Statistical Signal Processing

Final Exam

Wednesday, 6 July 2016, 8h15 - 11h15

You will hand in this sheet together with your solutions.

Write your personal data (please make it readable!).

Seat Number:

Family Name:

Name:

Read Me First!

Only the 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.

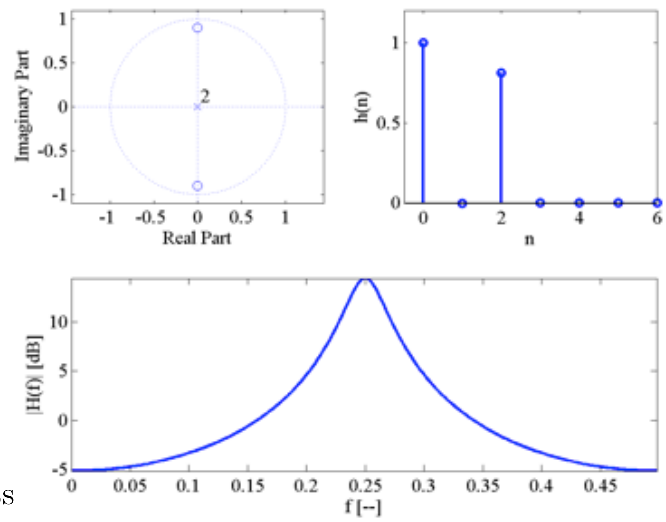
Return the text of the exam.

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. ONE SYSTEM OR MORE THAN ONE SYSTEM? (4 PTS)
Considering the following 3 plots:

- A z-transform in the z-plane (upper left corner), where `o` denotes the zeros and `x` the poles (the poles can in this framework be neglected);
- An impulse response $h(n)$ in the time domain (upper right corner);
- Magnitude of the frequency response $|H(e^{j2\pi f})|$ in normalized frequencies (bottom).



According to the plots

1) Do the z-transform and the impulse response $h(n)$ correspond to the same system (that is, is the plot of the z-plane the plot of the z-transform of $h(n)$)?

2) Do the z-transform and the magnitude of the frequency response $|H(e^{j2\pi f})|$ correspond to the same system (that is, is $|H(e^{j2\pi f})|$ the absolute value on the unit circle of the z-transform represented in the z-plane plot)?

Solution 1.

1) From the z-plane plot we see that the two zeros are $z_1 = j\alpha$ and $z_1 = -j\alpha$, where $0 < \alpha < 1$ (real). Consequently the z-transform has the form $\tilde{H}(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) = (1 - j\alpha z^{-1})(1 + j\alpha z^{-1}) = 1 + j\alpha z^{-1} - j\alpha z^{-1} + \alpha^2 z^{-2} = 1 + \alpha^2 z^{-2}$. The corresponding impulse response is therefore $\tilde{h}(0) = 1$, $\tilde{h}(1) = 0$, $\tilde{h}(2) = \alpha^2$, and $\tilde{h}(k) = 0$ for $k \geq 3$. Hence, $\tilde{h}(n) = h(n)$ and $\tilde{H}(z) = H(z)$.

2) The z-transforms shows two zeros near the unit circle at normalized frequencies $f_1 = 0.25$ and $f_2 = -0.25$. Consequently the magnitude of the corresponding frequency response $|\tilde{H}(e^{j2\pi f})|$ should show a minimum of the frequency response at $f_1 = 0.25$ and a minimum at $f_2 = -0.25$. The plot of $|H(e^{j2\pi f})|$ (bottom) shows a maximum at $f_1 = 0.25$ (and for the symmetry of the spectrum, a maximum at $f_2 = -0.25$). Consequently, $|H(e^{j2\pi f})| \neq |H(e^{j2\pi f})|$ and the plot of the z-transform and the plot of the magnitude of the frequency response $|H(e^{j2\pi f})|$ do not correspond to the same system.

Exercise 2. COLORED NOISE (4 PTS)

Let $W[n]$ be a white noise, that is, a i.i.d. sequence of random variables, with variance σ^2 and mean 0. Let $H(z) = 1 - az^{-1}$ be strictly minimum phase filter.

Consider the stochastic process $X[n]$ defined as $X[n]H(z) = W[n]$,

1) Prove that $X[n]$ is a w.s.s. process (without computing mean and correlation!).

2) Compute the expression of its correlation.

3) Compute the expression of its power spectral density.

4) Assuming a to be positive and real, sketch the magnitude of the transfer function $H(e^{j2\pi f})$ in normalized frequencies.

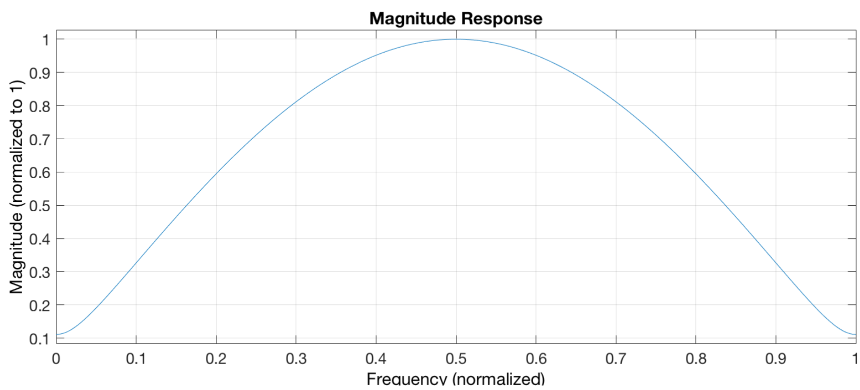
Solution 2.

