Statistical Signal & Data Processing - COM500
Final Exam

Friday June 30th, 2023, 9h15-12h15.

Hand in this sheet together with
the exam text, your solutions, and your cheat sheet.

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

Seat Number:

Family Name:

Name:

Read Me First!

You are allowed to use:

- A handwritten cheatsheet (two A4 sheets, double sided) summarizing the most important formulas (no exercise text or exercise solutions);
- A pocket calculator.

You are definitively not allowed to use:

- Any kind of support not mentioned above;
- Your neighbour; Any kind of communication systems (smartphones etc.) or laptops;
- Printed material; Text and Solutions of exercises/problems; Lecture notes or slides.

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.

All the best for your exam!!

Warmup Exercise

This is a warm up problem .. do not spend too much time on it. Please provide justified, rigorous, and simple answers. If needed, you can add assumptions to the problem setup.

Exercise 1. AVERAGING PERIODOGRAM (2PTS)
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).
Now we split the measured signal $(X[1], X[2], \ldots, X[N])$ into two parts

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

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 $\mathrm{Var}(Q(\omega))$?

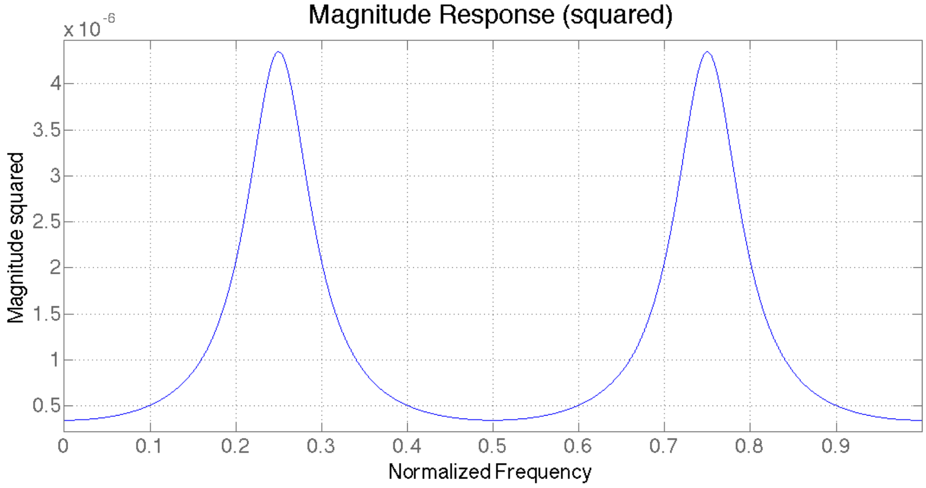
Solution 1.
The variables Y_1 and Y_2 are independent. Therefore the variance of their sum, or of the sum of a function of them, equals the sum of the variances. Indeed

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

Given that, as seen in class, the variance of the periodogram of a white noise equals the variance of the white noise, by calling σ_X^2 such variance we have

$$\mathrm{Var}\left(Q(\omega)\right) = \frac{1}{4}\left(\mathrm{Var}\left((P_{Y_1}(\omega)\right) + \mathrm{Var}\left(P_{Y_2}(\omega)\right)\right) = \frac{1}{4}\left(\sigma_X^2 + \sigma_X^2\right) = \frac{1}{2}\sigma_X^2.$$

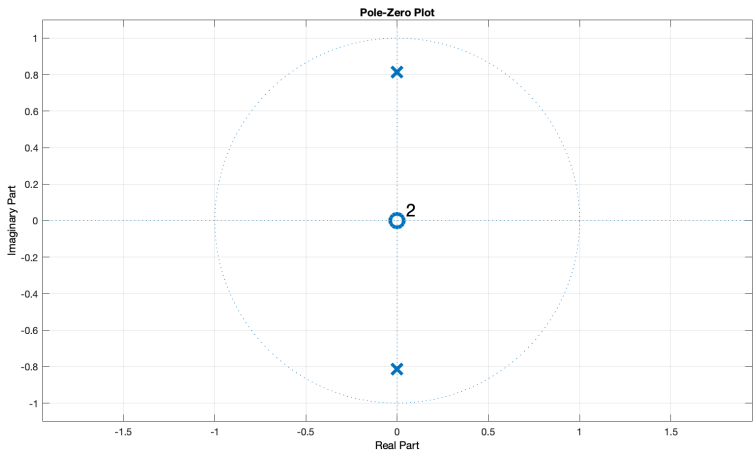
Exercise 2. A FILTER (4PTS)
Consider the following square module of the transfer function $H(e^{j\omega})$, *i.e.*, the square module of the discrete time Fourier transform of the impulse response $h[n]$. Call $H(z)$ the corresponding z-transform.



1) Draw poles and zeros of $H(z)$.
2) Write the expression of $H(z)$ by putting values for whatever can be easily determined from the above plot.
3) Can $H(z)$ represent the synthesis filter of an AR process? Please justify precisely.
4) Can $H(z)$ represent the analysis filter of an AR process? Please justify precisely.

Solution 2.

1)



2) From the Pole-Zero plot we have $z_1 = z_2 = 0$, $p_1 = \alpha e^{i\pi/2}$, $p_2 = \alpha e^{-i\pi/2}$, where $0 < \alpha < 1$. Hence

$$H(z) = \frac{z^2}{(z - p_1)(z - p_2)} = \frac{1}{(1 - p_1 z^{-1})(1 - p_2 z^{-1})} = \frac{1}{(1 - \alpha e^{i\pi/2} z^{-1})(1 - \alpha e^{-i\pi/2} z^{-1})}$$

3) The synthesis filter is an all poles stable filter. Hence, $H(z)$ can represent the synthesis filter of an AR process.
4) The analysis filter is an all zeros filter. Hence, $H(z)$ cannot represent an analysis filter of an AR process.

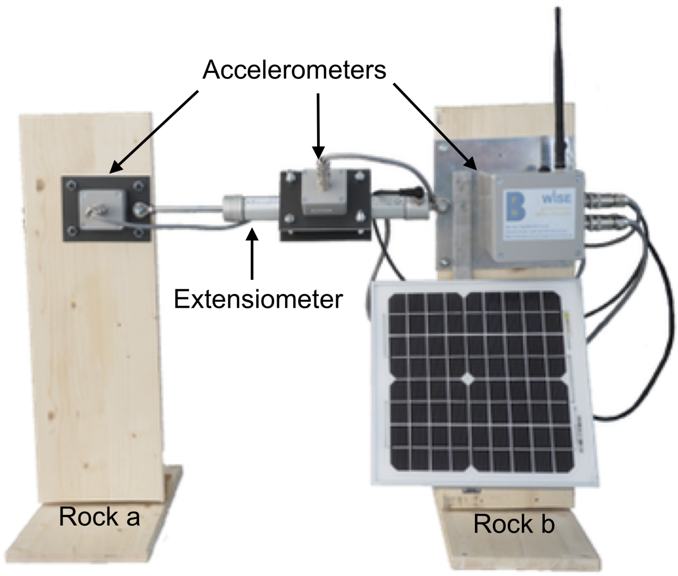
Main Problems

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. Remember that you are not simply asked to describe statistical signal processing tools, but rather to describe how to apply such tools to the given problem. If needed, you can add assumptions to the problem setup. You are asked to comply with the notation given in the problem.

Exercise 3. ROCK MONITORING IN CLIMATE CHANGE AND GLOBAL WARMING CONTEXT
With climate change and global warming we are experiencing high variation of temperature in short times with exceptionally high temperature at high altitudes. The direct consequence of such abnormal temperature pattern is a high stress on rock structures (mountains, cliffs, rock faces) and the melting of the permafrost, causing severe rock falls. Key examples are:

- The major rock fall in Pizzo Cengalo (Graubunden) in August 2017, that obliged to evacuate 100 inhabitants of Valle Bregaglia;
- The rock slide in Forclaz pass in January 2018, that interrupt the road connection of two villages for a couple of months;
- The significant rockfall on Trident du Tacul in Mont Blanc massif in September 2018;
- The major rock fall in Matterhorn in August 2019, that killed two alpinists.

