Solutions 11

2022.05.20

Solution 1.

(a) We want to minimize the error

$$E[n] \quad = \quad D[n] - Y[n] = D[n] - \sum_i f_n[i]X[n-i]$$

where

$$X[n] \quad = \quad \sum_k h[k]D[n-k] + S[n].$$

If we define the cost function as

$$J_n = \mathbb{E}[E[n]^2],$$

then the *normal* equations imply:

$$\sum_{j=0}^{L-1} f_n[j]R_X[n-j,n-i] = R_{DX}[n,n-i] \quad i = 0,\ldots,L-1, \quad \forall n \in \mathbb{Z}.$$

Now, we compute $R_X[n-j,n-i]$ and $R_{DX}[n,n-i]$ for the case where $D[n] = V_0[n]$.

$$\begin{aligned}
R_X[n-j,n-i] &= \mathbb{E}[(h[0]V_0[n-j] + h[1]V_0[n-j-1] + S[n-j]) \\
&\quad (h[0]V_0[n-i] + h[1]V_0[n-i-1] + S[n-i])] \\
&= 2R_{V_0}[i-j] + R_{V_0}[i-j+1] + R_{V_0}[i-j-1] + \sigma_S^2\delta[i-j] \\
&= 2\rho_0^{|i-j|} + \rho_0^{|i-j+1|} + \rho_0^{|i-j-1|} + \sigma_S^2\delta[i-j],
\end{aligned}$$

$$\begin{aligned}
R_{DX}[n,n-i] &= \mathbb{E}[V_0[n] \cdot (h[0]V_0[n-i] + h[1]V_0[n-i-1] + S[n-i])] \\
&= R_{V_0}[i] + R_{V_0}[i+1] \\
&= \rho_0^{|i|} + \rho_0^{|i+1|},
\end{aligned}$$

We can see that the processes $D[n]$ and $X[n]$ are stationary BUT the filter f_n is not a Wiener filter since we limit its length to $L = 3$.

From the Yule-Walker equation we have

$$f_n = R_{X,n}^{-1} R_{DX,n}.$$

$$\begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} = \begin{bmatrix} 2+2\rho_0+1 & 2\rho_0+1+\rho_0^2 & 2\rho_0^2+\rho_0+\rho_0^3 \\ 2\rho_0+1+\rho_0^2 & 2+2\rho_0+1 & 2\rho_0+\rho_0^2+1 \\ 2\rho_0^2+\rho_0+\rho_0^3 & 2\rho_0+1+\rho_0^2 & 2+2\rho_0+1 \end{bmatrix}^{-1} \begin{bmatrix} 1+\rho_0 \\ \rho_0+\rho_0^2 \\ \rho_0^2+\rho_0^3 \end{bmatrix}.$$

Replacing $\rho_0 = 1/2$, we get

$$f_n = [0.3944 \quad -0.0361 \quad 0.0031]^T.$$

We compute $\mathbb{E}[|E[n]|^2]$ from the formula:

$$\begin{aligned}\mathbb{E}[|E[n]|^2] &= \mathbb{E}[(D[n] - f^T X_n)^2] \\ &= \sigma_D^2 + f^T R_X f - 2 f^T R_{DX} = 0.4343.\end{aligned}$$

When the switch is in position "1", all the steps are the same and we need to change V_0 to V_1. In that case we have,

$$\begin{aligned}\begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} &= \begin{bmatrix} 2+2\rho_1+1 & 2\rho_1+1+\rho_1^2 & 2\rho_1^2+\rho_1+\rho_1^3 \\ 2\rho_1+1+\rho_1^2 & 2+2\rho_1+1 & 2\rho_1+\rho_1^2+1 \\ 2\rho_1^2+\rho_1+\rho_1^3 & 2\rho_1+1+\rho_1^2 & 2+2\rho_1+1 \end{bmatrix}^{-1} \begin{bmatrix} 1+\rho_1 \\ \rho_1+\rho_1^2 \\ \rho_1^2+\rho_1^3 \end{bmatrix} \\ &= \begin{bmatrix} 0.3999 \\ -0.0797 \\ 0.0144 \end{bmatrix}\end{aligned}$$

and

$$\mathbb{E}[|E[n]|^2] = \mathbb{E}[(D[n] - f^T X_n)^2] = 0.4912.$$

(b) - We need to distinguish the four cases:

$$R_D[n,m] = \begin{cases} \rho_0^{|n-m|} & \text{n, m even,} \\ \rho_1^{|n-m|} & \text{n, m odd,} \\ 0 & \text{n even, m odd,} \\ 0 & \text{n odd, m even.} \end{cases}$$

Clearly, the process $D[n]$ is not stationary.

