Statistical Signal Processing

Final Exam

Monday, 15 June 2015

You will hand in this sheet together with your solutions.

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

Seat Number:

Family Name:

Name:

Read Me First!

Only a personal cheatsheet of 4 pages (2 A4 sheets) is allowed.
No class notes, no exercise text or exercise solutions, no laptop, no communication devices.

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

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

Return the text of the exam.

Warmup exercises

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

Please provide justified, rigorous, and simple answers.

Exercise 1. Stochastic Process (4 pts)
Consider the process $X[n]$ given by the following expression

$$X[n] = X[n-2] - 0.2X[n-3] + W[n]\,,$$

where $W[n]$ is a centered white noise with variance $\sigma_W^2 = 1$.

Please provide justified, rigorous, and simple answers.

1) Is the process $X[n]$ w.s.s.?
Hint: Be careful to rewrite the process in a form that enables you to directly prove that is w.s.s.; Do not proceed to a direct computation of the mean and the correlation!

2) Give the expression for the optimal one-step predictor (in the MSE sense) of the process $X[n]$.

3) Give the expression for the optimal two-step predictor (in the MSE sense) of the process $X[n]$.

Solution 1.

1) $X[n] - X[n-2] + 0.2X[n-3] = W[n]$. Let $A(z) = 1 - z^{-2} + .2z^{-3}$, then $X[n]A(z) = W[n]$ The roots of the polynomial (that can be found using a scientific calculator) are $z_1 = 0.20915$, $z_2 = 0.87889$, and $z_3 = -1.08803$. The filter $1/A(z)$ is therefore not stable. It is not possible to apply the fundamental filtering formula and to state that the w.s.s. of $X[n]$ is a consequence of the w.s.s. of $W[n]$.

   If it was not possible to compute the roots, then one can make either the assumption that $A(z)$ is not stable (then the solution is as stated above), or the assumption that $A(z)$ is stable. In the latter case, it is then possible to apply the fundamental filtering formula and state the w.s.s. of $X[n]$ as a consequence of the w.s.s. of $W[n]$.

2) Since the filter is not stable, it is not possible to conclude that $X[n]$ is w.s.s. and therefore it is not possible to use the projection theorem. If the filter was assumed stable, the one step predictor reads
$$\widehat{X}[n] = X[n-2] - 0.2X[n-3]\,.$$

3) Since the filter is not stable, it is not possible to conclude that $X[n]$ is w.s.s. and therefore it is not possible to use the projection theorem. If the filter was assumed stable, the one step predictor reads
$$\widehat{X}[n] = X[n-2] - 0.2X[n-3]\,.$$

Exercise 2. How many parameters? (3 pts)

Let $X[n]$, be a stochastic process taking 10 possible values, and such that

- $\mathrm{P}(X[n] = i_0 | X[n-1] = i_1, \ldots, X[n-k] = i_k) = \mathrm{P}(X[n] = i_0 | X[n-1] = i_1)$, for every n, k
- $\mathrm{P}(X[n] = i) = \mathrm{P}(X[n+l] = i)$ for every l

We would like to describe the process $X[n]$ for $n = 1, \ldots, 1000$, that is, we would like to characterize the law (probability) of $X[1], \ldots, X[1000]$. How many parameters are necessary to characterize $X[1], \ldots, X[1000]$?

Solution 2.

$X[n]$ is a Markov chain, characterised by its initial and transition probabilities.

Now, we have 10 possible values (10 states), therefore

- 10 initial probabilities π_i, $i = 1, \ldots, 10$;
- 10×10 transition probabilities p_{ij}, $i = 1, \ldots, 10$, $j = 1, \ldots, 10$.

and

- 1 probability constraint for the initial probabilities $\sum_{i=1}^{10} \pi_i = 1$;
- 10 probability constraints for the transition probabilities $\sum_{j=1}^{10} p_{ij} = 1$, $i = 1, \ldots, 10$.

Finally, the law of the process is characterized by $10 - 1 + 10 \times 10 - 10 = 99$ parameters.