The monitoring of rock instability is therefore of foremost importance, both for safety and research purposes, since it enables to estimate the danger of rock falls in key locations (e.g., roads, populated valleys, frequented alpine routes). A rock monitoring device for extreme environments has been recently developed. It is composed an extensometer, to measure a crack widening or shrinking, and three accelerometers, to mesure rock movements (tilt, displacement). The picture below depicts the system on a demo setup (where the side wood represents the moving rocks).



The accelerometers can also be used to mesure rock vibrations. Measuring the vibrations in response to impulses enables the study of the rock structure, while measuring the vibrations in response to oscillations enables the study of rock characteristics.

In both cases, an acoustic signal y is propagated into the rock from an actuator positioned on the surface of the rock. The accelerometer is also positioned on the surface, but in another point, and it measures a signal x.

So the developed device enables for monitoring of rock instability as well as analysis of rock structure and characteristics. We shall focus here on the analysis features.

Part A, B, C, and D are independent.

Part A: Study of the Rock Structure (8pts)

We consider $y(t)$ to be an impulse *i.e.*, a signal that can be symbolically modelled in continuous time domain as a Dirac delta $\delta(t)$ (we shall work here in continuous time).

The impulse is partially transmitted and partially reflected by each layer. Consequently, the signal $x(t)$ recorded at the surface is the sum of all the reflections

$$x(t) = \sum_{k=1}^{\infty} \alpha_k \delta(t - \tau_k)\,,$$

where τ_k, corresponds to the propagation delay between the surface and the boundary between layer k and $k+1$, and α_k is linked to the type of material (air or permafrost) between layer k and $k+1$.

The estimation of τ_k is therefore of foremost importance since it provides an estimation of the thickness of the layers. Similarly, the estimation of α_k provide insights ont the material between two layers.

We record the signal $x(t)$ over an interval of $\tau = 0.1$ s.

We can assume the maximum number of layers of interest to be 5 and, therefore, a maximum of 5 reflected spikes. Assuming each layer to have an average thickness of 2 m, and the signal (sound) propagation to be 300 $\frac{\text{m}}{\text{s}}$, we can take as a recording interval $\tau = 0.1$ s. Therefore

$$x(t) = \sum_{k=1}^{5} \alpha_k \delta(t - \tau_k)\,, \quad t \in [0, 0.1]\ \text{s}\,.$$

We start by assuming that the signal $x(t)$ is recorded in a noise-free environment.

A.1) Assuming the absence of noise, propose a parametric method to estimate the positions τ_k, $k = 1, \ldots, 5$, and the amplitudes α_k, $k = 1, \ldots, 5$ of the spikes. Precisely describe such method. You are given $\tau = 0.1$ s of the signal $x(t)$ assuming a maximum of 5 spikes. You are asked to detail each step as if you have to implement the method in a computer. Precisely indicate the input and output of each step.

In practice $x(t)$ is recorded in a quite noisy environment. Therefore,

$$x(t) = \sum_{k=1}^{5} \alpha_k \delta(t - \tau_k) + \text{noise}\,, \quad t \in [0, 0.1]\ \text{s}\,.$$

A.2) Considering now the presence of a non-negligible noise, propose a parametric method (presented in class) to estimate the positions τ_k, $k = 1, \dots, 5$, **and** the amplitudes α_k, $k = 1, \dots, 5$ of the spikes. Precisely describe such method. You are given $\tau = 0.1$ s of the signal $x(t)$ assuming a maximum of 5 spikes. You are asked to detail each step as if you have to implement the method in a computer. Precisely indicate the input and output of each step.

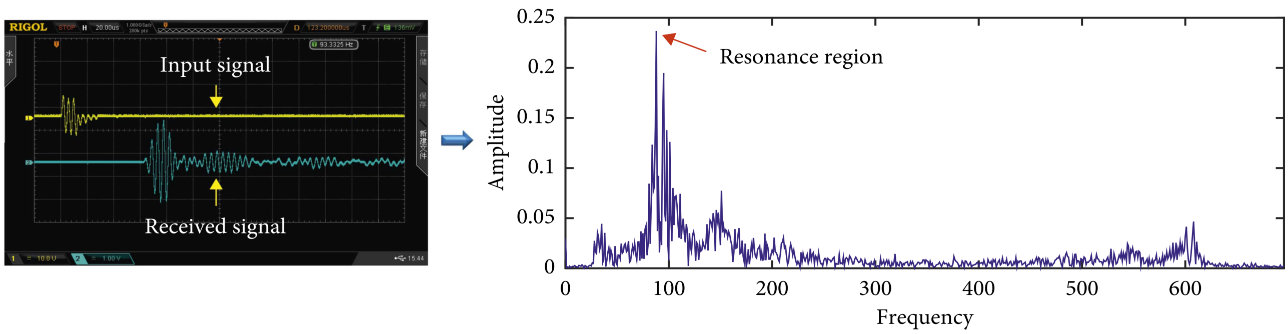
Part B: Study of the Rock Characteristics (12 points)

We consider the signal $y[n]$ to correspond to a short oscillation at 50 kHz, *i.e.*, a truncated real sinusoid with frequency 50 kHz (we shall work here in discrete time). The advantage of using a truncated sinusoid is that it generates all the harmonics (here we will consider the first harmonic to be the fundamental).

B.1) Write the signal $y[n]$ as a w.s.s. process $Y[n]$ (you can limit yourself to the 10th harmonic).

B.2) Prove that $Y[n]$ is indeed a w.s.s. process (you can limit yourself to the 3rd harmonic for such computation).

The response of the rock to a short oscillation $y[n]$ is measured using the accelerometer, obtaining a signal $x[n]$. Rocks have a specific resonance frequency depending on its characteristics. Therefore, the response signal $x[n]$ is NOT a harmonic signal but rather a signal with energy mostly distributed around a specific frequency. The estimation of such energy distribution, and the successive identification of the resonance frequency, provides insight into the characteristics of the rock. An example of the spectrum $S_x(\omega)$ of the signal $x[n]$ is depicted in the figure below



B.3) Propose a parametric method (presented in class) to estimate the spectrum of the signal $x[n]$ (which is NOT a harmonic signal), assuming to have measured $x[1], \dots, x[1000]$. The number of parameters (of the parametric method) should be automatically estimated by the method itself. Precisely describe such method. You are asked to detail each step as if you have to implement the method in a computer. Precisely indicate the input and output of each step.

Part C: Permafrost Analysis (10 points)

The study of the rock structure provided the parameters α_k, $k = 1, \dots, 5$, which are linked to the type of material (air or permafrost) between layer k and $k + 1$.

In particular, the permafrost can be classified based on its (relative) temperature. We can consider 4 classes of permafrost. Consequently, the type of material can be divided into 5 classes: 1 of air and 4 of permafrost.

We can model the type of material as a realization of a discrete valued process $V[n]$, taking 5 possible values. In addition, we can assume that there is a dependency in the material between successive layers. Consequently we model $V[n]$ as a Markov chain with 5 states, that is, the 5 possible types of material.

In practice, due to the inhomogeneity of the material between layers and to the measurement noise, what we obtain is rather a **noisy Markov chain** $S[n] = V[n] + W[n]$, where $W[n]$ is a Gaussian white noise. That is, we have a hidden Markov model.

We suppose to have measured the realizations $s[1], \ldots, s[5000]$ of the process $S[n]$.

C.1) Develop the expression of the cumulative distribution and the density of process $S[n]$.

C.2) Propose a method to de-noise $s[1], \ldots, s[5000]$ in order to estimate $v[1], \ldots, v[5000]$. Precisely describe such method. You are asked to detail each step as if you have to implement the method in a computer. Precisely indicate the input and output of each step. Remember that first you need to estimate the parameters of the model and then de-noise the signal.