1) Since $H(z)$ is strictly minimum phase, we can write $X[n] = W[n]/H(z)$. $W[n]$ is w.s.s. and $1/H(z)$ is stable (since $H(z)$ is strictly minimum phase). Consequently, by the fundamental filtering formula, $X[n]$ is w.s.s..

2) The process reads $X[n] - aX[n-1] = W[n]$. Then, for $k \geq 0$,
$$(X[n+k] - aX[n+k-1])X[n]^* = W[n+k]X[n]^* .$$
By taking the expectation of the left side
$$\mathrm{E}\left[(X[n+k] - aX[n+k-1])X[n]^*\right] = R_X(k) - aR_X(k-1) ,$$
and on the right side
$$\mathrm{E}\left[W[n+k]X[n]^*\right] = \sigma^2\delta_k .$$
Finally we obtain the expression of the correlation (structure)
$$R_X(k) - aR_X(k-1) = \sigma^2\delta_k .$$

3) By the fundamental filtering formula $S_x(f) = \sigma^2/|H(e^{j2\pi f})|$.

4) By assumption $a > 0$ and, since $H(z)$ is minimum phase, $a < 1$. The system has only one zero, on the positive real axis. Consequently, the magnitude of the transfer function has the following shape:



Main exercises

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

Please provide justified, rigorous, and simple answers.

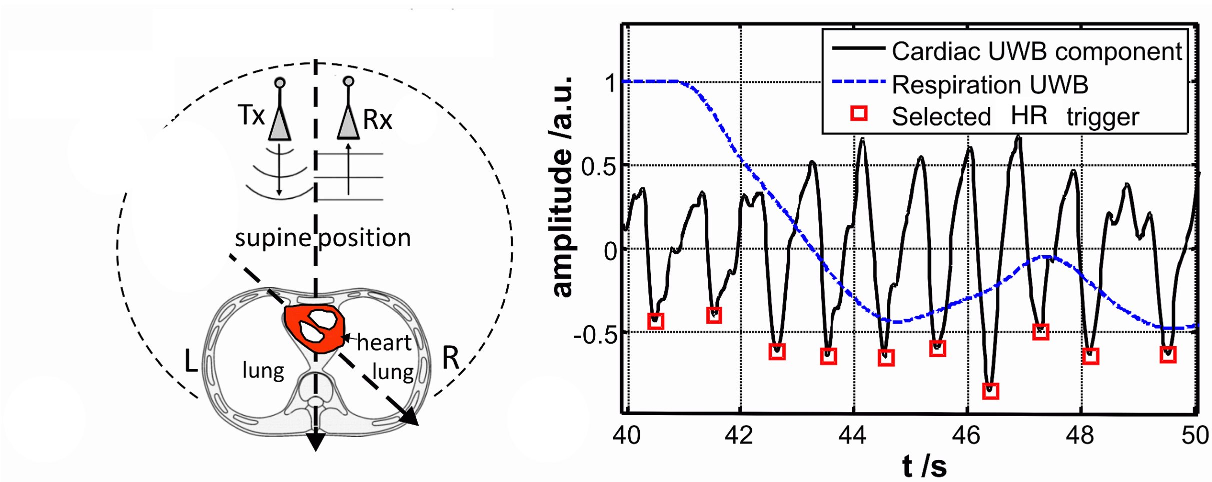
Exercise 3. Ultra Wide Band Radar for Heart Monitoring (28pts)

The principle of a radar is to emit a signal and then to receive its reflection. Measuring the delay between the emission and the reception of the reflection, as well as how the signal has been distorted, it is possible to identify the position of objects as well as to get information on their nature.

Ultra Wide Band Radars has recently found many applications in non contact physiological monitoring, including heart rate monitoring. The working principle is simple:

- The radar emits a sequence of periodic pulses (constant interval between two pulses), at positions $T_n = KT$, $K = 0, 1, \ldots$ where T is the constant interval between pulses;
- The pulses are reflected by the surface (membrane) of the heart;
- The delay between an emitted pulse and the received one indicates the distance between the radar and the heart membrane;
- Given that the heart beats, *i.e.* it changes its size, the distance between the radar and the heart membrane changes over time. Consequently the delay between a transmitted pulse and the corresponding received pulse changes over time;
- Given that the transmitted pulses are regularly spaced and that the delay between transmitted and received pulses changes over time, the interval between two consequent received pulses changes over time according to the heart beat.

The picture below describes the setup of a UWB radar for hear rate monitoring.



We suppose that the pulses are of very short duration and the they can be approximated to Dirac deltas. We call $y(t)$ the received sequence of pulses, where $y(t) = \sum_{k=1}^{\infty} \delta(t - \tau_n)$, and τ_n, $n = 1, \ldots$ are the positions in time of the received pulses (Dirac deltas).

The characteristics of the UWB radar for hear rate monitoring are the following:

- Pulses transmitted at a frequency of $f = 1/T = 500$Hz;
- Targeted heart rate between 30 and 220 beats per minute.