- Let us call $D_h[n] = \sum_k h[k]D[n-k] = D[n] + D[n-1]$. Then

$$\begin{aligned}R_X[n,m] &= \mathbb{E}[X[n]X[m]] = \mathbb{E}[(D_h[n] + S[n])(D_h[m] + S[m])] \\ &= R_{D_h}[n,m] + R_S[n,m] = R_{D_h}[n,m] + \sigma_S^2 \delta[n-m]\end{aligned}$$

and

$$\begin{aligned}R_{D_h}[n,m] &= \mathbb{E}[(D[n] + D[n-1])(D[m] + D[m-1])] \\ &= R_D[n,m] + R_D[n,m-1] + R_D[n,m-1] \\ &\quad + R_D[n-1,m] + R_D[n-1,m-1] \\ &= \begin{cases} \rho_0^{|n-m|} + 0 + 0 + \rho_1^{|n-m|} & \text{n, m even,} \\ \rho_1^{|n-m|} + 0 + 0 + \rho_0^{|n-m|} & \text{n, m odd,} \\ 0 + \rho_0^{|n-m+1|} + \rho_1^{|n-m-1|} + 0 & \text{n even, m odd,} \\ 0 + \rho_1^{|n-m+1|} + \rho_0^{|n-m-1|} + 0 & \text{n odd, m even.} \end{cases}\end{aligned}$$

The correlation R_{DX} is equal to

$$\begin{aligned} R_{DX}[n,m] &= \mathbb{E}[D[n]X[m]] = \mathbb{E}[D[n](D[m]+D[m-1]+S[m])] \\ &= R_D[n,m]+R_D[n,m-1] \\ &= \begin{cases} \rho_0^{|n-m|}+0 & \text{n, m even,} \\ \rho_1^{|n-m|}+0 & \text{n, m odd,} \\ 0+\rho_0^{|n-m+1|} & \text{n even, m odd,} \\ 0+\rho_1^{|n-m+1|} & \text{n odd, m even.} \end{cases} \end{aligned}$$

Clearly, the process is not stationary.

- We have the normal equation

$$\sum_{j=0}^{L-1} f_n[j] R_X[n-j,n-i] = R_{DX}[n,n-i] \quad i=0,\ldots,L-1, \quad \forall n \in \mathbb{Z}$$

and to evaluate R_X and R_{DX}, we need to consider the cases where n is even and n is odd.

Let us, for example, consider the case where n is even. Then,

$$\begin{aligned} \begin{bmatrix} f_n[0] \\ f_n[1] \\ f_n[2] \end{bmatrix} &= \begin{bmatrix} R_X[n,n] & R_X[n-1,n] & R_X[n-2,n] \\ R_X[n,n-1] & R_X[n-1,n-1] & R_X[n-2,n-1] \\ R_X[n,n-2] & R_X[n-1,n-2] & R_X[n-2,n-2] \end{bmatrix}^{-1} \begin{bmatrix} R_{DX}[n,n] \\ R_{DX}[n,n-1] \\ R_{DX}[n,n-2] \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1+\rho_0^2 & \rho_0^2+\rho_1^2 \\ \rho_0^2+1 & 3 & 1+\rho_1^2 \\ \rho_0^2+\rho_1^2 & \rho_1^2+1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \rho_0^2 \\ \rho_0^2 \end{bmatrix} \end{aligned}$$

and we compute

$$f_n = [0.3644 \quad -0.0963 \quad 0.0752]^T.$$

Applying the same formula for computing the error as in the previous part, we get:

$$\mathbb{E}[|E[n]|^2] = 0.6409$$

The process $D[n]$ is not stationary and this explains why the error is larger than in the part a) for both $D[n] = V_0[n]$ and $D[n] = V_1[n]$.