Part D: Rock Instability Characterization (14 points)

A total of $K = 2000$ rock faces has been analysed using the rock monitoring device presented above. In order to characterize the rock instability, for each rock face k, $k = 1, \ldots, 2000$, 5 variables have been taken into account.

- The type of permafrost between the first and second layer $p_1[k]$;
- The parameter α_1 between the first and second layer $a_1[k]$;
- The type of permafrost between the second and third layer $p_2[k]$;
- The parameter α_2 between the second and third layer $a_2[k]$;
- The resonance frequency $f[k]$.

D.1) Describe in detail, step by step, how to compute the principal components given the variables $\boldsymbol{v}[k] = [p_1[k], a_1[k], p_2[k], a_2[k], f[k]]$, $k = 1, \ldots, 2000$. We shall denote the principal components as $c_1[k], \ldots, c_5[k]$, $k = 1, \ldots, 2000$. Each step should be able 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. Also, **clearly indicate the dimensions of the matrices and vectors**.

D.2) What do the principal components represent?

After analyzing the variance of the principal components, it clearly appears that 3 principal components, namely $c_1[k]$, $c_2[k]$, and $c_3[k]$, $k = 1, \ldots, 2000$ account for most of the total sample variation.

D.3) What does it mean, in relation to the variables $\boldsymbol{v}$, that only 3 principal components account for most of the total sample variation?

By looking at the 3D plot of the 3 principal components we can isolate 3 different clusters.

D.4) Provide a Gaussian Mixture model based on the principal components $c_1[k]$, $c_2[k]$, and $c_3[k]$, $k = 1, \ldots, 2000$, describing the 3 clusters. More precisely, provide its cumulative distribution function and do not forget that the cluster plot has a total of 2000 points!

Call m_1, m_2, and m_3, respectively, the centers of the 3 clusters.

D.5) What do these centers represent, in relation to the variables $\boldsymbol{v}$?

D.6) Now that you have the centers of the 3 clusters, let's go back to the variable $\boldsymbol{v}$ space. How can you do so? Write the corresponding equations.

Solution 3.

Part A: Study of the Rock Structure

A.1) Given the absence of noise, the optimal parametric method to estimate the positions τ_k is the annihilating filter method (simple and computationally efficient).

- We have $x(t)$ for $0 \leq t \leq 0.1$ s with a maximum number of received pulses given by 5.
- 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 $x(t)$ (considered periodic with a period of $\tau = 0.1$ s)

$$\widehat{x}[n] = \frac{1}{\tau} \int_0^{\tau} x(t) e^{-j2\pi n \frac{t}{\tau}} = \frac{1}{0.1} \sum_{k=1}^{5} \alpha_k e^{-j2\pi n \frac{\tau_k}{0.1}} .$$

- Targeting the estimation of the position of 5 spikes, we need an annihilating filter with impulse response of length 5. The corresponding system reads

$$\begin{bmatrix} \widehat{x}[4] & \ldots & \widehat{x}[0] \\ \vdots & \ddots & \vdots \\ \widehat{x}[8] & \ldots & \widehat{x}[4] \end{bmatrix} \begin{bmatrix} h[1] \\ \vdots \\ h[5] \end{bmatrix} = - \begin{bmatrix} \widehat{x}[5] \\ \vdots \\ \widehat{x}[9] \end{bmatrix} .$$

  By solving the system we obtain $h[1], \ldots, h[5]$ (Toeplitz system, requiring 5^2 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

$$H(z) = \sum_{n=0}^{5} h[n] = 1 + h[1]z^{-1} + \ldots + h[5]z^{-5} .$$

– Compute the zeros of the z-transform $H(z)$, that we shall call $z_1, \ldots, z_5$, that is

$$H(z) = \prod_{k=1}^{5}(1 - z_k z^{-k}) .$$

– By taking the argument of the zeros we obtain the positions τ_k, $k = 1, \ldots, 5$, with the following formula

$$\tau_k = \tau \frac{\arg(z_k)}{2\pi} = 0.1\frac{\arg(z_k)}{2\pi} ,$$

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

– Now that we have the values of τ_k, $k = 1, \ldots, 5$, by exploiting the expression of the Fourier transform

$$\widehat{x}[n] = \frac{1}{0.1}\sum_{k=1}^{5}\alpha_k e^{-j2\pi n\frac{\tau_k}{0.1}} ,$$

we can write the linear system

$$\begin{bmatrix} 1 & \ldots & 1 \\ e^{-j2\pi\frac{\tau_1}{0.1}} & \ldots & e^{-j2\pi\frac{\tau_5}{0.1}} \\ \vdots & & \vdots \\ e^{-j2\pi 4\frac{\tau_1}{0.1}} & \ldots & e^{-j2\pi 4\frac{\tau_5}{0.1}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_5 \end{bmatrix} = 0.1 \begin{bmatrix} \hat{x}[0] \\ \hat{x}[1] \\ \vdots \\ \hat{x}[4] \end{bmatrix}$$

which solution provides the amplitudes of the spikes α_k, $k = 1, \ldots, 5$.

A.2) Given the presence of the noise, we need to adopt a different parametric method than the annihilating filter method, since the latter does not work in the presence of noise. The optimal choice is Music. As for the annihilating filter method, Music applies to harmonic signals. Therefore the signal $x(t)$, for $0 \leq t \leq 0.1$ s with a maximum number of received pulses given by 5, first needs to be transformed into a harmonic signal by taking the Fourier transformation (Fourier series) of $x(t)$ (considered periodic with a period of $\tau = 0.1$ s). Music in then applied in the Fourier domain.

$$\widehat{x}[n] = \frac{1}{\tau}\int_0^{\tau} x(t)e^{-j2\pi n\frac{t}{\tau}} = \frac{1}{0.1}\sum_{k=1}^{5}\alpha_k e^{-j2\pi n\frac{\tau_k}{0.1}} .$$

To be noticed that the data in the Fourier domain is complex! So Music is applied on complex data!

Here we have a 5 component harmonic signal and we consider $N >> 5$ samples, *i.e.*, $\widehat{x}[1], \ldots, \widehat{x}[N]$.

– We shall center the signal and use the biased empirical correlation. That is, $m_{\widehat{X}} = \frac{1}{N}\sum_{k=1}^{N}\widehat{x}[k]$, $\widetilde{\widehat{x}}[n] = \widehat{x}[n] - m_{\widehat{X}}$, and

$$\widehat{R}_{\widetilde{\widehat{X}}}[k] = \frac{1}{N}\sum_{n=1}^{N-k}\widetilde{\widehat{x}}[n+k]\widetilde{\widehat{x}}^*[n] , \quad k = 0, \ldots, N-1 , \quad \widehat{R}_{\widetilde{\widehat{X}}}[-k] = \widehat{R}^*_{\widetilde{\widehat{X}}}[k] .$$

Please notice the complex conjugate operator relating the correlation with positive indexes to the one with negative indexes!

Set $5 << M << N$, The empirical correlation matrix is then given by

$$\widehat{\boldsymbol{R}}_{\widehat{\widetilde{X}}}^{M\times M} = \begin{bmatrix} \widehat{R}_{\widehat{\widetilde{X}}}[0] & \widehat{R}_{\widehat{\widetilde{X}}}[1] & \cdots & \widehat{R}_{\widehat{\widetilde{X}}}[M-1] \\ \widehat{R}_{\widehat{\widetilde{X}}}[-1] & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \widehat{R}_{\widehat{\widetilde{X}}}[-M+1] & \cdots & \cdots & \widehat{R}_{\widehat{\widetilde{X}}}[0] \end{bmatrix} .$$

Notice that we set M bigger than the number of positions we are looking for, so to exploit redundancy for the estimation of the frequencies, and smaller than the number of samples, so to reduce the extreme lag errors of the correlation.