Bonus

If we also take into account that the Markov chain is stationary (*i.e.*, $\mathrm{P}(X[n] = i) = \mathrm{P}(X[n+l] = i)$ for every l), the initial distribution satisfy

$$\pi_j = \sum_{i=1}^{10} \pi_i p_{ij} \, .$$

In such a case the law of the Markov chain is completely characterized by its transition probability, that is, by 90 parameters.

Main exercises

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

Please provide justified, rigorous, and simple answers.

Exercise 3. STRIDE ANALYSIS (35PTS)
In order to improve their performances, runners are submitted to the analysis of their stride. By placing an accelerometer on a running shoe one obtains measurements of the stride.

A: Simplified Model for Stride Frequency Analysis
We consider here a simplified model of the stride, where the measured accelerometer signal is modeled as a harmonic **real** signal $X[n]$, composed of

- A main sinusoidal signal, which frequency f_1 corresponds to the main stride frequency (typical 150-200 strides, *e.g.* periods, per minute)
- Two secondary sinusoidal signals, mostly due to the form of the stride and the foot mouvement. These components have frequencies f_2 and f_3 that typically differ from the main stride frequency f_1 by $\pm f_{\Delta 1}$, $f_{\Delta 1} \in [.3, .4]Hz$ and $\pm f_{\Delta 2}$, $f_{\Delta 2} \in [.6, .8]Hz$, respectively.

The accelerometer signal is sampled at $f_s = 200Hz$.

1) Assuming the accelerometer signal $X[n]$ to be a w.s.s., write the mathematical expression for the simplified model

As a first approach we use the Periodogram $P_X(\omega)$.

2) How many samples are necessary to be able to distinguish the three sinusoidal components of the signal?

Weakness of the Periodogram are well known. We therefore use a parametric method.

3) Propose a parametric spectral estimation method than enables to estimate the frequencies f_1, f_2, and f_3 of the three components of the signal, as well as the corresponding amplitudes a_1, a_2, and a_3.
4) Given that we have measured $x[1], \ldots, x[100000]$, describe in detail the method, step by step, from the measured samples to the estimated spectrum, like if each step has to be interpreted and executed by a computer (in particular the input, the executed operation with corresponding equations, and the output of each step has to be clear)

B: More Realistic Model for Stride Characteristics Analysis

The waveform corresponding to a stride is actually more complex that just the sum of three sinusoids. While for frequency analysis the simplified three sinusoids model is sufficient, when we tackle the stride characteristics analysis we need to consider the realistic stride waveform. We still denote the realistic signal with $X[n]$.

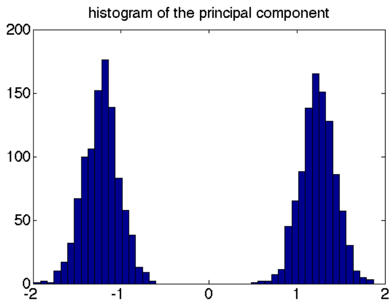
For stride characteristic analysis we collect 20000 stride waveforms measured by different runners. Each waveform is characterized by

- Mean, Median, Mode (time instant where the shape has its maximum value), Variance, Amplitude.

In order to reduce the complexity of the analysis and to verify whether it is really necessary to consider all these characteristics, we use principal component analysis.

5) Show in detail how to apply the principal component analysis to this problem. Check if the necessary conditions to apply the proposed method are satisfied or eventually make appropriate assumptions. Write the equations describing the method, **clearly indicating the dimensions of the matrices**. Here too, precisely describe the steps necessary to implement the method (if we plug in such steps into a computational software we must then obtain the desired result .. so don't miss any step! In particular the input, the executed operation with corresponding equations, and the output of each step has to be clear).

After analyzing the variance of the principal components, it clearly appears that 1 principal component accounts for most of the total sample variation. When plotting all the 20000 values of the principal component, the plots appears as a collection of 2 clusters, as depicted in the figure below.



