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

Boubacar Sow boubacar.sow@ens-paris-saclay.fr
Theïlo Terrisse theilo.terrisse@eleves.enpc.fr

November 7, 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 $\lambda > 0$.

For all $\beta \in \mathbb{R}^p$, let $f(\beta) = \frac{1}{2}\|y - X\beta\|_2^2$ and $g(\beta) = \|\beta\|_1$, so that the objective function writes:

$$\mathcal{L}(\beta) = f(\beta) + \lambda g(\beta). \tag{2}$$

f is differentiable on $\mathbb{R}^p$ with gradient $\nabla f(\beta) = X^T(X\beta - y)$.

However, g is not differentiable in 0, but we can consider its sub-differential:

$$\forall \beta \in \mathbb{R}^p, \partial g(\beta) = \left\{ \alpha \,\middle|\, \forall i \in \{1, \dots, p\},\ \alpha_i \in \begin{cases} \{1\} & \text{if } \beta_i > 0 \\ [-1, 1] & \text{if } \beta_i = 0 \\ \{-1\} & \text{if } \beta_i < 0 \end{cases} \right\} \tag{3}$$

Now, it is easy to check that f is convex, since its Hessian $X^T X$ is positive semi-definite on $\mathbb{R}^p$. Therefore, $\mathcal{L}$ is convex and its sub-differential is :

$$\forall \beta \in \mathbb{R}^p,\ \partial\mathcal{L}(\beta) = \left\{ X^T(X\beta - y) + \lambda\alpha \,\middle|\, \forall i \in \{1, \dots, p\},\ \alpha_i \begin{cases} = 1 & \text{if } \beta_i > 0 \\ \in [-1, 1] & \text{if } \beta_i = 0 \\ = -1 & \text{if } \beta_i < 0 \end{cases} \right\} \tag{4}$$

Then, by Fermat's rule:

$$\begin{aligned} 0 \in \arg\min_{\beta \in \mathbb{R}^p} \mathcal{L}(\beta) &\iff 0 \in \partial\mathcal{L}(0) \\ &\iff 0 \in \left\{ -X^T y + \lambda\alpha \,\middle|\, \forall i \in \{1, \dots, p\},\ \alpha_i \in [-1, 1] \right\} \\ &\iff \forall i \in \{1, \dots, p\},\ \exists \alpha_i \in [-1, 1],\ \lambda\alpha_i = (X^T y)_i \end{aligned}$$

Then, let $\lambda_{\max} = \|X^T y\|_\infty$. For all $\lambda > \lambda_{\max}$, $\forall i \in \{1, \dots, p\}$:

$$\lambda \geq |X^T y|_i \geq -(X^T y)_i \quad \Rightarrow \quad \exists \alpha_i \in [-1, 1],\ \lambda\alpha_i = (X^T y)_i. \tag{5}$$

Therefore, for all $\lambda > \lambda_{\max}$, $0 \in \arg\min_{\beta \in \mathbb{R}^p} \mathcal{L}(\beta)$, that is, 0 is a minimizer of (1).

$$\boxed{\lambda_{\max} = \left\|X^T y\right\|_\infty} \tag{6}$$

Remark: To get the unicity of $0 \in \mathbb{R}^p$ as *the* minimizer of (1), we can give another proof with a larger $\lambda_{\max}$ as follows:

Let's look for $\lambda > 0$ such that $\forall \beta \in \mathbb{R}^p,\ \beta \neq 0,\quad \frac{1}{2}\|y - X\beta\|_2^2 \quad + \quad \lambda\|\beta\|_1 \quad > \quad \frac{1}{2}\|y\|_2^2$.

For all $\beta \in \mathbb{R}^p,\ \beta \neq 0$,

$$\lambda > |y^T X \frac{\beta}{\|\beta\|_1}|$$

$$
\begin{aligned}
&\Rightarrow \ -y^T X\beta + \lambda \|\beta\|_1 > 0 \\
&\Rightarrow \ -y^T X\beta + \frac{1}{2}\|X\beta\|_2^2 + \lambda \|\beta\|_1 > 0 \\
\iff & \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \|\beta\|_1 > \frac{1}{2}\|y\|_2^2
\end{aligned}
$$

Now, $\forall j \in \mathbb{R}$, $\frac{\beta_j}{\|\beta\|_1} \leq 1$.

So, $|y^T X \frac{\beta}{\|\beta\|_1}| \quad \leq \quad \sum_{i,j} |X_{i,j} y_i \frac{\beta_j}{\|\beta\|_1}| \quad \leq \quad \sum_{i,j} |X_{i,j} y_i| \quad = \quad \left\|X^T y\right\|_1$.

Therefore, for $\lambda > \left\|X^T y\right\|_1$, $\quad \frac{1}{2}\|y - X\beta\|_2^2 \quad + \quad \lambda \|\beta\|_1 \quad > \quad \frac{1}{2}\|y\|_2^2$.

So, a $\lambda_{\max}$ such that for all $\lambda > \lambda_{\max}$, *the* minimizer of (1) is 0, is given by:

$$
\boxed{\lambda_{\max} = \left\|X^T y\right\|_1} \tag{7}
$$

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:

$$
\min_{(\mathbf{d}_k)_k, (\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\|\mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k\right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{8}
$$

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 $(\mathbf{d}_k)_{k \in \{1,\ldots,K\}}$ be given such that $\forall k$, $\|\mathbf{d}_k\|_2^2 \leq 1$.

