Assignment 1 (ML for TS) - MVA 2022/2023

Clémence Grislain clemence.grislain@eleves.enpc.fr
Roman Plaud roman.plaud@eleves.enpc.fr

May 30, 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.

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

Answer 1

The minimizer of (1) verifies

$$X^T(y - X\beta) = \lambda u$$

with $u \in \partial L_1(\beta) = \{u : \forall i \in [1, p], u_i \in \{sign(\beta_i)\} \text{ if } \beta_i \neq 0 \text{ and } u_i \in [-1, 1] \text{ if } \beta_i = 0\}$

- If $\beta = 0_p$ is a minimizer of (1)

  $$\implies X^T y = \lambda u \implies ||X^T y||_\infty = \lambda ||u||_\infty$$

  And we have $u \in \partial L_1(\beta) \implies ||u||_\infty \leq 1$.

  $$\implies ||X^T y||_\infty \leq \lambda$$

- If $\lambda > ||X^T y||_\infty$

  $$\implies \max_i |(X^T y)_i| < \lambda \implies \forall i \in [1, p], \begin{cases} (X^T y)_i < \lambda \\ -\lambda < (X^T y)_i \end{cases} \implies \begin{cases} X^T y < \lambda \mathbb{1} \\ -\lambda \mathbb{1} < X^T y \end{cases}$$

Let β be a minimizer, β verifies $X^T y = \lambda u + X^T X\beta$

$$\implies \begin{cases} \lambda u + X^T X\beta < \lambda \mathbb{1} \\ -\lambda \mathbb{1} < \lambda u + X^T X\beta \end{cases} \implies \begin{cases} X^T X\beta < \lambda(\mathbb{1} - u) \\ -X^T X\beta < \lambda(\mathbb{1} + u) \end{cases}$$

And $X^T X \geq 0$

- if $\beta_i > 0 \implies u_i = 1 \implies (X^T X\beta)_i < \lambda(1 - u_i) = 0$
- if $\beta_i < 0 \implies u_i = -1 \implies -(X^T X\beta)_i < \lambda(1 + u_i) = 0$

$$\implies \beta_i \times (X^T X\beta)_i \begin{cases} < 0 \text{ if } \beta_i \neq 0 \\ = 0 \text{ if } \beta_i = 0 \end{cases}$$

Thus, if $\exists i : \beta_i \neq 0 \implies \|X\beta\|^2 = \sum_{i=1}^{p} \beta_i \times (X^T X\beta)_i < 0 \implies \perp$

Thus $\forall i \in [1, p], \beta_i = 0$, $\beta = 0_p$ is the only minimizer of (1).

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

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:

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

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 express this problem in the form of the previous question.

We have, $z_k * d_k \in \mathbb{R}^N, \forall j \in [1, N]$

$$(z_k * d_k)_j = \sum_{i=1}^{N} \tilde{z_{ki}} \times \tilde{d}_{ki}$$

with

- $\tilde{z_k} \in \mathbb{R}^N, \tilde{z_{ki}} = \begin{cases} z_{ki} \text{ if } 1 \leq i \leq N - L + 1 \\ 0 \text{ otherwise} \end{cases}$

- $\tilde{d}_k^{[j]} \in \mathbb{R}^N, \tilde{d_{ki}} = \begin{cases} d_{k(j-i)} \text{ if } 1 \leq j-i \leq N \\ 0 \text{ otherwise} \end{cases}$

$$\implies \forall j \in [1, N], (z_k * d_k)_j = \tilde{d}_k^{[j]T} \tilde{z}_k$$

$$\implies \forall k \in [1, K], z_k * d_k = \underbrace{\begin{pmatrix} \tilde{d}_k^{[1]T} \\ \vdots \\ \tilde{d}_k^{[N]T} \end{pmatrix}}_{\in \mathbb{R}^{N \times N}} \times \tilde{z}_k$$

$$\sum_{k=1}^{K} z_k * d_k = \underbrace{\begin{pmatrix} \tilde{d}_1^{[1]T} & \dots & \tilde{d}_K^{[1]T} \\ \vdots & \vdots & \vdots \\ \tilde{d}_1^{[N]T} & \dots & \tilde{d}_K^{[N]T} \end{pmatrix}}_{D \in \mathbb{R}^{N \times KN}} \times \underbrace{\begin{pmatrix} \tilde{z}_1 \\ \vdots \\ \tilde{z}_K \end{pmatrix}}_{Z \in \mathcal{R}^{KN}}$$

We can note that $\sum_{k=1}^{K} \|\mathbf{z}_k\|_1 = \lambda \|Z\|_1$

Thus we have for a fixed dictionary $(d_k)_{[1,K]}$, thus a fixed matrix D:

$$\min_{(\mathbf{z}_k)_k} \quad \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 = \min_{Z \in \mathbb{R}^{KN}} \|x - DZ\|_2^2 + \lambda \|Z\|_1$$

And from the previous result we obtain, $\forall \lambda > \lambda_{max}$, $Z = 0_{KN}$ is the minimizer, with

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

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.