6) Explain how to use mixture models to **automatically** identify the clusters and in particular their central point and their variance. In particular

   6.1) Set up precisely the mathematical model expressing the probability density function of the principal component

   6.2) Describe the method in details. Here too, precisely describe the steps necessary to implement the method (if we plug in such steps into a computational software we must then obtain the desired result .. so don't miss any step! In particular the input, the executed operation with corresponding equations, and the output of each step has to be clear).

Solution 3.

1) $X[n]$ is a harmonic signal

$$\begin{aligned}X[n] =& \frac{a_1}{2}\left(e^{i(2\pi\frac{f_1}{f_s}n+\Theta_1)} + e^{-i(2\pi\frac{f_1}{f_s}n+\Theta_1)}\right)\\ &+\frac{a_2}{2}\left(e^{i(2\pi\frac{f_2}{f_s}n+\Theta_2)} + e^{-i(2\pi\frac{f_2}{f_s}n+\Theta_2)}\right)\\ &+\frac{a_3}{2}\left(e^{i(2\pi\frac{f_3}{f_s}n+\Theta_3)} + e^{-i(2\pi\frac{f_3}{f_s}n+\Theta_3)}\right)\end{aligned}$$

where Θ_i, $i = 1, 2, 3$ are i.i.d. random variables uniformly distributed over $[0, 2\pi]$.

Notice that:

- Θ_i, $i = 1, 2, 3$, play a key role since without them the process is not w.s.s.;
- The signal is specified to be real, therefore we have 6 complex exponentials (6 spectral lines), pairwise complex conjugated;
- The argument is in normalized frequencies, that is $2\pi\frac{f}{f_s}n$.

2) Given that $f_1 \in [2.5, 3.3]Hz$ The worst case, that is the situation where two spectral lines are the closest, is when $f_2 = f_1 + 0.4$ and $f_3 = f_1 + 0.6$, or, similarly, $f_2 = f_1 - 0.4$ and $f_2 = f_1 - 0.6$. In such a case, $|\Delta f| = 0.2$.

Consequently,

$$N > \frac{f_s}{|\Delta f|} = \frac{200}{0.2} = 1000\,.$$

3) We have:

- A harmonic signal, which spectrum is therefore a line spectrum;
- no noise.

Consequently, among the parametric methods we have seen in class, the most appropriate one is the Annihilating Filter Line Spectrum Estimation Method.

4) The available measured data is $x[1], \ldots, x[100000]$ and the spectrum we would like to estimate is composed of 6 spectral lines (pairwise symmetric).

The steps necessary to estimate such spectrum from $x[1], \ldots, x[100000]$ are:

1. Write the annihilating filter equation $(x * h)[n] = 0$ in linear form. Notice that
   - ∗ The order of the the annihilating filter is 6, *i.e.*, $h[0], h[1], \ldots, h[6]$, where $h[0] = 1$;
   - ∗ The first sample of the available measured data is $x[1]$and NOT $x[0]$.

   Then

$$\begin{bmatrix} x[6] & \ldots & x[1] \\ x[7] & \ldots & x[2] \\ \vdots & & \vdots \\ x[11] & \ldots & x[6] \end{bmatrix} \begin{bmatrix} h[1] \\ h[2] \\ \vdots \\ h[6] \end{bmatrix} = - \begin{bmatrix} x[7] \\ x[8] \\ \vdots \\ x[12] \end{bmatrix}$$