You are asked to develop the signal processing algorithms of the UWB heart rate monitoring system. Notice that parts A, B and C are independent.

Part A: Received Pulses

First of all, it is necessary to detect the positions τ_n of the received pulses.

1) In class we have seen a computationally efficient parametric method for the estimation of UWB pulse positions. Precisely describe how to apply such method on a 10 seconds recording of the received signal. As usual you are required to detail each step of the methods as if you have to implement it in a computer. Recall that the input of the method is a 10 seconds recording of $y(t)$ and that the output are the positions τ_n of the pulses.

Part B: Heart Rate

Secondly, we compute the interval between successive received pulses, that is, the difference between successive positions τ_n. Such signal, that we consider regularly sampled and we denote as $x[n]$, is the cardiac UWB component (see figure above) and it represents the beating of the heart.

One can select the maximum or the minimum values of such signal as reference for computing the HR (the figure above depicts the choice of the minimum values as HR trigger).

Unfortunately the signal $x[n]$ you have obtained appears to be very noisy (additive white noise), complicating the detection/choice of the HR trigger. Such a signal is NOT w.s.s., but it can assumed to be w.s.s. over short intervals of maximum 2 seconds.

In an interval of 10 seconds, we have computed (measured) $x[1], \ldots, x[5000]$

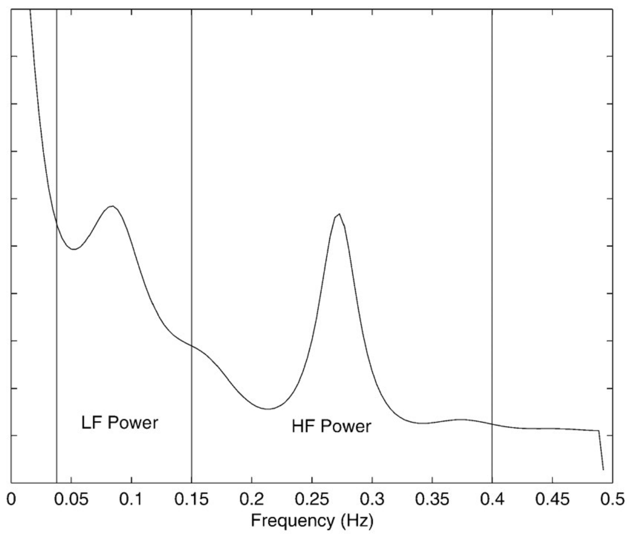
2) Precisely describe how to apply an optimal de-noising method for w.s.s. signals to the signal $x[n]$. Describe every step of the methods. Please notice that the input of the method is a sequence of samples of $x[n]$ (length of the sequence to be determined) and that the output is the denoised signal $\tilde{x}[n]$.

Once the signal denoised, it is simple to detect the HR trigger and compute the HR, instantaneous heart rate and averaged heart rate: The instantaneous heart rate is simply the inverse of the interval between to heart beats (measured in beats per seconds) while the averaged heart rate is the average of the instantaneous heart rate over a defined interval of time.

Part C: RR Intervals

The signal composed of the values of the time intervals between successive heart beats is called RR signal. The analysis of the spectrum of the RR signal is of foremost importance since it provides insight on the health status and well being (stress, fatigue, resting capability).

More precisely, it is interesting to analyze the spectral power (integral of the spectrum) over two frequency intervals. The picture below depicts a typical RR spectrum and the two frequency interval of interest, namely LF and HF.



Call $RR[n]$ the RR signal and $rr[1], \ldots, rr[10000]$ the samples obtained via the heart beat asurement. We suppose such signal to be regularly sampled and to be a w.s.s. process.

3) Propose a parametric spectral estimation method to estimate the spectrum of the RR signal in view of the analysis of the spectral power of the LF and HF frequency intervals. Justify precisely your answer.

4) Given that we have measured $rr[1], \ldots, rr[10000]$, describe in detail the method, step by step, from the measured samples to the estimated spectrum, like if each step has to be interpreted and executed by a computer (in particular the input, the executed operation with corresponding equations, and the output of each step has to be clear). Recall that the input of the method is the measured RR signal $rr[1], \ldots, rr[10000]$ and that the output is the spectrum $S_{RR}(f)$, where f is in normalized frequencies.

Solution 3.

1) The optimal parametric method to estimate the position of pulses (Diracs) is the annihilating filter method.

   Prior to apply the method we need to determine the maximum number of received pulses that might fall in a 10 second interval.

   Given that a pulse is transmitted every $T = 1/f = 1/500$, in a 10 seconds interval the transmitter will emit a maximum of $500 \times 10 = 5000$ pulses. Due to the heart beating, the received signal will have intervals between pulses that are sometimes slightly bigger than T and sometimes slightly smaller than T. Over a period of $\tau = 10$ seconds, the maximum number of received pulses can be assumed to also be 5000.