Let $(\mathbf{z}_k)_k \in (\mathbb{R}^{n-L+1})^K$. Let us introduce some notations:

- $\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \\ \vdots \\ \mathbf{z}_K \end{pmatrix} \in \mathbb{R}^{K(n-L+1)}$

- To find the design matrix for the problem, we want to find a matrix associated to the sum of the convolutions $\mathbf{z}_k * \mathbf{d}_k$, $\forall k \in \{1, \dots, K\}$. Remembering the interpretation of the convolution operation as sliding the inverted $\mathbf{d}_k$ along $\mathbf{z}_k$, we may introduce:

$$D = \left(D_1 \middle| \cdots \middle| D_k \middle| \cdots \middle| D_K\right) \in \mathbb{R}^{n \times K(n-L+1)} \tag{9}$$

where for all $k \in \{1, \dots, K\}$,

$$D_k = \begin{pmatrix} \boxed{\mathbf{d}_k} & \boxed{\mathbf{d}_k} & \boxed{\mathbf{d}_k} & \ddots & \boxed{\mathbf{d}_k} \end{pmatrix}, \tag{10}$$

that is, $\forall k \in \{1, \dots, K\}$, $\forall i \in \{1, \dots, n\}$,

$$(D_k)_i = \left(\tilde{\mathbf{d}}_k(i) \middle| \tilde{\mathbf{d}}_k(i-1) \middle| \cdots \middle| \tilde{\mathbf{d}}_k(i-j) \middle| \cdots \middle| \tilde{\mathbf{d}}_k(i-(n-L+1))\right) \tag{11}$$

where $\tilde{\mathbf{d}}_k$ is the vector $\mathbf{d}_k$ padded with zeros for indices where it is not defined.

With these definitions in place, we claim that equation (8) amounts to a LASSO regression problem with variable $\mathbf{Z}$, response vector $\mathbf{x}$ and design matrix D. Indeed:

- $\|\mathbf{Z}\|_1 = \sum_{l=1}^{K(n-L+1)} |\mathbf{Z}_l| = \sum_{k=1}^{K} \sum_{i=1}^{n-L+1} |\mathbf{Z}_{(k-1)(n-L+1)+i}| = \sum_{k=1}^{K} \left(\sum_{i=1}^{n-L+1} |\mathbf{z}_k(i)| \right) = \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 .$

- and $D\mathbf{Z} = \sum_{k=1}^{K} D_k \mathbf{z}_k$. Now, for all $k \in \{1, \dots, K\}$,

$$\forall i \in \{1, \dots, n\},\ (D_k\,\mathbf{z}_k)_i = \sum_{j=1}^{n-L+1} (D_k)_{i,j}\,\mathbf{z}_k(j) = \sum_{j=1}^{n-L+1} \tilde{\mathbf{d}}_k(i-j)\mathbf{z}_k(j) = \sum_{j=-\infty}^{\infty} \tilde{\mathbf{z}}_k(j)\tilde{\mathbf{d}}_k(i-j)$$

where $\tilde{\mathbf{z}}_k$ is the vector $\mathbf{z}_k$ padded with zeros for indices where it is not defined. So, $\forall\, i \in \{1, \dots, n\}$, $(D_k\,\mathbf{z}_k)_i = (\mathbf{z}_k * \mathbf{d}_k)_i$, therefore $D_k\,\mathbf{z}_k = \mathbf{z}_k * \mathbf{d}_k$.

Hence,

$$D\mathbf{Z} = \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k. \tag{12}$$

Assembling these results, and knowing that $(\mathbf{d}_k)_k$ is given, problem (8) rewrites:

$$\min_{\mathbf{Z} \in \mathbb{R}^{K(n-L+1)}} \quad \|\mathbf{x} - D\mathbf{Z}\|_2^2 \quad + \quad \lambda\,\|\mathbf{Z}\|_1 . \tag{13}$$

Last, using the result of Question 1, we know that there exists $\lambda_{\max}$ such that $\mathbf{Z}$ (and thus the sparse codes) is 0 for any $\lambda > \lambda_{\max}$, given by:

$$\boxed{\lambda_{\max} = \left\| D^T \mathbf{x} \right\|_1} \tag{14}$$

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \tag{15}$$

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

Gaussian white noise is a random process where each value in the sequence is independant and identically distributed, and each sample follows a normal distribution with mean 0 and variance σ^2. This means that there is no correlation between different time points:

$$\forall \tau \neq 0, \quad \gamma(\tau) = \mathbb{E}(X_n X_{n+\tau}) = \mathbb{E}[X_n]\mathbb{E}[X_{n+\tau}] = 0$$

In the case where $\tau = 0$, $\gamma(0)$ is the variance of the white noise process because it is the covariance of the variable with itself at the same time point.

$$\gamma(0) = \mathbb{E}[X_n^2] = Var(X_n) = \sigma^2$$

Finally we have:

$$\gamma(\tau) = \begin{cases} 0 & \text{if } \tau \neq 0 \\ \sigma^2 & \text{if } \tau = 0 \end{cases} \tag{16}$$

Therefore, the power spectrum associated to the Gaussian white noise is a constant:

$$S(f) = \sigma^2 \tag{17}$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{18}$$

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

$$Bias(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau)$$

Let's show that the estimator $\hat{\gamma}(\tau))$ is biased
We must distinguish 3 cases:

- If $\tau > 0$, we have:

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\tau)] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}\right] \\
&= \frac{1}{N}\sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}] \\
\mathbb{E}[\hat{\gamma}(\tau)] &= \frac{N-\tau}{N}\gamma(\tau)
\end{aligned} \tag{19}$$

$$\begin{aligned}
\Rightarrow Bias(\hat{\gamma}(\tau)) &= \frac{N-\tau}{N}\gamma(\tau) - \gamma(\tau) \\
&\neq 0 \qquad (\text{if } \gamma(\tau) \neq 0)
\end{aligned} \tag{20}$$

- If $\tau < 0$, since $\hat{\gamma}(\tau) = \hat{\gamma}(-\tau)$, we have:

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\tau)] &= \mathbb{E}[\hat{\gamma}(-\tau)] \\
&= \frac{N+\tau}{N}\gamma(-\tau) \\
&= \frac{N+\tau}{N}\gamma(\tau) \quad (\text{since } \tau \text{ is even})
\end{aligned} \tag{21}$$

$$\begin{aligned}
\Rightarrow Bias(\hat{\gamma}(\tau)) &= \frac{N+\tau}{N}\gamma(\tau) - \gamma(\tau) \\
&\neq 0 \qquad (\text{if } \gamma(\tau) \neq 0)
\end{aligned} \tag{22}$$

- If $\tau = 0$, we have:

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\tau)] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} X_n X_{n+\tau}\right] \\
&= \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}[X_n X_n] = \frac{1}{N}\sum_{n=0}^{N-1} \sigma^2 = \frac{N}{N}\sigma^2 \\
\mathbb{E}[\hat{\gamma}(\tau)] &= \sigma^2
\end{aligned} \tag{23}$$

$$
\begin{aligned}
\Rightarrow Bias(\hat{\gamma}(\tau)) &= \sigma^2 - \sigma^2 \\
&= 0
\end{aligned} \tag{24}
$$

Since $Bias(\hat{\gamma}(\tau)) \neq 0 \; \forall \tau \neq 0$, we conclude that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$ in the general case.