- Compute the M eigenvalues $\boldsymbol{\lambda}$ and M eigenvectors $\boldsymbol{g}$ of $\widehat{\boldsymbol{R}}_{\widehat{\widetilde{X}}}^{M\times M}$.
- Call $\boldsymbol{G}^{M\times(M-5)}$ the matrix of the $M-5$ eigenvectors corresponding to the $M-5$ smaller eigenvalues.
- Define the vector $\boldsymbol{e}^{M\times 1}(\omega) = \begin{bmatrix}1 & e^{-j\omega} & \dots & e^{-j(M-1)\omega}\end{bmatrix}^T$ as a function of the variable ω.
- Find the 5 values of ω minimizing the equation

$$\boldsymbol{e}^{M\times 1}(\omega)^H \widehat{\boldsymbol{G}}^{M\times(M-5)} \widehat{\boldsymbol{G}}^{M\times(M-5)^H} \boldsymbol{e}^{M\times 1}(\omega) .$$

- Given that $\omega_k = 2\pi\frac{\tau_k}{5}$, compute the corresponding values of τ_k, $k = 1, \dots, 5$.
- Like for the annihilating filter, now that we have the values of τ_k, $k = 1, \dots, 5$, by exploiting the expression of the Fourier transform

$$\widehat{x}[n] = \frac{1}{5}\sum_{k=1}^{5} \alpha_k e^{-j2\pi n\frac{\tau_k}{5}} ,$$

we can write the linear system

$$\begin{bmatrix} 1 & \dots & 1 \\ e^{-j2\pi\frac{\tau_1}{5}} & \dots & e^{-j2\pi\frac{\tau_5}{5}} \\ \vdots & & \vdots \\ e^{-j2\pi 4\frac{\tau_1}{5}} & \dots & e^{-j2\pi 4\frac{\tau_5}{5}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_5 \end{bmatrix} = 5 \begin{bmatrix} \hat{x}[0] \\ \hat{x}[1] \\ \vdots \\ \hat{x}[4] \end{bmatrix}$$

which solution provides the amplitudes of the spikes α_k, $k = 1, \dots, 5$.

B.1) A truncated sinusoids with frequency $f_1 = 50$ kHz, observed over a finite period of time, can be approximated as the sum of infinite harmonics. Notice that here the signal has been sampled and, therefore, low passed before sampling to avoid aliasing. Limiting ourself to the10th harmonic automatically implies that the sampling frequency f_s is bigger than twice the 10th harmonics, *i.e.*, $f_s > 20f_1$

$$y[n] = \sum_{k=1}^{10} \frac{a_k}{2}\left(e^{i2\pi k\frac{f_1}{f_s}n} + e^{-i2\pi k\frac{f_1}{f_s}n}\right)$$

The latter, can be written as a harmonic stochastic process.

$$Y[n] = \sum_{k=1}^{10} \frac{a_k}{2}\left(e^{i(2\pi k\frac{f_1}{f_s}n+\Theta_k)} + e^{-i(2\pi k\frac{f_1}{f_s}n+\Theta_k)}\right) .$$

B.2) We consider the harmonic stochastic process limited to the 3rd harmonic

$$Y[n] = \sum_{k=1}^{3} \frac{a_k}{2} \left(e^{i(2\pi k \frac{f_1}{f_s} n + \Theta_k)} + e^{-i(2\pi k \frac{f_1}{f_s} n + \Theta_k)} \right) .$$

In order to prove that the model represents a w.s.s. process we need to prove the following:

1) **The mean is constant**. For the computation of the mean we exploit the linearity, and, without loss of generality, we can focus on one single complex exponential:

$$\begin{aligned} \mathrm{E}\left[e^{i(2\pi k \frac{f_1}{f_s} n + \Theta_k)}\right] =& \mathrm{E}\left[e^{i2\pi k \frac{f_1}{f_s} n} e^{i\Theta_k}\right] = e^{i2\pi k \frac{f_1}{f_s} n} \mathrm{E}\left[e^{i\Theta_k}\right] \\ =& e^{i2\pi k \frac{f_1}{f_s} n} \int_0^{2\pi} \frac{1}{2\pi} e^{i\theta} d\theta = e^{i2\pi k \frac{f_1}{f_s} n} \frac{1}{j2\pi} \left[e^{i2\pi} - e^{i0}\right] = 0 \end{aligned}$$

The mean is constant, equal zero.

2) **The correlation only depends on the difference of the time lags**. For the computation of the correlation notice that

$$\mathrm{E}\left[Y[n+l]Y[n]^*\right] =$$

$$\mathrm{E}\left[\sum_{k=1}^{3}\sum_{m=1}^{3} \frac{a_k}{2}\frac{a_m^*}{2} \left(e^{i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)} + e^{-i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)}\right)\left(e^{-i(2\pi m \frac{f_1}{f_s} n + \Theta_k)} + e^{+i(2\pi m \frac{f_1}{f_s} n + \Theta_k)}\right)\right]$$

leads to cross products of the type

$$\mathrm{E}\left[e^{\pm i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)} e^{\pm i(2\pi m \frac{f_1}{f_s} n + \Theta_m)}\right] .$$

If $k \neq m$ then the above expectation reads

$$\mathrm{E}\left[e^{\pm i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)} e^{\pm i(2\pi m \frac{f_1}{f_s} n + \Theta_m)}\right] = \mathrm{E}\left[e^{\pm i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)}\right] \mathrm{E}\left[e^{\pm i(2\pi m \frac{f_1}{f_s} n + \Theta_m)}\right] = 0 .$$

If $k = m$ then the above expectation reads

$$\mathrm{E}\left[e^{\pm i(2\pi k \frac{f_1}{f_s}(n+l) + \Theta_k)} e^{\pm i(2\pi m \frac{f_1}{f_s} n + \Theta_m)}\right] = \begin{cases} \mathrm{E}\left[e^{+i(2\pi k \frac{f_1}{f_s}(2n+l) + 2\Theta_k)}\right] = 0 , \\ \mathrm{E}\left[e^{-i(2\pi k \frac{f_1}{f_s}(2n+l) - 2\Theta_k)}\right] = 0 , \\ \mathrm{E}\left[e^{i2\pi k \frac{f_1}{f_s} l}\right] = e^{i2\pi k \frac{f_1}{f_s} l} , \\ \mathrm{E}\left[e^{-i52\pi k \frac{f_1}{f_s} l}\right] = e^{-i2\pi k \frac{f_1}{f_s} l} . \end{cases}$$

Therefore these cross products only depend on the difference l of the time lags.

3) **The variance is finite**. A straightforward consequence of the fact that the amplitudes of the sinusoids are finite and that the number of the sinusoids are finite. Indeed:

$$\begin{aligned} \mathrm{Var}\left(Y[n]\right) =& \mathrm{E}\left[|Y[n]|^2\right] - \left|\mathrm{E}\left[Y[n]\right]\right|^2 = \mathrm{E}\left[|Y[n]|^2\right] \\ =& \mathrm{E}\left[\sum_{k=1}^{3}\sum_{m=1}^{3} \frac{a_k}{2}\frac{a_m^*}{2} \left(e^{i(2\pi k \frac{f_1}{f_s} n + \Theta_k)} + e^{-i(2\pi k \frac{f_1}{f_s} n + \Theta_k)}\right)\left(e^{-i(2\pi m \frac{f_1}{f_s} n + \Theta_k)} + e^{+i(2\pi m \frac{f_1}{f_s} n + \Theta_k)}\right)\right] \end{aligned}$$

By exploiting the results of the correlation computation, with $l = 0$, we have

$$\mathrm{Var}\left(Y[n]\right) = \sum_{k=1}^{3} \left|\frac{a_k}{2}\right|^2 < \infty .$$

B.3) The signal $x[n]$ is NOT a harmonic signal. It has the energies concentrated around the resonance frequency and, from the spectrum shown in Fig 2, it can be seen as signal with smooth spectrum.

We assume the observed measurements $x[1], \ldots, x[1000]$ to be the realisation of a w.s.s. process $X[n]$

The spectrum can be approximated as a smooth spectrum that, in turn, can be approximated as a rational spectrum

$$S_X(\omega) = \frac{1}{C(z)} = \left.\frac{\sigma_W^2}{|P(z)|^2}\right|_{z=e^{j\omega}} .$$

We therefore have a parametric spectrum, where the parameters $p_1, \ldots, p_M$, and σ_W.

