Statistical Signal Processing

Final Exam

Monday, June 18, 2012

You will hand in this sheet together with your solutions.

Write your personal data.

Family Name:

Name:

E-mail:

Statistical Signal Processing

Final Exam

Monday, June 18, 2012

Read Me First!

Only a personal cheat sheet is allowed.
No class notes, no exercise text or exercise solutions.

Write solutions on separate sheets,
***i.e.* no more than one solution per paper sheet.**

Return your sheets ordered according to problem (solution) numbering.

Return the text of the exam.

Warmup exercises

This is a warm up problem .. do not spend too much time on it.

Please provide justified, rigorous, and simple answers.

Exercise 1. AVERAGING PERIODOGRAM (3PTS)

The signal $X[n]$ is a zero mean Gaussian white noise with variance σ^2. We have measured N points of $X[n]$, and would like to use the periodogram $P_X^N(\omega)$ to estimate the power spectrum density (PSD). We know that the variance of this estimator is

$$\mathrm{Var}\left(P_X^N(\omega)\right) = \sigma^2 ,$$

i.e., a constant variance, no matter how many points N you have measured!

Now we split the measured signal $(X[1], X[2], \cdots, X[N])$ into two parts

$$Y_1 = (X[1], \cdots, X[N/2]) , \quad Y_2 = (X[N/2], \cdots, X[N]) .$$

We denote the periodograms of these two parts as $P_{Y_1}(\omega)$ and $P_{Y_2}(\omega)$, respectively. Then we compute the average of these two periodograms

$$Q(\omega) = \frac{1}{2}(P_{Y_1}(\omega) + P_{Y_2}(\omega)).$$

$Q(\omega)$ provides a new estimator of the PSD of $X[n]$. What is the variance of this estimator $\mathrm{Var}(Q(\omega))$? (*Hint: Y_1 and Y_2 are independent with each other.*)

Solution 1.

$$\begin{aligned}
\mathrm{Var}(Q(\omega)) &= \mathbb{E}\left|Q(\omega) - \mathbb{E}Q(\omega)\right|^2 \\
&= \frac{1}{4}\mathbb{E}\left|P_{Y_1}(\omega) - \mathbb{E}P_{Y_1}(\omega) + P_{Y_2}(\omega) - \mathbb{E}P_{Y_2}(\omega)\right|^2 \\
&= \frac{1}{4}(\mathrm{Var}(P_{Y_1}(\omega)) + \mathrm{Var}(P_{Y_2}(\omega))) \\
&= \frac{1}{2}\sigma^2
\end{aligned}$$

The third equality is due to the independence between Y_1 and Y_2 (and thus $P_{Y_1}(\omega)$) and $P_{Y_2}(\omega)$)).

Exercise 2. PROBABILITIES AND STOCHASTIC PROCESSES (3PTS)

These are simple true/false questions, each counting 0.6 points.
NOTE: Don't answer randomly: Each wrong answer will count for -0.6 points! (score of the exercise: 0 to 3)

Let X, Y, Z be continuous-valued random variables. Without further conditions, the following statements are true or false?

1) If X admits a probability density function $f_X(a)$, then $\mathbb{P}(X = a) = f_X(a)$.

2) If $Z = X + Y$, the cumulative distribution function of Z can be derived by convolution, i.e., $F_Z(a) = F_X(a) * F_Y(a)$.

3) If $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then X and Y are independent.

Let $X[n]$, $Y[n]$ and $Z[n]$ be a stochastic processes. Without further conditions, the following statements are true or false?

4) If $\mathbb{E}[X[n]] = 1$ and $\mathbb{E}[X[k]X[l]^*] = k - 2l$, then $X[n]$ is wide sense stationary.

5) If $X[n]$ and $Y[n]$ are i.i.d. Poisson distributed with mean λ, $Z[n] = X[n]+Y[n]$ has Poisson distribution with mean 2λ.

Solution 2.

1) False. $P(X = a) = 0$.

2) False. The condition that X and Y are independent is required.