1. Solve the linear system (the matrix is Toeplitz, therefore its solution requires 6^2 computations), obtaining $h[1], \ldots, h[6]$.
2. Compute the z transform $H[z] = 1 + h[1]z^{-1} + \ldots + h[6]z^{-6}$.
3. Find the zeroes of $H[z]$, *i.e.*, the solutions (roots) of the equation $1 + h[1]z^{-1} + \ldots + h[6]z^{-6} = 0$. Call $a_1, \ldots, a_6$ the zeroes, giving $H(z) = (1 - a_1 z^{-1}) \ldots (1 - a_6 z^{-1})$.
4. Through the relations $a_1 = e^{i2\pi\widetilde{f}_1}, \ldots, a_5 = e^{i2\pi\widetilde{f}_5}$ we can compute the normalized frequencies $\widetilde{f}_1, \ldots, \widetilde{f}_6$: The frequency is given by the argument of the complex exponential divided by 2π, *i.e.*,
$$\widetilde{f}_n = \arg(a_n)/i2\pi$$
To account for numerical errors providing non unitary roots, we can use $\widetilde{f}_n = \arg(a_n/|a_n|)/i2\pi$. Notice that the 6 frequencies $\widetilde{f}_1, \ldots, \widetilde{f}_6$ are pairwise symmetric, and they corresponds to the model
$$\begin{aligned} X[n] &= \sum_{k=1}^{6} \alpha_1 e^{j(2\pi \widetilde{f}_k n + \widetilde{\Theta}_k)} \\ &= \frac{a_1}{2}\left(e^{i(2\pi \frac{f_1}{f_s} n + \Theta_1)} + e^{-i(2\pi \frac{f_1}{f_s} n + \Theta_1)}\right) \\ &+ \frac{a_2}{2}\left(e^{i(2\pi \frac{f_2}{f_s} n + \Theta_2)} + e^{-i(2\pi \frac{f_2}{f_s} n + \Theta_2)}\right) \\ &+ \frac{a_3}{2}\left(e^{i(2\pi \frac{f_3}{f_s} n + \Theta_3)} + e^{-i(2\pi \frac{f_3}{f_s} n + \Theta_3)}\right) \end{aligned}$$
From $\widetilde{f}_1, \ldots, \widetilde{f}_6$ we can identify $f_1, \ldots, f_3$, where $f = \widetilde{f} f_s$.
5. Once estimated the frequencies, we can compute the amplitudes of the line spectrum via the linear system
$$\begin{bmatrix} e^{j2\pi\widetilde{f}_1} & \ldots & e^{j2\pi\widetilde{f}_6} \\ e^{j2\pi\widetilde{f}_1 2} & \ldots & e^{j2\pi\widetilde{f}_6 2} \\ \vdots & & \vdots \\ e^{j2\pi\widetilde{f}_1 6} & \ldots & e^{j2\pi\widetilde{f}_6 6} \end{bmatrix} \begin{bmatrix} \alpha_1 e^{j\widetilde{\Theta}_1} \\ \alpha_2 e^{j\widetilde{\Theta}_2} \\ \vdots \\ \alpha_6 e^{j\widetilde{\Theta}_3} \end{bmatrix} = \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[6] \end{bmatrix}$$
obtaining $\alpha_k e^{j\widetilde{\Theta}_k}$, $k = 1, \ldots, 6$, and more usefully $|\alpha_k| = |\alpha_k e^{j\widetilde{\Theta}_k}|$. The $|\alpha_k|$ are pairwise equal, corresponding to $|a_l|$, $l = 1, \ldots, 3$. The amplitudes of the 6 spectra lines are pairwise equal and they corresponds to $\frac{|a_l|^2}{4}$, $l = 1, \ldots, 3$.
Notice that
   * Here again, the first available sample is $x[1]$ and NOT $x[0]$;
   * The system provides $\alpha_k e^{j\widetilde{\Theta}_k}$ and not directly a_l;
   * The amplitudes we want to obtain are the amplitudes of the spectral lines, that is $\frac{|a_l|^2}{4}$.

5) From the measurements of the signal $X[n]$ we have obtained $M = 20000$ stride waveforms. For each waveform we compute $N = 5$ variables Mean, Median, Mode (time instant where the shape has its maximum value), Variance, Amplitude. We shall denote the such variables as
$$\boldsymbol{c}_m = [c_m[1], \ldots, c_m[5]]^T, \quad m = 1, \ldots, 20000, \quad N \times 1,$$

where m is the index of the waveform, and define

$$\boldsymbol{C} = \left[\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\right], \quad N \times M.$$

From the wide sense stationarity of $X[n]$ we assume that the variables $c_m[k]$ are, with respect to the index m, jointly w.s.s.. Notice that wide sense stationarity of the variables is a necessary condition in order to apply the PCA.

