_{n-1}|X_n = 1) \left(\frac{1}{2}\right)^{n+1} = \left(\frac{1}{2}\right)^{n+1}.
\end{aligned}$$

Similarly, $P(X_n = -1, Y^n = y^n) = (1/2)^{n+1}$. Thus,

$$\mathrm{E}(X^n|Y^n = y^n) = \frac{1\cdot(1/2)^{n+1} + (-1)\cdot(1/2)^{n+1}}{(1/2)^{n+1} + (1/2)^{n+1}} = 0$$

for $y^n \in \mathcal{S}_{n,0}$.

Next, suppose that $y^n \in \mathcal{S}_{n,k}^+$. Then

$$\begin{aligned}
P(X_n = x_n|Y^n = y^n) &= P(X_n = x_n|Y^n = y^n, X_k = 1) \\
&= P(X_n = x_n|Y_{k+1}^n = y_{k+1}^n, X_k = 1).
\end{aligned}$$

Consider

$$
\begin{aligned}
P(X_n = 1, X_k = 1, Y_{k+1}^n = y_{k+1}^n) \\
&= \sum_{x_{k+1}^{n-1}} P(X_n = 1, X_{k+1}^{n-1} = x_{k+1}^{n-1}, X_k = 1, Y_{k+1}^n = y_{k+1}^n) \\
&= \sum_{x_{k+1}^{n-1}} P(X_k = 1) P(X_n = 1 | X_k = 1) P(X_{k+1}^{n-1} = x_{k+1}^{n-1} | X_k = 1, X_n = 1) \\
&\qquad \times P(Y_n = 0 | X_n = 1) \prod_{i=k+1}^{n-1} P(Y_i = 0 | X_i = x_i) \\
&= \sum_{x_{k+1}^{n-1}} P(X_{k+1}^{n-1} = x_{k+1}^{n-1} | X_k = 1, X_n = 1) \left(\frac{1 - (1-2q)^{n-k}}{2} \right) \left(\frac{1}{2} \right)^{n-k+1} \\
&= \left(\frac{1 + (1-2q)^{n-k}}{2} \right) \left(\frac{1}{2} \right)^{n-k+1}.
\end{aligned}
$$

Similarly,

$$
P(X_n = 0, X_k = 1, Y_{k+1}^n = y_{k+1}^n) = \left(\frac{1 - (1-2q)^{n-k}}{2} \right) \left(\frac{1}{2} \right)^{n-k+1}.
$$

Therefore,

$$
\begin{aligned}
\mathrm{E}(X_n | Y^n = y^n) &= \mathrm{E}(X_n | Y_{k+1}^n = y_{k+1}^n, X_k = 1) \\
&= 1 \cdot \left(\frac{1 + (1-2q)^{n-k}}{2} \right) + (-1) \cdot \left(\frac{1 - (1-2q)^{n-k}}{2} \right) \\
&= (1-2q)^{n-k}
\end{aligned}
$$

for all $y^n \in \mathcal{S}_{n,k}^+$.
By symmetry, for all $y^n \in \mathcal{S}_{n,k}^-$,

$$
\mathrm{E}(X_n | Y^n = y^n) = -(1-2q)^{n-k}.
$$

Therefore,

$$
\mathrm{E}(X_n | Y^n = y^n) = \begin{cases} 0 & \text{if } y^n \in \mathcal{S}_{n,0} \\ (3/5)^{n-k} & \text{if } y^n \in \mathcal{S}_{n,k}^+ \\ -(3/5)^{n-k} & \text{if } y^n \in \mathcal{S}_{n,k}^-. \end{cases}
$$

(g) Suppose that $y^n \in \mathcal{S}_{n,0}$, the conditional MSE is

$$
\mathrm{Var}(X_n | Y^n = y^n) = \frac{(1-0)^2 \cdot (1/2)^{n+1} + (-1-0)^2 \cdot (1/2)^{n+1}}{(1/2)^{n+1} + (1/2)^{n+1}} = 1.
$$

For $y^n \in \mathcal{S}^+_{n,k}$, the conditional MSE is

$$\begin{aligned}
\mathrm{Var}(X_n|Y^n = y^n) &= \left(1-(1-2q)^{n-k}\right)^2 \cdot \left(\frac{1+(1-2q)^{n-k}}{2}\right) \\
&\quad + \left(-1-(1-2q)^{n-k}\right)^2 \cdot \left(\frac{1-(1-2q)^{n-k}}{2}\right) \\
&= 1-(1-2q)^{2(n-k)}.
\end{aligned}$$

By symmetry,

$$\mathrm{Var}(X_n|Y^n = y^n) = 1-(1-2q)^{2(n-k)}$$

for all $y^n \in \mathcal{S}^-_{n,k}$.