(c) Since the position of the switch is randomly chosen we can introduce the random variable $S_W[n]$ that describe the position of the switch. Positions $p0$ and $p1$ appear with the same probability of $1/2$. To compute $R_D[n,m] = \mathbb{E}[D[n]D[m]]$, we can use the following formula:

$$\begin{aligned} \mathbb{E}[f(D)] &= \mathbb{E}[\mathbb{E}[f(D)|S_W]] \\ &= \frac{1}{4}\mathbb{E}[f(D)|s_W=(0,0)] + \frac{1}{4}\mathbb{E}[f(D)|s_W=(0,1)] \\ &\quad + \frac{1}{4}\mathbb{E}[f(D)|s_W=(1,0)] + \frac{1}{4}\mathbb{E}[f(D)|s_W=(1,1)]\,. \end{aligned}$$

Then

$$\begin{aligned} R_D[n,m] &= \mathbb{E}[D[n]D[m]] \\ &= \frac{1}{4}\mathbb{E}[D[n]D[m]|s_W=(0,0)] + \frac{1}{4}\mathbb{E}[D[n]D[m]|s_W=(1,1)] \\ &= \frac{1}{4}\rho_0^{|n-m|} + \frac{1}{4}\rho_1^{|n-m|} \end{aligned}$$

The process is stationary.

To compute the optimal filter we need:

$$\begin{aligned} R_X[n,m] &= \mathbb{E}[X[n]X[m]] = \mathbb{E}[(D[n]+D[n-1]+S[n])(D[m]+D[m-1]+S[m])] \\ &= 2R_D[n-m]+R_D[n-m+1]+R_D[n-m-1]+\sigma_S^2\delta[n-m], \end{aligned}$$

and

$$\begin{aligned} R_{DX}[n,m] &= \mathbb{E}[D[n]X[m]] = \mathbb{E}[D[n](D[m]+D[m-1]+S[m])] \\ &= R_D[n-m]+R_D[n,m-1]. \end{aligned}$$

Changing in the normal equation we get:

$$\begin{bmatrix} f[0] \\ f[1] \\ f[2] \end{bmatrix} = \begin{bmatrix} 2.4167 & 1.0069 & 0.4294 \\ 1.0069 & 2.4167 & 1.0069 \\ 0.4294 & 1.0069 & 2.5167 \end{bmatrix}^{-1} \begin{bmatrix} 0.7083 \\ 0.2986 \\ 0.1308 \end{bmatrix} = \begin{bmatrix} 0.2923 \\ 0.0010 \\ 0.0018 \end{bmatrix}$$

Solution 2.

1) The signal has the form $x(t) = \sum_{k=1}^{100} \alpha_k \delta(t-\tau_k)\,, \quad t \in [0,T]$, where $\tau_1,\ldots,\tau_{100}$ are the positions of the spikes. The corresponding Fourier transform reads

$$\widehat{x}[n] = \frac{1}{T}\sum_{k=1}^{100} \alpha_k e^{-j2\pi n\tau_k/T} = \frac{1}{T}\sum_{k=1}^{100} \alpha_k e^{-j\omega_k n}\,, \quad \omega_k = 2\pi\tau_k/T\,.$$

The position of the spikes can be estimated using the annihilating filter approach. That is, we look for the filter h such that $(\widehat{x} * h)[n] = 0$, where h has 100 coefficients. In Matrix form we obtain

$$\begin{bmatrix} \widehat{x}[99] & \ldots & \widehat{x}[0] \\ \widehat{x}[100] & \ldots & \widehat{x}[1] \\ \vdots & & \vdots \\ \widehat{x}[198] & \ldots & \widehat{x}[99] \end{bmatrix} \begin{bmatrix} h[1] \\ h[2] \\ \vdots \\ h[100] \end{bmatrix} = - \begin{bmatrix} \widehat{x}[100] \\ \widehat{x}[101] \\ \vdots \\ \widehat{x}[199] \end{bmatrix}$$

The solution of such a linear system provides $h[1],\ldots,h[100]$ and therefore $H[z] = 1 + h[1]z^{-1} + \ldots + h[100]z^{-100}$. Call the $a_1,\ldots,a_{100}$ the roots of the latter equation, then the equalities $a_1 = e^{i2\pi\tau_1/T}$,...,$a_{100} = e^{i2\pi\tau_{100}/T}$ yield positions $\tau_1,\ldots,\tau_{100}$.

