Statistical Signal & Data Processing - COM500
Final Exam

Thursday August 20, 2020, 8h15-11h45.

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.

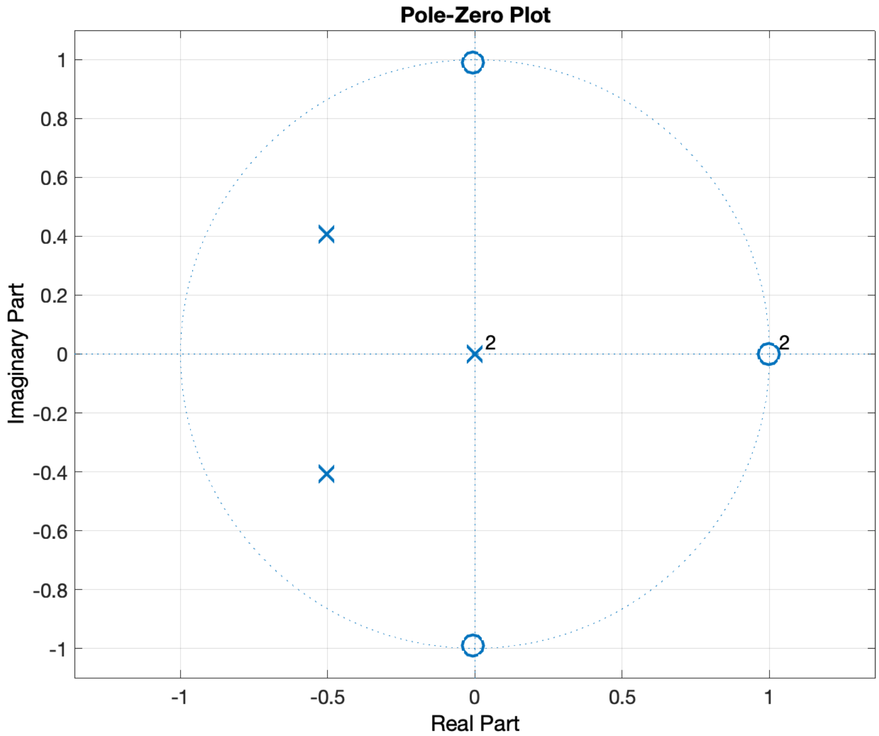
All the best for your exam!!

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. A SIMPLE SYSTEM (6 PTS)

Consider the z-plane plot below, where the two poles (not in the origin) have magnitude 0.65 and phase $\pm 3/4\pi$.



1) Write the z transform $H(z)$ corresponding to the z-plane plot.

2) Sketch the magnitude of transfer function $H(e^{j\omega})$.

3) Does it represent a FIR or a IIR system? (justify).

4) Consider the filtering interpretation of an AR process, can $H(z)$ represent the analysis filter? Can $H(z)$ represent the synthesis filter? Justify

Main Problems

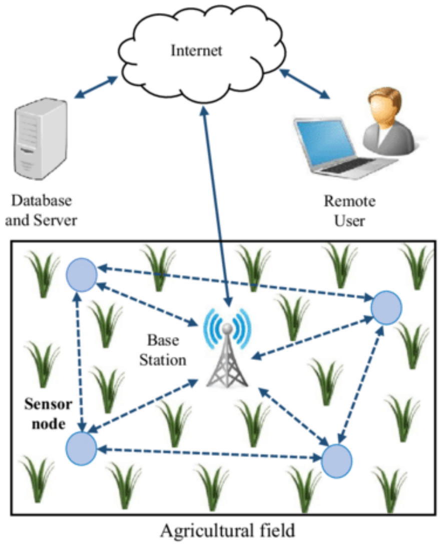
Here comes the core part of the exam .. take time to read the introduction and each problem statement. Please provide justified, rigorous, and simple answers. Remember that you are not simply asked to describe statistical signal processing tools, but rather to describe how to apply such tools to the given problem. If needed, you can add assumptions to the problem setup.

Exercise 2. SENSOR NETWORKS FOR VITICULTURE

In viticulture the knowledge of several environmental parameters and of the soil condition is of foremost importance. This is even more crucial when the viticulture is conducted using an organic or a biodynamic approach.

