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

Ben KABONGO ben.kabongo_buzangu@ens-paris-saclay.fr
Martin BROSSET martin.brosset@student-cs.fr

November 8, 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 7th 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

Let $f_\lambda(\beta)$ be the objective function to minimize in the Lasso problem.

$$\begin{aligned} f_\lambda(\beta) &= \frac{1}{2}\left\|y - X\beta\right\|_2^2 + \lambda\left\|\beta\right\|_1 \\ &= \frac{1}{2}y^T y - \beta^T X^T y + \frac{1}{2}\beta^T X^T X\beta + \lambda\left\|\beta\right\|_1 \end{aligned}$$

Since $\mathbb{R}^p$ is convex, we have the following property with the subgradient of f $\partial_\beta f$:

$$\hat{\beta} \text{ solution of (1)} \iff 0 \in \partial_\beta f(\hat{\beta})$$

The subgradient for the i^{th} coordinate of β is

$$\partial_{\beta_i} f(\beta) = -(X^T y)_i + (X^T X\beta)_i + \lambda\partial_{\beta_i}\left\|\beta\right\|_1$$

And

$$\partial_{\beta_i}\left\|\beta\right\|_1 = \begin{cases} 1 & \text{if } \beta_i > 0 \\ [0,1] & \text{if } \beta_i = 0 \\ -1 & \text{if } \beta_i < 0 \end{cases}$$

Then for $\beta_i = 0$,

$$\partial_{\beta_i} f(\beta) = \{\lambda t - (X^T y)_i \mid t \in [0,1]\}$$

Then $\beta = \mathbf{0}_p$ is solution of (1) if and only if $\forall i \in \mathbb{R}^p$:

$$\begin{aligned} 0 \in \partial_{\beta_i} f(\beta) &\iff -\lambda - (X^T y)_i \leq 0 \leq \lambda - (X^T y)_i \\ &\iff -\lambda \leq (X^T y)_i \leq \lambda \\ &\iff |(X^T y)_i| \leq \lambda \end{aligned}$$

We then have :

$$\lambda_{\max} = \max_{i=1\ldots p} |(X^T y)_i| \tag{2}$$

And since the norm 1 of all of the solutions of the lasso problem are equal, we have the unicity, so $\forall\lambda > \lambda_{\max}$, $\mathbf{0}_p$ is the only solution of (1).

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:

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

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

Let (d_k) be a fixed dictionary such that $\|\mathbf{d}_k\|_2^2 \leq 1$.
Let's clarify the convolution product. For $0 \leq m \leq n$:

$$\begin{aligned} z_k * d_k[m] &= \sum_{n=-\infty}^{+\infty} z_k[n]\ d_k[m-n] \\ &= \sum_{n=0}^{m} z_k[n]\ d_k[m-n] \end{aligned}$$

So let $\mathbf{D}_k \in \mathcal{M}_{n,L}$ such that :

$$(\mathbf{D}_k)_{i,j} = \begin{cases} d_k[i-j] & \text{if } 0 \leq i-j \leq L-1 \\ 0 & \text{otherwise} \end{cases}$$

We then have $z_k * d_k = \mathbf{D}_k z_k$.
Let D and z be :

$$\mathbf{D}^T = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_K \end{bmatrix} \text{ and } z = \begin{bmatrix} z_1 \\ \vdots \\ z_K \end{bmatrix}.$$

Then :

$$\begin{aligned} \mathbf{D}z &= \sum_{k=1}^{K} \mathbf{D}_k z_k = \sum_{k=1}^{K} z_k * d_k \\ \|z\|_1 &= \sum_{k=1}^{K} \|z_k\| \end{aligned}$$

So we can rewrite the convolutional dictionary learning problem as :

$$\min_z \frac{1}{2}\|x - \mathbf{D}z\|_2^2 + \frac{\lambda}{2}\|z\|_1$$

And using the previous question we find :

$$\lambda_{\max} = 2 \max_{0 \leq i \leq L} |(D^T x)_i|$$

Such that $\forall \lambda > \lambda_{\max}, z_k = \mathbf{0}$ $(k = 1, ..., K)$.