2) The likelihood function for a 20-sample observation $x[1],\ldots,x[20]$ of a Markov chain $X[n]$ reads

$$\begin{aligned} h(x[1],\ldots,x[20];\Theta) =& \mathbb{P}\left(X[1]=x[1],\ldots,X[20]=x[20]\right) \\ =& \pi_{x[1]} p_{x[1]x[2]} p_{x[2]x[3]} \cdots p_{x[19]x[20]}\,, \end{aligned}$$

and in our case 00011010110110111011

$$\begin{aligned} h(x[1],\ldots,x[20];\Theta) &= \mathbb{P}\left(X[1]=0,\ldots,X[20]=1\right) \\ &= \pi_0 p_{00} p_{00} p_{01} p_{11} p_{10} p_{01} p_{10} p_{01} p_{11} p_{10} p_{01} p_{11} p_{10} p_{01} p_{11} p_{11} p_{10} p_{01} p_{11} \\ &= \pi_0 p_{00}^2 p_{01}^6 p_{11}^6 p_{10}^5\,. \end{aligned}$$

3) We need to maximize $h(x[1],\ldots,x[20];\Theta)$ or $\log h(x[1],\ldots,x[20];\Theta)$ with respect to π_i and p_{ij}, $i,j=0,1$ under the constraints $\pi_0+\pi_1=1$, $p_{00}+p_{01}=1$, and $p_{10}+p_{11}=1$. Using Lagrange maximizers and the log likelihood, we must maximize

$$\log(h(x[1],\ldots,x[20];\Theta)) - \lambda_1(\pi_0+\pi_1-1) - \lambda_2(p_{00}+p_{01}-1) - \lambda_3(p_{10}+p_{11}-1)$$

that is

$$\log\pi_0 + 2\log p_{00} + 6\log p_{01} + 6\log p_{11} + 5\log p_{10} - \lambda_1(\pi_0+\pi_1-1) - \lambda_2(p_{00}+p_{01}-1) - \lambda_3(p_{10}+p_{11}-1)\,.$$

Maximization w.r.t.:

- π_0 and π_1 gives $\pi_0 = 1$
- p_{00} gives $2/p_{00} - \lambda_2 = 0$, *i.e.* $p_{00} = 2/\lambda_2$, and w.r.t. p_{01} gives $6/p_{01} - \lambda_2 = 0$, *i.e.* $p_{01} = 6/\lambda_2$. Considering the constraint $p_{00}+p_{01}=1$ we obtain $\lambda_2 = 8$ and therefore $p_{00} = 1/4$ and $p_{01} = 3/4$.
- p_{10} gives $5/p_{10} - \lambda_3 = 0$,*i.e.* $p_{10} = 5/\lambda_3$, and w.r.t. p_{11} gives $6/p_{11} - \lambda_3 = 0$, *i.e.* $p_{11} = 6/\lambda_3$. Considering the constraint $p_{10}+p_{11}=1$ we obtain $\lambda_3 = 11$ and therefore $p_{10} = 5/11$ and $p_{11} = 6/11$.

4)

$$\begin{aligned}
F_{Y[n]}(y) = \mathbb{P}\left(Y[n] \le y\right) &= \sum_{x\in 0,1} \mathbb{P}\left(Y[n]\le y, X[n]=x\right) \\
&\overset{\text{Bayes}}{=} \sum_{x\in 0,1} \mathbb{P}\left(Y[n]\le y \mid X[n]=x\right)\mathbb{P}\left(X[n]=x\right) \\
&= \sum_{x\in 0,1} \mathbb{P}\left(W[n]+x\le y\right)\mathbb{P}\left(X[n]=x\right) \\
&= \mathbb{P}\left(W[n]\le y\right)\mathbb{P}\left(X[n]=0\right) + \mathbb{P}\left(W[n]+1\le y\right)\mathbb{P}\left(X[n]=1\right) \\
&= \mathbb{P}\left(W[n]\le y\right)\pi_0 + \mathbb{P}\left(\widetilde{W}[n]\le y\right)\pi_1\,,
\end{aligned}$$

where $W[n]$ is a centered Gaussian process with variance σ_W^2 and $\widetilde{W}[n]$ is a Gaussian process with mean 1 and variance σ_W^2. The probability density function then reads

$$f_{Y[n]}(y) = \mathcal{G}_{0,\sigma_W^2}(y)\pi_0 + \mathcal{G}_{1,\sigma_W^2}(y)\pi_1\,.$$

where $\mathcal{G}_{m,\sigma^2}(y)$ denotes a Gaussian probability density function with mean m and variance σ^2.

5) Notice that, as seen in class, in order to compute the probability density function we first need to compute the cumulative distribution function.

Call $\mathcal{X} = \{(x[1],\ldots,x[10]) \mid x[i]\in\{0,1\}\}$ the set of all possible combinations of 0 and 1 so to form a vector of length 10.