3) False. It is true only when X and Y are jointly Gaussian.

4) False. In this case, the autocorrelation is not a function of time difference $k-l$.

5) True.

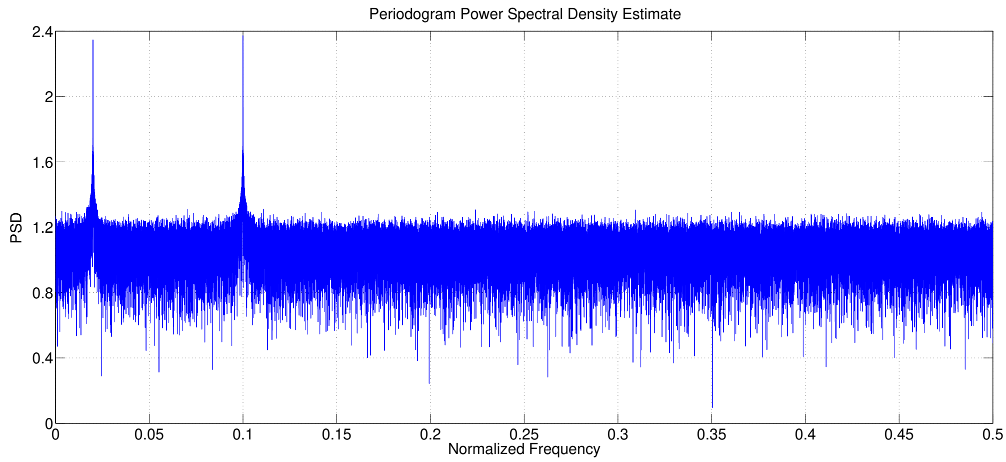
Main exercises

Here comes the core part of the exam .. take time to read the introduction and each problem statement.

Please provide justified, rigorous, and simple answers.

Exercise 3. ALP HORN CONCERT (20PTS)

During a Concert, an alp horn orchestra plays for 5 seconds a combination of tones (each tone is played for 5 seconds), so to provide a signal that can be considered stationary. We have recorded 100000 samples $[X[1], \ldots, X[100000]]$ at 20KHz and the corresponding periodogram S_X reads



Notice that such a spectrum presents two spikes at 0.02 and 0.1 (normalized frequencies) and a constant baseline.

1) What kind of signal has generated such spectrum? (it can be the sum of different signals)

The spectrum suggests that the alp horn orchestra have played **at least** 2 tones

2) Recalling that the sampling frequency is 20KHz, and that the two spikes are, in normalized frequencies, at 0.02 and 0.1, provide the frequencies in Hz of the two tones.

3) Recalling that we have recorded 100000 samples, is it possible that the alp horn orchestra played more than two tones? Is yes, what are the possible frequencies of the additional tones? if no, why?

We make another recording of the alp horn orchestra. The orchestra plays another combination of tones, but this time, we record only 1 second instead of 5 (each tone is played for 1 seconds). We keep the same sampling frequency as before (20KHz).

We do not know exactly how many tones the orchestra has played but we assume that at maximum is has played 5 different tones. We assume that the recording in made with noise-free microphones.

4) Propose a parametric method to estimate the frequencies (only the frequencies!) of the played tones and justify you choice.

5) Precisely describe such method: You are given the samples corresponding to a recording of 1 second at 20KHz and you are asked to detail each step like if you have to implement the method in a computer.

6) What happens if the recording is corrupted by noise? Which of the steps that you have precisely described is affected?

Solution 3.

1) The signal that has generated the spectrum is the sum of a two sinusoids (spikes at 0.02 and 0.1) and a white noise (constant baseline), that is

$$X[n] = e^{i2\pi n0.02+\Theta_1} + e^{-i2\pi n0.02+\Theta_1} + e^{i2\pi n0.1+\Theta_2} + e^{-i2\pi n0.1+\Theta_2} + W[n]\,.$$

where Θ_1 and Θ_2 are independent random variables uniformly distributed over $[0, 2\pi]$.