These parameters are monitored using dedicated sensors. It has to be noticed that such sensors are placed in the field, that is, far away from power outlets and internet connections. Therefore sensor are battery / solar panel powered and a low power transmission technique needs to be used in order to reach a powered base station (or sink node) and, from there, the internet.

A Swiss vineyard.



Typical sensor network scenario.

Two common long range low power transmission solutions are the LoRA based and the SigFox network.

LoRA is a **spread spectrum technique**, *i.e.*, the spectral energy is spread over a large frequency band. Although it uses a proprietary chirp modulation technique, the deployment of LoRA networks is open and regulated by a non profit association, the LoRA Alliance.

SigFox network is based on an **ultra narrow band transmission**, *i.e.*, the spectral energy is concentrate around specific frequencies. It is a proprietary modulation technique and a proprietary network, *i.e.*, for a SigFox based transmission one must subscribe to the SigFox network and pay a monthly fee.

In the following we will not necessarily refer to LoRA or SigFox technologies but rather to a generic spread spectrum transmission and to a generic ultra narrow band transmission.

We start by considering the monitoring of air and soil temperature, where the temperature samples are taken every 30 minutes. For simplicity, we assume that the relevant temperature values have increments of 0.5°C. For instance, we will measure temperatures like 24°C, 24.5°C, 25°C, and so on. Assuming the temperature range limited between few negative degrees and values around 40°C, the temperature signal $x[n]$ can be considered to be a **discrete value signal**.

More precisely, we shall denote $x_a[n]$ the measurement of the air temperature, and $x_s[n]$ the measurement of the soil temperature.

The following four parts A, B, C, and D, are independent.

Part A: Temperature Monitoring & Ultra Narrow Band Transmission (16 points)

We consider here a narrow band transmission and assume that the the different values of temperature are coded into sinusoids of different frequencies. More precisely, for a given n, $x_a[n]$ and $x_s[n]$ are coded into sinusoidal signals, $y_a[1], \ldots, y_a[M]$, and $y_s[1], \ldots, y_s[M]$, respectively. The frequency of each sinusoidal signal is related to the value of the temperature.

The measurement of the air temperature $x_a[n]$ and the measurement of the soil temperature $x_s[n]$ are transmitted simultaneously. Consequently, the received signal is $y[m] = y_a[m] + y_s[m]$, $m = 1, \ldots, M$, *i.e.*, the sum of two real sinusoids.

A.1) Provide a w.s.s. stochastic model $Y[m]$ for the received signal $y[m]$, and precisely prove that it is indeed a w.s.s. process. Please provide justified, rigorous, and simple proof.

We would like now to develop a method to estimate the frequencies of the two sinusoids, and therefore the two temperatures.

A.2) Assuming the absence of noise, propose a parametric method to estimate the frequencies of the two sinusoids. Precisely describe such method. You are given $M = 100000$ samples. 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 a real scenario we need to consider that the transmission is affected by noise (just think of the case of a storm and the electromagnetic interferences it causes). The receive signal therefore becomes $Y[m] = Y_a[m] + Y_s[m] + W[m]$, $m = 1, \ldots, M$, where $W[m]$ is a centred Gaussian white noise with variance σ_W^2.

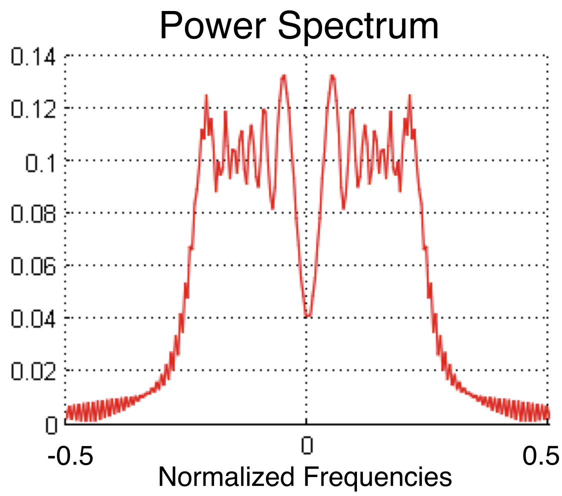
A.3) Accounting for the presence of noise, propose a parametric method to estimate the frequencies of the two sinusoids. Precisely describe such method. You are given $M = 100000$ samples. You are asked to detail each step as if you have to implement the method in a computer. Precisely indicate the input and output of each step.

