Assignment 1 (ML for TS) - MVA 2023/2024

Benjamin Lapostolle benjamin.jean.lapostolle@gmail.com

December 15, 2023

1 Introduction

Objective. This assignment has three parts: questions about the convolutional dictionary learning, the spectral features and a data study using the DTW.

Warning and advice.

- Use code from the tutorials as well as from other sources. Do not code yourself well-known procedures (e.g. cross validation or k-means), use an existing implementation.
- The associated notebook contains some hints and several helper functions.
- Be concise. Answers are not expected to be longer than a few sentences (omitting calculations).

Instructions.

- Fill in your names and emails at the top of the document.
- Hand in your report (one per pair of students) by Tuesday 7$^{\text{th}}$ November 23:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname2.pdf` and
  `FirstnameLastname1_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: docs.google.com/forms/d/e/1FAIpQLSdTwJEyc6QIoYTknjk12kJMtcKllFvPlWLk5LbyugW0YO7K6Q/viewform?usp=sf_link.

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|y - X\beta\|_2^2 \quad + \quad \lambda \|\beta\|_1 \tag{1}$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n \times p}$ the design matrix, $\beta \in \mathbb{R}^p$ the vector of regressors and $\lambda > 0$ the smoothing parameter.

Show that there exists $\lambda_{\max}$ such that the minimizer of (1) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

The objective function of this problem is convex and is infinite at infinity. Therefore, we know that the problem is feasible. Since the norm l_1 is not differentiable on $\mathbb{R}^p$, we can compute the subgradient of the objective function to determine the minimum. A point x^* is a minimizer of a function f (not necessarily convex) if and only if f is subdifferentiable at x^* and $0 \in \partial f(x^*)$.

Let's name g the objective function.

For a given $\beta \in \mathbb{R}^p$,

$\partial g(\beta) = -X^T(y - X\beta) + \lambda z$ with z such that $\forall i \in \{1, 2, \ldots, p\}$, $z_i = \begin{cases} \text{sign}(\beta_i) & \text{if } \beta_i \neq 0 \\ z_i \in [-1, 1] & \text{otherwise} \end{cases}$

0_p is a minimizer $\Leftrightarrow 0 \in \partial g(0_p) \Leftrightarrow \exists z \in [-1, 1]^p, X^T y = \lambda z$

The last equality can only be true for a $z \in [-1, 1]^p$ if and only if $\lambda >= \|X^T y\|_\infty$, hence we have the result.

$$\lambda_{\max} = \|X^T y\|_\infty \tag{2}$$

Question 2

For a univariate signal $\mathbf{x} \in \mathbb{R}^n$ with n samples, the convolutional dictionary learning task amounts to solving the following optimization problem:

$$\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{3}$$

where $\mathbf{d}_k \in \mathbb{R}^L$ are the K dictionary atoms (patterns), $\mathbf{z}_k \in \mathbb{R}^{N-L+1}$ are activations signals, and $\lambda > 0$ is the smoothing parameter.

Show that

- for a fixed dictionary, the sparse coding problem is a lasso regression (explicit the response vector and the design matrix);
- for a fixed dictionary, there exists $\lambda_{\max}$ (which depends on the dictionary) such that the sparse codes are only 0 for any $\lambda > \lambda_{\max}$.

Answer 2

We want to write 3 in the form $\|x - XZ\|_2^2 + \lambda \|Z\|_1$, with $X \in \mathbb{R}^{N \times (N-L+1)K}$, $Z \in\in \mathbb{R}^{(N-L+1)K}$.

Let's write $Z = (z_1, \dots, z_K)^T$ and $X = (X_1, \dots, X_K)$, with $X_i = \begin{pmatrix} d_i^1 & 0 & \dots & 0 \\ d_i^2 & d_i^1 & & \vdots \\ \vdots & d_i^2 & & 0 \\ d_i^L & \vdots & & d_i^1 \\ 0 & d_i^L & & \vdots \\ \vdots & \vdots & & d_i^{L-1} \\ 0 & 0 & \dots & d_i^L \end{pmatrix}$

By doing the matrix product, we see that we obtain the same signal in the both formulations of the problem. By similarity with question 1, we can compute the $\lambda_{\max}$.

$$\lambda_{\max} = \|2X^T x\|_\infty \tag{4}$$

3 Spectral feature

Let X_n ($n = 0, \dots, N-1$) be a weakly stationary random process with zero mean and autocovariance function $\gamma(\tau) := \mathbb{E}(X_n X_{n+\tau})$. Assume the autocovariances are absolutely summable, i.e. $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$, and square summable, i.e. $\sum_{\tau \in \mathbb{Z}} \gamma^2(\tau) < \infty$. Denote by f_s the sampling frequency, meaning that the index n corresponds to the time instant n/f_s and for simplicity, let N be even.