   - We have $y(t)$ for $0 \leq t \leq 10$ (seconds)
   - Recalling that the annihilating filter works on harmonic signals, we first need to transform the sequences of Deltas into a harmonic signal by taking the Fourier transformation (Fourier series) of $y(t)$ (considered periodic with a period of $\tau = 10$ seconds)

$$\widehat{y}[n] = \frac{1}{\tau} \int_0^\tau y(t) e^{-j2\pi n \frac{t}{\tau}} = \frac{1}{10} \sum_{k=1}^{5000} e^{-j2\pi n \frac{\tau_k}{10}} .$$

– Targeting the estimation of the position of 5000, we need an annihilating filter with impulse response of length 5000. The corresponding system reads

$$\begin{bmatrix} \widehat{y}[4999] & \dots & \widehat{y}[0] \\ \vdots & \ddots & \vdots \\ \widehat{y}[9998] & \dots & \widehat{y}[4999] \end{bmatrix} \begin{bmatrix} h(1) \\ \vdots \\ h(5000) \end{bmatrix} = - \begin{bmatrix} \widehat{y}[5000] \\ \vdots \\ \widehat{y}[9999] \end{bmatrix} .$$

By solving the system we obtain $h(1), \dots, h(5000)$ (Toeplitz system, requiring 5000^2 multiplications), and therefore, with $h(0) = 1$, we have the impulse response of the annihilating filter.

– Having the impulse response $h(n)$ we compute the z-transform

$$H(z) = \sum_{n=0}^{5000} h(n) = 1 + h(1)z^{-1} + \dots + h(5000)z^{-5000} .$$

– Compute the zeros of the z-transform $H(z)$, that we shall call $z_1, \dots, z_{5000}$, that is

$$H(z) = \prod_{k=1}^{5000} (1 - z_k z^{-k}) .$$

– By taking the argument of the zeros we obtain the positions τ_k, $k = 1, \dots, 5000$, with the following formula

$$\tau_k = \tau \frac{\arg(z_k)}{2\pi} = 10 \frac{\arg(z_k)}{2\pi} ,$$

where $\arg(z_k)$ is constrained in $[0, 2\pi]$.

2) In order to consider the measured signal a w.s.s. signal, we need to divide it into blocks of maximum2 seconds. Having measured 5000 samples in 10 seconds, by taking the block length equal to 2 seconds we have $5000 \times 2/10 = 1000$ samples in each block. Call $x_l[1], \dots, x_l[1000]$ the 1000 samples of the k-th block (where in our case $k = 1, \dots, 5$).

Having a w.s.s. signal corrupted by an additive white noise, the optimal denoising technique is given by the filtering of the noisy signal with the corresponding Wiener filter.

By calling $\widetilde{x}_l[n]$ the original signal, $w[n]$ the realizations of the noise, we can write $x_l[n] = \widetilde{x}_l[n] + w[n]$ and the expression of the transfer function of the Wiener filter reads

$$H(e^{j2\pi f}) = \frac{S_{X_l}(f) - \sigma_W^2}{S_{X_l}(f)}$$

where $S_{X_l}(f)$ is the power spectral density of the measured signal (block) and σ_W^2 is the power spectral density of the noise, and f is a normalized frequency $\in [0, 1]$.

By considering discrete frequencies (we are implementing the formulas in a computer!), and replacing the power spectra densities with their estimation based on the samples $x_l[1], \dots, x_l[1000]$, the transfer function of the Wiener filter reads

$$H(e^{j2\pi \frac{k}{1000}}) = \frac{\widehat{S}_{X_l}(k) - \widehat{\sigma}_W^2}{\widehat{S}_{X_l}(k)} ,$$

where $k = 0, \dots, 999$.

The for each block (for $l = 1 : 5$) we do:

- Compute the estimation of the power spectral density $\widehat{S}_{X_l}(k)$ of the noisy signal $x_l[1], \ldots, x_l[1000]$, for instance using the periodogram. Call $\widehat{X}_l$ the Fourier transform of the noisy signal $x_l[1], \ldots, x_l[1000]$

$$\widehat{X}_l(k) = \frac{1}{1000} \sum_{n=1}^{1000} x_l[n] e^{-j2\pi \frac{k(n-1)}{1000}},$$

then, the periodogram reads

$$\widehat{S}_{X_l}(k) = \frac{1}{1000} \left| \sum_{n=1}^{1000} x_l[n] e^{-j2\pi \frac{k(n-1)}{1000}} \right|^2 = 1000 \left| \widehat{X}_l(k) \right|^2,$$

where f is a normalized frequency $\in [0, 1]$.

- Estimate the power of the noise σ_W^2. As discussed in class, this can be done by computing the energy of the uncorrelated part of the signal, for instance via the diagonalization of the empirical correlation matrix of the noisy signal block $x_l[1], \ldots, x_l[1000]$.
- Compute the transfer function of the Wiener filter

$$H(e^{j2\pi \frac{k}{1000}}) = \frac{\widehat{S}_{X_l}(k) - \widehat{\sigma}_W^2}{\widehat{S}_{X_l}(k)}, \quad k = 0, \ldots, 999.$$

- To compute the block of the denoised signal $\widetilde{x}_l[1], \ldots, \widetilde{x}_l[1000]$ we can proceed in two ways:

  a) Convolution in time domain

  * Compute the impulse response of the Wiener filter as the inverse discrete Fourier transform of the transfer function

$$h(n) = \frac{1}{1000} \sum_{k=0}^{999} H(e^{j2\pi \frac{k}{1000}}) e^{j2\pi \frac{kn}{1000}}, \quad n = 0, \ldots, 999.$$