Let's show that the estimator $\hat{\gamma}(\tau))$ is asymptotically unbiased:
When $\tau \to \infty$

$$
\begin{aligned}
\lim_{N\to\infty} \frac{N-\tau}{N}\gamma(\tau) &= \gamma(\tau) \\
\Rightarrow Bias(\hat{\gamma}(\tau)) &= \gamma(\tau) - \gamma(\tau) \\
&= 0
\end{aligned} \tag{25}
$$

We conclude that the estimator is asymptotically unbiased.

Debiasing the estimator: The estimator $\hat{\gamma}(\tau)$ for the autocovariance function is biased because it divides by N instead of the actual number of terms in the sum, which is $N-\tau$. This bias can be corrected by dividing by $N-\tau$ instead of N. The unbiased estimator of the autocovariance is then given by:

$$
\hat{\gamma}'(\tau) = \frac{1}{N-\tau} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{26}
$$

Question 5

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

$$
J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{27}
$$

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

Writing $|J(f_k)|^2$ as a function of the sample autocovariances

$$
J(f) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n \exp^{-2\pi i f n/f_s}
$$

$$\Rightarrow |J(f_k)| = \frac{1}{N} | \sum_{n=0}^{N-1} X_n cos\,(2\pi kn/N) - i \sum_{n=0}^{N-1} X_n sin\,(2\pi kn/N)\,|$$

$$\Rightarrow |J(f_k)|^2 = \frac{1}{N}\left[(\sum_{n=0}^{N-1} X_n cos\,(2\pi kn/N))^2 + (\sum_{n=0}^{N-1} X_n sin\,(2\pi kn/N))^2 \right]$$

$$= \frac{1}{N}\left[\sum_{n=0}^{N-1} X_n^2 cos\left(\frac{2\pi kn}{N}\right)^2 + 2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n cos\left(\frac{2\pi kn}{N}\right) X_m cos\left(\frac{2\pi km}{N}\right) \right.$$

$$\left. + \sum_{n=0}^{N-1} X_n^2 sin\left(\frac{2\pi kn}{N}\right)^2 + 2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n sin\left(\frac{2\pi kn}{N}\right) X_m sin\left(\frac{2\pi km}{N}\right) \right]$$

$$= \frac{1}{N}\left[\sum_{n=0}^{N-1} X_n^2 + 2\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n X_m \left(cos\left(\frac{2\pi km}{N}\right) cos\left(\frac{2\pi km}{N}\right) + sin\left(\frac{2\pi km}{N}\right) sin\left(\frac{2\pi km}{N}\right)\right)\right]$$

$$= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{m=n+1}^{N-1} X_n X_m cos\left(\frac{2\pi k(m-n)}{N}\right)$$

Setting $\tau = m - n$, we have:

$$= \hat{\gamma}(0) + \frac{2}{N}\sum_{n=0}^{N-1}\sum_{\tau=1}^{N-n-1} X_n X_{n+\tau} \cos\left(\frac{2\pi k\tau}{N}\right)$$

Here, the outer sum is over n from 0 to $N-1$, representing all time points in our data. For each time point n, we then sum over τ from 1 to $N-n-1$. This represents all possible lags that we can calculate from time point n. We then calculate the product $X_n X_{n+\tau}$, which is the τ-lagged product of our data at time point n. This product is then weighted by a cosine function that depends on τ, and all these weighted products are summed up. Inverting the sums on n and τ, extracting the cos from the second sum and inserting $\frac{1}{N}$ into the first sum, we get:

$$\begin{aligned} |J(f_k)|^2 &= \hat{\gamma}(0) + 2\sum_{\tau=1}^{N-1}\frac{1}{N}\left(\sum_{n=1}^{N-\tau-1} X_n X_{n+\tau}\right)\cos\left(\frac{2\pi k\tau}{N}\right) \\ &= \hat{\gamma}(0) + 2\sum_{\tau=1}^{N-1}\hat{\gamma}(\tau)\cos\left(\frac{2\pi k\tau}{N}\right) \\ &= \sum_{\tau=-N+1}^{N-1} \exp^{-2i\pi k\frac{\tau}{N}}\hat{\gamma}(\tau) \end{aligned} \tag{28}$$

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

We start by defining the Fourier frequencies f_k.

$f_k = \frac{kf_s}{N}$ for $k = 0, 1, \ldots, N-1$

We then define $f^{(N)}$ as the Fourier frequency that is closest to a given frequency f.

$$f_k = \frac{kf_s}{N} \quad \text{for } k = 0, 1, \ldots, N-1$$
$$f^{(N)} = \arg\min_{f_k} |f - f_k|$$

The maximum difference between f and $f^{(N)} = \frac{f_s}{2N}$

Let $f_0 = \frac{f_s}{2} \Rightarrow \frac{f_0}{N} = \frac{f_s}{2N}$, therefore $|f^{(N)} - f| \leq \frac{f_0}{N}$

We have:

$$\lim_{N\to\infty} f^{(N)} = f \Rightarrow \lim_{N\to\infty} \hat{\gamma}(\tau)e^{\frac{-2\pi f^{(N)}\tau}{f_s}}$$

We have already proved that the estimator $\hat{\gamma}(\tau)$ is consistent, i.e. $\lim_{N\to\infty}\hat{\gamma}(\tau) = \gamma(\tau)$

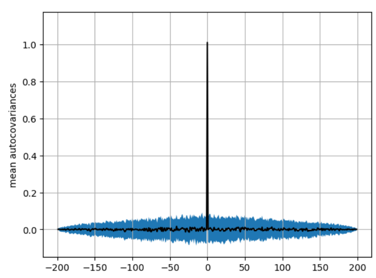
By the Dominated Convergence Theorem, we have $\lim_{N\to\infty}\hat{\gamma}(\tau)e^{\frac{-2\pi f^{(N)}\tau}{f_s}} = \gamma(\tau)e^{\frac{-2\pi f\tau}{f_s}}$. Therefore, the periodogram $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of the power spectrum $S(f)$, i.e., $\lim_{N\to\infty}\mathbb{E}[|J(f^{(N)})|^2] = S(f)$.