$P(z) = 1 + p_1 z^{-1} + \ldots + p_M z^{-M}$ is assumed strictly minimum phase. Then $H(z) = \frac{1}{P(z)}$ is a stable filter. Given that $W[n]$ is *w.s.s.*, by the fundamental filtering formula, the process $X[n] = H(z)W[n]$ is w.s.s., and, given its expression, it corresponds to an autoregressive process.

The parameters of $H(z)$ can be then estimated using the Yule-Walker method for the estimation of the parameters of an autoregressive process.

It is required that the number of parameters $M+1$ should be automatically estimated by the method itself. The order M of the autoregressive model can be estimated by using the Levinson's algorithm to solve the Yule Walker equations.

- Fix an error variation threshold ε;
- Center the measurements $x[n]$, $n = 1, \ldots, 1000$, obtaining the cantered values $\bar{x}[n]$, $n = 1, \ldots, 1000$ (notice that *a priori* an AR is zero mean. Nevertheless, since we work by approximation, we should then first center the process $x[n]$);
- Apply the Levinson's algorithm on the $N = 1000$ samples $x[n]$, $n = 1, \ldots, 1000$, that is, iteratively estimate the parameters using the Levinson's algorithm equations by starting with an AR of order $M = 1$ and by increasing the order at each iteration.
- Check the value of the reflection coefficients/error at every iteration;
- Stop the iteration when the reflection coefficients/error remains stable, *i.e.* when the difference of the reflection coefficients/error between two iteration is less than ε. This will give the set o parameters $p_1, \ldots, p_M$, and σ_W for a determined M (the latter corresponds to the iteration number);
- Then $H(z) = \frac{1}{1+\widehat{p}_1 z^{-1}+\ldots+\widehat{p}_M z^{-M}}$

Part C: Permafrost Analysis (10 points)

The study of the rock structure provided the parameters α_k, $k = 1, \ldots, 5$, which are linked to the type of material (air or permafrost) between layer k and $k+1$.

In particular, the permafrost can be classified based on its (relative) temperature. We can consider 4 classes of permafrost. Consequently, the type of material can be divided into 5 classes: 1 of air and 4 of permafrost.

We can model the type of material as a realization of a discrete valued process $V[n]$, taking 5 possible values. In addition, we can assume that there is a dependency in the material between successive layers. Consequently we model $V[n]$ as a Markov chain with 5 states, that is, the 5 possible types of material.

In practice, due to the inhomogeneity of the material between layers and to the measurement noise, what we obtain is rather a **noisy Markov chain** $S[n] = V[n] + W[n]$, where $W[n]$ is a Gaussian white noise. That is, we have a hidden Markov model.

We suppose to have measured the realizations $s[1], \ldots, s[5000]$ of the process $S[n]$.

C.1) Develop the expression of the cumulative distribution and the density of process $S[n]$.

$$F_{S[n]}(s[n]) = \mathrm{P}\left(S[n] \leq s[n]\right) = \sum_{k=1}^{5} \mathrm{P}\left(S[n] \leq s[n], V[n] = k\right)$$
$$\overset{\text{Bayes}}{=} \sum_{k=1}^{5} \mathrm{P}\left(S[n] \leq s[n] | V[n] = k\right) \mathrm{P}\left(V[n] = k\right) .$$

Notice that Given $V[n] = v[n]$, $S[n] = W[n] + v[n]$ is a white noise, *i.e.* where $\mathrm{P}\left(S[n] \leq s[n] \,|\, V[n] = v\right.$ is a Gaussian cumulative distribution with mean $m = v[n]$ and variance σ_W^2. The latter admits a density, *i.e.*, a Gaussian probability density function $\mathcal{G}_{v[n],\sigma_W^2}(s[n])$. As a consequence of the last point, $S[n]$ also admits a density.

$$f_{S[n]}(s[n]) = \sum_{k=1}^{5} \mathcal{G}_{k,\sigma_W^2}(s[n]) \mathrm{P}\left(V[n] = k\right) .$$

When we consider all the 5000 time instants, and we indicate $\boldsymbol{S} = [S[1], \ldots, S[5000]]$, $\boldsymbol{s} = [s[1], \ldots, s[5000]]$, $\boldsymbol{V} = [V[1], \ldots, V[5000]]$, and $\boldsymbol{v} = [v[1], \ldots, v[5000]]$, we have

$$F_{\boldsymbol{S}}(\boldsymbol{s}) = \sum_{\boldsymbol{v} \in \mathcal{V}} \mathrm{P}\left(\boldsymbol{S} \leq \boldsymbol{s}, \boldsymbol{V} = \boldsymbol{v}\right) \overset{\text{Bayes}}{=} \sum_{\boldsymbol{v} \in \mathcal{V}} \mathrm{P}\left(\boldsymbol{S} \leq \boldsymbol{s} \,|\, \boldsymbol{V} = \boldsymbol{v}\right) \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right)$$

where $\mathcal{V}$ represents all the possible combinations of the 5 types of material over the 5000 observations. $V[n]$ is a Markov chain, therefore

$$\begin{aligned} \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right) =& \mathrm{P}\left(V[1] = v[1]\right) \mathrm{P}\left(V[2] = v[2] | V[1] = v[1]\right), \ldots \\ & \ldots, \mathrm{P}\left(V[5000] = v[5000] | V[4999] = v[4999]\right) \\ =& \pi_{v[1]} p_{v[1],v[2]}, \ldots, p_{v[4999],v[5000]} ; \end{aligned}$$

Nice that, $\boldsymbol{V} = \boldsymbol{v}$, $\boldsymbol{S} = \boldsymbol{W} + \boldsymbol{v}$ is a White noise process, ie, a sequence of i.i.d. random variables). Consequently

$$\mathrm{P}\left(\boldsymbol{S} \leq \boldsymbol{s} \,|\, \boldsymbol{V} = \boldsymbol{v}\right) = \prod_{n=1}^{5000} \mathrm{P}\left(S[n] \leq s[n] \,|\, V[n] = v[n]\right) .$$

Finally,

$$F_{\boldsymbol{S}}(\boldsymbol{s}) = \sum_{\boldsymbol{v} \in \mathcal{V}} \prod_{n=1}^{5000} \mathrm{P}\left(S[n] \leq s[n] \,|\, V[n] = v[n]\right) \pi_{v[1]} p_{v[1],v[2]}, \ldots, p_{v[4999],v[5000]} .$$

From the expression of the densityof $S[n]$, we have

$$f_{\boldsymbol{S}}(\boldsymbol{s}) = \sum_{\boldsymbol{v} \in \mathcal{V}} \prod_{n=1}^{5000} \mathcal{G}_{v[n],\sigma_W^2}(s[n]) \pi_{v[1]} p_{v[1],v[2]}, \ldots, p_{v[4999],v[5000]} .$$

C.2) We have a typical incomplete data problem, where $f_{\boldsymbol{S}}(\boldsymbol{s})$ is a Gaussian mixture model with Markovian classes. In order to denoise the signal $V[n]$, *i.e.*, to recover the Markov chain, we need to

- Estimate the parameters of the probability density function of the observed data $f_{\boldsymbol{S}}(\boldsymbol{s})$ using the maximum likelihood approach. That is, find the parameters $\hat{\theta}$ that maximise the corresponding likelihood function

$$h(\boldsymbol{s}\ ;\boldsymbol{\theta}) = \sum_{\boldsymbol{v}\in\mathcal{V}} \prod_{n=1}^{5000} \mathcal{G}_{v[n],\sigma_W^2}(s[n])\pi_{v[1]}p_{v[1],v[2]},\ldots,p_{v[4999],v[5000]}\,.$$

  Maximization is done with respect to the parameters

$$\boldsymbol{\theta} = \left\{v[1],\ldots,v[5],\sigma_W^2,\pi_i,p_{i,j},i,j=1,\ldots,5\right\}\,,$$

  and can be achieved by mean of the EM algorithm.