  * Compute the block of the denoised signal $\widetilde{x}_l[1], \ldots, \widetilde{x}_l[1000]$ as the convolution between the noisy signal $x_l[1], \ldots, x_l[1000]$ and the corresponding impulse response of the Wiener $h(0), \ldots, h(999)$ filter

$$\widetilde{x}_l[n] = h * x_l[n].$$

  b) Multiplication in frequency domain

  * Compute the inverse Fourier transform of the product between the transfer function of the Wiener filter and the Fourier transform of the noisy signal

$$\widetilde{x}_l[n] = \frac{1}{1000} \sum_{k=0}^{999} \widehat{X}_l(k) H(e^{j2\pi \frac{k}{1000}}) e^{j2\pi \frac{k(n-1)}{1000}}, \quad n = 1, \ldots, 1000.$$

- Once obtained all the denoised signal blocks $\widetilde{x}_l[1], \ldots, \widetilde{x}_l[1000]$, $l = 1, \ldots, 5$, these can sequentially combined. Alternatively, one can consider overlapping blocks (therefore more than 5), that are then to be recombined using a weighting window.

- Given the smooth form of the spectral density and the fact that we aim at computing and integral, the optimal spectral estimation method is the one based on a polynomial expression of the spectrum

$$S_{RR}(f) = \frac{\sigma^2}{|P(e^{j2\pi f})|^2} = \frac{\sigma^2}{|1 + p_1 e^{-j2\pi f} + \ldots + p_M e^{-j2\pi f M}|^2}$$

where the polynomial coefficients $p_1, \ldots, p_M$ and σ^2 are estimated using the Yule-Walker equations. Concerning the order M of the polynomial, it can be assigned empirically based on the plot of the typical spectrum of a RR signal. You can observe that the spectrum has at least 5 local maxima that can be associated to 5 poles. As seen in class, during the Matlab exercise of the estimation of an RR spectrum using AR model, a typical RR spectrum is associated to an AR of order 8, so let's take $M = 8$. Alternatively one can use the Levinson algorithm (not see in class but in the lecture notes).

- We have computed 10000 samples of the RR signal $rr[1], \ldots, rr[10000]$ (regularly sampled and w.s.s.), and from the typical spectrum of a RR signal (see figure) the order of the polynomial can be assumed to be $M = 8$ $(p_1, \ldots, p_8)$.

  In order to estimate the spectrum Yule Walker equations we proceed as follows

  - Compute the empirical correlation (biased of unbiased version)

    $$\widehat{R}_{RR}(k) = \frac{1}{10000} \sum_{n=1}^{10000-k} rr[n+k] * rr[n]\,, \quad k = 0, \ldots, 8\,,$$

    with $\widehat{R}_{RR}(-k) = \widehat{R}_{RR}(k)$, $k = 1 \ldots, 8$ (since the RR signal si real).
  - Write the Yule Walker equations for the estimation of $p_1, \ldots, p_8$ and σ^2

    $$\begin{bmatrix} \widehat{R}_{RR}(0) & \ldots & \widehat{R}_{RR}(7) \\ \vdots & \ddots & \vdots \\ \widehat{R}_{RR}(7) & \ldots & \widehat{R}_{RR}(0) \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_8 \end{bmatrix} = - \begin{bmatrix} \widehat{R}_{RR}(1) \\ \vdots \\ \widehat{R}_{RR}(8) \end{bmatrix}$$

    $$\sigma^2 = \widehat{R}_{RR}(0) + \widehat{R}_{RR}(1) p_1 + \ldots + \widehat{R}_{RR}(8) p_8\,.$$

  - Solve the equations (Toeplitz symmetric $8^2 + 8$ multiplications) in order to obtain $p_1, \ldots, p_8$ and σ^2
  - Compute the estimation of the spectrum as

    $$\widehat{S}_{RR}(f) = \frac{\sigma^2}{|1 + p_1 e^{-j2\pi f} + \ldots + p_8 e^{-j2\pi f 8}|^2}$$

    where the normalized frequency f can be discretized over N points: $f = k/N$, $k = 0, \ldots, N-1$

    $$\widehat{S}_{RR}(k) = \frac{\sigma^2}{\left|1 + p_1 e^{-j2\pi \frac{k}{N}} + \ldots + p_8 e^{-j2\pi \frac{k}{N} 8}\right|^2}\,, \quad k = 0, \ldots, N-1\,.$$

Exercise 4. LETTERS AND WORDS (28PTS)

We would like to model a speech signal. Notice that parts A and B are independent.

Part A: Only Vowels

We start with a very simple assumption: A human can only pronounce 3 vowels A, E, I, O, that is, the words are a sequence of these for vowels.

As an example you consider the word AOIIIEAIEOOIAAAEEOIE. Since the vowels are not pronounced independently, it is a good idea to model the sequence, as the realization of a Markov chain $X[n]$, where each vowel is associated to a numerical value so to obtain a numerical signal $x[1], \ldots, x[20]$.

1) Write the likelihood function associated to the observations $x[1], \ldots, x[20]$.
2) If you are to use the Lagrange multipliers to estimate the parameter of the Markov chain via the maximization of the likelihood function, which are the contraints? Which are the equations to be solved? How many multipliers you have to use?