Now we only need to find $P(\mathcal{S}_{n,0})$, $P(\mathcal{S}^+_{n,k})$, and $P(\mathcal{S}^-_{n,k})$. We first consider

$$\begin{aligned}
P(\mathcal{S}_{n,0}) &= P(Y_1 = 0, \ldots, Y_n = 0) \\
&= \sum_{x^n} P(X^n = x^n, Y_1 = 0, \ldots, Y_n = 0) \\
&= \sum_{x^n} P(X^n = x^n) \prod_{i=1}^{n} P(Y_i = 0 | X_i = x_i) \\
&= \sum_{x^n} P(X^n = x^n) \left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n.
\end{aligned}$$

Next, consider

$$\begin{aligned}
P(\mathcal{S}^+_{n,k}) &= P(Y_k = 2, Y_{k+1} = 0, \ldots, Y_n = 0) \\
&= P(X_k = 1, Y_k = 2, Y_{k+1} = 0, \ldots, Y_n = 0) \\
&= \sum_{x^{n\setminus k}} P(X^{n\setminus k} = x^{n\setminus k}, X_k = 1, Y_k = 2, Y_{k+1} = 0, \ldots, Y_n = 0) \\
&= \sum_{x^{n\setminus k}} P(X_k = 1) P(X^{n\setminus k} = x^{n\setminus k} | X_k = 1) \\
&\qquad \times P(Y_k = 2 | X_k = 1) \prod_{i=k+1}^{n} P(Y_i = 0 | X_i = x_i) \\
&= \sum_{x^{n\setminus k}} P(X^{n\setminus k} = x^{n\setminus k} | X_k = 1) \left(\frac{1}{2}\right)^{n-k+2} = \left(\frac{1}{2}\right)^{n-k+2}.
\end{aligned}$$

By symmetry,

$$P(\mathcal{S}^-_{n,k}) = \left(\frac{1}{2}\right)^{n-k+2}.$$

Therefore, the MSE is

$$\begin{aligned}
\mathrm{Var}(X_n|Y^n) &= P(\mathcal{S}_{n,0})\,\mathrm{Var}(X_n|Y^n \in \mathcal{S}_{n,0}) \\
&\quad + \sum_{k=1}^{n} P(\mathcal{S}_{n,k}^{+})\,\mathrm{Var}(X_n|Y^n \in \mathcal{S}_{n,k}^{+}) \\
&\quad + \sum_{k=1}^{n} P(\mathcal{S}_{n,k}^{-})\,\mathrm{Var}(X_n|Y^n \in \mathcal{S}_{n,k}^{-}) \\
&= \left(\frac{1}{2}\right)^{n} + 2\sum_{k=1}^{n}\left(\frac{1}{2}\right)^{n-k+2}\big(1-(1-2q)^{2(n-k)}\big) \\
&= 1 - \sum_{k=1}^{n}\left(\frac{1}{2}\right)^{n-k+1}(1-2q)^{2(n-k)}.
\end{aligned}$$

Taking $n \to \infty$, we have

$$\lim_{n\to\infty} \mathrm{Var}(X_n|Y^n) = 1 - \frac{1/2}{1-(1/2)(1-2q)^2} = \frac{16}{41},$$

which is lower than the MSE of the optimal linear causal filter.

3. (32 points + 10 bonus points) Let $\{f(t), 0 \le t \le T\}$ be a known, deterministic, continuous-time finite-energy signal. Let $X^T = \{X_t, 0 \le t \le T\}$ be the signal of interest given by $X_t = G \cdot f(t)$, where $G \sim \mathcal{N}(0,1)$. Let

$$Y_t = \sqrt{\gamma} X_t + N_t, \quad 0 \le t \le T,$$

where $\gamma > 0$ is a known number and N_t is Additive White Gaussian Noise of unit power spectral density, independent of X^T.