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

A Gaussian white noise satisfies the two following properties :

- $\forall n \in [0, N-1],\ X_n \sim \mathcal{N}(0,\ \sigma^2)$
- $\forall (i,j) \in [0, N-1]^2,\ i \neq j \Rightarrow \mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = 0$

We can then compute γ :

- $\gamma(0) = \mathbb{E}[X_0^2] = \sigma^2$
- $\forall \tau \in [1, N-1],\ \gamma(\tau) = \mathbb{E}[X_0 X_\tau] = 0$

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 $\tau = 0, 1, \dots, N-1$ 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

$$\mathbb{E}[\hat{\gamma}(\tau)] = \frac{1}{N} \sum_{n=0}^{N-\tau-1} \underbrace{\mathbb{E}[X_n X_{n+\tau}]}_{=\gamma(\tau)} = \frac{N-\tau}{N}\gamma(\tau)$$

$$\mathbb{E}[\hat{\gamma}(\tau)] \neq \gamma(\tau) \Rightarrow \hat{\gamma}(\tau) \text{ is biased}$$

$$\frac{N-\tau}{N} \xrightarrow{N \to +\infty} 1 \Rightarrow \mathbb{E}[\hat{\gamma}(\tau)] \xrightarrow{N \to +\infty} \gamma(\tau) \Rightarrow \hat{\gamma}(\tau) \text{ is asymptotically unbiased}$$

A simple way to debias this estimator would be to consider

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

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

- $$\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\left(\sum_{n=0}^{N-1} X_n e^{\frac{2i\pi kn}{N}}\right)\left(\sum_{p=0}^{N-1} X_p e^{\frac{-2i\pi kp}{N}}\right) \\
&= \frac{1}{N}\sum_{n,p} X_n X_p e^{\frac{-2i\pi k(n-p)}{N}} \\
&= \frac{1}{N}\sum_{p=0}^{N-1}\sum_{h=-p}^{N-1-p} X_{p+h}X_p e^{\frac{-2i\pi kh}{N}} \\
&= \frac{1}{N}\sum_{h=-N+1}^{N-1}\sum_{p=0}^{N-1-k} X_{p+h}X_p e^{\frac{-2i\pi kh}{N}} \\
&= \sum_{h=-N+1}^{N-1} e^{\frac{-2i\pi kh}{N}} \underbrace{\frac{1}{N}\sum_{p=0}^{N-1-k} X_{p+h}X_p}_{=\hat{\gamma}(h)} \\
&= \sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{\frac{-2i\pi k\tau}{N}}
\end{aligned}$$

  3-th equation holds by performing the change of index $n = p + h$. 4-th equation holds by interveting summing domains (which correspond to a triangle in a 2D space).

- Let f>0, $\varepsilon > 0$
  - $\sum_{\tau\in\mathbb{Z}} |\gamma(\tau)| < +\infty$:
  $$\exists N_1 \in \mathbb{N},\ \forall N \geq N_1,\ \sum_{|\tau|>N-1} |\gamma(\tau)| < \frac{\varepsilon}{2}$$
  - Let's consider $g_\tau(f) = \gamma(\tau) e^{-\frac{-2i\pi f\tau}{f_s}}$
  $$\forall f,\ |g_\tau(f)| \leq |\gamma(\tau)|$$

And

$$\sum_{\tau \in \mathbb{N}} |\gamma(\tau)| < +\infty$$

Thus, $\sum_{\tau \in \mathbb{N}} g_\tau$ is normally convergent therefore uniformly convergent. We have moreover that g_τ in continuous at point f for every τ. Then a classic continuity theorem allow us to say that $S = \sum_{\tau \in \mathbb{Z}} g_\tau$ is continuous at point f. And we also have $f^{(N)} \xrightarrow{N \to +\infty} f$ such that by continuity of S :

$$\exists N_2 \in \mathbb{N},\ \forall N \geq N_2,\ |S(f^{(N)}) - S(f)| \leq \frac{\varepsilon}{2}$$