The *power spectrum* S of the stationary random process X is defined as the Fourier transform of the autocovariance function:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{5}$$

The power spectrum describes the distribution of power in the frequency space. Intuitively, large values of $S(f)$ indicates that the signal contains a sine wave at the frequency f. There are many estimation procedures to determine this important quantity, which can then be used in a machine learning pipeline. In the following, we discuss about the large sample properties of simple estimation procedures, and the relationship between the power spectrum and the autocorrelation.

(Hint: use the many results on quadratic forms of Gaussian random variables to limit the amount of calculations.)

Question 3

In this question, let X_n ($n = 0, \dots, N-1$) be a Gaussian white noise.

- Calculate the associated autocovariance function and power spectrum. (By analogy with the light, this process is called "white" because of the particular form of its power spectrum.)

Answer 3

$(X_i)_{i \leq N-1}$ follow a Gaussian $\mathcal{N}(0, \sigma^2)$ and are i.i.d.. For a given $\tau \neq 0$ and n integer so that $n + \tau \leqslant N-1$, we have $\gamma(\tau) := \mathbb{E}(X_n X_{n+\tau}) = \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) = 0$, by independance of the $(X_i)_{i \leq N-1}$.

Hence we have: $\forall \tau, \gamma(\tau) = \begin{cases} \sigma^2 & \text{if } \tau = 0 \\ 0 & \text{if } \tau \neq 0 \end{cases}$. and $\forall f,\ S(f) = \gamma(0) = \sigma^2$.

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{6}$$

for $\tau = 0, 1, \ldots, N-1$ and $\hat{\gamma}(\tau) := \hat{\gamma}(-\tau)$ for $\tau = -(N-1), \ldots, -1$.

- Show that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$ but asymptotically unbiased. What would be a simple way to de-bias this estimator?

Answer 4

For a given τ,

$$\mathbb{E}(\gamma(\tau)) := \mathbb{E}(\frac{1}{N} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau}) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} \gamma(\tau) = \frac{N-\tau}{N} \gamma(\tau)$$

Hence, we see that our estimator is biased. However, when N goes to infinity it becomes unbiased. A simple way to de-bias the estimator would be to multiply our estimator by $\frac{N}{N-\tau}$.

Question 5

Define the discrete Fourier transform of the random process $\{X_n\}_n$ by

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \tag{7}$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \ldots, |J(f_{N/2})|^2$ where $f_k = f_s k/N$. (They can be efficiently computed using the Fast Fourier Transform.)

- Write $|J(f_k)|^2$ as a function of the sample autocovariances.
- For a frequency f, define $f^{(N)}$ the closest Fourier frequency f_k to f. Show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Answer 5

For a given k, we have: $J(f_k) := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-2\pi i \frac{f_k n}{f_s}} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{-2\pi i \frac{kn}{N}}$

$$
\begin{aligned}
|J(f_k)|^2 &= J(f_k)\overline{J(f_k)} \\
&= \frac{1}{N}(\sum_{n=0}^{N-1} X_n e^{-2\pi i \frac{kn}{N}})(\sum_{n=0}^{N-1} X_n e^{2\pi i \frac{kn}{N}}) \\
&= \frac{1}{N}((\sum_{n=0}^{N-1} X_n^2) + \sum_{n=0}^{N-1}\sum_{\substack{m=0\\ n\neq m}}^{N-1} X_n X_m e^{2\pi i \frac{k(m-n)}{N}}) \\
&= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n X_m e^{2\pi i \frac{k(m-n)}{N}} \\
&= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{\tau=1}^{N-1-n} X_n X_{n+\tau} e^{2\pi i \frac{k\tau}{N}} \\
&= \hat{\gamma}(0) + \frac{2}{N}\sum_{\tau=1}^{N-1}\sum_{n=0}^{N-1-\tau} X_n X_{n+\tau} e^{2\pi i \frac{k\tau}{N}} \\
&= \hat{\gamma}(0) + \frac{2}{N}\sum_{\tau=1}^{N-1}\sum_{n=0}^{N-1-\tau} X_n X_{n+\tau} e^{2\pi i \frac{k\tau}{N}} \\
&= \hat{\gamma}(0) + 2\sum_{\tau=1}^{N-1} \hat{\gamma}(\tau) e^{2\pi i \frac{k\tau}{N}} \\
&= \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{2\pi i \frac{k\tau}{N}} \\
&= \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) cos(2\pi \frac{k\tau}{N})
\end{aligned}
\tag{8}
$$