(a) (8 points) Find the optimal non-causal filter of X^T based on Y^T.
(Hint: note the Markov relation $X^T - \int_0^T Y_t f(t) dt - Y^T$.)

(b) (8 points) Find $\text{mmse}_f(\gamma)$, the mean square error of the filter from the previous part.

(c) (8 points) Find the optimal causal filter of X^T based on Y^T.

(d) (8 points) Find $\text{cmmse}_f(\gamma)$, the mean square error of the filter from the previous part, either directly or by use of the relationship $\text{cmmse}_f(\gamma) = \frac{1}{\text{snr}} \int_0^{\text{snr}} \text{mmse}_f(\gamma) d\gamma$.

Bonus part (extra credit):

Let now $\text{cmse}_{f,Q}(\gamma)$ denote the mean square error of the causal filter of X^T based on Y^T that would have been optimal if $X^T \sim Q$, when the noise-free signal is really given by $X_t = G \cdot f(t)$. Similarly, let $\{g(t), 0 \le t \le T\}$ be another function satisfying $\int_0^T f(t)^2 dt = \int_0^T g(t)^2 dt$ and $\int_0^T f(t) g(t) dt = 0$, and let $\text{cmse}_{g,Q}(\gamma)$ denote the mean square error of the causal filter of X^T based on Y^T that would have been optimal if $X^T \sim Q$, when the noise-free signal is really given by $X_t = G \cdot g(t)$. Suppose the underlying signal is known to either satisfy $X_t = G \cdot f(t)$ for all $0 \le t \le T$ or $X_t = G \cdot g(t)$ for all $0 \le t \le T$. We seek the prior distribution Q on X^T inducing the causal Bayesian filter which is optimal in the minimax sense. In other words, we seek the Q achieving the minimum in

$$\min_Q \max\{\text{cmse}_{f,Q}(\gamma) - \text{cmmse}_f(\gamma), \text{cmse}_{g,Q}(\gamma) - \text{cmmse}_g(\gamma)\}, \tag{2}$$

where, similarly to $\text{cmmse}_f(\gamma)$, $\text{cmmse}_g(\gamma)$ denotes the mean square error of the optimal causal filter when the noise-free signal is given by $X_t = G \cdot g(t)$.

(e) (Bonus: 5 points) Find the Q that achieves the minimum in (2).
(Hint: You can use the result on mismatched causal estimation in AWGN to express the differences in (2) as relative entropies, then use the 'redundancy-capacity' theorem to bring the problem to one of maximizing for a mutual information, and then find the maximizing Q using symmetry.)

(f) (Bonus: 5 points) Find the induced causal Bayesian filter, i.e., the causal filter which is optimal for $X^T \sim Q$, where Q is the distribution from the previous part.

III

Solution to 3

(a) Since, w.p. 1, X^T is in the span of the one "basis function" f, it follows that

$$X^T - \underbrace{\int_0^T Y_t \cdot f(t)\,dt}_{= \langle Y^T, f\rangle} - Y^T.$$

Now:

$$\langle Y^T, f\rangle \triangleq \int_0^T Y_t \cdot f(t)\,dt = \int_0^T \left(\sqrt{\gamma}\, X_t + N_t\right) f(t)\,dt$$

$$= \sqrt{\gamma} \cdot G \cdot \|f\|_2^2 + \int_0^T N_t \cdot f(t)\,dt$$

$$\Rightarrow \frac{\langle Y^T, f\rangle}{\sqrt{\gamma} \cdot \|f\|_2^2} = \underset{N(0,1)}{G} + \underbrace{\frac{\langle N^T, f\rangle}{\sqrt{\gamma} \cdot \|f\|_2^2}}_{N\left(0, \frac{1}{\gamma \|f\|^2}\right)}$$

indep.

$$\Rightarrow \hat{G}(Y^T) = E[G \mid Y^T] = E[G \mid \langle Y^T, f\rangle]$$

optimal estimate of G based on Y^T

$$= \frac{\langle Y^T, f\rangle}{\sqrt{\gamma} \cdot \|f\|^2} \cdot \frac{1}{1 + \frac{1}{\gamma \|f\|^2}}$$

$$= \langle Y^T, f\rangle \cdot \frac{\sqrt{\gamma}}{\gamma \|f\|^2 + 1}$$

$$\Rightarrow \underbrace{E[X_t \mid Y^T]}_{\hat{X}_t(Y^T) =} = E[G \cdot f(t) \mid Y^T] = f(t) \cdot E[G \mid Y^T] = \langle Y^T, f\rangle \frac{\sqrt{\gamma}\, f(t)}{\gamma \|f\|^2 + 1}.$$