2) $f_1 = 0.02 \times 20\text{KHz} = 400\text{Hz}$, and $f_1 = 0.1\ \times 20\text{KHz} = 2\text{KHz}$.

3) Yes, it depends on the spectral resolution. Here the spectral resolution is $1/100000$, therefore any tone (spectral line) in the intervals $[0.02-\frac{1}{100000}, 0.02+\frac{1}{100000}]$ and $[0.1-\frac{1}{100000}, 0.1+\frac{1}{100000}]$ will not be visible in the periodogram estimate of the power spectral density.

4) Here the spectral estimation method to be used is the annihilating filter. Indeed, we want to estimate the position of spectral lines (frequencies) and the data is assumed to be noiseless. Any other method will not directly, optimally, and precisely provide the position of the spectral lines.

5) We now have recorded 20000 samples, *i.e.*, $x[1], \ldots, x[20000]$, and we assume that the orchestra has played at most 5 different tones. Therefore, we look for a filter h with 10 coefficients such that $(x * h)[n] = 0$. The steps for estimating the frequencies with the annihilating filter approach are:

   1) Write $(x * h)[n] = 0$ in matrix form

$$\begin{bmatrix} x[10] & \ldots & x[1] \\ x[9] & \ldots & x[2] \\ \vdots & & \vdots \\ x[1] & \ldots & x[10] \end{bmatrix} \begin{bmatrix} h[1] \\ h[2] \\ \vdots \\ h[10] \end{bmatrix} = - \begin{bmatrix} x[11] \\ x[7] \\ \vdots \\ x[20] \end{bmatrix}$$

   obtaining a linear system. Notice that there the samples are $x[1], \ldots$ and NOT $x[0], \ldots$.

2) Solve the linear system (the matrix is Toeplitz, therefore its solution requires K^2 computations), obtaining $h[1], \ldots, h[5]$.

3) Compute the z transform $H[z] = 1 + h[1]z^{-1} + \ldots + h[10]z^{-10}$.

4) Find the zeroes of $H[z] = 1 + h[1]z^{-1} + \ldots + h[10]z^{-10}$, *i.e.*, the solutions (roots) of the equation $1 + h[1]z^{-1} + \ldots + h[10]z^{-10} = 0$. Call $a_1, \ldots, a_{10}$ the zeroes, giving $H(z) = (1 - a_1 z^{-1}) \ldots (1 - a_{10} z^{-1})$.

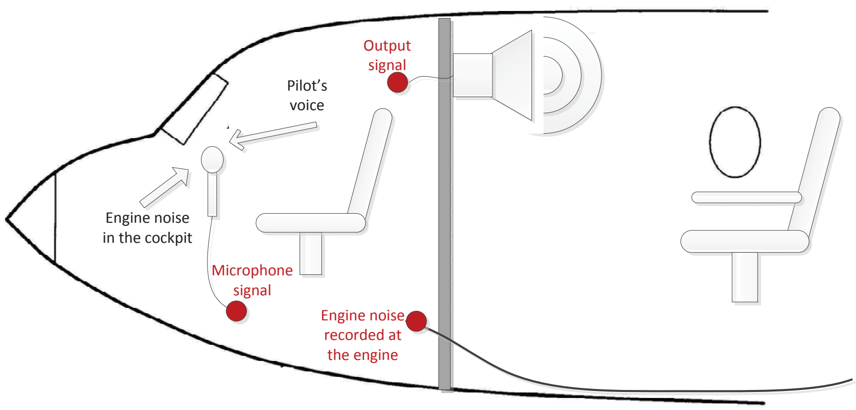
5) Through the relations $a_1 = e^{i2\pi f_1}$,...,$a_{10} = e^{i2\pi f_{10}}$ we can compute the (normalized) frequencies f_1,...,f_{10}: The frequency is given by the argument of the complex exponential divided by 2π, *i.e.*, $f_n = \ln a_n / i2\pi$ (to account for numerical errors providing non unitary roots, we can use $f_n = \arg a_n / 2\pi$.

