Statistical Signal & Data Processing - COM500
Final Exam

Tuesday, June 18, 2019, 12h15-15h45.

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!

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 neighbor; 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.

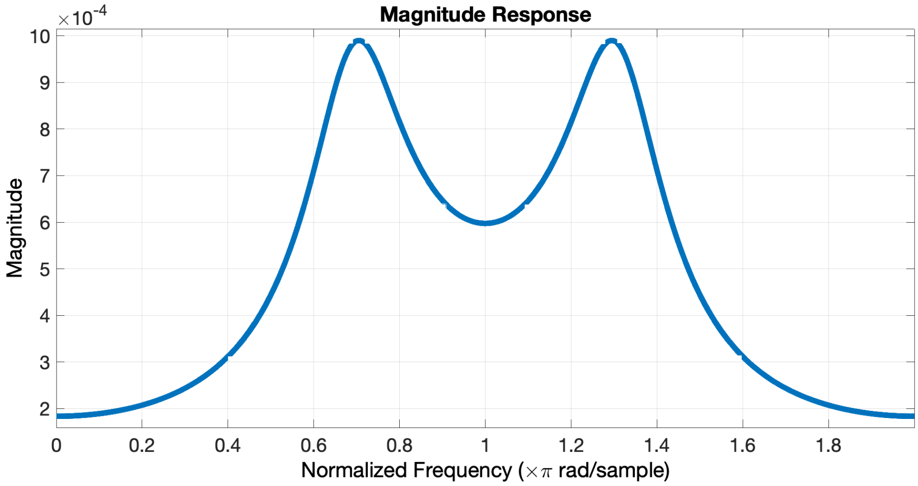
All the best for your exam!!

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. APPROXIMATING A SYSTEM (6 POINTS)

Consider a system with a transfer function $H_0(z)|_{z=e^{j\omega}}$ having the following magnitude



1) Approximate the system $H_0(z)$ with an all zeros system $H_1(z)$. Write the expression of $H_1(z)$, and plot the zeros on the z-plane.

2) Approximate the system $H_0(z)$ with an all poles system $H_2(z)$. Write the expression of $H_2(z)$, and plot the poles on the z-plane.

3) Which of the two systems $H_1(z)$ and $H_2(z)$ will best approximate $H_0(z)$? Justify your answer precisely.

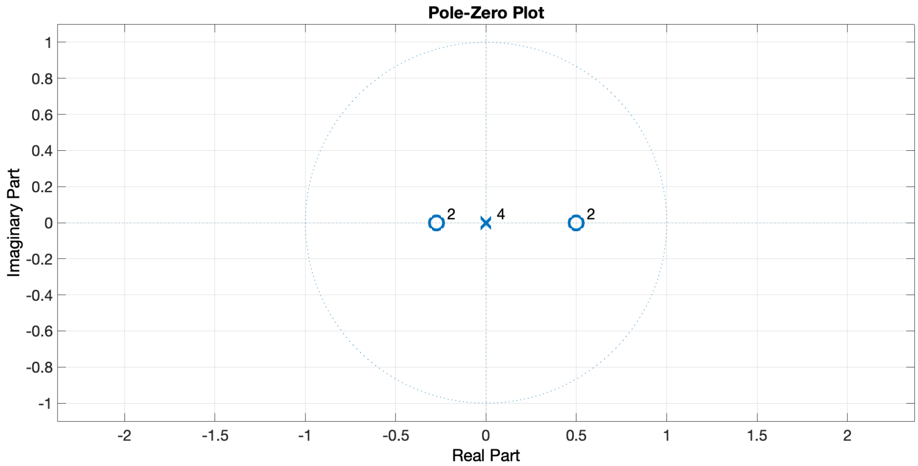
Solution 1.

Recall that poles and zeros at the origin do not play any and are used only for order normalization purpose.

1) A way to approximate $H_0(z)$ with an all zeros system is to place a zero z_1 in the positive real axis and a zero z_2 in the negative real axis, where z_1 is closer to the unit circle than z_2. Given that the effect of zeros is usually not very strong (except when placed on the unit circle), we might consider zeros with higher order. A possible all zero system approximation is the following

$$H_1(z) = \frac{(z - z_1)^2 (z - z_2)^2}{z^2} = (1 - z_1 z^{-1})^2 (1 - z_2 z^{-1})^2 \,,$$