IV

(b) $$E\left[(G-\hat{G})^2\right] = \frac{1 \cdot \frac{1}{\gamma \|f\|^2}}{1 + \frac{1}{\gamma \|f\|^2}} = \frac{1}{\gamma \|f\|^2 + 1}$$

$$\Rightarrow E\left[\left(X_t - \hat{X}_t(Y^T)\right)^2\right] = f(t)^2 \cdot E\left[(G-\hat{G})^2\right]$$

$$= \frac{f(t)^2}{\gamma \|f\|^2 + 1} \qquad \circledast$$

$$\Rightarrow \mathrm{mmse}_f(\gamma) = \int_0^T E\left[\left(X_t - \hat{X}_t(Y^T)\right)^2\right] dt$$

$$= \frac{\|f\|^2}{\gamma \|f\|^2 + 1}$$

(c) Using our findings from part (a) with t playing the role of T, we get:

$$\hat{X}_t(Y^t) = E[X_t \mid Y^t] = \int_0^t Y_s \cdot f(s)\, ds \cdot \frac{\sqrt{\gamma} \cdot f(t)}{\gamma \int_0^t f(s)^2 ds + 1}$$

(d) $$E\left[\left(X_t - \hat{X}_t(Y^t)\right)^2\right] = \frac{f(t)^2}{\gamma \int_0^t f(s)^2 ds + 1}$$

using $\circledast$ above with $T=t$

(V)

$$\Rightarrow \text{cmmse}_f(\gamma) = \int_0^T E\left[\left(X_t - \hat{X}_t(Y^t)\right)^2\right]$$

$$= \int_0^T \frac{f(t)^2}{\gamma \int_0^t f(s)^2 ds + 1} dt$$

~~done here.~~

$h(t) \triangleq \int_0^t f(s)^2 ds$

$$= \int_0^T \frac{h'(t)\, dt}{\gamma h(t) + 1}$$

$$= \frac{1}{\gamma} \ln\left(\gamma h(t) + 1\right) \Big|_{t=0}^{t=T}$$

$$= \frac{1}{\gamma} \ln\left(\gamma \|f\| + 1\right).$$

$h(0) = 0$
$h(T) = \|f\|^2$

VI

~~Not~~ An alternative would have been:

$$\mathrm{cmmse}_f(\mathrm{snr}) = \frac{1}{\mathrm{snr}} \int_0^{\mathrm{snr}} \mathrm{mmse}_f(\gamma)\, d\gamma$$

$$= \frac{1}{\mathrm{snr}} \int_0^{\mathrm{snr}} \frac{\|f\|^2}{\gamma \|f\|^2 + 1}\, d\gamma$$

$$= \frac{1}{\mathrm{snr}} \ln(\gamma \|f\|^2 + 1) \Big|_{\gamma=0}^{\mathrm{snr}}$$

$$= \frac{1}{\mathrm{snr}} \ln(\mathrm{snr} \cdot \|f\|^2 + 1).$$

VII

(e) ~~Our~~ The results we've seen on mismatched causal estimation in continuous-time AWGN imply

$$\text{cmse}_{f,Q}(\gamma) - \text{cmmse}_f(\gamma) = \frac{2}{\gamma} D(Q_{Y^T} \| P^f_{Y^T})$$

where $P^f_{Y^T}$ denotes the distribution of Y^T when $X_t = G \cdot f(t)$.

Similarly $\text{cmse}_{g,Q}(\gamma) - \text{cmmse}_g(\gamma) = \frac{2}{\gamma} D(Q_{Y^T} \| P^g_{Y^T})$

Thus we seek the Q (dist. on X^T) achieving:

$$\min_Q \max \{ D(Q_{Y^T} \| P^f_{Y^T}) \| D(Q_{Y^T} \| P^g_{Y^T}) \}$$

which, by the redundancy-capacity theorem, is ~~given by~~ equal to:

$$\max_{P_U} I(U; Y^T)$$

where $U \in \{0,1\}$ is a binary variable and:

$Y^T | U=0 \sim Y^T$ when the underlying process is $X_t = G f(t)$

$Y^T | U=1 \sim Y^T$ " " " " " $X_t = G \cdot g(t)$

VIII

Since f and g have equal energy, symmetry implies that $I(U;Y^T)$ is the same if $P_U(1)=q$ or $P_U(1)=1-q$.