For a given f, let's defined $f^{(N)}$. $k_f := \lfloor \frac{fN}{f_s} \rfloor$, we defined $f^{(N)} = \begin{cases} f_{k_f} \text{ if } f^N \in [f_{k_f} + \frac{f_s}{2N}] \\ f_{k_f+1} \text{if } f^N \in [\frac{f_s}{2N}, f_{k_f+1}] \end{cases}$ By definition, we have $f^{(N)} \xrightarrow{N\to\infty} f$. For a fixed N, we have: $|J(f^{(N)})|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{2\pi i \frac{k^{(N)}\tau}{N}}$. Moreover,

$$
\begin{aligned}
\mathbb{E}(|J(f^{(N)})|^2) &= \mathbb{E}(|J(f^{(N)})|^2) \\
&= \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}(\hat{\gamma}(\tau)) e^{-2\pi i \frac{k^{(N)}\tau}{N}} \\
&= \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}(\hat{\gamma}(\tau)) e^{-2\pi i \frac{k^{(N)}\tau}{N}} \\
&= \sum_{\tau=-(N-1)}^{N-1} \frac{N-\tau}{N} \gamma(\tau) e^{-2\pi i \frac{k^{(N)}\tau}{N}} \\
&= \sum_{\tau=-\infty}^{\infty} I_{[-(N-1),N-1]}(\tau) \frac{N-\tau}{N} \gamma(\tau) e^{-2\pi i \frac{k^{(N)}\tau}{N}} \\
&= \sum_{\tau=-\infty}^{\infty} a_\tau^N
\end{aligned}
\tag{9}
$$

We have: $\forall \tau \in \mathbb{R}, \forall N \leqslant 1, |a_\tau^N| \leqslant \gamma(\tau)$ and $a_\tau^N \xrightarrow{\text{N}\to\infty} \gamma(\tau) e^{-2\pi i \frac{f\tau}{f_s}}$.

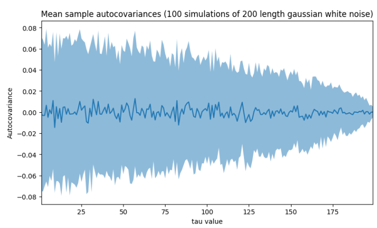
Because $(\gamma(\tau))_\tau$ is sommable, when can use the dominated convergence theorem to conclude that $\mathbb{E}(|J(f^{(N)})|^2) \xrightarrow{\text{N}\to\infty} S(f)$.

Question 6

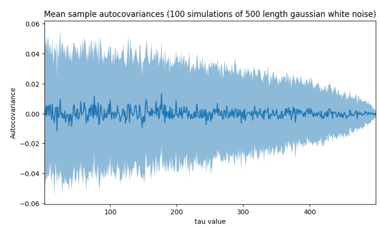
In this question, let X_n $(n = 0, \dots, N-1)$ be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs τ) for 100 simulations of X. Plot the average value as well as the average $\pm$ the standard deviation. What do you observe?
- For $N \in \{200, 500, 1000\}$, compute the *periodogram* ($|J(f_k)|^2$ vs f_k) for 100 simulations of X. Plot the average value as well as the average $\pm$ the standard deviation. What do you observe?

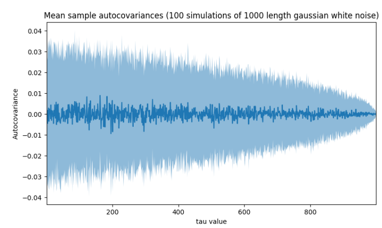
Add your plots to Figure 1.