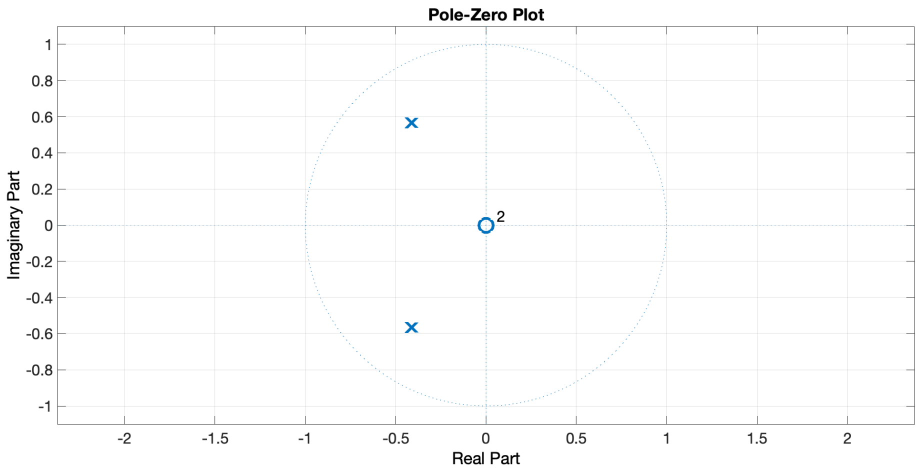
where $z_1 = 0.5$ and $z_2 = -0.25$.



2) A way to approximate $H_0(z)$ with an all pole system is to place a pair of conjugate poles z_1 and z_2 with phase $\approx \pm 0.7\pi$ and magnitude < 1.. A possible all pole system approximation is the following

$$H_1(z) = \frac{z^2}{(z - z_1)(z - z_2)} = \frac{1}{(1 - z_1 z^{-1})(1 - z_2 z^{-1})},$$

where $z_1 = 0.75 e^{j0.7\pi}$ and $z_2 = 0.75 e^{-j0.7\pi}$.



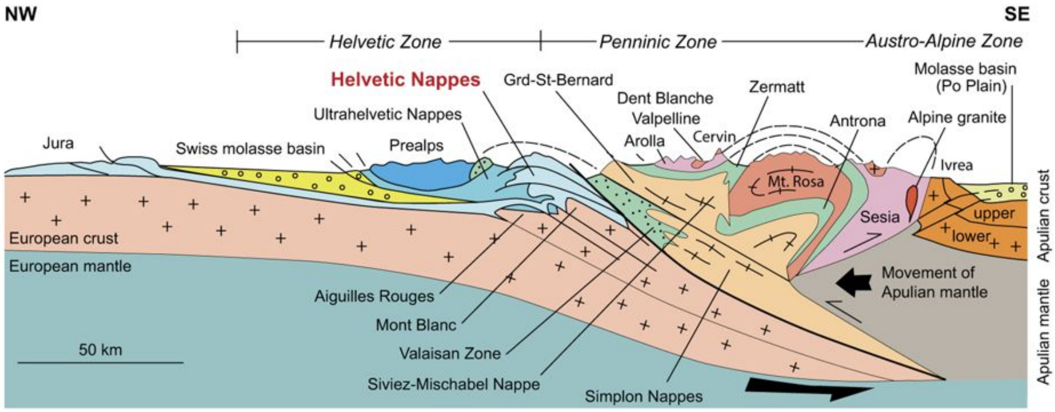
3) When the magnitude of the transfer function is characterized by "peaks", it is better represented/approximated using poles. A single pole can have a strong effect, easily defining the sharpness of the peak.

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.

Exercise 2. Reflection Seismology

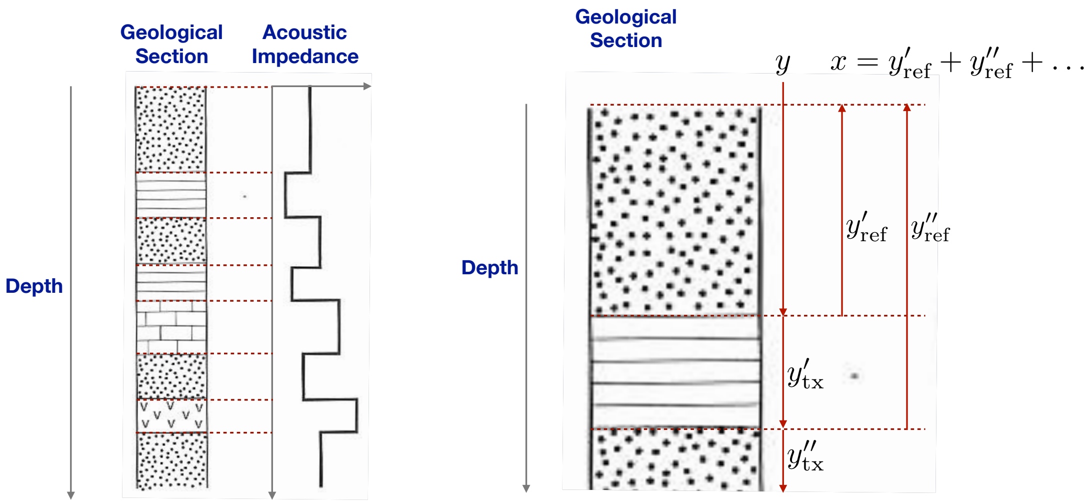
The study of the earth's subsurface is of great interest, enabling the identification of the different **nappes** and the understanding of the **plate tectonics**.