3 Spectral feature

Let X_n $(n = 0, \ldots, N-1)$ 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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{4}$$

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

We have by independence of the X_n and denoting σ the standart deviation of the white noise :

$$\gamma(\tau) = \mathbb{E}(X_n X_{n+\tau}) = \begin{cases} \mathbb{E}(X_n^2) = \sigma^2 & \text{if } \tau = 0 \\ \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) = 0 & \text{otherwise} \end{cases}$$

And then :

$$\begin{aligned} S(f) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s} \\ &= \sigma^2 \end{aligned}$$

It's called white noise since its module is equal for all of the frequencies.

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{5}$$

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

- 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

We have :

$$\begin{aligned}\mathbf{E}[\hat{\gamma}] &= \frac{1}{N}\sum_{n=0}^{N-\tau-1} \mathbf{E}[X_n X_{n+\tau}] \\ &= \frac{1}{N}\sum_{n=0}^{N-\tau-1} \gamma(\tau) \\ &= \frac{N-\tau}{N}\gamma(\tau)\end{aligned}$$

So we can express the bias :

$$\begin{aligned}B(\hat{\gamma}(\tau)) &= \mathbf{E}[\hat{\gamma}] - \gamma(\tau) \\ &= -\frac{\tau}{N}\gamma(\tau) \neq 0\end{aligned}$$

So the estimator is indeed biased but since $\lim_{N\to\infty} \frac{\tau}{N}\gamma(\tau) = 0$ it's asymptotically unbiased.
If we take $\tilde{\gamma}(\tau) = \frac{N}{N-\tau}\hat{\gamma}(\tau)$, we then find that its bias is null.

Question 5

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

$$J(f) := (1/\sqrt{N})\sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{6}$$

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

We can write since the process is real :

$$|J(f_k)|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m e^{-2i\pi k(m-n)/N}$$

Let's pose $\tau = m - n$

$$|J(f_k)|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{\tau=-n}^{N-1-n} X_n X_{n+\tau} e^{-2i\pi k\tau/N}$$

Considering that $X_k = 0$ for $k < 0$ or $k > N-1$

$$\begin{aligned}&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{\tau=-(N-1)}^{N-1} X_n X_{n+\tau} e^{-2i\pi k\tau/N} \\ &= \sum_{\tau=-(N-1)}^{N-1} e^{-2i\pi k\tau/N}\frac{1}{N}\sum_{n=0}^{N-1} X_n X_{n+\tau}\end{aligned}$$

Using the same argument we can write

$$
\begin{aligned}
&= \sum_{\tau=-(N-1)}^{N-1} e^{-2i\pi k\tau/N} \frac{1}{N} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \\
&= \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{-2i\pi k\tau/N}
\end{aligned}
$$

Now, let's prove the convergence.
For $N > 0$, we have :

$$
\begin{aligned}
\left| S(f) - \mathbf{E}\left[|J(f^{(N)})|^2 \right] \right| &= | \sum_{\tau=-N}^{+\infty} \gamma(\tau) \exp\left(-2\pi i\tau \frac{f}{f_s} \right) \\
&\quad + \sum_{\tau=-\infty}^{N} \gamma(\tau) \exp\left(-2\pi i\tau \frac{f}{f_s} \right) \\
&\quad + \sum_{\tau=-(N-1)}^{N-1} \gamma(\tau) \exp\left(-2\pi i\tau \frac{f}{f_s} \right) - \mathbf{E}\left[\gamma(\tau)\right] \exp\left(-2\pi i\tau \frac{f^{(N)}}{f_s} \right) | \\
&\leq \sum_{\tau=-N}^{+\infty} |\gamma(\tau)| + \sum_{\tau=-\infty}^{-N} |\gamma(\tau)| \\
&\quad + \sum_{\tau=-N}^{N} |\gamma(\tau)| \left(2\sin\left(\pi\tau \frac{f^{(N)} - f}{f_s} \right) \right) + \frac{|\tau|}{N}
\end{aligned}
$$

Using question 4 to get the last inequality.
Let's focus on the last sum that we denote as S_n.
We have:

$$
S_n = 2\sum_{\tau=-N}^{N} |\gamma(\tau)| \sin\left(\pi\tau \frac{f_N}{f_s} \right) + \frac{|\tau|}{N}
$$