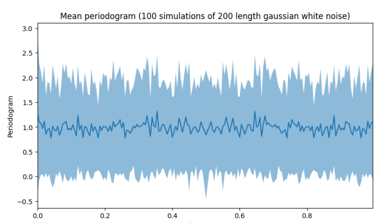
Autocovariance ($N = 200$)



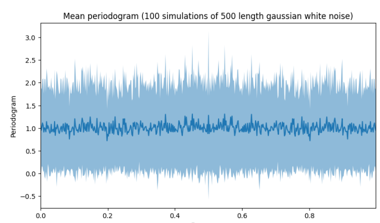
Autocovariance ($N = 500$)



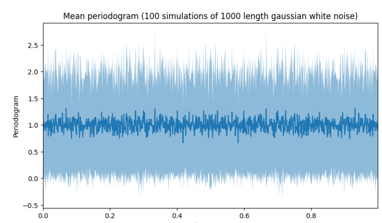
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

Answer 6

In Figure 1, we can observe the sample autocovariance plots and periodogram graphs for Gaussian white noise of length 200, 500 and 1000 with variance 1 and sampling frequency 1 Hz.

- On the first row of figures, we observe that the autocovariance of our samples is quasi-constant, and equals to 0. This was to be expected as our Gaussian white noise defines a process for which every sample is independent. We can also notice that the standard deviation of the autocovariance decreases with the value of τ. This is because, when τ increases, the sum becomes used to calculate $\hat{\gamma}(\tau)$ becomes smaller, leading to less intrinsic standard deviation.

- On the second row of figures, we observe that the periodogram is constant with value 1. This was to be expected, as it is a good approximation of the Fourier transform of the signal, and that as the name suggests, a Gaussian white noise has a Fourier signature similar to the one of a white light source, ie a constant spectrum across all light frequencies.

Question 7

We want to show that the estimator $\hat{\gamma}(\tau)$ is consistent, i.e. it converges in probability when the number N of samples grows to ∞ to the true value $\gamma(\tau)$. In this question, assume that X is a wide-sense stationary *Gaussian* process.

- Show that for $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{10}$$

  (Hint: if $\{Y_1, Y_2, Y_3, Y_4\}$ are four centered jointly Gaussian variables, then $\mathbb{E}[Y_1Y_2Y_3Y_4] = \mathbb{E}[Y_1Y_2]\mathbb{E}[Y_3Y_4] + \mathbb{E}[Y_1Y_3]\mathbb{E}[Y_2Y_4] + \mathbb{E}[Y_1Y_4]\mathbb{E}[Y_2Y_3]$.)

- Conclude that $\hat{\gamma}(\tau)$ is consistent.

Answer 7

For a given τ,

$$\begin{aligned}
\mathbb{V}(\gamma(\hat{\tau})) &= \mathbb{V}(\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}) \\
&= \mathbb{E}((\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2) - \mathbb{E}(\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2 \\
&= \mathbb{E}((\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2) - \frac{1}{N^2}(\sum_{n=0}^{N-\tau-1} \gamma(\tau))^2 \\
&= \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau} X_m X_{m+\tau})) - A, A = \frac{1}{N^2}(\sum_{n=0}^{N-\tau-1} \gamma(\tau))^2 \\
&= \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} (\mathbb{E}(X_n X_{n+\tau})\mathbb{E}(X_m X_{m+\tau}) + \mathbb{E}(X_n X_m)\mathbb{E}(X_{n+\tau} X_{m+\tau}) + \mathbb{E}(X_n X_{m+\tau})\mathbb{E}(X_m X_{n+\tau}))) - A \\
&= \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \gamma(\tau)^2) + \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \gamma(m-n)^2) + \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \gamma(m+\tau-n)\gamma(n+\tau-m)) - A \\
&= \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \gamma(m-n)^2) + \frac{1}{N^2}(\sum_{n,m=0}^{N-\tau-1} \gamma(m+\tau-n)\gamma(n+\tau-m))
\end{aligned} \tag{11}$$

Furthermore, we have:

$$\sum_{n,m=0}^{N-\tau-1} \gamma(m-n)^2 = (N-\tau)\gamma(0) + 2\sum_{k=0}^{N-\tau-1} (N-\tau-|k|)\gamma(k)^2 = \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|)\gamma(k)^2$$

$$\sum_{n,m=0}^{N-\tau-1} \gamma(m+\tau-n)\gamma(n+\tau-m) = \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|)\gamma(\tau-k)\gamma(\tau+k)$$

One can see that the equations above are true by placing each term of the sum computed in a $(N-\tau-1)\times(N-\tau-1)$ matrix, and see that there are $N-\tau$ terms where $m=n$, $2(N-1-\tau)$ terms where $|m-n|=1$ — the diagonal above and under the main diagonal of the matrix — and so on. Doing the same, you can also compute the second sum above.