Nappes in the Swiss Alps.

Reflection Seismology (or seismic reflection) is a method of exploration geophysics that estimates the properties of the earth's subsurface from reflected seismic waves.

The basic idea is to propagate an acoustic signal y into the ground. Each subsurface layer has a different acoustic impedance and when the signal encounters a boundary between two layers some of the energy in the wave will be reflected at the boundary, while some of the energy will be transmitted through the boundary (impedance mismatch). The signal x recorded at the surface, will be the sum of all the reflections.



Principle of Reflection Seismology.

The goal is to properly use y to get information on the different geological layers (nappes), their position, and on their acoustic impedance. In particular, the acoustic impedance provides information on the nappe's geological characteristics.

We shall first explore different types of signals y (Part A & B) to get such information, and then model and estimate the acoustic impedance (Part C). The three parts A, B, and C are independent.

Part A: The signal y is a spike (18 points)

We consider y to be a spike, *i.e.*, a signal that can be symbolically modeled in continuous time domain as a Dirac delta $\delta(t)$ (we shall work here in continuous time domain). In practice this can be obtained with an explosion at the surface.

The signal $x(t)$ recorded at the surface is therefore

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

where τ_k, corresponds to propagation delay between the surface and the boundary of the k-th layer and the k-th $+1$ layer, and α_k is linked to the reflection coefficient of the boundary.

The estimation of τ_k is therefore of foremost importance since it provides an estimation of the depth of each boundary between two layers. Similarly, the estimation of α_k enables the computation of the acoustic impedance of the layers.

We record the signal $x(t)$ over an interval of $\tau = 20$ s. Assuming each nappe (layer) to have an average height of 800 m, and the signal (sound) propagation to be 300 $\frac{\text{m}}{\text{s}}$, we can take the maximum number of layers to be 10 and, therefore, a maximum of 10 reflected spikes, therefore obtaining

$$x(t) = \sum_{k=1}^{10} \alpha_k \delta(t - \tau_k) \,, \quad t \in [0, 20] \text{ s} \,.$$

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, $k = 1, \ldots, 10$, and the amplitudes α_k, $k = 1, \ldots, 10$ of the spikes. Precisely describe such method. You are given $\tau = 20$ s of the signal $x(t)$ assuming a maximum of 10 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 (outside, on the ground)

A.2) Considering now the presence of a non-negligible noise, propose a parametric method (presented during the lectures) to estimate the positions τ_k, $k = 1, \ldots, 10$, **and** the amplitudes α_k, $k = 1, \ldots, 10$ of the spikes. Precisely describe such method. You are given $\tau = 20$ s of the signal $x(t)$ assuming a maximum of 10 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: The signal y is a white noise (18 points)

We consider y to be a Gaussian white noise in discrete time, *i.e.*, $y[n] = w[n]$, with variance σ^2 (known).

The signal $x[n]$ recorded at the surface can be modeled as

$$x[n] = \sum_{k=1}^{\infty} h_k w[n-k]\,.$$

Notice that now, k does not represent the k-th layer. Here k is linked to the sampling period T_s. Assuming the signal (sound) propagation to be 300 $\frac{\text{m}}{\text{s}}$, k is linked to a propagation distance of kT_s300 m. Consequently, for those values kT_s300 m corresponding to the depth of a boundary between two layers, the coefficient h_k will be linked to the reflection coefficient of the boundary.

Finally, the estimation of the coefficients h_k, is of foremost importance since it enables the computation of the acoustic impedance of the layers and, indirectly, the depth of each boundary between two layers.

We rewrite the coefficients h_k as $h[k]$, and consider the stochastic processes associated to the measurements, obtaining

$$X[n] = \sum_{k=1}^{\infty} h[k]W[n-k]\,.$$

Assume $h[k] \in \ell_1$ (absolutely summable) and that the corresponding z-transform $H(z)$ has all the zeros strictly inside the unit circle.

B.1) Is $X[n]$ a w.s.s. process? Justify precisely your answer (notice there is no need to make any computation here!).

B.2) What kind of process is $X[n]$? Justify precisely your answer and, if needed, add some assumptions.

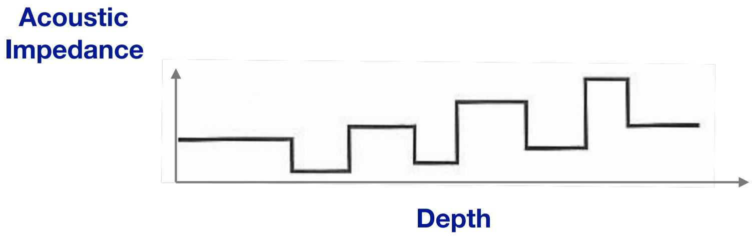
Assume the sampling frequency to be $f_s = 1000$ Hz and to have measured $N = 20000$ samples $x[1], \ldots, x[20000]$. We would like to analyze the coefficients $h[k]$, $k = 1, \ldots$, by analyzing the module of their Fourier transform, *i.e.*, $|H(z)|_{z=e^{j2\pi k/N}}$.