Put together with the concavity of $I(U;Y^T)$ in P_U, this implies that $\max I(U;Y^T)$ is achieved by $P_U(1)=\frac{1}{2}$.

Thus, Q^* is the law of a process given by $G\cdot f(t)$ w.p. $\frac{1}{2}$ and $G\cdot g(t)$ w.p. $\frac{1}{2}$

In other words, the law of:

$$X_t = G\left(B\cdot f(t) + (1-B)\cdot g(t)\right)$$

where $G\sim N(0,1)$ $B\sim \text{Bernoulli}\left(\frac{1}{2}\right)$ and G, B are independent.

(f) Under the Q^* found in the previous part, X^t is, w.p. 1, in the span of the functions $f^t \triangleq \{f(s),\ 0\leq s\leq t\}$ and $g^t \triangleq \{g(s),\ 0\leq s\leq t\}$.

IX

Thus:

$$X_t - \left(\langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle \right) - Y^t$$

so the optimal causal filter is:

$$\hat{X}_t(Y^t) = E_{Q^*}\left[X_t \mid \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle \right].$$

It remains to compute $E_{Q^*}[X_t \mid \langle\ \rangle, \langle\ \rangle]$ explicitly.

Towards this end note that under Q^*, conditioned on $B=1$,

$$\left(X_t, \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle \right)$$

$$\sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} f(t)^2 & \sqrt{\gamma} f(t) \|f^t\|^2 & \sqrt{\gamma} f(t) \langle f^t, g^t \rangle \\ & \gamma \|f^t\|^4 + \|f^t\|^2 & \gamma \|f^t\|^2 \langle f^t, g^t \rangle + \langle f^t, g^t \rangle \\ & & \gamma \langle f^t, g^t \rangle^2 + \|g^t\|^2 \end{pmatrix} \right)$$

denote the associated density by:

$$\phi_{t,\gamma}(x, y_f, y_g)$$

Similarly, ~~cond.~~ under Q^*, conditioned on $B=0$:

$$\left(X_t,\ \langle Y^t, f^t\rangle,\ \langle Y^t, g^t\rangle\right)$$

$$\sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\ \begin{pmatrix} g(t)^2 & \sqrt{\gamma}\, g(t)\langle f^t, g^t\rangle & \sqrt{\gamma}\, g(t)\|g^t\|^2 \\ & \gamma\langle f^t, g^t\rangle^2 + \|f^t\|^2 \ \cancel{\gamma\|g^t\|^4 + \|g^t\|^2} & \gamma\|g^t\|^2\langle f^t, g^t\rangle + \langle f^t, g^t\rangle \\ & & \gamma\|g^t\|^4 + \|g^t\|^2 \end{pmatrix}\right)$$

denote the associated density by

$$\varphi_{t,\gamma}(x, y_f, y_g).$$

Thus, the ~~co~~ joint density of $\left(X_t, \langle Y^t, f^t\rangle, \langle Y^t, g^t\rangle\right)$ (unconditioned on B) under Q^* is:

$$\frac{1}{2}\left(\Phi_{t,\gamma}(x, y_f, y_g) + \varphi_{t,\gamma}(x, y_f, y_g)\right)$$

~~and the joint density of~~

implying:

$$\textcircled{XI}$$

$$\hat{X}_t(Y^t) = E_{Q^*}\left[X_t \,/\, \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle \right]$$

$$= \frac{\int_{-\infty}^{\infty} dx \cdot x \left[\phi_{t,\gamma}\left(x, \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle\right) + \varphi_{t,\gamma}\left(x, \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle\right) \right]}{\int_{-\infty}^{\infty} \left[\phi_{t,\gamma}\left(x', \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle\right) + \varphi_{t,\gamma}\left(x', \langle Y^t, f^t \rangle, \langle Y^t, g^t \rangle\right) \right] dx'}$$