Hence, we obtain the result:

$$\text{var}(\hat{\gamma}(\tau)) = \frac{1}{N}\sum_{n=-(N-\tau-1)}^{N-\tau-1}\left(1-\frac{\tau+|n|}{N}\right)\left(\gamma^2(n)+\gamma(n-\tau)\gamma(n+\tau)\right)$$

$$\text{var}(\hat{\gamma}(\tau)) = \sum_{n=-\infty}^{\infty} a_n^N$$

with $a_n^N = I_{[-(N_\tau-1),N-\tau-1]}(n)\frac{1}{N}\left(1-\frac{\tau+|n|}{N}\right)\left(\gamma^2(n)+\gamma(n-\tau)\gamma(n+\tau)\right)$. We have $\forall n, \forall N |a_n^N| \leqslant \frac{1}{N}(\gamma^2(n)+\gamma(n-\tau)\gamma(n+\tau))$, which is summable. Since $a_n^N \xrightarrow{\text{N}\to\infty} 0$, we can conclude that $\text{var}(\hat{\gamma}(\tau)) \xrightarrow{\text{N}\to\infty} 0$.

$$\begin{aligned}
\mathbb{P}[|\hat{\gamma}(\tau)-\gamma(\tau)|>\epsilon] &= \mathbb{P}[|\hat{\gamma}(\tau)-\mathbb{E}(\hat{\gamma}(\tau))+\mathbb{E}(\hat{\gamma}(\tau))+\gamma(\tau)|>\epsilon] \\
&\leqslant \mathbb{P}[|\gamma\hat{(}\tau)-\mathbb{E}(\gamma\hat{(}\tau))|+|\mathbb{E}(\gamma\hat{(}\tau))-\gamma(\tau)|>\epsilon] \\
&= \mathbb{P}[|\hat{\gamma}(\tau)-\mathbb{E}(\hat{\gamma}(\tau))|>\epsilon-|\mathbb{E}(\hat{\gamma}(\tau))-\gamma(\tau)|] \\
&\leqslant \mathbb{P}[|\hat{\gamma}(\tau)-\mathbb{E}(\hat{\gamma}(\tau))|>\epsilon-\frac{\epsilon}{2}]
\end{aligned} \tag{12}$$

The last inequality is true for $N \geqslant N_0$, since $|\mathbb{E}(\hat{\gamma}(\tau))-\gamma(\tau)| \xrightarrow{\text{N}\to\infty} 0$.

By using the Chebyshev inequality for $\epsilon > 0$, we have:

$$\mathbb{P}\left[|\hat{\gamma}(\tau)-\mathbb{E}[\hat{\gamma}(\tau)]|>\frac{\epsilon}{2}\right] \leq 2\text{Var}(\hat{\gamma}(\tau))\cdot\frac{1}{\epsilon}$$

Therefore, we can conclude that our estimator is consitent because we have shown that $\text{var}(\hat{\gamma}(\tau)) \xrightarrow{\text{N}\to\infty} 0$.

Contrary to the correlogram, the periodogram is not consistent. It is one of the most well-known estimators that is asymptotically unbiased but not consistent. In the following question, this is proven for a Gaussian white noise but this holds for more general stationary processes.

Question 8

Assume that X is a Gaussian white noise (variance σ^2) and let $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n/f_s$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n/f_s)$. Observe that $J(f) = (1/\sqrt{N})(A(f)+iB(f))$.

- Derive the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, \ldots, f_{N/2}$ where $f_k = f_s k/N$.
- What is the distribution of the periodogram values $|J(f_0)|^2$, $|J(f_1)|^2, \ldots, |J(f_{N/2})|^2$.
- What is the variance of the $|J(f_k)|^2$? Conclude that the periodogram is not consistent.
- Explain the erratic behavior of the periodogram in Question 6 by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

For the expectation we have:

$$\begin{aligned}
\mathbb{E}(J(f)) &= \mathbb{E}(\frac{1}{N}(A(f) + iB(f))) \\
&= \mathbb{E}(\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s}) \\
&= \frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} \mathbb{E}(X_n) e^{-2\pi i f n/f_s} \\
&= 0
\end{aligned} \tag{13}$$

As such, $\mathbb{E}(A(f)) = 0$ and $\mathbb{E}(B(f)) = 0$ for all f.