The we proceed as follows

1. Center the w.s.s. process associated to each variable, that is

$$\boldsymbol{\mu}_c = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{c}_m \quad N \times 1,$$

   and set $\boldsymbol{c}_m = \boldsymbol{c}_m - \boldsymbol{\mu}_c$, $m = 1, \ldots, M$.
2. Compute the empirical correlation matrix

$$\widehat{\boldsymbol{R}}_{\boldsymbol{c}} = \frac{1}{M}\boldsymbol{C}\boldsymbol{C}^H = \frac{1}{M}\sum_{m=1}^{M} \boldsymbol{c}_m \boldsymbol{c}_m^H, \quad N \times N.$$

3. Compute the eigenvalues and eigenvectors of the empirical correlation matrix, that is solve the equation

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

   where $\boldsymbol{V}$ is the unitary matrix of eigenvectors and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ is the diagonal matrix of eigenvectors. Then $\boldsymbol{V}^H\widehat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{V} = \boldsymbol{\Lambda}$.
4. Analyze the eigenvalues to select those that accounts for most of the energy of the correlation matrix (first K eigenvalues) and call $\boldsymbol{V}_K$ the matrix of the corresponding eigenvectors.
5. Compute the principal components

$$\boldsymbol{z}_m = \boldsymbol{V}^T\boldsymbol{c}_m, \quad \boldsymbol{Z} = \boldsymbol{V}^T\boldsymbol{C}, \quad N \times M.$$

   and in particular those corresponding to the most significant eigenvalues of the correlation matrix.

$$\boldsymbol{z}_{Km} = \boldsymbol{V}_K^T\boldsymbol{c}_m, \quad K \times 1, \quad \boldsymbol{Z}_K = \boldsymbol{V}_K^T\boldsymbol{C}, \quad K \times M.$$

6. The analysis of these K principal components enables to characterize the M stride waveforms with a reduced set of variables.

6) Our goal here is to automatically identify the central point of each cluster.

6.1) The histogram represents an empirical estimator of the probability density function of the principal component accounting for most of the total sample variation. The histogram shows that the probability density function can be modeled as the mixture of two gaussian distributions, that is

$$f_Z(z) = \pi_1\mathcal{G}_{\mu_1,\sigma_1}(z) + \pi_2\mathcal{G}_{\mu_2,\sigma_2}(z),$$

where

$$\mathcal{G}_{\mu,\sigma}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) ,$$

and π_1, π_2, such that $\pi_1 + \pi_2 = 1$, are the mixture proportions.
When considering all the values of the principal components, that is $\boldsymbol{z} = [z[1], \ldots, z[M]]$, by assuming that these are independently and identically sampled, we have

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{m=1}^{M} \left(\pi_1 \mathcal{G}_{\mu_1,\sigma_1}(z[m]) + \pi_2 \mathcal{G}_{\mu_2,\sigma_2}(z[m])\right) .$$

6.2) 1. The parameters of the mixture model are $\theta = \{\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2\}$, where μ_1 and μ_2 correspond to the center of the two clusters. Consequently, the identification of the central point of each cluster is obtained by estimating the parameters of the mixture model based on the data $z[1], \ldots, z[M]$.

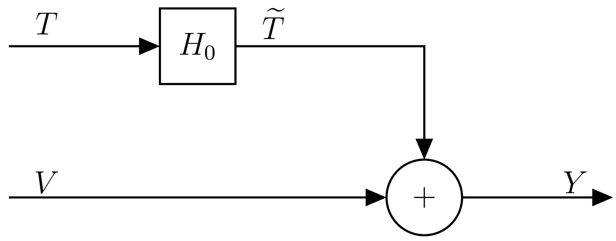
2. Mixture model parameter estimation is achieved by maximization of the likelihood function $h(\theta; \boldsymbol{z})$ which in our case reads

$$h(\theta; \boldsymbol{z}) = f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{m=1}^{M} \left(\pi_1 \mathcal{G}_{\mu_1,\sigma_1}(z[m]) + \pi_2 \mathcal{G}_{\mu_2,\sigma_2}(z[m])\right) .$$