Then the joint cumulative distribution of $\boldsymbol{Y} = [Y[1],\ldots,Y[10]]$ reads

$$\begin{aligned}
F_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathbb{P}\left(\boldsymbol{Y}\le\boldsymbol{y}\right) &= \sum_{\boldsymbol{x}\in\mathcal{X}} \mathbb{P}\left(\boldsymbol{Y}\le\boldsymbol{y}, \boldsymbol{X}=\boldsymbol{x}\right) \\
&\overset{\text{Bayes}}{=} \sum_{\boldsymbol{x}\in\mathcal{X}} \mathbb{P}\left(\boldsymbol{Y}\le\boldsymbol{y} \mid \boldsymbol{X}=\boldsymbol{x}\right)\mathbb{P}\left(\boldsymbol{X}=\boldsymbol{x}\right)\,.
\end{aligned}$$

$\mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x})$ is the distribution of $\boldsymbol{Y} = \boldsymbol{x} + \boldsymbol{W}$ (distribution of $\boldsymbol{Y}$ given that $\boldsymbol{X}$ is known), *i.e.*, distribution of i.i.d. Gaussian random variables, with the same variance σ_W^2 and means given by the vector $\boldsymbol{x}$.

$$\mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}) = F_{\boldsymbol{W}+\boldsymbol{x}}(\boldsymbol{y}) = \prod_{n=1}^{10} F_{W[n]+x[n]}(y[n]),$$

which in terms of probability density reads

$$f_{\boldsymbol{W}+\boldsymbol{x}}(\boldsymbol{y}) = \prod_{n=1}^{10} \mathcal{G}_{x[n],\sigma_W^2}(y[n]), \text{ and } f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \prod_{n=1}^{10} \mathcal{G}_{x[n],\sigma_W^2}(y[n]) \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}).$$

Given that

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \pi_{x[1]} p_{x[1]x[2]} \cdots p_{x[9]x[10]},$$

we have

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \left(\prod_{n=1}^{10} \mathcal{G}_{x[n],\sigma_W^2}(y[n]) \right) \pi_{x[1]} p_{x[1]x[2]} \cdots p_{x[9]x[10]}$$

6) We need here to apply PCA. Call

$$\boldsymbol{y}_m = [y[1]_m, \ldots, y[30]_m]^T, \quad m = 1, \ldots, 20000, \quad (M = 20000, N = 30).$$

the vector containing the 30 samples of the m-th shape. Then

- Create zero mean data by averaging all the M=20000 shapes, that is

$$\boldsymbol{y}_{\text{mean}} = \frac{1}{20000} \sum_{m=1}^{20000} \boldsymbol{y}_m.$$

  For every m, center the m-th shape by subtracting the mean

$$\widetilde{\boldsymbol{y}}_m = \boldsymbol{y}_m - \boldsymbol{y}_{\text{mean}}.$$

- Compute the empirical correlation matrix (using the centered data $\widetilde{\boldsymbol{y}}_m$)

$$\widehat{\mathbf{R}}_{\boldsymbol{y}} = \frac{1}{20000} \sum_{m=1}^{20000} \widetilde{\boldsymbol{y}}_m * \widetilde{\boldsymbol{y}}_m^H$$

- Compute the unitary matrix V of eigenvectors of $\widehat{\mathbf{R}}$ and the eigenvalues by solving the equation

$$\widehat{\mathbf{R}}_{\boldsymbol{y}} \boldsymbol{V} = \boldsymbol{V} \boldsymbol{\Lambda}$$

  where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is the eigenvalues diagonal matrix and $\boldsymbol{V}^H \widehat{\mathbf{R}}_{\boldsymbol{y}} \boldsymbol{V} = \boldsymbol{\Lambda}$.
- Look at the eigenvalues and select those with highest values, for instance, $\lambda_1, \ldots \lambda_k$, where $k < N$ (usually $k << N$).
- Compute the principal components

$$\boldsymbol{z}_m = \boldsymbol{V}^T \boldsymbol{y}_m, \text{ for } m = 1, \ldots, 20000.$$

- Select the principal components corresponding to the eigenvalues with highest values, for instance, $\boldsymbol{z}[1]_m, \ldots, \boldsymbol{z}[k]_m$, $m = 1, \ldots, 20000$, and plot them in a k-dimensional plot so to recognize the clusters and therefore the number of different shapes.