For the variance we have:

$$\begin{aligned}
\mathbb{V}(B(f_k)) &= \mathbb{V}(\sum_{n=0}^{N-1} X_n \sin(-2\pi f_k n/f_s)) \\
&= \sigma^2 \sum_{n=0}^{N-1} \sin(2\pi f_k n/f_s)^2
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mathbb{V}(A(f_k)) &= \mathbb{V}(\sum_{n=0}^{N-1} X_n \cos(-2\pi f_k n/f_s)) \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos(2\pi f_k n/f_s)^2
\end{aligned} \tag{15}$$

As such, we can show that:

$$\begin{aligned}
\mathbb{V}(A(f_k)) + \mathbb{V}(B(f_k)) &= \sigma^2 \sum_{n=0}^{N-1} \cos(2\pi f_k n/f_s)^2 + \sigma^2 \sum_{n=0}^{N-1} \sin(2\pi f_k n/f_s)^2 \\
&= N\sigma^2 \\
\mathbb{V}(A(f_k)) - \mathbb{V}(B(f_k)) &= \sigma^2 \sum_{n=0}^{N-1} \cos(2\pi f_k n/f_s)^2 - \sigma^2 \sum_{n=0}^{N-1} \sin(2\pi f_k n/f_s)^2 \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos(4\pi k n/N)
\end{aligned} \tag{16}$$

Therefore:

$$\begin{aligned}\mathbb{V}(A(f)) &= \frac{\sigma^2}{2}(N + \sum_{n=0}^{N-1}\cos(4\pi kn/N)) \\ \mathbb{V}(B(f)) &= \frac{\sigma^2}{2}(N - \sum_{n=0}^{N-1}\cos(4\pi kn/N))\end{aligned} \tag{17}$$

When $\frac{N}{2} > k > 0$, since $\sum_{n=0}^{N-1}\cos(4\pi kn/N)) = Re(\sum_{n=0}^{N-1}(e^{\frac{i4\pi kn}{N}})^n) = 0$, we have that:

$$\mathbb{V}(A(f_k)) = \mathbb{V}(B(f_k)) = \frac{N\sigma^2}{2}$$

Moreover, the covariance between A_{f_k} and B_{f_k} is 0. Because they are gaussians, it means that these random variables are independant.

$$\begin{aligned}cov(A_{f_k}, B_{f_k}) &= \mathbb{E}(A_{f_k}B_{f_k}) \\ &= \mathbb{E}(\sum_{n=0}^{N-1} X_n^2 \cos(\frac{-2\pi kn}{N})\sin(\frac{-2\pi kn}{N}) \quad = \frac{\sigma^2}{2}\sum_{n=0}^{N-1}\sin(\frac{-4\pi kn}{N}) = 0\end{aligned} \tag{18}$$

Hence, A and B are two independant gaussians of same mean and variance. When $k = 0$ and $k = \frac{N}{2}$, then $B(f_{\frac{N}{2}}) = B(f_0) = 0$ and $A(f_{\frac{N}{2}}) = A(f_0) = \sum_{n=0}^{N-1} X_n$.

Distribution of $|J(f_k)|^2|$

For $f = f_k$ we have:

$$\begin{aligned}|J(f)|^2 &= J(f)\overline{J(f)} \\ &= \frac{1}{N}(A(f) + iB(f))(A(f) - iB(f)) \\ &= \frac{1}{N}(A(f)^2 + B(f)^2) \\ &= \begin{cases} \frac{\mathbb{V}(A(f)}{N}(\frac{A(f)^2}{\mathbb{V}(A(f))} + \frac{B(f)^2}{\mathbb{V}(B(f))}) \text{ if k > 0} \\ \frac{\mathbb{V}(A(f)}{N}\frac{A(f)^2}{\mathbb{V}(A(f))} \text{ if k = 0}\end{cases}\end{aligned} \tag{19}$$

$A(f)$ is a gaussian distribution (as it is a sum of independent gaussian variables $X_n\cos(-2\pi fn/f_s)$), and so is $B(f)$ and they are independant. As such, $\frac{N}{\mathbb{V}(A(f))}|J(f_k)|^2 = \frac{2}{\sigma^2}|J(f_k)|^2$ follow a chi-squared law of order 2.

The variance of a chi-squared law of order 2 is 4 (and 2 for order 1 in the f_0 case). Therefore, the variance of $|J(f_k)|^2$ is σ^4. Since for every N, the variance of the estimators and their distributions do not change, the periodogram is not consistent.