The words are corrupted by additive Gaussian white noise $W[n]$ (centered, with variance σ_W^2), obtaining a (numerical) signal $Y[n] = X[n] + W[n]$.

3) Propose a method to denoise the words. Justify precisely your answer.
4) Describe such method in detail, step by step, like if each step has to be interpreted and executed by a computer. Recall that the input of the method are the measured samples $y[1], \ldots, y[20]$ and that the output are the denoised samples $\tilde{x}[1], \ldots, \tilde{x}[20]$

Part B: A Continuous Signal Model

The only vowels assumption is way too simple, and yet the model for it way too complicate.

We switch to a complete different approach. We now assume to record a speech (composed as several words) as a continuous signal $s[n]$. In a complex speech we have identified 20000 undistinguished words, each of a maximum of 100 samples. To characterize each word, we process the corresponding samples of continuous signal by computing the following variables

- $l[1]$ = The (normalized) frequency with highest energy;
- $l[2]$ = The mean value;
- $l[3]$ = The maximum value of the absolute amplitude;
- $l[4]$ = The number of signal samples composing the word;
- $l[5]$ = The total power (sum of the square value of the samples).

It is now question to analyze how many different words are in the recorded speech.

First of all, we proceed by checking if the analysis can be simplified by reducing the number of variables.

5) Given the received data (20000 words of a maximum of 100 samples each) precisely describe how to apply the PCA method to reduce the number of variables to be analyzed. You are asked to detail each step like if you have to implement the method in a computer. Recall that the input of the method are the variables (for each signal segment), and that the output are the number of reduced variables K and their expression $\boldsymbol{Z}$.

   In order to provide the number of reduced variables and their expression, you can assume that two eigenvalues of the variable correlation matrix are distinguishably much greater than the others.

Secondly we determine the number of different words and the values of their characteristic variables.

6) Given the reduced set of variables $\boldsymbol{Z}$, explain in detail, step by step, how we can determine the number of different words P and the values of the corresponding characteristics variables $\tilde{l}_n[1], \ldots, \tilde{l}_n[5]$, $n = 1, \ldots, P$. You are asked to detail each step like if you have to implement the method in a computer. Recall that the input of the method are the reduced set of variables $\boldsymbol{Z}$, and that the output are the number of words P and the values of the corresponding characteristic variables $\tilde{l}_n[1], \ldots, \tilde{l}_n[5]$, $n = 1, \ldots, P$.

Solution 4.

We associate numerical values to the vowels as follows $A = 1$, $E = 2$, $I = 3$, $O = 4$, and we have observed a word of 20 vowels AOIIIEAIEOOIAAAEEOIE, therefore $x[1]x[2] \ldots x[20] =$ 14333213244311122432

1) The parameter of the likelihood functions are the initial probabilities $p_i = \mathrm{P}\left(x[n] = i\right)$, $i = 1, \ldots, 4$ and the transition probabilities $p_{ij} = \mathrm{P}\left(x[n] = j \mid x[n-1] = i\right)$, $i, j = 1, \ldots, 4$, therefore $\boldsymbol{\theta} = \{p_i \ (i = 1, \ldots, 4), p_{ij} \ (i, j = 1, \ldots, 4)\}$. The likelihood associated to a Markov chain realization reads

$$
\begin{aligned}
h(x[1], \ldots, x[20] \ ; \ \boldsymbol{\theta}) &= p_{x[1]} p_{x[1]x[2]} p_{x[2]x[3]} \cdots p_{x[19]x[20]} \\
&= p_1 p_{14} p_{43} p_{33} p_{33} p_{32} p_{21} p_{13} p_{32} p_{24} p_{44} p_{43} p_{31} p_{11} p_{11} p_{12} p_{22} p_{24} p_{43} p_{32} \\
&= p_1 p_{14} (p_{43})^3 (p_{33})^2 (p_{32})^3 p_{21} p_{13} (p_{24})^2 p_{44} p_{31} (p_{11})^2 p_{12} p_{22}
\end{aligned}
$$

   with its logarithmic version given by

$$
\begin{aligned}
\log h(x[1], \ldots, x[20] \ ; \ \boldsymbol{\theta}) = \log p_1 + \log p_{14} + 3 \log p_{43} + 2 \log p_{33} + 3 \log p_{32} \\
+ \log p_{21} + \log p_{13} + 2 \log p_{24} + \log p_{44} + \log p_{31} + 2 \log p_{11} + \log p_{12} + \log p_{22}
\end{aligned}
$$

2) The constraints on the parameters $\boldsymbol{\theta} = \{p_i \ (i = 1, \ldots, 4), p_{ij} \ (i, j = 1, \ldots, 4)\}$ are

$$
\sum_{i=1}^{4} p_i = 1 \,, \quad \sum_{i=1}^{4} p_{ji} = 1 \,, j = 1, \ldots, 4 \,,
$$

   to be associated to 5 multipliers: $\lambda_1, \ldots, \lambda_5$.