Question 6

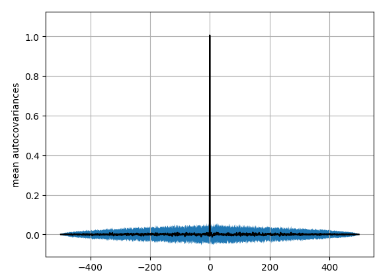
In this question, let X_n ($n = 0, \ldots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

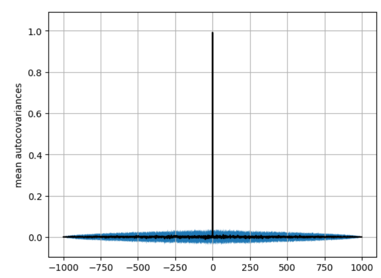
Add your plots to Figure 1.



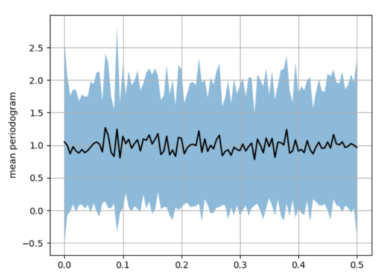
Autocovariance ($N = 200$)



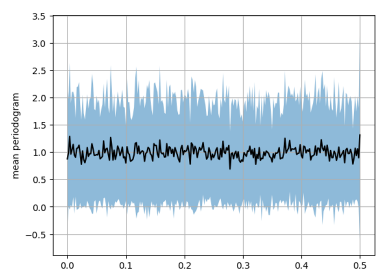
Autocovariance ($N = 500$)



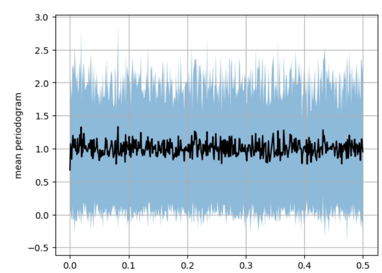
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

- **Autocovariances** In our simulations, we found that the mean autocovariance at lag 0 is close to 1, which aligns with our expectation for a Gaussian white noise process. This is because the autocovariance at lag 0 is essentially the variance of the process. The standard deviation is high at lag 0, indicating a greater spread in the autocovariance estimates. As the lag increased, the autocovariances approaches 0, reflecting the uncorrelated nature of the white noise process, then the standard deviation decreases. This decrease in standard deviation with increasing lag suggests that our estimates are more consistent for larger lags. As we increase the number of data points N, the autocovariance estimates become more accurate, and their variability across different simulations decreases. This observations confirms our previous result of the estimator $\hat{\gamma}(\tau)$ being asymptotically unbiased.
- **Periodograms** In our analysis, we found that the periodogram of the Gaussian white noise process oscillates around 1. This is consistent with the theoretical power spectrum of a white noise process, which is constant across all frequencies. Interestingly, unlike many statistical estimates, the variance of the periodogram does not decrease as we increase N. This observation is in accordance with our result in 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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{29}$$

  (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

Show that for $\tau > 0$,

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

We have:

$$var(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)^2] - \mathbb{E}[\hat{\gamma}(\tau)]^2$$

where:

- $\mathbb{E}[\hat{\gamma}(\tau)]^2 = \left(\frac{N-\tau}{N}\gamma(\tau)\right)^2$
- and:

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

$$= \frac{1}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=0}^{N-\tau-1} \mathbb{E}\Big[X_i X_{i+\tau} X_j X_{j+\tau}\Big]$$

$$= \frac{1}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=0}^{N-\tau-1} \Big[\mathbb{E}[X_i X_{i+\tau}]\mathbb{E}[X_j X_{j+\tau}] + \mathbb{E}[X_i X_j]\mathbb{E}[X_{i+\tau} X_{j+\tau}] + \mathbb{E}[X_i X_{j+\tau}]\mathbb{E}[X_{i+\tau} X_j]\Big]$$

(where we used the hint)

$$= \frac{1}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=0}^{N-\tau-1} \Big[\gamma(\tau)^2 + \gamma(j-i)^2 + \gamma(j-i+\tau)\gamma(j-i-\tau))\Big].$$

Now,

$$\frac{1}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=0}^{N-\tau-1} \gamma(\tau)^2 = \left(\frac{N-\tau}{N}\gamma(\tau)\right)^2 = \mathbb{E}[\hat{\gamma}(\tau)]^2.$$

So,

$$\begin{aligned} var(\hat{\gamma}(\tau)) &= \frac{1}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=0}^{N-\tau-1} \Big[\gamma(j-i)^2 + \gamma(j-i+\tau)\gamma(j-i-\tau))\Big] \\ &= \frac{1}{N^2} \sum_{k=0}^{N-\tau-1} \Big[\gamma(0)^2 + \gamma(\tau)^2\Big] + \frac{2}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{j=i+1}^{N-\tau-1} \Big[\gamma(j-i)^2 + \gamma(j-i+\tau)\gamma(j-i-\tau))\Big] \\ &= \frac{N-\tau}{N^2}(\gamma(0)^2 + \gamma(\tau)^2) + \frac{2}{N^2} \sum_{i=0}^{N-\tau-1} \sum_{n=1}^{N-\tau-i-1} \Big[\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau))\Big] \end{aligned}$$

where we changed variables $n = j - i$ in the second sum. Reversing the sums on i and n, we get that:

$$\begin{aligned} var(\hat{\gamma}(\tau)) &= \frac{N-\tau}{N^2}(\gamma(0)^2 + \gamma(\tau)^2) + \frac{2}{N^2} \sum_{n=1}^{N-\tau-1} \sum_{i=0}^{N-\tau-n-1} \Big[\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau))\Big] \\ &= \frac{N-\tau}{N^2}(\gamma(0)^2 + \gamma(\tau)^2) + \frac{2}{N^2} \sum_{n=1}^{N-\tau-1} (N-\tau-n)\Big[\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau))\Big] \end{aligned}$$