- Estimate the most probable realisation $v[1],\ldots,v[5000]$ of the Markov chain $V[n]$, *i.e.*, denoise the hidden Markov chain. This can be done by finding the realisation $\hat{\boldsymbol{v}}$ that maximises the *a posteriori* distribution

$$\hat{\boldsymbol{v}} = \arg\max_{\boldsymbol{v}} \mathrm{P}\left(\boldsymbol{V}=\boldsymbol{v}|\boldsymbol{s}\right)\,,\quad \text{under the constraints}\quad v[n]\in\{1,2,3,4,5\}\,,\forall n\,,$$

  where, by definition

$$\mathrm{P}\left(\boldsymbol{V}=\boldsymbol{v}|\boldsymbol{s}\right) = \frac{h(\boldsymbol{s},\boldsymbol{v}\ ;\hat{\boldsymbol{\theta}})}{h(\boldsymbol{s}\ ;\hat{\boldsymbol{\theta}})} = \frac{f_{\boldsymbol{S}}(\boldsymbol{s}|\boldsymbol{V}=\boldsymbol{v})\mathrm{P}\left(\boldsymbol{V}=\boldsymbol{v}\right)}{f_{\boldsymbol{S}}(\boldsymbol{s})}\,.$$

  Maximization can be achieved by mean of the Viterbi algorithm.

Part D: Rock Instability Characterization (14 points)

A total of $K = 2000$ rock faces has been analysed using the rock monitoring device presented above. In order to characterize the rock instability, for each rock face k, $k = 1,\ldots,2000$, 5 variables have been taken into account.

- The type of permafrost between the first and second layer $p_1[k]$;
- The parameter α_1 between the first and second layer $a_1[k]$;
- The type of permafrost between the second and third layer $p_2[k]$;
- The parameter α_2 between the second and third layer $a_2[k]$;
- The resonance frequency $f[k]$.

D.1) Computing the principal components

- Given $\boldsymbol{v}[k] = [p_1[k],a_1[k],p_2[k],a_2[k],f[k]]$ (dimensions 1×5), , $k = 1,\ldots,2000$, compute the mean of the variables

$$\boldsymbol{m}_v = \frac{1}{2000}\sum_{k=1}^{2000}\boldsymbol{v}[k]\,,\quad \text{(dimensions } 1\times 5)\,.$$

– Given the mean $\boldsymbol{m}_v$, center the variables

$$\bar{\boldsymbol{v}}[k] = \boldsymbol{v}[k] - \boldsymbol{m}_v\,, \quad (\text{dimensions } 1 \times 5)\,, \quad k = 1, \ldots, 2000\,.$$

– Given the centred variables $\bar{\boldsymbol{v}}[k]$, $k = 1, \ldots, 2000$, compute the empirical correlation matrix of the centred variables

$$\widehat{\boldsymbol{R}}_v = \frac{1}{2000} \sum_{k=1}^{2000} \bar{\boldsymbol{v}}[k]^T \bar{\boldsymbol{v}}[k]\,, \quad (\text{dimensions } 5 \times 5)\,.$$

– Given empirical correlation matrix $\widehat{\boldsymbol{R}}$ diagonalize it

$$\boldsymbol{V}^t \widehat{\boldsymbol{R}}_v \boldsymbol{V} = \boldsymbol{\Lambda}\,, \quad (\text{dimensions } 5 \times 5)\,,$$

where $\boldsymbol{V}$ is the matrix of eigenvectors (dimensions 5×5).

– Given the the centred variables $\bar{\boldsymbol{v}}[k]$, $k = 1, \ldots 2000$ and the matrix of eigenvectors $\boldsymbol{V}$, compute the principal components

$$\boldsymbol{c}[k] = \bar{\boldsymbol{v}}[k]\boldsymbol{V}\,, \quad (\text{dimensions } 1 \times 5)\,, \quad k = 1, \ldots, 2000\,,$$

where $\boldsymbol{c}[k] = [c_1[k], \ldots, c_5[k]]$, $k = 1, \ldots, 2000$.

D.2) What do the principal components represent? For every $k = 1, \ldots, 2000$, the 5 principal components $\boldsymbol{v}[k]$ represents the one-to-one projection of the 5 centred variable $\bar{\boldsymbol{v}}[k]$ into a 5 dimensional orthogonal space (orthogonalisation of the centred variable)

D.3) It means that for each $k = 1, \ldots, 2000$ the centred version $\bar{\boldsymbol{v}}[k]$ of the 5 variables $\boldsymbol{v}[k]$ are a linear combination of only 2 of the 5 principal components $\boldsymbol{p}[k]$. That is, for each $k = 1, \ldots, 100'000$, the projection of the 5 centred version $\bar{\boldsymbol{v}}[k]$ of the variables $\boldsymbol{v}[k]$ into the principal component space results in only two components.

By looking at the 3D plot of the 3 principal components we can isolate 3 different clusters.

D.4) Provide a Gaussian Mixture model based on the principal components $c_1[k]$, $c_2[k]$, and $c_3[k]$, $k = 1, \ldots, 2000$, describing the 3 clusters. More precisely, provide its cumulative distribution function and do not forget that the cluster plot has a total of 2000 points!

Call m_1, m_2, and m_3, respectively, the centers of the 3 clusters.

D.5) What do these centers represent, in relation to the variables $\boldsymbol{v}$?

D.6) Now that you have the centers of the 3 clusters, let's go back to the variable $\boldsymbol{v}$ space. How can you do so? Write the corresponding equations.

We have 100'000 realizations (data size) of 5 variables, *i.e.*, $\boldsymbol{v}[k] = [i_a[k], i_d[k], t_i[k], w_a[k], w_d[k]]$ (dimensions 1×5), $k = 1, \ldots, 100'000$.

The principal components are denoted as $\boldsymbol{p}[k] = [p_1[k], \ldots, p_5[k]]$, $k = 1, \ldots, 100'000$.

Principal Components

– Given $\boldsymbol{v}[k] = [i_a[k], i_d[k], t_i[k], w_a[k], w_d[k]]$ (dimensions 1×5), $k = 1, \ldots, 100'000$, compute the mean of the variables

$$\boldsymbol{m}_v = \frac{1}{100'000} \sum_{k=1}^{100'000} \boldsymbol{v}[k] \,, \quad (\text{dimensions } 1 \times 5) \,.$$

– Given the mean $\boldsymbol{m}_v$, center the variables

$$\bar{\boldsymbol{v}}[k] = \boldsymbol{v}[k] - \boldsymbol{m}_v \,, \quad (\text{dimensions } 1 \times 5) \,, \quad k = 1, \ldots, 100'000 \,.$$

– Given the centred variables $\bar{\boldsymbol{v}}[k]$, $k = 1, \ldots, 100'000$, compute the empirical correlation matrix of the centred variables

$$\widehat{\boldsymbol{R}}_v = \frac{1}{100'000} \sum_{k=1}^{100'000} \bar{\boldsymbol{v}}[k]^T \bar{\boldsymbol{v}}[k] \,, \quad (\text{dimensions } 5 \times 5) \,.$$

– Given empirical correlation matrix $\widehat{\boldsymbol{R}}$ diagonalize it

$$\boldsymbol{V}^t \widehat{\boldsymbol{R}}_v \boldsymbol{V} = \boldsymbol{\Lambda} \,, \quad (\text{dimensions } 5 \times 5) \,,$$

where $\boldsymbol{V}$ is the matrix of eigenvectors (dimensions 5×5).

– Given the the centred variables $\bar{\boldsymbol{v}}[k]$, $k = 1, \ldots, 100'000$ and the matrix of eigenvectors $\boldsymbol{V}$, compute the principal components

$$\boldsymbol{p}[k] = \bar{\boldsymbol{v}}[k] \boldsymbol{V} \,, \quad (\text{dimensions } 1 \times 5) \,, \quad k = 1, \ldots, 100'000 \,.$$

B.2)

B.3) It means that for each $k = 1, \ldots, 100'000$ the centred version $\bar{\boldsymbol{v}}[k]$ of the 5 variables $\boldsymbol{v}[k]$ are a linear combination of only 2 of the 5 principal components $\boldsymbol{p}[k]$. That is, for each $k = 1, \ldots, 100'000$, the projection of the 5 centred version $\bar{\boldsymbol{v}}[k]$ of the variables $\boldsymbol{v}[k]$ into the principal component space results in only two components.