B.3) Propose a parametric method to estimate $|H(z)|_{z=e^{j2\pi k/N}}$ given $x[1], \ldots, x[20000]$. 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.

B.4) In the above parametric method, you need to assume that the number of parameters is known (this is linked to the assumption needed in B.2)). It there a way to estimate the number of parameters? Explain.

Part C: Modeling and estimation of the acoustic impedance (16 points)

Once estimated the depth of each boundary between two layers and the reflection coefficient of the boundary it is possible to obtain the acoustic impedance. Theoretically, we expect something like this



Acoustic Impedance.

As you can see, the acoustic impedance is a realization of a discrete valued process, that we shall denote with $V[n]$. In addition, we can assume that there is a dependency in the sequence of layers, and therefore we model such process as a Markov chain, where we can assume the Markov chain $V[n]$ to have 10 states (discrete values of $V[n]$).

In practice, due to the inhomogeneity of each layer, 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 computed the realizations $s[1], \ldots, s[6000]$ of the process $S[n]$.

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

C.2) Propose a method to de-noise $s[1], \ldots, s[6000]$ in order to estimate $v[1], \ldots, v[6000]$. 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 parameter of the model and then de-noise the signal.

Solution 2.

Part A: The signal y is a spike

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).

Over a period of $\tau = 20$ s, the given maximum number of received pulses is 10.

- We have $x(t)$ for $0 \leq t \leq 20$ s
- 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 $\tau = 20$ s)

$$\widehat{x}[n] = \frac{1}{\tau} \int_0^{\tau} x(t) e^{-j2\pi n \frac{t}{\tau}} = \frac{1}{20} \sum_{k=1}^{10} \alpha_k e^{-j2\pi n \frac{\tau_k}{20}} .$$

– Targeting the estimation of the position of 10 spikes, we need an annihilating filter with impulse response of length 10. The corresponding system reads

$$\begin{bmatrix} \widehat{x}[9] & \dots & \widehat{x}[0] \\ \vdots & \ddots & \vdots \\ \widehat{x}[18] & \dots & \widehat{x}[9] \end{bmatrix} \begin{bmatrix} h[1] \\ \vdots \\ h[10] \end{bmatrix} = - \begin{bmatrix} \widehat{x}[10] \\ \vdots \\ \widehat{x}[19] \end{bmatrix} .$$

By solving the system we obtain $h[1], \dots, h[10]$ (Toeplitz system, requiring 10^2 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

$$H(z) = \sum_{n=0}^{10} h[n] = 1 + h[1]z^{-1} + \dots + h[10]z^{-10} .$$

– Compute the zeros of the z-transform $H(z)$, that we shall call $z_1, \dots, z_{10}$, that is

$$H(z) = \prod_{k=1}^{10} (1 - z_k z^{-k}) .$$

– By taking the argument of the zeros we obtain the positions τ_k, $k = 1, \dots, 10$, with the following formula

$$\tau_k = \tau \frac{\arg(z_k)}{2\pi} = 20 \frac{\arg(z_k)}{2\pi} ,$$

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

– Now that we have the values of τ_k, $k = 1, \dots, 10$, by exploiting the expression of the Fourier transform

$$\widehat{x}[n] = \frac{1}{20} \sum_{k=1}^{10} \alpha_k e^{-j2\pi n \frac{\tau_k}{20}} ,$$

we can write the linear system

$$\begin{bmatrix} 1 & \dots & 1 \\ e^{-j2\pi \frac{\tau_1}{20}} & \dots & e^{-j2\pi \frac{\tau_{10}}{20}} \\ \vdots & & \vdots \\ e^{-j2\pi 9 \frac{\tau_1}{20}} & \dots & e^{-j2\pi 9 \frac{\tau_{10}}{20}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10} \end{bmatrix} = 20 \begin{bmatrix} \hat{x}[0] \\ \hat{x}[1] \\ \vdots \\ \hat{x}[9] \end{bmatrix}$$

which solution provides the amplitudes of the spikes α_k, $k = 1, \dots, 10$.

A.2) In the presence of a non-negligible noise the annihilating filter method cannot be apply. The parametric method seen in class that fits the problem is MUSIC. Recall that, like the annihilating filter, MUSIC applies to a harmonic signal. Therefore, in our case, we apply MUSIC on the data in the Fourier domain

$$\widehat{x}[n] = \frac{1}{20} \sum_{k=1}^{10} \alpha_k e^{-j2\pi n \frac{\tau_k}{20}} .$$

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

Here we have a 10 component harmonic signal and we consider $N >> 10$ samples, *i.e.*, $\widehat{x}[1], \ldots, \widehat{x}[M]$.