We note that the second sum is multiplied by 2, so we can expand this pair and change variables $n' = -n$ in one of the two sums. Remembering that γ is even, we thus get that:

$$var(\hat{\gamma}(\tau)) = \frac{N-\tau}{N^2}(\gamma(0)^2 + \gamma(\tau)^2) + \frac{1}{N^2} \sum_{\substack{n=-(N-\tau-1) \\ n \neq 0}}^{N-\tau-1} (N-\tau-|n|)\Big[\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau))\Big],$$

that is:

$$\boxed{var(\hat{\gamma}(\tau)) = \frac{1}{N} \sum_{n=-(N-\tau-1)}^{N-\tau-1} (1 - \frac{\tau+|n|}{N})\Big[\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau))\Big]} \tag{30}$$

Show that $\hat{\gamma}(\tau)$ is consistent: We can show that when $N \to \infty$, $\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| \geq \epsilon) = 0$. According to the Bienaymé-Chebychev inequality, for any $\epsilon > 0$, we have: $\mathbb{P}(|\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq$

$\frac{\epsilon}{2}) \leq \frac{4Var(\hat{\gamma}(\tau))}{\epsilon^2}$. Now, for $N > \tau$:

$$
\begin{aligned}
var(\hat{\gamma}(\tau)) &= \frac{1}{N^2} \sum_{n=-(N-\tau-1)}^{N-\tau-1} (N - \tau - |n|) \left(\gamma(n)^2 + \gamma(n+\tau)\gamma(n-\tau)\right) \\
&\leq \frac{1}{N} \sum_{n=-(N-\tau-1)}^{N-\tau-1} \left(|\gamma(n)|^2 + |\gamma(n+\tau)\gamma(n-\tau)|\right) \\
&\leq \frac{1}{N} \sum_{n=-(N-\tau-1)}^{N-\tau-1} \left(|\gamma(n)|^2 + \frac{1}{2}(|\gamma(n+\tau)|^2 + |\gamma(n-\tau)|^2)\right) \\
&\leq \frac{2}{N} \sum_{n=-\infty}^{\infty} |\gamma(n)|^2
\end{aligned}
$$

Since the autocovariances are absolutely summable, i.e. $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$, and square summable, i.e. $\sum_{\tau \in \mathbb{Z}} \gamma^2(\tau) < \infty$, it follows that the series in the expression above is finite. Therefore:

$$
\text{as } N \to \infty, \ \frac{1}{N} \text{ will dominate and go to zero}
$$
$$
\Rightarrow var(\hat{\gamma}(\tau)) \xrightarrow[N \to \infty]{} 0
$$

Now,

$$
\begin{aligned}
&|\gamma(\tau) - \hat{\gamma}(\tau)| \leq |\gamma(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| + |\mathbb{E}[\hat{\gamma}(\tau)] - \hat{\gamma}(\tau)| \\
\iff &|\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq |\gamma(\tau) - \hat{\gamma}(\tau)| - |\gamma(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]|
\end{aligned}
$$

And since $\hat{\gamma}(\tau)$ is asymptotically unbiased, for N large enough, $|\gamma(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \leq \frac{\epsilon}{2}$.

So, $|\gamma(\tau) - \hat{\gamma}(\tau)| \geq \epsilon \Rightarrow |\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq \frac{\epsilon}{2}$. Therefore:

$$
\mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| \geq \epsilon) \leq \mathbb{P}(|\gamma(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq \frac{\epsilon}{2}) \leq \frac{4Var(\hat{\gamma}(\tau))}{\epsilon^2} \xrightarrow[N \to \infty]{} 0.
$$

This shows that the probability that $\hat{\gamma}(\tau)$ deviated from its expected value by more than an arbitrarily small $\epsilon > 0$ goes to zero as $N \to \infty$. Therefore 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 $f = f_0, f_1, \ldots, f_{N/2}$ 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

Deriving the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, ..., f_{N/2}$ where $f_k = f_s k/N$

Given that X is a gaussian white noise with variance σ^2, we know that the mean of X is 0. Therefore the mean of both $A(f)$ and $B(f)$ is 0

$$\begin{aligned}\mathbb{E}[A(f_k)] &= \sum_{n=0}^{N-1} \mathbb{E}[X_n] cos\left(\frac{-2\pi kn}{N}\right) = 0 \\ \mathbb{E}[B(f_k)] &= \sum_{n=0}^{N-1} \mathbb{E}[X_n] sin\left(\frac{-2\pi kn}{N}\right) = 0\end{aligned} \tag{31}$$

The variance of a sum of independent random variables is the sum of their variances. Therefore, we have:

$$\begin{aligned}var(A(f)) &= var\left(\sum_{n=0}^{N-1} X_n cos(\frac{-2\pi fn}{f_s}\right) \\ &= \sum_{n=0}^{N-1} var(X_n cos\left(\frac{-2\pi fn}{f_s}\right)) \\ &= \sum_{n=0}^{N-1} cos^2\left(\frac{-2\pi fn}{f_s}\right) var(X_n) \\ \Rightarrow var(A(f_k)) &= \sigma^2 \sum_{n=0}^{N-1} cos^2\left(\frac{-2\pi kn}{N}\right) \\ &= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} \left(1 + cos(\frac{-4\pi kn}{N})\right) \text{ using } cos^2(\theta) = \frac{1+cos(2\theta)}{2} \\ &= \begin{cases} \frac{\sigma^2}{2}(N+N) & \text{since } \sum_{n=0}^{N-1} cos(\frac{-4\pi kn}{N}) = N \quad \text{if } k = 0 \\ \\ \frac{\sigma^2}{2}(N+0) & \text{since } \sum_{n=0}^{N-1} cos(\frac{-4\pi kn}{N}) = 0, \quad \text{otherwise} \end{cases} \\ \Rightarrow var(A(f_k)) &= \begin{cases} \sigma^2 N, & \text{if } k = 0 \\ \\ \frac{\sigma^2}{2} N, & \text{otherwise} \end{cases}\end{aligned}$$