Let's consider $n = max(N_1, N_2)$. Let $N \geq n$. Then :

$$\begin{aligned}
\left| \mathbb{E}\left[\sum_{\tau=-N+1}^{N-1} \hat{\gamma}(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} \right] - S(f) \right| &= \left| \sum_{\tau=-N+1}^{N-1} \mathbb{E}\left[\hat{\gamma}(\tau)\right] e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} - S(f) \right| \\
&= \left| \sum_{\tau=-N+1}^{N-1} \gamma(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} - S(f^{(N)}) + S(f^{(N)}) - S(f) \right| \\
&\leq \left| \sum_{\tau=-N+1}^{N-1} \gamma(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} - S(f^{(N)}) \right| + \left| S(f^{(N)}) - S(f) \right| \\
&= \left| \sum_{|\tau| > N-1} \gamma(\tau) e^{-\frac{2i\pi f^{(N)}\tau}{f_s}} \right| + \frac{\varepsilon}{2} \\
&\leq \sum_{|\tau| > N-1} |\gamma(\tau)| + \frac{\varepsilon}{2} \\
&\leq \varepsilon
\end{aligned}$$

Thus we have :

$$\mathbb{E}\left[\left| J(f^{(N)}) \right|^2 \right] \xrightarrow{N \to +\infty} = S(f)$$

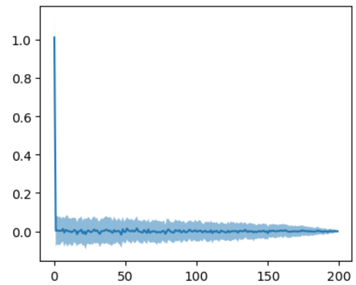
. We can thus conclude that $\left| J(f^{(N)}) \right|^2$ is an is an asymptotically unbiased estimator of $S(f)$ for $f > 0$

Question 6

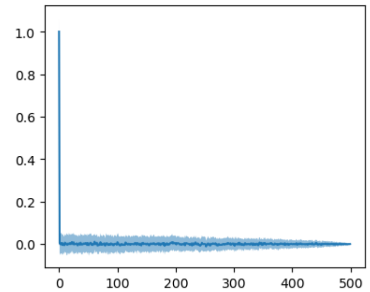
In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

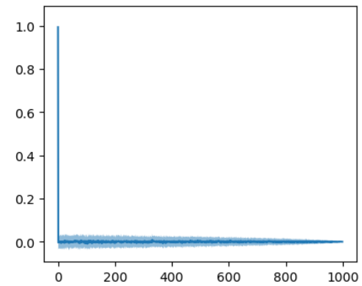
Add your plots to Figure **??**.



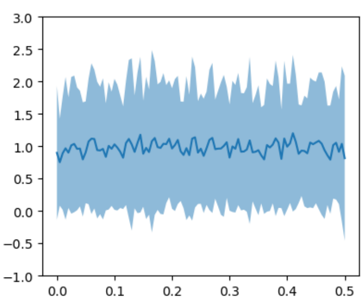
Autocovariance ($N = 200$)



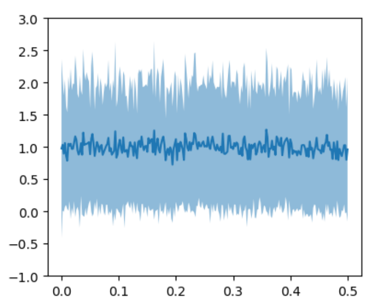
Autocovariance ($N = 500$)



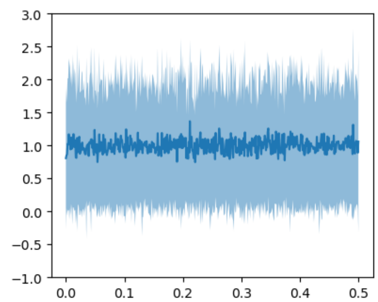
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

- The value of the autocovariance for $\gamma = 0$ is constant and the variance of the estimator $\hat{\gamma}(\tau)$ decreases as the number N of samples increases. This indicates that the estimator is consistent, like proven in the next question.
- The variance of $|J(f_k)|^2$ is constant over the different number of samples N (this observation fits with the results of question 8).