- We shall center the signal and use the biased empirical correlation. That is, $m_{\widehat{X}} = \frac{1}{M}\sum_{k=1}^{M}\widehat{x}[k]$, $\widetilde{\widehat{x}}[n] = \widehat{x}[n] - m_{\widehat{X}}$, 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 $10 << M << N$, The empirical correlation matrix is then given by

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

  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 $\widehat{\boldsymbol{R}}^{M\times M}_{\widetilde{\widehat{X}}}$.
- Call $\boldsymbol{G}^{M\times(M-10)}$ the matrix of the $M-10$ eigenvectors corresponding to the $M-10$ smaller eigenvalues.
- Define the vector $\boldsymbol{e}^{M\times 1}(\omega) = \begin{bmatrix}1 & e^{-j\omega} & \ldots & e^{-j(M-1)\omega}\end{bmatrix}^T$ as a function of the variable ω.
- Find the 10 values of ω minimizing the equation

$$\boldsymbol{e}^{M\times 1}(\omega)^H \widehat{\boldsymbol{G}}^{M\times(M-10)}\widehat{\boldsymbol{G}}^{M\times(M-10)^H}\boldsymbol{e}^{M\times 1}(\omega)\,.$$

- Given that $\omega_k = 2\pi\frac{\tau_k}{20}$, compute the corresponding values of τ_k, $k = 1, \ldots, 10$.
- As for the annihilating filter, now that we have the values of τ_k, $k = 1, \ldots, 10$, by exploiting the expression of the Fourier transform

$$\widehat{x}[n] = \frac{1}{20}\sum_{k=1}^{10}\alpha_k e^{-j2\pi n\frac{\tau_k}{20}}\,,$$

  we can write the linear system

$$\begin{bmatrix} 1 & \cdots & 1 \\ e^{-j2\pi\frac{\tau_1}{20}} & \cdots & e^{-j2\pi\frac{\tau_{10}}{20}} \\ \vdots & & \vdots \\ e^{-j2\pi 9\frac{\tau_1}{20}} & \cdots & e^{-j2\pi 9\frac{\tau_{10}}{20}} \end{bmatrix}\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{10}\end{bmatrix} = 20\begin{bmatrix}\hat{x}[0] \\ \hat{x}[1] \\ \vdots \\ \hat{x}[9]\end{bmatrix}$$

  which solution provides the amplitudes of the spikes α_k, $k = 1, \ldots, 10$.

Part B: The signal y is a white noise

B.1) By the fundamental filtering formula, given that $h[k] \in \ell^1$ and that $W[n]$ is w.s.s., $X[n]$ is also w.s.s..

B.2) Using z as a delay operator, the process can be (symbolically) written

$$X[n] = \sum_{k=1}^{\infty} h[k]W[n-k] = \sum_{k=1}^{\infty} h[k]W[n]z^{-k} = W[n]\sum_{k=1}^{\infty} h[k]z^{-k} = W[n]H(z)\,.$$

We assume the infinite order polynomial $H(z)$ to be generated by a fractional finite polynomial, *i.e.*,

$$H(z) = \frac{1}{P(z)}\,,$$

where $P(z)$ is indeed a finite order polynomial. Notice that, given the assumptions on $H(z)$, $P(z)$ is a stable strictly minimum phase polynomial.

Then, by definition, $X[n]$ is an autoregressive process

$$X[n]P(z) = W[n]\,.$$

B.3) Given the assumption

$$H(z) = \frac{1}{P(z)}\,,$$

the estimation of $|H(z)|_{z=e^{j2\pi k/N}}$, $N = 20000$, is equivalent to the estimation of the spectrum of an autoregressive process. Therefore, the Yule-Walker method is the optimal parametric estimation method.

We assume the order of the autoregressive to be $M << N = 20000$. The value of M can be deduced by from the periodogram of $x[1], \ldots, x[20000]$ or by using the Levinson's algorithm (see point B.4).

The M order autoregressive process reads

$$X[n]P(z) = X[n] + \sum_{k=1}^{M} p_k X[n-k] = W[n]\,,$$

where $W[n]$ is a centered Gaussian white noise with known variance σ^2.

It is characterized by the M coefficients $p_1, \ldots, p_M$ (a total of M parameters).