Similarly we have:

$$\begin{aligned}var(B(f)) &= var\left(\sum_{n=0}^{N-1} X_n sin(\frac{-2\pi fn}{f_s}\right) \\ &= \sum_{n=0}^{N-1} var(X_n sin\left(\frac{-2\pi fn}{f_s}\right)) \\ &= \sum_{n=0}^{N-1} sin^2\left(\frac{-2\pi fn}{f_s}\right) var(X_n) \\ \Rightarrow var(B(f_k)) &= \sigma^2 \sum_{n=0}^{N-1} sin^2\left(\frac{-2\pi kn}{N}\right) \\ &= \frac{\sigma^2}{2} \sum_{n=0}^{N-1} \left(1 - cos(\frac{-4\pi kn}{N})\right) \text{ using } sin^2(\theta) = \frac{1-cos(2\theta)}{2}\end{aligned}$$

$$= \begin{cases} \frac{\sigma^2}{2}(N-N) & \text{since } \sum_{n=0}^{N-1} cos(\frac{-4\pi kn}{N}) = N \quad \text{if } k=0 \\ \\ \frac{\sigma^2}{2}(N+0) & \text{since } \sum_{n=0}^{N-1} cos(\frac{-4\pi kn}{N}) = 0, \quad \text{otherwise} \end{cases}$$

$$\Rightarrow var(B(f_k)) = \begin{cases} 0, & \text{if } k=0 \\ \\ \frac{\sigma^2}{2}N, & \text{otherwise} \end{cases}$$

Distribution of the periodogram values $|J(f_0)|^2, |J(f_1)|^2, ..., |J(f_{N/2})|^2$

$$\begin{aligned} J(f) &= \frac{1}{\sqrt{N}}(A(f) + iB(f)) \\ \Rightarrow |J(f_k)|^2 &= \frac{1}{N}(A(f_k)^2 + B(f_k)^2) \end{aligned} \tag{32}$$

- For $k \neq 0$:
  $A(f_k)$ and $B(f_k)$ have the same mean and variance. Let us admit that they are independent. The quantity $|J(f_k)|^2$ is indeed the sum of the squares of two independent Gaussian random variables. Therefore, it follows a Chi-Squared (χ^2) distribution with two degrees of freedom.

  $$|J(f_k)|^2 \sim \chi^2(2) \quad \textbf{This results seems wrong}$$

- For $k = 0$:
  $B(f_0) = 0 \Rightarrow |J(f_0)|^2 = \frac{1}{N}A(f_0)^2$
  Hence, $|J(f_0)|^2$ follows a scaled χ^2 distribution with 1 degree of freedom:
  $|J(f_0)|^2 \sim \frac{1}{N}\chi^2(1)$

Variance of the $|J(f_k)|^2$:

We know that $|J(f_k)|^2 = \frac{1}{N}(A(f_k)^2 + B(f_k)^2)$.

- For $k \neq 0$, we have:

$$\begin{aligned} var(|J(f_k)|^2) &= \frac{1}{N^2}(Var(A(f_k)^2) + Var(B(f_k)^2)) \qquad \text{(because } A(f_k)^2 \text{ and } B(f_k)^2 \text{ are independent)} \\ &= \frac{1}{N^2}\left(2(\sigma\sqrt{\frac{N}{2}})^4 + 2(\sigma\sqrt{\frac{N}{2}})^4\right) \\ \Rightarrow var(|J(f_k)|^2) &= \sigma^4 \end{aligned}$$

- For $k = 0$:

$$\begin{aligned} var(|J(f_k)|^2) &= \frac{1}{N^2}(Var(A(f_k)^2)) \\ &= \frac{1}{N^2}\left(2(\sigma\sqrt{N})^4\right) \\ \Rightarrow var(|J(f_k)|^2) &= 2\sigma^4 \end{aligned}$$

So, the variance of $|J(f_k)|^2$ is σ^4 when $k \neq 0$ and $2\sigma^4$ when $k = 0$. This implies that the periodogram is not <u>**consistent**</u> because the variance does not decrease as the number of observations increases or simply because the variance does not depend on N.

<u>**Covariances between the $|J(f_k)|^2$**</u>

First, note that for $k \neq 0$, $\mathbb{E}[|J(f_k)|^2] = \frac{1}{N}(Var(A(f_k)) + Var(B(f_j))) = \sigma^2$. Then:

$$\begin{aligned}
cov(|J(f_k)|^2, |J(f_l)|^2) &= \mathbb{E}[|J(f_k)|^2|J(f_l)|^2] - \mathbb{E}[|J(f_k)|^2]\mathbb{E}[|J(f_l)|^2] \\
&= \mathbb{E}[|J(f_k)|^2|J(f_l)|^2] - \sigma^4 \\
&= \frac{1}{N^2}\mathbb{E}[A(f_k)^2A(f_l)^2 + B(f_k)^2B(f_l)^2 + A(f_k)^2B(f_l)^2 + B(f_k)^2A(f_l)^2] - \sigma^4 \\
&= \frac{1}{N^2}\left(\mathbb{E}[A(f_k)^2A(f_l)^2] + \mathbb{E}[B(f_k)^2B(f_l)^2] + \mathbb{E}[A(f_k)^2B(f_l)^2] + \mathbb{E}[B(f_k)^2A(f_l)^2]\right) - \sigma^4
\end{aligned}$$

- let's compute $\mathbb{E}[A(f_k)^2B(f_l)^2]$:

$$\begin{aligned}
\mathbb{E}[A(f_k)^2B(f_l)^2] &= \mathbb{E}[A(f_k)^2]\mathbb{E}[B(f_l)^2] \\
&= Var(A(f_k))Var(B(f_l)) \\
\mathbb{E}[A(f_k)^2B(f_l)^2] &= \begin{cases} \frac{\sigma^4N^2}{4}, & \text{if } k \neq 0, l \neq 0 \\ \frac{\sigma^4N^2}{2}, & \text{if } k = 0, l \neq 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