6) If the recording is corrupted by noise the zero finding $(1 + h[1]z^{-1} + \ldots + h[10]z^{-10} = 0)$ becomes a very critical operation, since highly affected by the noise. The 10 roots we obtain using noisy data present a large error when compared to the roots we would have obtained using noiseless data. Consequently, the 10 frequencies we obtain from the 10 roots might be quite different form the true frequencies.

Exercise 4. Noise cancelation

Consider a pilot in an airplane giving an announcement to the passengers. When the pilot speaks into a microphone, the engine noise in the cockpit combines with his voice signal. This additional noise makes the resultant signal heard by passengers of low quality.

The goal of this exercise is to design a system that reduced the engine noise from the microphone signal and tries to output the pilot's voice only. To this end, we are also given a sample of the engine noise directly from the engine itself, see Figure below.



1) Design an adaptive filtering system that will perform engine noise cancellation and output the pilot's voice.

   1.1) Draw a schematic of this system and mark clearly the input processes, the output process, and the error process E[n] of the adaptive filter.

   1.2) What signal statistics your system need? Propose a method to learn about these statistics.

   1.3) Propose an adaptation algorithm and show that your noise cancellation scheme will indeed cancel the engine noise. Justify the answer by writing the expression for the error process E[n].

2) Assume you choose the LMS adaptation algorithm. Write a filter coefficient update procedure.

3) Assuming that you are not given any information about the statistics of the noise signal in advance, propose a method to estimate an appropriate range form the step size μ of the LMS algorithm.

4) Compare the performance of LMS with RLS in terms of the speed and computational complexity.

5) Propose a variation of LMS algorithm and explain what the advantages of this variation are. At what cost we have these advantages? (Hint: Remember the computer exercise.)

Solution 4. NOISE CANCELATION

1) 1.1) Let us denote the signal recorded at the microphone by $D[n]$ and the engine noise recorded directly at the engine as $X[n]$.

   The input processes of the adaptive filter are: 1) the signal $X[n]$; and 2) the error process $E[n] = D[n] - f * X[n]$.

   The output of the adaptive filter is $Y[n] = f * X[n]$ and it represents the estimated engine noise that should be subtracted from the microphone signal.

   1.2) We need the correlation matrix R_X and R_{DX}, which we can estimate empirically. For the ease of notation we assume a stationary case:

$$\begin{aligned}\hat{R}_X[k] &= \frac{1}{M-k}\sum_{i=0}^{M-k-1} x[n-i]x[(n-i)-k] \\ \hat{R}_{DX}[k] &= \frac{1}{M-k}\sum_{i=0}^{M-k-1} d[n-i]x[(n-i)-k].\end{aligned} \tag{1}$$

   1.3) Let us denote the pilot's voice by $S[n]$ and the engine noise that is recorded by microphone inside cockpit by $V[n]$. Assuming that the pilot's voice is uncorrelated with the engine noise we propose an adaptation algorithm that will minimize the following error in the mean squared sense:

$$\begin{aligned}\mathbb{E}[E[n]^2] &= \mathbb{E}[(D[n]-Y[n])^2] = \mathbb{E}[(S[n]+V[n]-Y[n])^2] \\ &= \mathbb{E}[S[n]^2] + 2\mathbb{E}[S[n](V[n]-Y[n])] + \mathbb{E}[(V[n]-Y[n])^2] \\ &= \mathbb{E}[S[n]^2] + \mathbb{E}[(V[n]-Y[n])^2]\end{aligned} \tag{2}$$

   As we can see, by minimizing $E[n]$ we are going to minimize $\mathbb{E}[(V[n] - Y[n])^2$, that is, our adaptive filter will estimate the engine noise recorded at the microphone and we will subtract that from the microphone signal.