Covariance between the $|J(f_k)|^2$

For $k_1 \neq k_2$:

First, let's prove that $A(f_{k_1})$, $B(f_{k_1})$, $A(f_{k_2})$, $B(f_{k_2})$ are independant. We prove it for one couple, but the proof for the others uses the same argument.

$$\mathbb{E}(A(f_{k_1})A(f_{k_2})) = \frac{\sigma^2}{2}\sum_{n=0}^{N-1}\cos(\frac{2\pi k_1 n}{N})\cos(\frac{2\pi k_2 n}{N}) = \frac{\sigma^2}{2}(\sum_{n=0}^{N-1}\cos(\frac{(k_1+k_2)2\pi n}{N}) + \sum_{n=0}^{N-1}\cos(\frac{(k_1-k_2)2\pi n}{N})) = 0$$

Hence, by recalling that if A and B are independant, A^2 and B^2 are also independant, we can use this to compute the covariance:

$$\begin{aligned}
\mathrm{Cov}(|J(f_{k_1})|^2, |J(f_{k_2})|^2) &= \mathbb{E}\left[\left(\frac{1}{N}[A(f_{k_1})^2 + B(f_{k_1})^2] - \mathbb{E}\left[\frac{1}{N}(A(f_{k_1})^2 + B(f_{k_1})^2)\right] \times \right.\right. \\
&\left.\left.\left(\frac{1}{N}(A(f_{k_2})^2 + B(f_{k_2})^2] - \mathbb{E}\left[\frac{1}{N}(A(f_{k_2})^2 + B(f_{k_2})^2)\right]\right)\right] = \mathbb{E}\left[\frac{1}{N^2}\right. \\
&(A(f_{k_1})^2A(f_{k_2})^2 + B(f_{k_1})^2B(f_{k_2})^2 + A(f_{k_1})^2B(f_{k_2})^2 + B(f_{k_1})^2A(f_{k_2})^2) - \sigma^2 \\
&\times\left(\frac{1}{N}(A(f_{k_2})^2 + B(f_{k_2})^2 + A(f_{k_1})^2 + B(f_{k_1})^2) - \sigma^2\right) = \frac{1}{N^2} \\
&\times\left(\frac{\sigma^4N^2}{4} + \frac{\sigma^4N^2}{4} + \frac{\sigma^4N^2}{4} + \frac{\sigma^4N^2}{4}\right) - \sigma^2 \\
&\times\left(\frac{1}{N}(\frac{\sigma^2N}{2} + \frac{\sigma^2N}{2} + \frac{\sigma^2N}{2} + \frac{\sigma^2N}{2}) - \sigma^2\right) = \sigma^4 - \sigma^2(\sigma^2) = 0
\end{aligned}$$

Hence, the covariance between the $|J(f_k)|$ is 0 explains the randomness we observe in the graph.

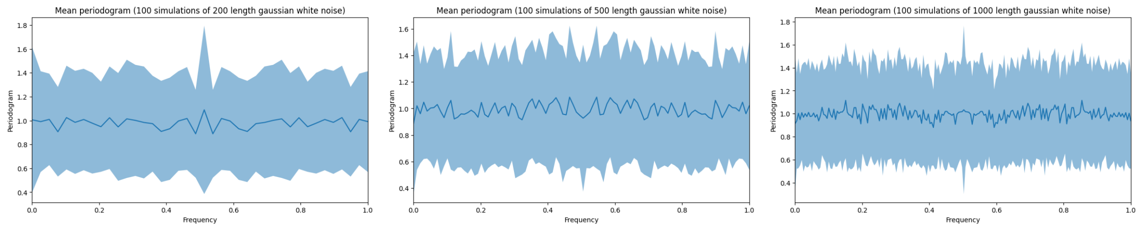
Question 9

As seen in the previous question, the problem with the periodogram is the fact that its variance does not decrease with the sample size. A simple procedure to obtain a consistent estimate is to divide the signal in K sections of equal durations, compute a periodogram on each section and average them. Provided the sections are independent, this has the effect of dividing the variance by K. This procedure is known as Bartlett's procedure.

- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe.

Add your plots to Figure 2.

Answer 9



Periodogram ($N = 200$) Periodogram ($N = 500$) Periodogram ($N = 1000$)

Figure 2: Barlett's periodograms of a Gaussian white noise (see Question 9).

4 Data study

4.1 General information

Context. The study of human gait is a central problem in medical research with far-reaching consequences in the public health domain. This complex mechanism can be altered by a wide range of pathologies (such as Parkinson's disease, arthritis, stroke,...), often resulting in a significant loss of autonomy and an increased risk of fall. Understanding the influence of such medical disorders on a subject's gait would greatly facilitate early detection and prevention of those possibly harmful situations. To address these issues, clinical and bio-mechanical researchers have worked to objectively quantify gait characteristics.