B.4) It means that the $100'000$ car passages can be divided in 3 groups, as if there are only 3 types of car.

B.5) Provided that each cluster can be modelled as a 2 dimension Gaussian distribution, the 3 2D clusters can be modelled as a 2 dimension Gaussian mixture model. The variables of the clusters are the two principal components $\boldsymbol{p}[k] = [p_1[k], p_2[k]]$, $k = 1, \ldots, 100'000$. They can be considered at the realization of a 2D stochastic process $\boldsymbol{P}[k] = [P_1[k], P_2[k]]$. For each $k = 1, \ldots, 100'000$, the cumulative distribution reads

$$F_{\boldsymbol{P}[k]}(\boldsymbol{p}[k]) = \sum_{c[k]=1}^{3} \mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k], C[k] = c[k]\right) \overset{\text{Bayes}}{=} \sum_{c[k]=1}^{3} \mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k] \,|\, C[k] = c[k]\right) \mathrm{P}\left(C[k] = c[k]\right)$$

where $C[k]$ is the random variable indicating the cluster, *i.e.*, indicating which one of the 3 types of car (groups) has generated the 2 pressure events at the passage number k.

When we consider all the $100'000$ passages and we indicate $\boldsymbol{P} = [\boldsymbol{P}[1], \ldots, \boldsymbol{P}[100'000]]$, $\boldsymbol{p} = [\boldsymbol{p}[1], \ldots, \boldsymbol{p}[100'000]]$ and $\boldsymbol{C} = [C[1], \ldots, C[100'000]]$, we have

$$F_{\boldsymbol{P}}(\boldsymbol{p}) = \sum_{\boldsymbol{c} \in \mathcal{C}} \mathrm{P}\left(\boldsymbol{P} \leq \boldsymbol{p}, \boldsymbol{C} = \boldsymbol{c}\right) \stackrel{\text{Bayes}}{=} \sum_{\boldsymbol{c} \in \mathcal{C}} \mathrm{P}\left(\boldsymbol{P} \leq \boldsymbol{p} \,|\, \boldsymbol{C} = \boldsymbol{c}\right) \mathrm{P}\left(\boldsymbol{C} = \boldsymbol{c}\right)$$

where $\mathcal{C}$ represents all the possible combinations of types of car (groups) in the $100'000$ passages.

Notice that:

- We can assume that the types of car passing by, form an independent sequence (no reason to suppose the contrary), *i.e.*

$$\mathrm{P}\left(\boldsymbol{C} = \boldsymbol{c}\right) = \prod_{k=1}^{100'000} \mathrm{P}\left(C[k] = c[k]\right) ;$$

- Given the type of car that passed by, the two pressure events can be considered to be equal to a pair of constant values plus noise, *i.e.*

$$\mathrm{P}\left(\boldsymbol{P} \leq \boldsymbol{p} \,|\, \boldsymbol{C} = \boldsymbol{c}\right) = \prod_{k=1}^{100'000} \mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k] \,|\, C[k] = c[k]\right) ,$$

  where $\mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k] \,|\, C[k] = c[k]\right)$ is a Gaussian (bivariate) cumulative distribution. The latter admits a Gaussian (bivariate) density $\mathcal{G}_{\boldsymbol{m}[c[k]], \Sigma^2[c[k]]}(\boldsymbol{p}[k])$ with mean $\boldsymbol{m}[c[k]] = [m_1[c[k]], m_2[c[k]]]$ and covariance matrix $\Sigma^2[c[k]]$;
- Given that the principal components are independent, the covariance matrix is diagonal

$$\Sigma^2[c[k]] = \begin{bmatrix} \sigma_1^2[c[k]] & 0 \\ 0 & \sigma_2^2[c[k]] \end{bmatrix} .$$

  Therefore, calling $\boldsymbol{\sigma}^2[c[k]] = [\sigma_1^2[c[k]], \sigma_w^2[c[k]]]$

$$\mathcal{G}_{\boldsymbol{m}[c[k]], \Sigma^2[c[k]]}(\boldsymbol{p}[k]) = \mathcal{G}_{\boldsymbol{m}[c[k]], \boldsymbol{\sigma}^2[c[k]]}(\boldsymbol{p}[k]) ,$$

  and

$$\begin{aligned} \mathcal{G}_{\boldsymbol{m}[c[k]], \boldsymbol{\sigma}^2[c[k]]}(\boldsymbol{p}[k]) = & \frac{1}{\sqrt{2\pi\sigma_1^2[c[k]]}} \exp\left(\frac{(p_1[k] - m_1[c[k]])^2}{\sigma_1^2[c[k]]}\right) \\ & \frac{1}{\sqrt{2\pi\sigma_2^2[c[k]]}} \exp\left(\frac{(p_2[k] - m_2[c[k]])^2}{\sigma_2^2[c[k]]}\right) \end{aligned}$$

Finally, the cumulative distribution of the (random variables associated to the) two principal components $p_1[k]$ and $p_2[k]$, $k = 1, \ldots, 100'000$, reads

$$\begin{aligned} F_{\boldsymbol{P}}(\boldsymbol{p}) = & \prod_{k=1}^{100'000} \sum_{c[k]=1}^{3} \mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k] \,|\, C[k] = c[k]\right) \mathrm{P}\left(C[k] = c[k]\right) \\ & \prod_{k=1}^{100'000} \sum_{l=1}^{3} \mathrm{P}\left(\boldsymbol{P}[k] \leq \boldsymbol{p}[k] \,|\, C[k] = l\right) \mathrm{P}\left(C = l\right) , \end{aligned}$$

where $\mathrm{P}\left(C = l\right])$, $l, 1, \ldots, 3$, that we shall denote π_l, $l, 1, \ldots, 3$, are the mixture proportions. Such a cumulative distribution admits the density

$$f_{\boldsymbol{P}}(\boldsymbol{p}) = \prod_{k=1}^{100'000} \sum_{l=1}^{3} \mathcal{G}_{\boldsymbol{m}[l], \boldsymbol{\sigma}^2[l]}(\boldsymbol{p}[k]) \mathrm{P}\left(C = l\right)$$

B.6) The center of the 3 clusters are given my the 3 means of the bivariate Gaussian distribution, *i.e.*, $\boldsymbol{m}[l] = [m_1[l], m_2[l]]$, $l = 1, \ldots, 3$. Given the realizations of the two principal components $\boldsymbol{p} = [\boldsymbol{p}[1], \ldots, \boldsymbol{p}[100'000]]$, the means can be obtained by estimating the parameters of the density $f_{\boldsymbol{P}}(\boldsymbol{p})$ using the maximum likelihood method.

- Set the likelihood function

$$h(\boldsymbol{p}\,;\,\boldsymbol{\theta}) = \prod_{k=1}^{100'000} \sum_{l=1}^{3} \mathcal{G}_{\boldsymbol{m}[l], \boldsymbol{\sigma}^2[l]}(\boldsymbol{p}[k]) \pi_l\,,$$

  function of the parameters $\boldsymbol{\theta} = \{m_1[l], m_2[l], \sigma_1^2[l], \sigma_2^2[l], \pi_l, l = 1, \ldots, 3\}$
- Given the realizations of the two principal components $\boldsymbol{p} = [\boldsymbol{p}[1], \ldots, \boldsymbol{p}[100'000]]$, by mean of the EM algorithm, compute the set of the parameters that maximises the likelihood function

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} h(\boldsymbol{p}\,;\,\boldsymbol{\theta})\,.$$

B.7) Each center represent the value of the principal components p_1 and p_2 corresponding the centred version of the variables i_a, i_d, t_i, w_a, w_d describing a characteristic 2 pressure event.