But since $\gamma(\tau)$ is absolutely summable, there exists for $c > 0$, N_0 such that

$$
\sum_{\tau=-\infty}^{-N_0} |\gamma(\tau)| < c \quad \text{and} \quad \frac{1}{N} \sum_{\tau=-N_0}^{N_0} |\gamma(\tau)||\tau| < c \quad \text{for} \quad N > N_0
$$

Since we also have for $x \in \mathbb{R}$, $\sin(x)$ is less than or equal to $\min(1, x)$, it results in:

$$
\begin{aligned}
S_n &\leq 2 \sum_{|\tau| \leq N_0} |\gamma(\tau)| \frac{\pi|\tau|}{f_s|N_0|} |f^{(N)} - f| + 3\frac{1}{N} \sum_{|\tau| > N_0} |\gamma(\tau)||\tau| \\
&\leq \frac{2\pi}{f_s} \frac{1}{|N_0|} \sum_{|\tau| \leq N_0} |\gamma(\tau)||\tau||f^{(N)} - f| + 4\epsilon
\end{aligned}
$$

But $f^{(N)}$ converges towards f, so we can conclude that there exists N_1, such that if $N > N_1$, $S_n < 5\epsilon$. So we obtain that for $N > N_1$,

$$
|S(f) - \mathbf{E}[|J(f^{(N)})|^2]| < 7\epsilon
$$

We can conclude that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

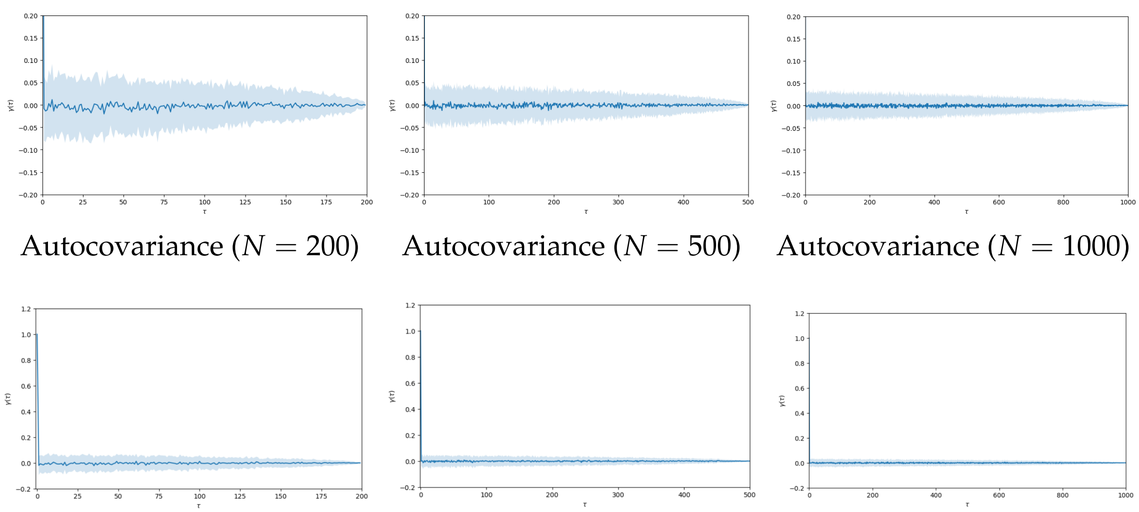
Question 6

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 accordance to the previous theoretical results, the autocovariance approaches $\sigma^2 = 1$ in 0 and oscillate around 0 after.
Moreover, the standard deviation decreases as τ and N increase.
The periodogram's standard deviation however doesn't depend on theses factors.

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 $\tau > 0$

$$\mathrm{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \qquad (7)$$

(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

Let's clarify the variance of our estimator. For $\tau > 0$:

$$\begin{aligned}
\mathrm{Var}(\hat{\gamma}(\tau)) &= \frac{1}{N^2}\mathrm{Var}\left(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right) \\
&= \frac{1}{N^2}\left(\mathbb{E}[(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2] - \mathbb{E}\left[\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right]^2\right) \\
&= \frac{1}{N^2}\left(\mathbb{E}[\left[\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} X_n X_{n+\tau} X_m X_{m+\tau}\right] - \left(\sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}]\right)^2\right) \\
&= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] - \left(\sum_{n=0}^{N-\tau-1} \gamma(\tau)\right)^2\right)
\end{aligned}$$