   The equations to be solved are

– Maximization of the constrained log likelihood with respect to the parameters $\boldsymbol{\theta}$

$$
\begin{aligned}
g(\boldsymbol{\theta}) = \log h(x[1],\ldots,x[20]\ ;\ \boldsymbol{\theta}) &- \lambda_1(\sum_{i=1}^{4} p_{1i} - 1) - \lambda_2(\sum_{i=1}^{4} p_{2i} - 1) \\
&- \lambda_3(\sum_{i=1}^{4} p_{3i} - 1) - \lambda_4(\sum_{i=1}^{4} p_{4i} - 1) - \lambda_5(\sum_{i=1}^{4} p_{i} - 1)
\end{aligned}
$$

obtaining a set of parameters depending on the Lagrange multipliers;

– 5 Contraints, obtaining the values of the Lagrange multipliers.

3) The signal can be modeled as the realization of a discrete value stochastic process corrupted by noise. Consequently, the noisy signal can seen as the outcome of a mixture of Gaussian distributions, where the means of the distributions corresponds to the discrete values of the process. In order to reconstruct the signal, we first need to estimate the parameters of the mixture of Gaussian distributions (MLE) and then maximise the a posteriori distribution (MAP). To be noticed that the mixture of Gaussian distribution is Markovian and not i.i.d..

4) We proceed as follows

– First of all we need to define the mixture model. Given the measured samples $y[1],\ldots,y[20]$, call $\mathcal{C} = \{(c[1],\ldots,c[20]) \mid c[i] \in \{m_1,\ldots,m_4\}\}$ the set of all possible 20 combinations of the numerical values $m_1,\ldots,m_4$ associated to the vowels, and $\boldsymbol{X} = [X[1],\ldots,X[20]]$ the Markov chain modeling the original signal $\boldsymbol{x} = [x[1],\ldots,x[20]]$. Then the mixture of distribution reads

$$
\begin{aligned}
f_{\boldsymbol{Y}}(y[1],\ldots,y[20]) &= \sum_{\boldsymbol{c}\in\mathcal{C}} \prod_{n=1}^{20} \mathcal{G}_{c[n],\sigma^2}(y[n]) \mathrm{P}\left(\boldsymbol{X} = \boldsymbol{c}\right) \\
&= \sum_{\boldsymbol{c}\in\mathcal{C}} \prod_{n=1}^{20} \mathcal{G}_{c[n],\sigma^2}(y[n]) p_{c[1]} p_{c[1]c[2]} \cdots p_{c[19]c[20]}\,.
\end{aligned}
$$

The parameters of the models are

$$
\boldsymbol{\eta} = \left\{\sigma^2, m_i\ (i=1,\ldots,4), p_i\ (i=1,\ldots,4), p_{ij}\ (i,j=1,\ldots,4)\right\}
$$

Describe such method in detail, step by step, like if each step has to be interpreted and executed by a computer. Recall that the input of the method are the measured samples $y[1],\ldots,y[20]$ and that the output are the denoised samples $\tilde{x}[1],\ldots,\tilde{x}[20]$

– Secondly we need to estimate the parameters of the model using the maximum likelihood approach. The likelihood function is given by

$$
h(y[1],\ldots,y[20]\ ;\ \boldsymbol{\eta}) = \sum_{\boldsymbol{c}\in\mathcal{C}} \prod_{n=1}^{20} \mathcal{G}_{c[n],\sigma^2}(y[n]) p_{c[1]} p_{c[1]c[2]} \cdots p_{c[19]c[20]}\,.
$$

As seen in class, such a maximization needs to be performed using an iterative algorithm such as the EM algorithm. Call $\widehat{\boldsymbol{\eta}}$ the parameter estimation.

- Finally, having the observed samples $\boldsymbol{y} = [y[1], \dots, y[20]]$ and having computed the model parameter estimation $\widehat{\boldsymbol{\eta}}$, we need to find the most probable sequence of the denoised signal $\tilde{\boldsymbol{x}} = [\tilde{x}[1], \dots, \tilde{x}[20]]$ by maximizing the a posteriori distribution (MAP)

$$\begin{aligned}
\mathrm{P}\left(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{y}\right) =& \frac{\text{complete data likelihood function}}{\text{marginal likelihood function}} \\
=& \frac{h(\boldsymbol{y}, \boldsymbol{x}; \widehat{\boldsymbol{\eta}})}{h(\boldsymbol{y}; \widehat{\boldsymbol{\eta}})} = \frac{f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}) \mathrm{P}\left(\boldsymbol{X} = \boldsymbol{x}\right)}{f_{\boldsymbol{Y}}(\boldsymbol{y})} \\
=& \alpha \prod_{n=1}^{20} \mathcal{G}_{x[n], \widehat{\sigma}^2}(y[n]) \widehat{p}_{x[1]} \widehat{p}_{x[1]x[2]} \cdots \widehat{p}_{x[19]x[20]} \, .
\end{aligned}$$

  The maximization is done with respect to $\boldsymbol{x}$ constrained to the discrete values $\widehat{m}_i$, $i = 1, \dots, 4$ (that is $\sum_{n=1}^{20} \prod_{k=1}^{4} (x[n] - \widehat{m}_i) = 0$). Therefore

$$\tilde{\boldsymbol{x}} = \arg\max_{\boldsymbol{x} \in \widehat{\mathcal{C}}} \mathrm{P}\left(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{y}\right) \, .$$

  Such maximization is numerically carried using iterative algorithm such as the Viterbi algorithm.