B.8) Set $p_1[l] = m_1[l]$, $p_2[l] = m_2[l]$, $l = 1, \ldots, 3$. Then the 3 set of variables $\boldsymbol{v}[l] = [i_a[l], i_d[l], t_i[l], w_a[l], w_d[l]]$, $l = 1, \ldots, 3$, describing the 3 characteristic 2 pressure events are given by

$$\boldsymbol{v}[l] = \boldsymbol{p}[l]\boldsymbol{V}^T + \boldsymbol{m}_v\,, \quad (\text{dimensions } 1 \times 5)\,, \quad l = 1, \ldots, 3\,.$$

B.9) The variable given in B.8), *i.e.*,

$$\boldsymbol{v}[l] = \boldsymbol{p}[l]\boldsymbol{V}^T + \boldsymbol{m}_v\,, \quad (\text{dimensions } 1 \times 5)\,, \quad l = 1, \ldots, 3\,,$$

represent the three characteristic 2 pressure events. In other words, the $100'000$ sets of variables $\boldsymbol{v}[k] = [i_a[k], i_d[k], t_i[k], w_a[k], w_d[k]]$, $k = 1, \ldots, 100'000$ can be seen as only 3 sets of values $\boldsymbol{v}[l] = [i_a[l], i_d[l], t_i[l], w_a[l], w_d[l]] = \boldsymbol{p}[l]\boldsymbol{V}^T + \boldsymbol{m}_v$, $l = 1, \ldots, 3$, plus noise.

Exercise 4. THE MOVIE AND THE TRAIN ANNOUNCEMENT (10 PTS)

You are watching a movie on your iPad during a journey in a train and you have forgotten your noise cancelling headphones. The coach loudspeaker keeps giving announcements in German at every train station "Wir treffen in Lausanne ein, Ihre nächste Anschluss ..." and they really annoy you.

Having freshly finished the Statistical Signal and Data Processing course, you decide to design an adaptive system to cancel the announcement so to be able to focus on the movie.

A microphone M pics up the sound of the iPad and of the announcement loudspeaker, and we denote its signal by $M[n]$. The microphone signal has two components. The first and most important one is the sound of your iPad denoted by $W[n]$. This is what we would like to listen to, and thus considered as the desired signal. In addition, there is the sound coming from the announcement loudspeaker, denoted by $V_1[n]$. We have

$$M[n] = W[n] + V_1[n] = W[n] + (h_1 * V)[n].$$

Note that the announcement loudspeaker emits the signal $V[n]$, but it gets picked up by the microphone after going through the acoustic propagation channel of the train coach, denoted by $h_1[n]$.

To make things simpler, assume that the sound of the announcement loudspeaker is a single sinusoid with a random initial phase ϕ uniformly distributed in $[0, 2\pi)$ and unknown frequency ω_V:

$$V[n] = \sin(\omega_V n + \phi).$$

Assume also that the sound of your iPad $W[n]$ is a zero-mean i.i.d. Gaussian process with variance $\sigma_W^2 = 1$ (nice movie!).

1) Show a scheme with an *adaptive filter* $f_n[n]$ that can be used to remove the announcement loudspeaker sound from $M[n]$, without using any other reference signals. Indicate the signal that is used for filter coefficients' adaptation and the quantity $J(f_n)$ that the adaptive filter minimizes. By expanding the expression for $J(f_n)$, show that the denoised signal is indeed the best approximation of $W[n]$ when $J(f_n)$ is minimized. *Hint: use the properties of the signals* $W[n]$ *and* $V[n]$.

2) Now assume that another microphone N is positioned in the close proximity of the announcement loudspeaker and it picks up the sound $V_2[n]$, that is only due to the announcement loudspeaker (it completely suppresses the sounds from the iPad). In other words, the signal recorded by the second microphone is

   $$N[n] = V_2[n] = (h_2 * V)[n],$$

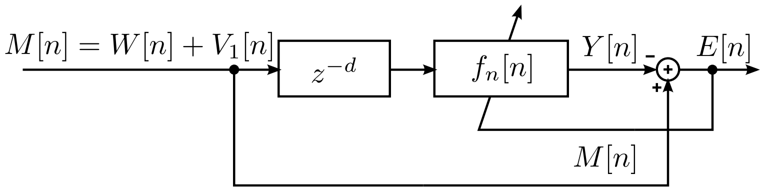
   where $h_2[n]$ models a propagation path from the announcement loudspeaker to the microphone N.

   The signal $N[n]$ is now available to you in order to improve your announcement loudspeaker removal from $M[n]$ with an adaptive filter. Devise a scheme for announcement loudspeaker sound removal with an *adaptive filter* $g_n[n]$ that uses both $M[n]$ and $N[n]$.

   Again, draw a block diagram, show the signal that is used for adaptation and give the quantity $J(g_n)$ that the adaptive filter minimizes. Also, by expanding the expression for $J(g_n)$, show that the denoised signal optimally approximates $W[n]$ when $J(g_n)$ is minimized.

Solution 4.

1) Since $\mathrm{E}[W[n]W[n-k] = 0, \quad k \geq 1$, one could use the denoising scheme without reference depicted in the figure below.



Given that $W[n]$ is i.i.d., the delay d can be as short as a single sample. The signal used for filter coefficients' adaptation is

$$E[n] = M[n] - Y[n] = M[n] - (f_n * M')[n]\,,$$

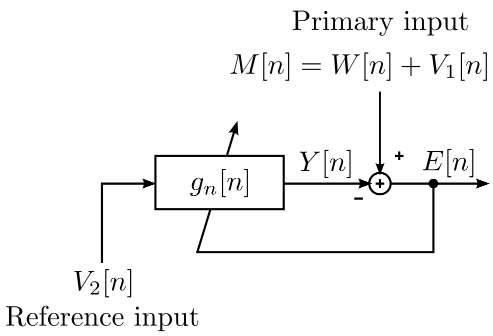
with $M'[n] = M[n-d]$, $d \geq 1$.

The cost function that gets minimized by the adaptive filter is given by:

$$J(f_n) = \mathrm{E}\left[E[n]^2\right] = \mathrm{E}\left[W[n]^2\right] + \mathrm{E}\left[(V_1[n] - Y[n])^2\right]\ .$$

One can see that when the filter minimizes $J(f_n)$, it actually minimizes the error in the subspace orthogonal to $W[n]$. Additionally, the output $Y[n]$ of the adaptive filter $f_n[n]$ is the best approximation of the process $V_1[n]$ (announcement sound picked up by the mike) given its past. Thus, removing the MSE-optimal prediction of the announcement sound given its past will give the optimally denoised sound.

2) This is an easier case, as it represents the standard adaptive noise cancellation with a reference depicted in the figure below.



The signal used for filter coefficients' adaptation is

$$E[n] = M[n] - Y[n] = M[n] - (g_n * V_2)[n]\,.$$

The cost function that gets minimized by the adaptive filter is given by

$$J(g_n) = \mathrm{E}\left[E[n]^2\right] = \mathrm{E}\left[W[n]^2\right] + \mathrm{E}\left[(V_1[n] - Y[n])^2\right].$$

Once again, one can see that the filter operates in the subspace orthogonal to the subspace generated by $W[n]$, giving the best MSE estimate of $V_1[n]$ in the subspace generated by the reference announcement sound $N[n] = V_2[n]$. Thus, removing the optimally predicted annoucement sound from $N[n] = V_2[n]$ gives the optimally denoised sound in this setup.

Grade Scale.

The exams accounts for a total of 60 points (exact response to each question).

The grading has been done on a 50 points scale (50 points = 6/6), according to the following formula

$$\text{grade over } 6 = 1 + (5 * \text{points}/50)$$

The result is then constrained to be at maximum 6. The maximum obtained score was 53.5/50.

The grade of the final exam has been added to the midterm grade and to the mini-project grades and then rounded to .25 steps, that is

$$\text{rounded grade over } 6 = (\text{round-to-0-digit}(4 * \text{grade over } 6))/4\,.$$