Part B: Temperature Monitoring & Spread Spectrum Transmission (16 points)

We now turn our attention to the transmission of the temperature using a spread spectrum technology with chirp modulation.

The measurement of the air temperature $x_a[n]$ and the measurement of the soil temperature $x_s[n]$ are transmitted using different chirp patterns, obtaining a transmitted signal $y[m]$, $m = 1, \ldots, M$ (most probably affected by noise).

A typical power spectrum (computed using a periodogram) of a chirp modulated transmitted signal is depicted in the figure below.



We can consider that the "shaky behaviour" of the spectrum is due to the variance of the periodogram and the presence of a white noise $w[m]$, and not to the signal $y[m]$ itself.

B.1) Based on the spectrum above, provide a parametric w.s.s. stochastic model $Y[m]$ for the received signal $y[m]$, justify precisely your choice. Precisely prove that it is indeed a w.s.s. process. Please provide justified, rigorous, and simple proof.

B.2) Based on the spectrum above suggest the required number of parameters of the w.s.s. model $Y[m]$.

B.3) Propose an algorithm to estimate the number of parameters of the w.s.s. model $Y[m]$. Precisely describe how to implement the algorithm and how to estimate the number of parameters given $M = 100000$ samples of $y[m]$.

Part C: Multi-Parameter Monitoring & Analysis (16 points)

After a successful testing of the monitoring of the air and soil temperatures, the winemaker got really excited by sensor network technology and decided to monitor 8 different parameters (air and soil temperature, wind, air and soil humidity, solar radiation, dew point, rainfall).

We shall denote the corresponding measurements with $x_i[n]$, $i = 1, \ldots, 8$, $n = 1, \ldots, N$.

The 8 parameters are monitored over several days, obtaining a total of $N = 17500$ samples for each parameter.

You would like to analyze such variables in order to determine whether there are specific recurring correlated patterns of values.

You decide to apply the principal component analysis method to first reduce the dimension of the parameter space.

C.1) Describe in detail, step by step, how to compute the 8 principal components given the parameter samples $x_i[n]$, $i = 1, \ldots, 8$, $n = 1, \ldots, 17500$. We shall denote the 8 principal components as $c_1[n], \ldots, c_8[n]$, $n = 1, \ldots, 17500$.

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**.

After analyzing the variance of the principal components, it clearly appears that 3 principal components, namely $c_1[n]$, $c_2[n]$, $c_3[n]$, $n = 1, \ldots, 17500$ account for most of the total sample variation.

By looking at the 3D plot of the 3 principal components we can isolate 5 different clusters.

C.2) Provide a Gaussian Mixture model describing the 5 clusters based on the principal components, namely $c_1[n]$, $c_2[n]$, $c_3[n]$, $n = 1, \ldots, 17500$ values of the 3.

C.3) Write the corresponding likelihood function clearly indicating which are its parameters.

Part D: Denoising the Received Temperature Measurements (16 points)

The measurement of the air temperature $x_a[n]$ are sent using a long range transmission system to a receiver located in the main building of the wine growing estate. Due to the channel noise, the received values $x_a^r[n]$ are themselves affected by noise. We can write the corresponding stochastic processes as

$$X_a^r[n] = X_a[n] + W[n]\,.$$

As initially mentioned, $x_a[n]$ can be assumed to be a discrete value signal. We shall consider a maximum of 80 possible values. In addition, given the measurement interval of 30 minutes, we can consider that the temperature value at the instant n is dependent on the temperature value at the instant $n-1$.

You are asked to propose a denoising method that optimally exploit the discrete value nature of the original signal, and the one-step dependency of the temperature values.

D.1) Provide a stochastic model for $X_a^r[n]$ expressing the one-step dependency of $X_a[n]$.

D.2) Compute the cumulative distribution function of $X_a^r[n]$.

D.3) Assuming you have received $x_a^r[1], \ldots, x_a^r[10000]$ values, describe in detail a denoising approach enabling to estimate the corresponding $x_a[1], \ldots, x_a[10000]$ values.