- let's compute $\mathbb{E}[B(f_k)^2A(f_l)^2]$:

$$\begin{aligned}
\mathbb{E}[A(f_k)^2B(f_l)^2] &= \mathbb{E}[A(f_k)^2]\mathbb{E}[B(f_l)^2] \\
&= Var(B(f_k))Var(A(f_l)) \\
\mathbb{E}[A(f_k)^2B(f_l)^2] &= \begin{cases} \frac{\sigma^4N^2}{4}, & \text{if } k \neq 0, l \neq 0 \\ \frac{\sigma^4N^2}{2}, & \text{if } k \neq 0, l = 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

- let's compute $\mathbb{E}[A(f_k)^2A(f_l)^2]$:

$$\begin{aligned}
\mathbb{E}[A(f_k)^2A(f_l)^2] &= \mathbb{E}\left[\sum_{n,m,p,q}^{N-1} X_nX_mX_pX_q \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)\cos\left(\frac{2\pi lp}{N}\right)\cos\left(\frac{2\pi lq}{N}\right)\right] \\
&= \sum_{n,m,p,q}^{N-1} \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)\cos\left(\frac{2\pi lp}{N}\right)\cos\left(\frac{2\pi lq}{N}\right) \\
&\quad (\mathbb{E}[X_nX_m]\mathbb{E}[X_pX_q] + \mathbb{E}[X_nX_p]\mathbb{E}[X_mX_q] + \mathbb{E}[X_nX_q]\mathbb{E}[X_mX_p])
\end{aligned}$$

Since the X_n, X_m, X_p, X_q are all independent, we will study only the cases where

1. $n = m, p = q$

2. $n = p, m = q$

3. $n = q, m = p$

by splitting $(\mathbb{E}[X_n X_m]\mathbb{E}[X_p X_q] + \mathbb{E}[X_n X_p]\mathbb{E}[X_m X_q] + \mathbb{E}[X_n X_q]\mathbb{E}[X_m X_p])$ in 3 parts.

– When $n = m, p = q$:

$$\sum_{n,p}^{N-1} \cos^2\left(\frac{2\pi kn}{N}\right)\cos^2\left(\frac{2\pi lp}{N}\right)\mathbb{E}[X_n^2]\mathbb{E}[X_p^2] = \sigma^4 \sum_{n,p}^{N-1} \cos^2\left(\frac{2\pi kn}{N}\right)\cos^2\left(\frac{2\pi lp}{N}\right)$$

$$= \frac{\sigma^4}{4}\sum_{n,p}^{N-1}\left(1 + cos(\frac{-4\pi kn}{N})\right)\left(1 + cos(\frac{-4\pi lp}{N})\right)$$

using $cos^2(\theta) = \dfrac{1 + cos(2\theta)}{2}$

$$= \begin{cases} \frac{\sigma^4}{4}(N+N)(N+N) & \text{if } k = 0, l = 0 \\ \frac{\sigma^2}{4}(N+0)(N+0) & \text{if} k \neq 0, l \neq 0 \\ \frac{\sigma^2}{4}(N+N)(N+0) & otherwise \end{cases}$$

$$= \begin{cases} \sigma^4 N^2 & \text{if } k = 0, l = 0 \\ \frac{\sigma^2}{4}N^2 & \text{if } k \neq 0, l \neq 0 \\ \frac{\sigma^2}{2}N^2 & otherwise \end{cases}$$

– When $n = p, m = q$:

$$\sum_{n,m} \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)$$
$$\cos\left(\frac{2\pi ln}{N}\right)\cos\left(\frac{2\pi lm}{N}\right)\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] = \sigma^4\sum_n \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi ln}{N}\right)$$
$$\sum_m \cos\left(\frac{2\pi km}{N}\right)\cos\left(\frac{2\pi lm}{N}\right)$$
$$= \frac{\sigma^4}{2}\sum_n \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi ln}{N}\right)$$
$$\sum_m\left(\cos\left(\frac{2\pi(l-k)m}{N}\right) + \cos\left(\frac{2\pi(l+k)m}{N}\right)\right)$$
$$= \begin{cases} \sigma^4 N^2 & \text{if } k = 0, l = 0 \\ \frac{\sigma^4}{4}N^2 & \text{if } k \neq 0, l \neq 0, k = l \\ 0 & otherwise \end{cases}$$

– When $n = q, m = p$, using the same process as the previous step we get: j

$$\sum_{n,m} \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)\cos\left(\frac{2\pi lm}{N}\right)\cos\left(\frac{2\pi lq}{N}\right)\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] = \begin{cases} \sigma^4 N^2 & \text{if } k = 0, l = 0 \\ \frac{\sigma^4}{4}N^2 & \text{if } k \neq 0, l \neq 0, k = l \\ 0 & otherwise \end{cases}$$

Finally we have:

$$\mathbb{E}[A(f_k)^2A(f_l)^2] = \begin{cases} 3\sigma^4N^2 & \text{if } k=0, l=0 \\ 3\times\frac{\sigma^4}{4}N^2 & \text{if } k\neq 0, l\neq 0, k=l \\ \frac{\sigma^4}{4}N^2 & \text{if } k\neq 0, l\neq 0, k\neq l \\ \frac{\sigma^2}{2}N^2 & \textit{otherwise} \end{cases}$$

- Let's compute $\mathbb{E}[B(f_k)^2B(f_l)^2]$

$$\mathbb{E}[B(f_k)^2B(f_l)^2] = \begin{cases} 0 & \text{if } k=0, l=0 \\ 3\times\frac{\sigma^4}{4}N^2 & \text{if } k\neq 0, l\neq 0, k=l \\ \frac{\sigma^4}{4}N^2 & \text{if } k\neq 0, l\neq 0, k\neq l \\ 0 & \textit{otherwise} \end{cases}$$

Combining all the results we get:

$$cov(|J(f_k)|^2, |J(f_l)|^2) = \begin{cases} 3\sigma^4-\sigma^4 & \text{if } k=0, l=0 \\ \frac{2\sigma^4}{4}+\frac{6\sigma^4}{4}-\sigma^4 & \text{if } k\neq 0, l\neq 0, k=l \\ \frac{4\sigma^4}{4}-\sigma^4 & \text{if } k\neq 0, l\neq 0, k\neq l \\ \frac{\sigma^4}{2}+\frac{\sigma^4}{2}-\sigma^4 & \textit{otherwise} \end{cases}$$

$$cov(|J(f_k)|^2, |J(f_l)|^2) = \begin{cases} 2\sigma^4 & \text{if } k=0, l=0 \\ \sigma^4 & \text{if } k\neq 0, l\neq 0, k=l \\ 0 & \textit{otherwise} \end{cases}$$

When $k=l$, this result is obvious as it is simply the variance, so we could have just looked at the case where $k\neq l$. When $k\neq l$, the covariance is 0, which explains the erratic behaviour observed in question 6.