Among the gait features that have proved their relevance in a medical context, several are linked to the notion of step (step duration, variation in step length, etc.), which can be seen as the core atom of the locomotion process. Many algorithms have therefore been developed to automatically (or semi-automatically) detect gait events (such as heel-strikes, heel-off, etc.) from accelerometer and gyrometer signals.

Data. Data are described in the associated notebook.

4.2 Step classification with the dynamic time warping (DTW) distance

Task. The objective is to classify footsteps then walk signals between healthy and non-healthy.

Performance metric. The performance of this binary classification task is measured by the F-score.

Question 10

Combine the DTW and a k-neighbors classifier to classify each step. Find the optimal number of neighbors with 5-fold cross-validation and report the optimal number of neighbors and the associated F-score. Comment briefly.

Answer 10

The method is clearly explained in the notebook. However the key points are that:

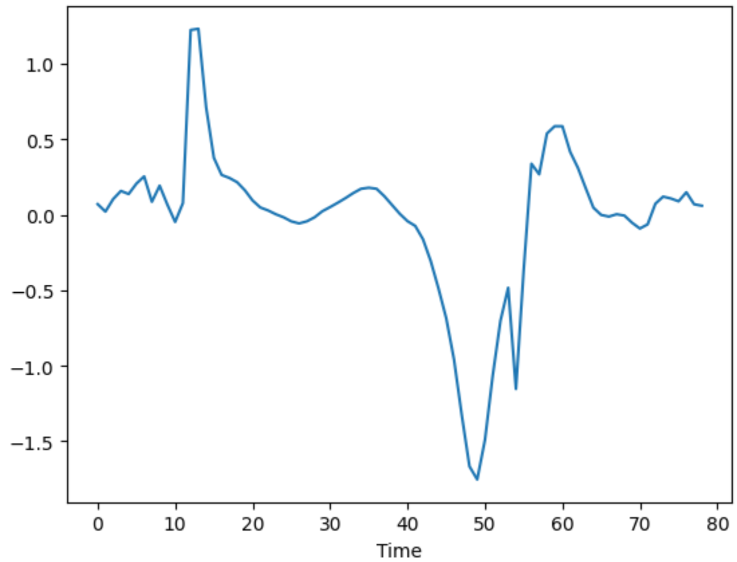
- We started by analysing the classification task, and saw that the train/test cut was poorly implemented, leading to a drastic distribution shift between the two sets. This hindered a lot the performances of our model.
- The best performing hyperparameter K was 1, and we report a F1-score of 0.48, with an accuracy of 0.34.

This performance is clearly bad, however we can't do much more with the dataset as it is. However, the exercice is still interesting, as it is an example of benign overfit. No matter how many neighbours we consider, the test performances are stable. The result is that an overfitting model (K=1) is not detrimental to the final metric.

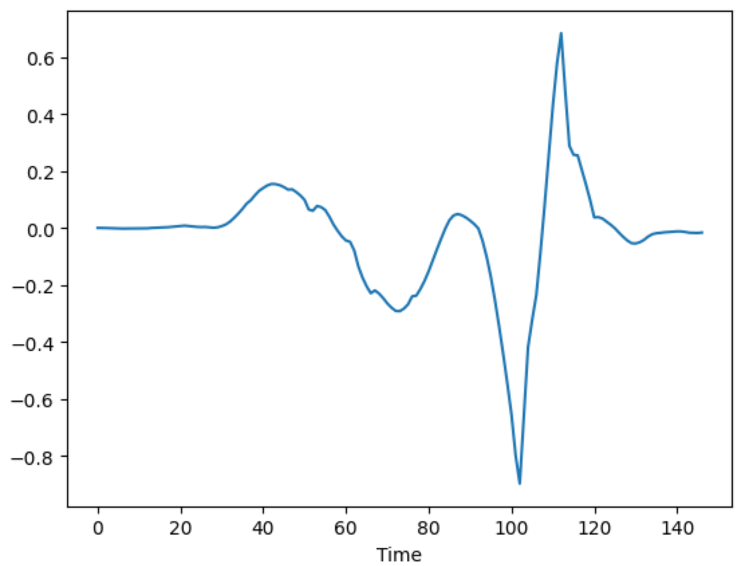
Question 11

Display on Figure 3 a badly classified step from each class (healthy/non-healthy).

Answer 11



Badly classified healthy step



Badly classified non-healthy step

Figure 3: Examples of badly classified steps (see Question 11).