And since X is gaussian and stationary :

$$\begin{aligned}
\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] &= \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}] + \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_m X_{n+\tau}] \\
&= \gamma(\tau)^2 + \gamma(n-m)\gamma(n+\tau-m-\tau) + \gamma(n-m-\tau)\gamma(n+\tau-m) \\
&= \gamma(\tau)^2 + \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n+\tau-m)
\end{aligned}$$

We can simplify the variance because the term in $\gamma(\tau)^2$ and the second sum of the variance cancel each other. We have then :

$$\mathrm{V}(\hat{\gamma}(\tau)) = \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \left[\gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n+\tau-m)\right]\right)$$

Let's pose $k = n - m$

$$\begin{aligned}
&= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{k=n-(N-\tau-1)}^{n} \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right]\right) \\
&= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\left(\sum_{k=n-(N-\tau-1)}^{0} \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] + \sum_{k=1}^{n}\left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right]\right)
\end{aligned}$$

Let's inverse the summation order

$$
\begin{aligned}
&= \frac{1}{N^2}\left(\sum_{k=-(N-\tau-1)}^{0} \sum_{n=0}^{k+N-\tau-1} \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] + \sum_{k=1}^{N-\tau-1}\sum_{n=k}^{N-\tau-1} \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] \right) \\
&= \frac{1}{N^2}\left(\sum_{k=-(N-\tau-1)}^{0} (N-\tau+k) \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] + \sum_{k=1}^{N-\tau-1} (N-\tau-k) \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] \right) \\
&= \frac{1}{N^2}\left(\sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|) \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] \right) \\
&= \frac{1}{N}\left(\sum_{k=-(N-\tau-1)}^{N-\tau-1} (1-\frac{\tau-|k|}{N}) \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] \right)
\end{aligned}
$$

Let's prove now that the estimator is consistent.
We have by bounding the terms :

$$
\begin{aligned}
\operatorname{Var}(\hat{\gamma}(\tau)) &\leq \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left[\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)\right] \\
&\leq \frac{1}{N}\left(\sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma(k)^2 + \max_k |\gamma(k-\tau)| \sum_{k=-(N-\tau-1)}^{N-\tau-1} |\gamma(k+\tau)| \right)
\end{aligned}
$$

Since $\gamma(\tau)$ is absolutely summable and square summable, the limits of the two sums when $N \to \infty$ are finished.
Thus $\lim_{N\to\infty} \operatorname{Var}(\hat{\gamma}(\tau)) = 0$ so according to Bienayme-Tchebytchev's inequality, it converges in probability to $\gamma(\tau)$.

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 $J(f) = (1/N)(A(f) + iB(f))$.

- Derive the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, \dots, f_{N/2}$ where $f_k = f_s k/N$.
- What is the distribution of the periodogram values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$.
- 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

- Mean and variance of A(f) and B(f) :

$$
\begin{aligned}
\mathbb{E}(A(f)) &= \mathbb{E}\left(\sum_{n=0}^{N-1} X_n \cos(-2\pi k f n/f_s)\right) \\
&= \sum_{n=0}^{N-1} \mathbb{E}(X_n)\mathbb{E}(\cos(-2\pi k f n/f_s)) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(B(f)) &= \mathbb{E}\left(\sum_{n=0}^{N-1} X_n \sin(-2\pi k f n/f_s)\right) \\
&= \sum_{n=0}^{N-1} \mathbb{E}(X_n)\mathbb{E}(\sin(-2\pi k f n/f_s)) \\
&= 0
\end{aligned}
$$

- For $k = 0$:

  $$\mathrm{Var}(A(f_0)) = \sum_{n=0}^{N-1} \mathrm{Var}[X_n] = N\sigma^2$$

  We have that $B(f_0) = 0$, so :

  $$\mathrm{Var}(B(f_0)) = 0$$

- For $k \neq 0$: We have that $\sum_{n=0}^{N-1} \cos(4\pi kn/N) = Re(\sum_{n=0}^{N-1} \exp(4\pi kn/N) = 0$, so :