3. Due to its form (product of a sum) the maximization of the likelihood does not admit a closed form and it is therefore necessary to perform it by means of an iterative algorithm. A well suited algorithm for the likelihood of Gaussian mixture models is the Expectation-Maximization algorithm.
   - Specify the order of the mixture (therefore its parameters)
   - Specify the estimation error (difference of the parameter values between two iterations);
   - Plug the data in;
   - Run the algorithm.

Exercise 4. THE TRAVELLING SALES REPRESENTATIVE (20 PTS)

Mr. Jones is a travelling sales representative traveling almost every day. While waiting for train connection at stations, he calls his son Bill to tell him a fairytale. The voice V of Mr Jones is disturbed with the background noise T of the trains arriving and leaving the station. Such a noise is filtered by the acoustic channel $H_0 = a_0 + a_1 z^{-1} + a_2 z^{-2} + ... + a_N z^{-N}$ (where H_0 is a stable filter) . The length of the filter is known, but the coefficients a_i, are unknown.



We assume that Mr. Jones' voice V, and the background train noise T, are wide sense stationary and independent. Also, T can be modeled by a centered Gaussian white noise with variance σ_T^2.

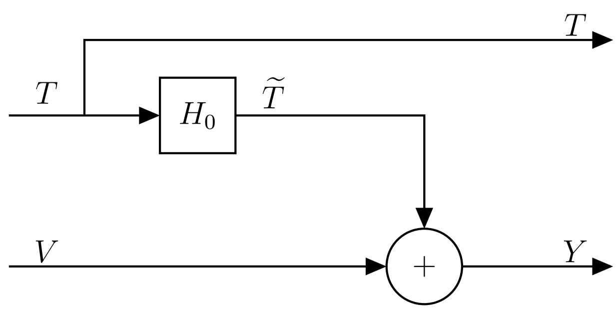
1) Compute the autocorrelation of $\widetilde{T}$.

2) Describe a method that will remove in a MSE sense the noise $\widetilde{T}$ from Y, step by step (if we plug in such steps into a computational software we must then obtain the desired result .. so don't miss any step! In particular the input, the executed operation with corresponding equations, and the output of each step has to be clear).

3) After implementing the previous method, you realize that T and V are stationary only for durations of 100 ms. How would you implement an optimal filter in this case?

4) The implementation of the previous question is too slow for real-time processing. Explain why this is the case.

5) Assume not know anything about the statistics of the train's noise T, but that we are able to measure it with a microphone. Describe how to use an adaptive filter to remove the noise $\widetilde{T}$ from the microphone signal Y. Draw where you would place the filter, what should the filter length be, and what signals you would use to determine the filter coefficients.



6) Now assume that the LMS algorithm is to be used for computing the adaptive filter.

   Write the expression for updating the filter coefficients. What is the possible range for the step-size, so that the algorithm converges?

Solution 4.

1) We have

$$\widetilde{T}[n] = \sum_{i=0}^{N} a_i T[n-i]\,.$$

T is wide sense stationary and H_0 is stable, then, by the fundamental filtering formula, $\widetilde{T}$ is also wide sense stationary. Its correlation reads

$$\begin{aligned}
R_{\widetilde{T}}[k] &= E[\widetilde{T}[n]\widetilde{T}^*[n-k]] \\
&= E\left[\sum_{i=0}^{N} a_i T[n-i] \sum_{j=0}^{N} a_j^* T^*[n-j-k]\right] \\
&= \sum_{i=0}^{N}\sum_{j=0}^{N} a_i a_j^* E\left[T[n-i]T^*[n-j-k]\right] \\
&= \sum_{i=0}^{N}\sum_{j=0}^{N} a_i a_j^* \delta_{i-j-k}\sigma_T^2 \\
&= \sum_{i=k}^{N} a_i a_{i-k}^* \sigma_T^2
\end{aligned}$$

The last equality follows from the fact that δ_{i-j-k} is non-zero for $j = i-k$, and $j \geq 0$, therefore $i \geq k$.