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 $\tau > 0$

$$\mathrm{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{8}$$

  (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

$$\mathrm{var}(\hat{\gamma}(\tau)) = \frac{1}{N^2}(\mathrm{var}(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})$$

$$= \frac{1}{N^2}\left(\mathbb{E}(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \sum_{m=0}^{N-\tau-1} X_m X_{m+\tau}) + \mathbb{E}(\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2\right)$$

$$= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau} X_m X_{m+\tau}) + (\sum_{n=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau}))^2\right)$$

$$= \frac{1}{N^2}\left(\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau} X_m X_{m+\tau}) + (\sum_{n=0}^{N-\tau-1} \gamma(\tau))^2\right)$$

And we have:

$\mathbb{E}(X_n X_{n+\tau} X_m X_{m+\tau})$
$= \mathbb{E}(X_n X_{n+\tau})\mathbb{E}(X_m X_{m+\tau}) + \mathbb{E}(X_n X_m)\mathbb{E}(X_{n+\tau} X_{m+\tau}) + \mathbb{E}(X_n X_{m+\tau})\mathbb{E}(X_m X_{n+\tau})$
$= \gamma(\tau)^2 + \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau)$

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

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

with $t = n - m$

$$= \frac{1}{N^2}\left(\sum_{t=-(N-\tau-1)}^{N-\tau-1}\sum_{m=0}^{N-\tau-|t|}(\gamma(t)^2 + \gamma(t-\tau)\gamma(t+\tau))\right)$$

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

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

- $\sum_{t=-\infty}^{+\infty}(\gamma(t)^2 + \gamma(t-\tau)\gamma(t+\tau)) = 2\sum_{n=0}^{+\infty}(\gamma(n)^2 + \gamma(t-\tau)\gamma(t+\tau)) \underset{+\infty}{\sim} 4\sum_{n=0}^{+\infty}\gamma(n)^2 < +\infty$ as γ is an even function.
- $\forall N,\ \forall n \in [0, N-1],\ (1 - \frac{n}{N}) \leq 1$

$$\implies 0 \leq \mathrm{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N}\left(\sum_{t=-(N-\tau-1)}^{N-\tau-1}(\gamma(t)^2 + \gamma(t-\tau)\gamma(t+\tau))\right)$$

$$\implies 0 \leq \mathrm{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N}\left(\sum_{t=-\infty}^{+\infty}(\gamma(t)^2 + \gamma(t-\tau)\gamma(t+\tau))\right)$$

Thus, $\mathrm{var}(\hat{\gamma}(\tau)) \xrightarrow{N\to+\infty} 0$.
By the Bienaymé-Tchebychev inequality,

$$0 \le P(|\hat{\gamma}(\tau) - \gamma(\tau)| > \alpha) \le \frac{\mathrm{var}(\hat{\gamma}(\tau))}{\alpha^2}$$

(as$\mathbb{E}(\hat{\gamma}(\tau)) = \gamma(\tau)$) which implies $P(|\hat{\gamma}(\tau) - \gamma(\tau)| > \alpha) \xrightarrow{N\to+\infty} 0$ so the estimator $\hat{\gamma}(\tau)$ is consistent.

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 $|J(f_0)|^2$, $|J(f_1)|^2$, ..., $|J(f_{N/2})|^2$.
- 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 **??** by looking at the covariance between the $|J(f_k)|^2$.

Answer 8

- For $f_k = f_s \frac{k}{N}$
  - $\mathbb{E}(A(f_k)) = \mathbb{E}(\sum_{n=0}^{N-1} X_n \cos(-2\pi f_k n/f_s)) = \sum_{n=0}^{N-1} \cos(-2\pi f_k n/f_s)\mathbb{E}(X_n) = 0$
  - $\mathbb{E}(B(f_k)) = \mathbb{E}(\sum_{n=0}^{N-1} X_n \sin(-2\pi f_k n/f_s)) = \sum_{n=0}^{N-1} \sin(-2\pi f_k n/f_s)\mathbb{E}(X_n) = 0$
  - $\mathrm{var}(A(f_k)) = \mathbb{E}(A(f_k)^2) - \mathbb{E}(A(f_k))^2$

    $= \mathbb{E}(\sum_{n=0}^{N-1} X_n \cos(-2\pi k \times n/N) \sum_{m=0}^{N-1} X_m \cos(-2\pi k \times n/N))$