  $$
  \begin{aligned}
  \mathrm{Var}(A(f_k)) &= \sum_{n=0}^{N-1} \mathrm{Var}[X_n]\cos^2(-2\pi f_k n/f_s) \\
  &= \frac{\sigma^2}{2}\sum_{n=0}^{N-1}(\cos(4\pi kn/N)+1) \\
  &= N\frac{\sigma^2}{2}
  \end{aligned}
  $$

  $$
  \begin{aligned}
  \mathrm{Var}(B(f_k)) &= \sum_{n=0}^{N-1} \mathrm{Var}[X_n]\sin^2(-2\pi f_k n/f_s) \\
  &= \frac{\sigma^2}{2}\sum_{n=0}^{N-1}(1-\cos(4\pi kn/N)) \\
  &= N\frac{\sigma^2}{2}
  \end{aligned}
  $$

  We have that $A(f_k), B(f_k) \sim \mathcal{N}(0, N\frac{\sigma^2}{2})$ because $A(f_k), B(f_k)$ are linear combinations of

Gaussian independent variables.

$$
\begin{aligned}
\mathrm{Cov}(A(f_k), B(f_k)) &= \frac{\sigma^2}{2}\sum_{n=0}^{N-1}\cos(-2\pi f_k n/f_s)\sin(-2\pi f_k n/f_s) \\
&= \frac{\sigma^2}{2}\sum_{n=0}^{N-1}\sin(-4\pi k n/N) \\
&= \frac{\sigma^2}{2}\mathrm{Im}(\sum_{n=0}^{N-1}\exp(4\pi k n/N) \\
&= 0
\end{aligned}
$$

- Distribution of the periodogram values and variance

  We have that $Var(\mathcal{X}^2(n)) = 2n$.

  - For $k = 0$:

$$\frac{1}{\sigma^2}|J(f_0)|^2 = \frac{1}{N\sigma^2}(A(f_0))^2 \sim \mathcal{X}^2(1)$$

$$|J(f_0)|^2 \sim \sigma^2\mathcal{X}^2(1)$$

$$Var(|J(f_0)|^2) = 2\sigma^4$$

  - For $k \neq 0$:

$$\frac{2}{\sigma^2}|J(f_k)|^2 = \frac{2}{\sigma^2}(A(f_k)^2 + B(f_k)^2) \sim \mathcal{X}^2(2)$$

$$|J(f_k)|^2 \sim \frac{\sigma^2}{2}\mathcal{X}^2(2)$$

$$Var(|J(f_k)|^2) = \sigma^4$$

  Since the variance of $Var(|J(f_k)|^2)$ is constant, the periodogram is not consistent.

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

Let k and l be such that $k \neq l$

$$\begin{aligned}
\mathbb{E}(|J(f_k)|^2|J(f_l)|^2) &= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\mathbb{E}(X_nX_mX_pX_q)\exp\left(\frac{2i\pi(k(n-m)+l(p-q))}{N}\right) \\
&= \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\Big(\mathbb{E}(X_nX_m)\mathbb{E}(X_pX_q)+\mathbb{E}(X_nX_p)\mathbb{E}(X_mX_q) \\
&\quad + \mathbb{E}(X_nX_q)\mathbb{E}(X_mX_p)\Big)\exp\left(\frac{2i\pi(k(n-m)+l(p-q))}{N}\right) \\
&= \sigma^4 + \frac{\sigma^4}{N^2}\left(\sum_{n=0}^{N-1}\exp\left(\frac{-2i\pi(n(k+l)}{N}\right)\times\exp\left(\frac{-2i\pi(-m(k+l)}{N}\right)\times\right. \\
&\quad \left.\exp\left(\frac{-2i\pi(n(k-l)}{N}\right)\times\exp\left(\frac{-2i\pi(n(k-l)}{N}\right)\right) \\
&= \sigma^4 \qquad \text{because} \quad \forall k \in \mathbb{Z}, \exp\left(\frac{-2i\pi nk}{N}\right) = 0
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(|J(f_k)|^2, |J(f_l)|^2) &= \mathbb{E}(|J(f_k)|^2|J(f_l)|^2) - \mathbb{E}(|J(f_k)|^2)\mathbb{E}(|J(f_l)|^2) \\
&= \mathbb{E}(|J(f_k)|^2|J(f_l)|^2) - \sigma^4 \\
&= \sigma^4 - \sigma^4 \\
&= 0
\end{aligned}$$

Since the correlation is zero, this implies that for $k \neq l$, $|J(f_k)|^2$ and $|J(f_l)|^2$ are not correlated with each other. This explains the result obtained for question 6.