- Given the samples $x[1], \ldots, x[20000]$, compute the (biased) empirical correlation

$$\widehat{R}_X[k] = \frac{1}{20000}\sum_{n=1}^{20000-k} x[n+k]x[n]^*\,, \quad k = 0, \ldots, M\,.$$

- Write the Yule-Walker equations for the estimation of the M parameters (do not forget that σ^2 is given)

$$\begin{aligned} \widehat{R}_X[0]p_1 + \ldots + \widehat{R}_X[M-1]p_M &= \widehat{R}_X[1] \\ \vdots \quad &= \quad \vdots \\ \widehat{R}_X[M-1]p_1 + \ldots + \widehat{R}_X[0]p_M &= \widehat{R}_X[M] \end{aligned}$$

– Solve the equations to obtain $\widehat{p}_1, \ldots, \widehat{p}_M$.
– Compute $|H(z)|_{z=e^{j2\pi k/N}}$ as

$$|H(z)|_{z=e^{j2\pi k/N}} = \frac{1}{|P(z)|_{z=e^{j2\pi k/N}}} = \frac{1}{1+\widehat{p}_1 e^{-j2\pi k/N} + \ldots + \widehat{p}_M e^{-j2\pi Mk/N}} , \quad k = 0, \ldots, N-1 .$$

B.4) As briefly mentioned in B.2), it is possible to use the Levinson's algorithm. The latter provides an iterative solution to the Yule Walker equations for increasing order. It is sufficient to observe the reflection coefficients of the Levinson's algorithm and stop the iterations (and therefore the increasing of the order) as soon as the reflection coefficients do not significantly change between two iterations (threshold to be fixed).

Part C: Modeling and estimation of the acoustic impedance

Given the process $S[n] = V[n] + W[n]$, call $\boldsymbol{S} = [S[1], \ldots, S[N]]$, $\boldsymbol{V} = [V[1], \ldots, V[N]]$, $\boldsymbol{s} = [s[1], \ldots, s[N]]$, $\boldsymbol{v} = [v[1], \ldots, v[N]]$, and denote with $\gamma_1, \ldots, \gamma_{10}$ the ten possible values of the acoustic impedance, *i.e.*, the ten states of the Markov chain.

C.1) The cumulative distribution of the process $S[n]$ is

$$\begin{aligned} F_{\boldsymbol{S}}(\boldsymbol{s}) &= \sum_{\boldsymbol{v}\in\mathcal{C}} \mathrm{P}\left(\boldsymbol{S} \le \boldsymbol{s}, \boldsymbol{V} = \boldsymbol{v}\right) \overset{\text{Bayes}}{=} \sum_{\boldsymbol{v}\in\mathcal{C}} \mathrm{P}\left(\boldsymbol{S} \le \boldsymbol{s} \,|\, \boldsymbol{V} = \boldsymbol{v}\right) \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right) \\ &= \sum_{\boldsymbol{v}\in\mathcal{C}} \left[\prod_{n=1}^{N} \mathrm{P}\left(S[n] \le s[n] \,|\, V[n] = v[n]\right)\right] \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right) , \quad \forall N \end{aligned}$$

where $\mathcal{C}$ represents all the possible combinations of the values of the realization $\boldsymbol{v}$ of the Markov chain, where $\boldsymbol{v} \in \{\gamma_1, \ldots, \gamma_{10}\}^N$. Notice that the above expression should not only defined by the time index n-uplet $1, \ldots, N$, $\forall N$, but also for all the shifts of such n-uplet $1+k, \ldots, N+k$, $\forall N, k$. Nevertheless, by assuming the Markov chain to be stationary, we can defined it simply $\forall N$.

The density corresponding to the cumulative distribution $\mathrm{P}\left(S[n] \le s[n] \,|\, V[n] = v[n]\right)$ is a Gaussian distribution with mean $v[n]$ and variance σ_W^2, *i.e.*, $\mathcal{G}_{v[n],\sigma_W^2}(s[n])$. Therefore the density associated to $F_{\boldsymbol{S}}(\boldsymbol{s})$ reads

$$f_{\boldsymbol{S}}(\boldsymbol{s}) = \sum_{\boldsymbol{v}\in\mathcal{C}} \left[\prod_{n=1}^{N} \mathcal{G}_{v[n],\sigma_W^2}(s[n])\right] \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right) ,$$

where, by subsequently applying Bayes' rule and exploiting the Markov property

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

with $v[n] \in \{\gamma_1, \ldots, \gamma_{10}\}$, $n = 1, \ldots, N$.

C.2) We need to adopt the de-noising approach via Mixture models.

The density model reads

$$f_{\boldsymbol{S}}(\boldsymbol{s}) = \sum_{\boldsymbol{v}\in\mathcal{C}} \left[\prod_{n=1}^{N} \mathcal{G}_{v[n],\sigma_W^2}(s[n])\right] p_{v[N-1]v[N]} \cdots p_{v[1]v[2]} \pi_{v[1]} ,$$

and, based on the measurements, we have computed $s[1], \ldots, s[6000]$.