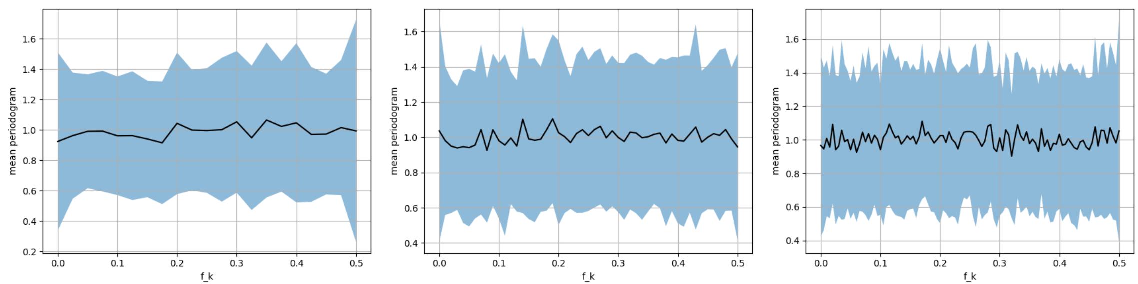
Question 9

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 see here that comparatively to the previous plots in question 6, the standard deviation is lower. In question 6, the standard deviation was around 0.9, now it is only around 0.4. This is in accordance with the Barlett's procedure because the variance is being divided by K which means that the standard deviation is divided by $\sqrt{K}$, i.e $\frac{0.9}{\sqrt{5}} = 0.4$.

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

A 5-fold cross validation gives an optimal number of neighbors equal to 5, with an associated F-score equal to 0.87. Note that we observe a high variance of this score over the 5 folds.

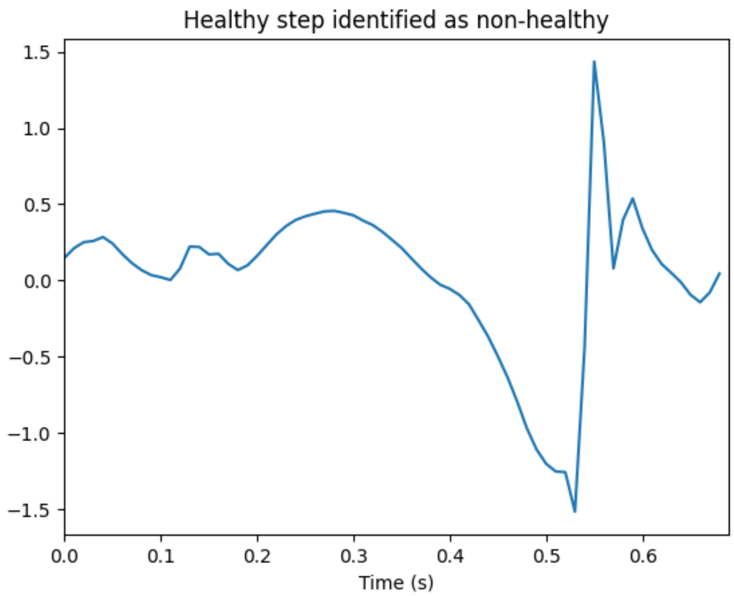
However, training a k-neigbors classifier with this choice of parameters on the full training set, and testing it on the test set gives an F-score of 0.48. This shows that the classifier generalizes very badly. This is probably worsened by the fact that the test set is quite different from the training set; in particular, in contains a majority of non-healthy steps that the classifier fails to recognize as non-healthy.
More generally, distance matrices built using the DTW distance on the training set reveals that this distance sometimes fails to distinguish the two classes of steps (healthy and non-healthy), as one step is not always closer to steps of the same class than to steps of the other class (in fact, it is sometimes the exact opposite). This explains why this classification problem is difficult.

In particular, taking both right and left steps (and not only left steps as we do here) might improve the results.

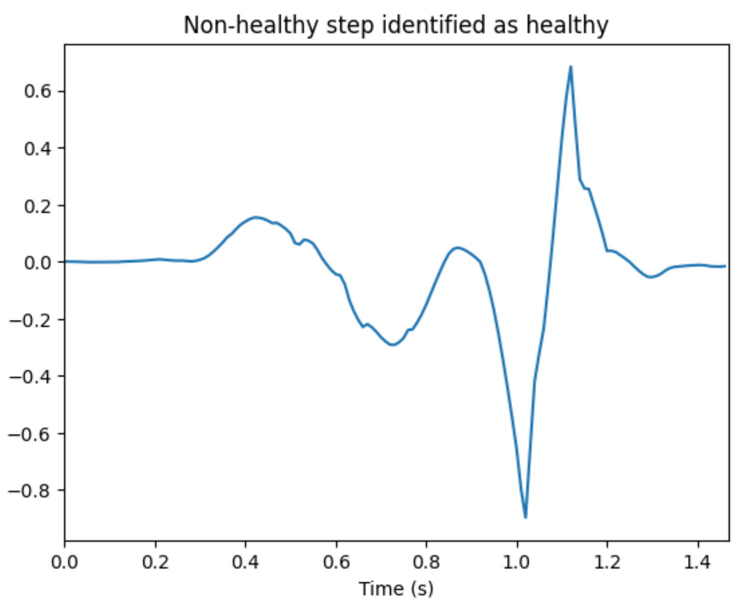
Question 11

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).