    $= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}(X_m X_n) \cos(-2\pi k \times n/N) \cos(-2\pi k \times n/N)$

    and $\mathbb{E}(X_n X_m) = \begin{cases} \sigma^2 \text{ if } n = m \\ 0 \text{ otherwise} \end{cases}$

    $$\mathrm{var}(A(f_k)) = \sum_{n=0}^{N-1} \sigma^2 \cos(-2\pi k \times n/N)^2 = \sum_{n=0}^{N-1} \sigma^2 \frac{1 + \cos(-4\pi k \times n/N)}{2}$$

  - Similarly,

    $$\mathrm{var}(B(f_k)) = \sum_{n=0}^{N-1} \sigma^2 \sin(-2\pi k \times n/N)^2 = \sum_{n=0}^{N-1} \sigma^2 \frac{1 - \cos(-4\pi k \times n/N)}{2}$$

    By noticing that $\forall k \neq 0$:

$\sum_{n=0}^{N-1} \cos(-4\pi k \times n/N)$

$= Re(\sum_{n=0}^{N-1} exp(-i \times 4\pi k \times n/N))$

$= Re(\frac{1-exp(-i\times 4\pi k\times n)}{1-exp(-i\times 4\pi k\times n/N)})$

$= 0$

$$\forall k \neq 0 \quad \text{var}(A(f_k)) = \text{var}(B(f_k)) = \sigma^2 \frac{N}{2}$$

$$k = 0 \quad \text{var}(A(f_k)) = \sigma^2 N, \quad \text{var}(B(f_k)) = 0$$

- $|J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2)$
  - For $k \neq 0$:

    $cov(A(f_k), B(f_k)) = cov(\sum_{n=0}^{N-1} X_n \cos(-2\pi k \times n/N), \sum_{m=0}^{N-1} X_m \sin(-2\pi k \times m/N))$

    $= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \cos(-2\pi k \times n/N) \sin(-2\pi k \times m/N) cov(X_n, X_m)$

    $= \sigma^2 \sum_{n=0}^{N-1} \cos(-2\pi k \times n/N) \sin(-2\pi k \times n/N)$

    $= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} \sin(-4\pi k \times n/N)$

    $= \frac{\sigma^2}{2} Im(\sum_{n=0}^{N-1} \exp(-4\pi k \times n/N))$

    $= 0$

    $A(f_k)$ and $B(f_k)$ are Gaussian vectors $\implies A(f_k) \perp\!\!\!\perp B(f_k)$
  - $A(f_k)$ and $B(f_k)$ are two independent Gaussian variables with same mean and variance

    $$\implies |J(f_k)|^2 \sim \chi^2(2)$$
  - For $k = 0$:
  - $B(f_0) = 0 \implies |J(f_0)|^2 = \frac{1}{N} A(f_0)^2$
  - $A(f_0)$ is a Gaussian vector

    $$\implies |J(f_0)|^2 \sim \chi^2(1)$$

- $S = X_1^2 + X_2^2 \sim \chi^2(2)$ with variance equals to 4 ($2 \times k$ with k the degree of freedom of the Chi-square law) if X_1 and X_2 are independent Gaussian variables of zero mean and variance 1.

  For $k \neq 0$

  $$\implies \text{var}\left(\left(\sqrt{\frac{2}{N}}\frac{1}{\sigma}A(f_k)\right)^2 + \left(\sqrt{\frac{2}{N}}\frac{1}{\sigma}B(f_k)\right)^2\right) = 4$$

  $$\implies \text{var}\left(\frac{2}{N}\frac{1}{\sigma^2}\left(A(f_k)^2 + B(f_k)^2\right)\right) = 4$$

$$\implies \frac{4}{\sigma^4}\text{var}\left(\frac{1}{N}\left(A(f_k)^2 + B(f_k)^2\right)\right) = 4$$

$$\implies \text{var}\left(|J(f_k)|^2\right) = \sigma^4$$

For $k = 0$

$$\implies \text{var}\left(\left(\sqrt{\frac{1}{N}}\frac{1}{\sigma}A(f_k)\right)^2\right) = 2$$

$$\implies \text{var}\left(\frac{1}{N}\frac{1}{\sigma^2}\left(A(f_k)^2\right)\right) = 2$$

$$\implies \frac{1}{\sigma^4}\text{var}\left(\frac{1}{N}\left(A(f_k)^2\right)\right) = 2$$

$$\implies \text{var}\left(|J(f_k)|^2\right) = 2 \times \sigma^4$$

Thus the variance cannot converge to 0 when $N \to +\infty$ and thus $|J(f_k)|^2$ is not consistent.