2)

$$f_{n+1} = f_n + \mu X_n E[n], \tag{3}$$

$$f_0 = \text{initial guess.} \tag{4}$$

3) A range of μ that is frequently used is given by

$$0 < \mu < \frac{2}{LS_{X,\max}}, \tag{5}$$

   where $S_{X,\max}$ is the maximum value of the power spectral density $S_X(\omega)$ of the process $X[n]$.

4) The RLS algorithm converges faster but it has higher computational complexity. The computational complexity of LMS is $O(L)$ and of RLM is $O(L^2)$ where L is the filter length.

5) We can propose a normalized LMS method, nLMS. The advantages of nLMS over LMS is that it has an adaptive step size that adjusts depending on the signal energy. However, It has a slightly higher complexity but still within the same order of $O(L)$ as LMS.

Exercise 5. TRANSMITTING PULSES (26PTS)

A transmitted signal is composed of a sequence of spikes. The interval between each spike codes the information to be transmitted. In particular, the interval is a when a bit 0 is transmitted, and $2a$ when a bit 1 is transmitted.

We observe the spikes over an interval [0,T].

1) Knowing that in such an interval there are at the most 100 spikes, precisely describe a method to estimate the positions of the spikes and then the interval between the spikes.

Markov Coding

Using the method you have described, we have been able to decode the information and we have obtained the following sequence of bits: 00011010110110111011, where $x[1] = 0$ and $x[20] = 1$. The coding is a Markovian one, that is, the process $X[n]$ that generates the bits $\{0, 1\}$ is a Markov chain.

2) Write the likelihood function associated to the observations $x[1], \dots, x[20]$.

3) Using the maximum likelihood and the Lagrange multipliers, estimate the parameters of the Markov chain based on the observations $x[1], \dots, x[20]$.

The channel corrupts the transmission with an additive Gaussian white noise $W[n]$ (centered, with variance σ_W^2). The decoded signal can be the written as $Y[n] = X[n] + W[n]$.

4) Write the probability distribution and then the probability density function of $Y[n]$.

5) Using Baye's rule, write the joint probability density function of $Y[1], \dots, Y[10]$.

Transforming Spikes into Pulses

Engineers are not sure if transmitting spikes with the information 0 1 coded into the spike distance is an optimal method. They decide to transmit pulses and to increase the coding symbols: instead of only 0 and 1, the information is coded into different symbols and for each symbol there is a different pulse shape.

At the receiver end each shape is received as a sequence of 30 samples, and a total of 20000 shapes are received. The problem now is that at the receiver end the number of possible different shapes (and therefore of coding symbols) is unknown and has to be determined. That is, the 20000 shapes received need to be analyzed so to understand of how many different shapes the transmission is composed.

6) Given the received data (20000 shapes of 30 samples each) precisely describe a method seen in class that enables to simply understand of how many different shapes the transmission is composed. You are asked to detail each step like if you have to implement the method in a computer.

Solution 5.

1) The signal has the form $x(t) = \sum_{k=1}^{100} \alpha_k \delta(t - \tau_k)$, $\quad t \in [0, T]$, where $\tau_1, \dots, \tau_{100}$ 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

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

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

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

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

and in our case 00011010110110111011

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

3) We need to maximize $h(x[1], \dots, x[20]; \Theta)$ or $\log h(x[1], \dots, x[20]; \Theta)$ 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 the Lagrange maximizers and the log likelihood we have to maximize

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

that is

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

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)

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

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

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

$\mathbb{P}\left(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}\right)$ 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 $\boldsymbol{x}$.

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

which in terms of probability density reads

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

Given that

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

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 $\widetilde{\boldsymbol{y}}_m$)

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

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

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

  where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_N)$ is the eigenvalues diagonal matrix and $\boldsymbol{V}^H \widehat{\boldsymbol{R}}_{\boldsymbol{y}} \boldsymbol{V} = \boldsymbol{\Lambda}$.
- 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.

Grading

The complete solution of the exam provides 76 points. To obtain a grade 6 is enough to obtain 68 points out of 76.

The grade of the final exam is computed using the following formula

$$1 + 5 * \frac{\text{points}}{68},$$

and then rounded to multiples of .5.