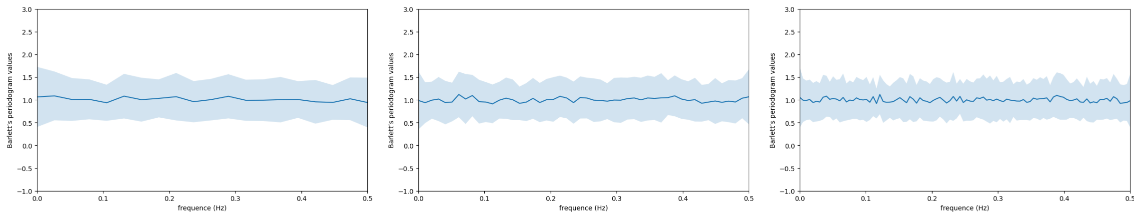
Question 9

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).

We can observe that the standart deviation has indeed been divided by $\sqrt{5}$.

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

Here we are looking for the optimal number of neighbors to classify the signals.

We begin by randomly mixing the initial training and test data. Then we reshuffle them according to the same initial proportions. *This manipulation enables us to obtain better performance in testing, compared with the initial distribution of the data.*

Applying 5-fold cross-validation on train set with DTW distance as metric and F1-Score for scoring, we obtained **3** as optimal number of neighbors with F1-Score equals to 86.88%.

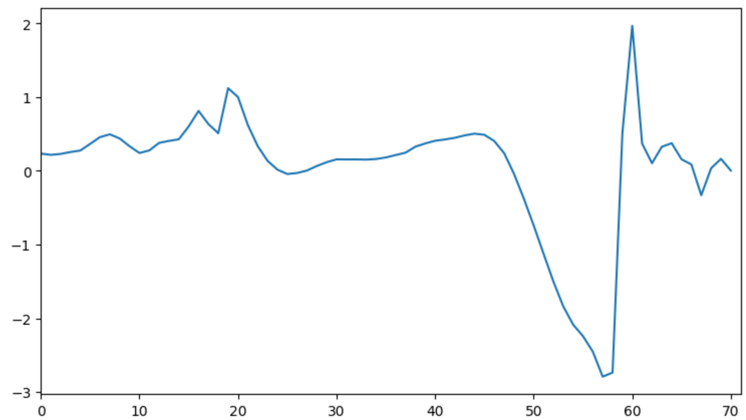
We then initialize a classifier with the optimal number of neighbors (in this case 3). We train the classifier on our new training data, then calculate the F1-Score in training and testing. We obtain 88.88% on the test data.

The prediction of a signal's class is based on the majority class of the k (here k=3) signals closest in the sense of the DTW distance.

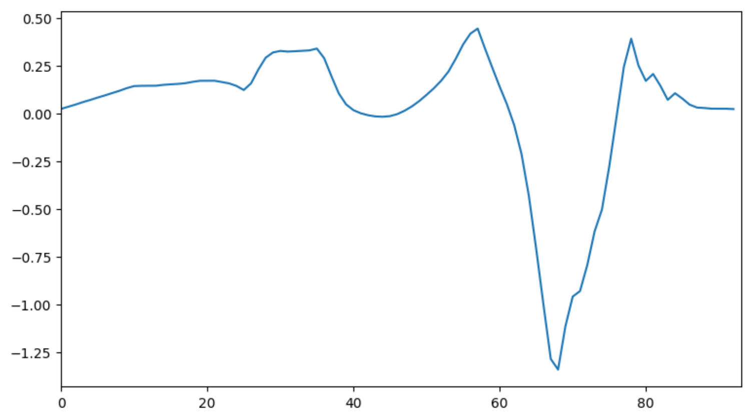
Question 11

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).

Based on the earlier graphs, it is evident that we can distinguish between the two signals, categorizing them as either *healthy* or *non-healthy*. Specifically, the healthy signal displays a more even and consistent pattern, whereas the non-healthy signal exhibits greater irregularities or disturbances. However, it's important to note that the fundamental shapes of these two signal types are quite similar, which could account for the misclassification.