- Parameter estimation of the model.
  The parameters of the models are

$$\boldsymbol{\theta} = \left\{ \sigma_W^2, \gamma_1, \dots, \gamma_{10}, \left\{ p_{\gamma_i \gamma_j} \right\}_{i,j=1,\dots,10}, \left\{ \pi_{\gamma_i} \right\}_{i=1,\dots,10} \right\} .$$

  Given $s[1], \dots, s[6000]$, the likelihood function reads

$$h([1], \dots, s[6000] \ ; \ \boldsymbol{\theta}) = \sum_{\boldsymbol{v} \in \mathcal{C}} \left[\prod_{n=1}^{6000} \mathcal{G}_{v[n], \sigma_W^2}(s[n]) \right] p_{v[N-1]v[N]} \cdots p_{v[1]v[2]} \pi_{v[1]} .$$

  By using an EM algorithm for Markovian mixture models, we can obtain an estimation of the parameter $\widehat{\boldsymbol{\theta}}$.
- A posteriori maximization.
  Once estimate the parameters $\widehat{\boldsymbol{\theta}}$, and given the observations $\boldsymbol{s} = [s[1], \dots, s[6000]]$ we can proceed to the estimation of the most probable realization of $\boldsymbol{v} = [v[1], \dots, v[6000]]$ as the argument maximizing the *a posteriori* function, where the *a posteriori* function is defined as

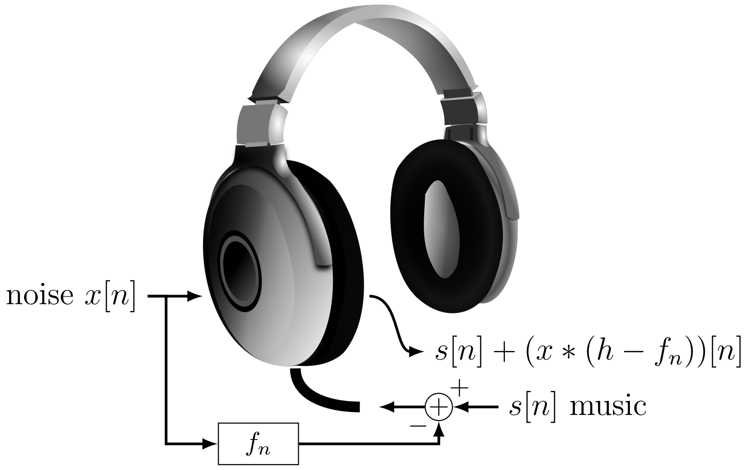
$$\begin{aligned} P(\boldsymbol{V} = \boldsymbol{v} \,|\, \boldsymbol{s}) = \frac{h(\boldsymbol{s}, \boldsymbol{v}; \widehat{\boldsymbol{\theta}})}{h(\boldsymbol{s}; \widehat{\boldsymbol{\theta}})} &= \frac{f_{\boldsymbol{S}}(\boldsymbol{s} \,|\, \boldsymbol{V} = \boldsymbol{v}) \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right)}{f_{\boldsymbol{S}}(\boldsymbol{s})} \propto f_{\boldsymbol{S}}(\boldsymbol{s} \,|\, \boldsymbol{V} = \boldsymbol{v}) \mathrm{P}\left(\boldsymbol{V} = \boldsymbol{v}\right) \\ &= \prod_{n=1}^{6000} \mathcal{G}_{v[n], \sigma_W^2}(s[n]) p_{v[N-1]v[N]} \cdots p_{v[1]v[2]} \pi_{v[1]} , \end{aligned}$$

  where $v[n] \in \{\widehat{\gamma}_1, \dots, \widehat{\gamma}_{10}\}$, $n = 1, \dots, 6000$.
  The maximization of the *a posteriori* function can be achieved using the Viterbi algorithm, with the constraints $v[n] \in \{\widehat{\gamma}_1, \dots, \widehat{\gamma}_{10}\}$, $n = 1, \dots, 6000$.

Exercise 3. Noise Cancelling Headphones Quality Control (14 points)

Your work for a company producing noise cancelling headphones. Being extremely concerned by providing a high quality product, your boss ask you to perform an analysis so to check if the headphones produced over a trimester have the same properties.

Recall that the principle of a noise cancelling headphone is



Noise Cancelling Headphones.

Every trimester the company produces 20000 headphones. To analyze the properties you play some music into the headphones (the Bach cello suites performed by Mstislav Rostropovich), obtaining the signal $s[n]$, and you add an external white Gaussian noise $x[n]$ with variance (power) σ^2. Notice that we have here the standard audio sampling frequency of 44 kHz.

For each headphone, you record the signal $e[n] = s[n]+(x*(h-f_n))[n]$, obtaining $e[1], \dots, e[220000]$ samples. For each recording you compute:

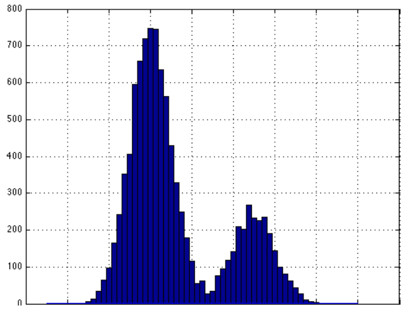
- The power of the signal $e[n]$;
- The frequency corresponding to the maximum value of the spectrum of $e[n]$ (we consider the frequencies between 0 and 22 kHz);
- The power at low frequencies ($f \in [0, 2]$ kHz) of the spectrum of ;
- The power at high frequencies ($f \in [2, 22]$ kHz).