5) We have $M = 20000$ words (dimension of the data). Each word is characterized by 5 variables $\boldsymbol{l}_k = [l_k[1], \dots, l_k[5]]$ (dimensions 1×5), $k = 1, \dots, 20000$, as clearly described in the problem. In order to apply PCA we need to proceed as follows:

   - Center the variables $\bar{\boldsymbol{l}}_k = \boldsymbol{l}_k - \frac{1}{20000} \sum_{n=1}^{20000} \boldsymbol{l}_n$ (dimension 1×5), $k = 1, \dots, 20000$.
   - Compute the correlations matrix of the centered variables $\boldsymbol{R}_L = \frac{1}{20000} \sum_{k=1}^{20000} \boldsymbol{l}_k^t \boldsymbol{l}_k$ (dimensions 5×5).
   - Diagonalize the correlation matrix $\boldsymbol{V}^t \boldsymbol{R}_L \boldsymbol{V} = \boldsymbol{\Lambda}$ (dimensions 5×5), where $\boldsymbol{V}$ (dimensions 5×5) is the matrix of the eigenvectors of $\boldsymbol{R}_L$.
   - As stated in the problem, we assume that two eigenvalues, λ_1 and λ_2, of the variable correlation matrix are distinguishably much greater than the others. Consequently, in order to characterize the data, it is sufficient to compute only 2 of the 5 principal components. Call $\boldsymbol{V}_2$ the 5×2 matrix containing the two eigenvector associated to the two eigenvalues λ_1 and λ_2. The two principal components are then

$$(\boldsymbol{z}_2)_k = [z[1]_k, z[2]_k] = \boldsymbol{l}_k \boldsymbol{V}_2 \text{ (dimensions } 1 \times 2) \, , \quad k = 1, \dots, 20000 \, .$$

6) From the previous question we have achieved, in the space of the principal components, a reduction of the variables from 5 to 2. The data is, in the space of the principal components, characterized by $(\boldsymbol{z}_2)_k = [z[1]_k, z[2]_k]$, $k = 1, \dots, 20000$.

   We proceed as follows

   - Analyze the two principal components by plotting their (20000) values in a 2 dimensional plot (each dimension corresponding to a principal component), that is, by plotting the couples $(z[1]_1, z[2]_1), \dots, (z[1]_{20000}, z[2]_{20000})$ on the $x - y$ plane. The data will appear in clusters. The number of the cluster corresponds to the number of different words P, while the center of each cluster represents the characteristic values of a word in the principal component space.

– The clusters can be modeled as a mixture of 2-dimensional i.i.d. Gaussian distributions, with independent dimensions

$$f\left((z[1]_1, z[2]_1), \dots, (z[1]_{20000}, z[2]_{20000})\right) = \prod_{n=1}^{20000} \sum_{k=1}^{P} \pi_k \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left(-\frac{(z[1]_n - m_{1k})^2}{2{\sigma_1}_k^2} - \frac{(z[2]_n - m_{2k})^2}{2{\sigma_2}_k^2}\right)$$

The number of clusters P can be determined using a model order estimation technique, while the center of the clusters corresponds to the couples $(m_{11}, m_{21}), \dots, (m_{1P}, m_{2P})$.

– Estimate the parameters of the mixture model (EM algortihm), obtaining an estimation of the the center of the clusters $(m_{11}, m_{21}), \dots, (m_{1P}, m_{2P})$.

– The values of the 2 principal components $\tilde{\boldsymbol{z}}_2$characterizing the P words (in the space of the principal components) are

$$(\tilde{\boldsymbol{z}}_2)_k = [m_{1k}, m_{2k}]\,, k = 1, \dots, P$$

or in matrix form

$$\tilde{\boldsymbol{Z}}_2 = \begin{bmatrix} m_{11} & m_{21} \\ \vdots & \vdots \\ m_{1P} & m_{2P} \end{bmatrix}, \quad (P \times 2)$$

– The values of the 5 characteristics (centered) corresponding to the P words are then

$$\bar{\tilde{\boldsymbol{L}}} = \tilde{\boldsymbol{Z}}_2 \boldsymbol{V}_2^t = \begin{bmatrix} \bar{\tilde{\boldsymbol{l}}}_1 \\ \vdots \\ \bar{\tilde{\boldsymbol{l}}}_P \end{bmatrix}, \quad (P \times 5)$$

and by adding the mean we obtain

$$\tilde{\boldsymbol{L}} = \bar{\tilde{\boldsymbol{L}}} + \begin{bmatrix} \frac{1}{20000} \sum_{n=1}^{20000} \boldsymbol{l}_n \\ \vdots \\ \frac{1}{20000} \sum_{n=1}^{20000} \boldsymbol{l}_n \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{l}}_1 \\ \vdots \\ \tilde{\boldsymbol{l}}_P \end{bmatrix},$$

Finally, $\tilde{\boldsymbol{l}}_k$ represents the values of the 5 characteristics associated to the $k-th$ work, with $k = 1, \dots, P$.

Grade Scale.

The exams accounts for a total of 64 points (exact response to each question).

The grading has been done on a 59 points scale (59 points = 6/6), according to the following formula

$$\text{grade over } 6 = 1 + (5 * \text{points}/59)$$

and then rounded to .5 steps, that is

$$\text{rounded grade over } 6 = (\text{round-to-0-digit}(2 * \text{grade over } 6))/2$$

The result is then constraint to be at maximum 6.