2)
- The signals are wide sense stationary, therefore, we are in the framework of Wiener filtering. The available signal is Y and the signal we want to estimate is V, where $Y[n] = V[n] + \widetilde{T}[n]$.
- The Wiener filter reads

  $$H_1(\omega) = \frac{S_{VY}(\omega)}{S_Y(\omega)}\,.$$

  where, given the independence of V and T, $S_{VY}(\omega) = S_V(\omega)$. Notice that we do not have access to V and therefore to S_V. Consequently, using $Y[n] = V[n] + \widetilde{T}[n]$, we express $S_V(\omega) = S_Y(\omega) - S_{\widetilde{T}}(\omega)$. Finally

  $$H_1(\omega) = \frac{S_Y(\omega) - S_{\widetilde{T}}(\omega)}{S_Y(\omega)}\,.$$

- $S_Y(\omega)$ is obtained directly form the samples of $Y[n]$ using a periodogram or another appropriate spectral estimator, while $S_{\widetilde{T}}(\omega)$ is approximated as the estimate of the power of the uncorrelated component of Y (whitening filter).

Alternatively, we can determine the filter as

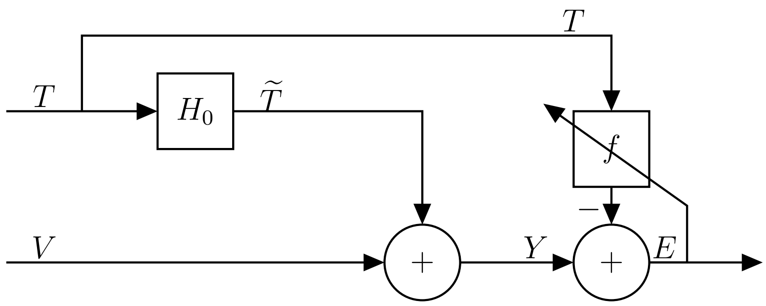
$$\mathbf{H_1} = \mathbf{R_Y}^{-1}\mathbf{R_{VY}} = \mathbf{R_Y}^{-1}(\mathbf{R_Y} - \mathbf{R_{\widetilde{T}}}).$$

$R_Y[k]$ can be obtained from the samples of Y, and $R_{\widetilde{T}}[k]$ was determined before.

3) The filter coefficients should be computed every 100 ms. One must be careful that now the maximum length of the filter is $0.1 \times Fs$, where Fs is the sampling frequency.

4) Estimating the power spectrum $S_Y(\omega)$ (or computing the inverse of R_Y) is too computationally demanding.

5) The filter order is N (the filter length $N+1$), and the error signal controlling the filter is $E[n] = Y[n] - (T * f)[n]$, where $*$ represents convolution.



6)

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \mathbf{T}_n E[n]$$

$\mathbf{f}_{n+1}$ and $\mathbf{f}_n$ are $(N+1) \times 1$ vectors containing the filter coefficients at time instants $n+1$ and n, μ is the step size, $E[n]$ is the error signal (to be minimized) and $\mathbf{T}$ is a $(N+1) \times 1$ vector containing the last $N+1$ samples of T.

To insure convergence, the step size should be $0 < \mu < \frac{2}{\lambda_{max}}$, where λ_{max} is the highest eigenvalue of R_T, and R_T can be estimated from the samples of T.

An easier-to-determine condition is $0 < \mu < \frac{2}{(N+1)\sigma_T^2}$, because it is easier to estimate σ_T^2 than the eigenvalues of R_T (where $N+1$ is the length of the filter).

Grade Scale.

The exams accounts for a total of 62 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

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

and then rounded to .5 steps, that is

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

The result is then constraint to be at maximum 6.