That is, you compute 4 characteristics, that we shall call $c[1], \dots, c[4]$

In order to verify whether it is really necessary to consider 4 characteristic and to reduce the complexity of the analysis, you can use principal component analysis.

1) Describe in detail the principal component analysis method, step by step, from the **4 characteristics** $c[1], \dots, c[4]$ **computed based on the recorded signal** $e[1], \dots, e[220000]$, to the principal components $z_m[k]$, 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). Also, **clearly indicate the dimensions of the matrices and vectors**.

After analyzing the variance of the principal components, it clearly appears that 1 principal component, namely $z_m[1]$, accounts for most of the total sample variation. When making a histogram of all the 20000 values of the principal component, the plots appears as a mixture of two distributions, as depicted in the figure below.



This suggests that the produced 20000 headphones can be divided into two classes, each being described by a Gaussian distribution.

2) Provide the **explicit** expression of a two Gaussian mixture model for the principal components having as realization $z_m[1]$, $m = 1, \ldots, 20000$.

3) Write the corresponding likelihood function.

4) How can we find the two values of the principal component each characterizing one of the two classes of the produced headphones?

Solution 3.

1) Each headphone is characterize by 4 variables $c[1], \ldots, c[4]$. We shall call $\boldsymbol{c}_m = [c_m[1], \ldots, c_m[4]]$ (dimension 1×4), $m = 1, \ldots, 20000$ the variable vectors of the $M = 20000$ produced headphones.

   In order to reduce the number of variables characterizing the headphones, we apply PCA as follows.

   – Compute the mean of the variables $\boldsymbol{m}_C = \dfrac{1}{20000} \displaystyle\sum_{m=1}^{20000} \boldsymbol{c}_m$.
   – Center the variables $\bar{\boldsymbol{c}}_m = \boldsymbol{c}_m - \boldsymbol{m}_C$ (dimension 1×4).
   – Compute the empirical correlation matrix of the variables $\widehat{\boldsymbol{R}}_C = \dfrac{1}{20000} \displaystyle\sum_{m=1}^{20000} \bar{\boldsymbol{c}}_m^T * \bar{\boldsymbol{c}}_m$ (dimension 4×4);
   – Diagonalize the correlation Matrix $\boldsymbol{V}^t \widehat{\boldsymbol{R}}_C \boldsymbol{V} = \boldsymbol{\Lambda}$ (dimensions 4×4), where $\boldsymbol{V}$ is the matrix of eigenvectors (dimensions 4×4);
   – Compute the principal components $\boldsymbol{z}_m = \bar{\boldsymbol{c}}_m \boldsymbol{V}$ (dimensions 1×4), $m = 1, \ldots, 20000$.

2) We have now reduced the complexity from 4 to 1 variable, in the principal components space. Notice that the reduction of variable is performed in the principal component space and NOT in the original space! At this point, is not interesting to go back to the original space, since the variables are a linear combination of a (reduced) set of principal components.

   The histogram shows that the principal component $z_m[1]$, $m = 1, \ldots, 20000$ can be modeled as the outcome of an i.i.d. mixture of two Gaussian distributions. Call $\boldsymbol{Z} = [Z_1[1], \ldots, Z_{20000}[1]]$ the sequence of random variables (stochastic process) associated to the outcomes $\boldsymbol{z} = [z_1[1], \ldots, z_{20000}[1]]$. The density (model) is then

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{m=1}^{20000} \left(\pi_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(z_m[1] - \mu_1)^2}{2\sigma_1^2} \right) + \pi_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(z_m[1] - \mu_2)^2}{2\sigma_2^2} \right) \right)$$

3) The parameters of the density are $\boldsymbol{\theta} = \{\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$, and the corresponding likelihood function reads

$$h(\boldsymbol{z}\,;\,\boldsymbol{\theta}) = \prod_{m=1}^{20000} \left(\pi_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(z_m[1] - \mu_1)^2}{2\sigma_1^2} \right) + \pi_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(z_m[1] - \mu_2)^2}{2\sigma_2^2} \right) \right)$$

4) The two values of the principal component correspond to the two means μ_1 and μ_2 of the mixture model. They can be estimated by estimating the parameters $\boldsymbol{\theta}$ via the maximum likelihood approach.

Grade Scale.

The exams accounts for a total of 72 points (exact response to each question).

The grading has been done on a 61 points scale (61 points = 6/6), according to the following formula

$$\text{grade over 6} = 1 + (5 * \text{points}/62)$$

and then rounded to .25 steps, that is

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

The result is then constraint to be at maximum 6.