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 **??**, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe.

Add your plots to Figure **??**.

Answer 9

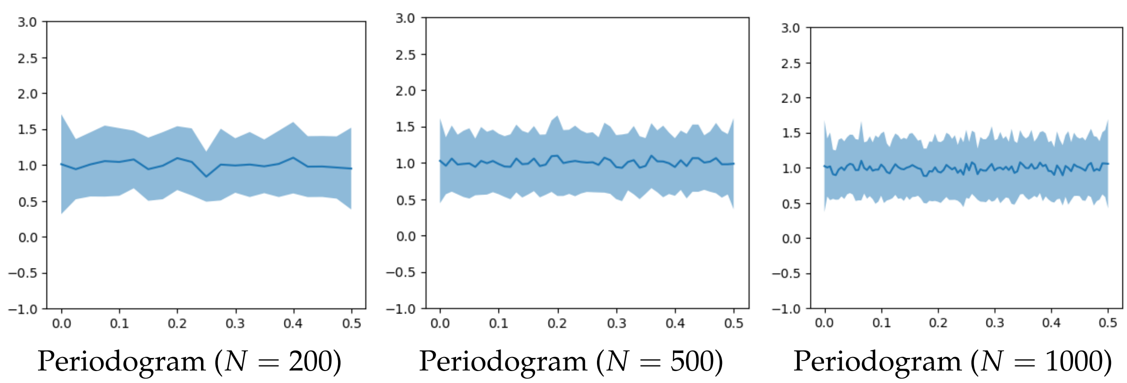


Figure 2: Barlett's periodograms of a Gaussian white noise (see Question **??**).

The variance is smaller that in the previous experiment and it seems to decrease as the number N of samples increases which may indicate that the Barlett's estimate is indeed consistent.

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.

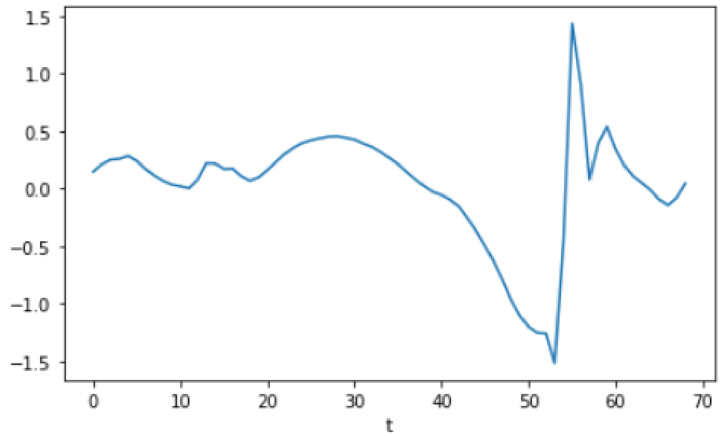
Answer 10

The best number of neighbors is $k = 1$ which achieves a f1 score of 0.61 on the validation sets (mean over the 5-folds) and 0.51 of f1 score on the test set (fitted KNN on the entire training set). This indicates that the model is not making accurate predictions, as it is only slightly better than a random predictor which achieves f1 score of 0.5.

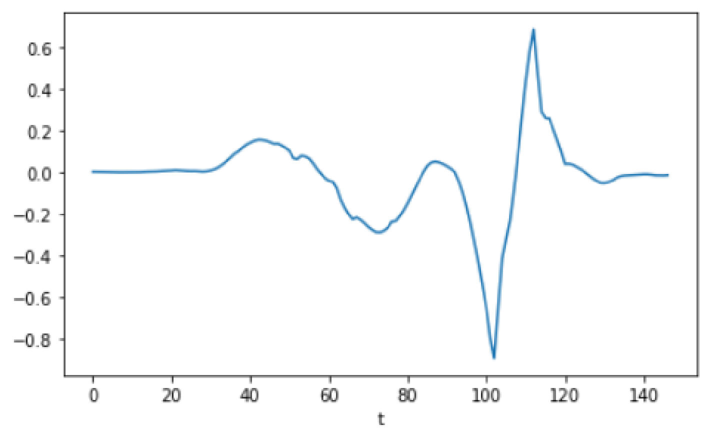
Question 11

Display on Figure **??** 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 **??**).