# Assignment 1 (ML for TS) - MVA

Elsa Doukhan elsadoukhan@gmail.com
Ilan Bacry ilan.bacry20@gmail.com

November 9, 2025

## 1 Introduction

**Objective.** This assignment has three parts: questions about convolutional dictionary learning, 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 Sunday 9$^{\text{th}}$ November 23:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname2.pdf` and
  `FirstnameLastname1_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_ValerioGuerrini.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: 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 us consider the following function

$$g(\beta) = \frac{1}{2}\,\|y - X\beta\|_2^2 + \lambda\|\beta\|_1, \quad \beta \in \mathbb{R}^p.$$

The function $g$ is convex because it's the sum of two convex functions.

Thus, according to the subgradient optimality condition for a convex function, we have:

$$\beta^\star \text{ minimises } g \iff 0_p \in \partial g(\beta^\star).$$

where $\partial g(\beta^\star)$ is the subdifferential of $g$ at $\beta^\star$

For $\beta^\star = 0_p$ we have :

$$0_p \text{ minimise } g \iff 0_p \in \partial g(0_p)$$

We know that for a LASSO regression we have :

$$\partial g(0_p) = -X^\top y + \lambda z, \quad \text{with } z \in \{\, z \in \mathbb{R}^p : \|z\|_\infty \le 1\,\}.$$

So there exists $z \in \mathbb{R}^p$ such that $\|z\|_\infty \le 1$ and

$$-X^\top y + \lambda z = 0_p.$$

$z$ must be $z = \frac{1}{\lambda}X^\top y$

This value of $z$ is admissible if and only if

$$\|z\|_\infty = \tfrac{1}{\lambda}\|X^\top y\|_\infty \le 1 \iff \lambda \ge \|X^\top y\|_\infty.$$

Therefore, 0 is a subgradient and $\beta$ minimizes $g$ whenever $\lambda \ge \|X^\top y\|_\infty$

We thus set

$$\boxed{\lambda_{\max} = \|X^\top y\|_\infty.}$$

## 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:

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

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

#### First point

We fix the $d_k \in \mathbb{R}^L$ and consider

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_K \end{bmatrix}, \quad \text{where } Z \in \mathbb{R}^{N \times K(N-L+1)}.$$

The response vector is

$$x \in \mathbb{R}^N.$$

The goal is to determine a matrix $D$ such that, for all $1 \leq m \leq N$,

$$(DZ)_m = \left( \sum_{k=1}^{K} z_k * d_k \right)_m$$

where $*$ denotes the discrete convolution.

For each $k \in \{1, \ldots, K\}$, we define the matrix $D_k \in \mathbb{R}^{N \times (N-L+1)}$. We also fix a padding rule: if $(m - j + 1)$ exceeds the length of $d_k$ or is negative, its value is set to 0.

We can then write

$$(z_k * d_k)_m = \sum_{j=1}^{N-L+1} d_k(m - j + 1)\, z_k(j) = (D_k z_k)_m.$$

Hence,

$$(DZ)_m = (D_1 z_1 + D_2 z_2 + \cdots + D_K z_K)_m = \begin{bmatrix} D_1 & D_2 & \cdots & D_K \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{bmatrix}.$$

By identification, the design matrix is

$$D = \begin{bmatrix} D_1 & D_2 & \cdots & D_K \end{bmatrix}, \quad \text{and we define } Z = \begin{bmatrix} z_1 \\ \vdots \\ z_K \end{bmatrix}.$$

Hence, the problem becomes a LASSO regression problem with

$$\boxed{\text{response vector: } x,} \qquad \boxed{\text{design matrix: } D,}$$

**Second point**

According to **Question 1** . This problem can be written as a LASSO regression with $y = x$ and $X = D$. Hence, we have

$$\boxed{\lambda_{\max} = \|D^\top x\|_\infty} \tag{3}$$

# 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 the sampling frequency by $f_s$, meaning that the index $n$ corresponds to the time $n/f_s$. 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 $S(f)$ indicate 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 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 number 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 consider $X_n$ $(n = 0, \dots, N-1)$ a Gaussian white noise process. Let $\sigma^2$ denote its variance.

**I. Computation of the autocovariance function**

Since the noise is white, the samples are uncorrelated with each other.

- When $t = 0$, we compare the sample $X_n$ with itself. Hence, $\gamma(t) = \mathbb{E}[X_n^2] = \sigma^2$.
- When $t \neq 0$, the two terms $X_n$ and $X_{n+t}$ are independent, so $\gamma(t) = \mathbb{E}[X_n]\,\mathbb{E}[X_{n+t}] = 0$.

Therefore,

$$\boxed{\gamma(t) = \begin{cases} \sigma^2, & \text{if } t = 0, \\ 0, & \text{if } t \neq 0 \end{cases}}$$

**II. Computation of the power spectrum**

By definition, the power spectrum is the Fourier transform of the autocovariance function $\gamma$. Let $S$ denote the power spectrum. We have

$$S(f) = \sum_{t=-\infty}^{+\infty} \gamma(t)\, e^{-2i\pi f t / f_s}$$

From the definition of $\gamma$, we get:

$$S(f) = \gamma(0) e^{-2i\pi f \cdot 0 / f_s} + \sum_{t \neq 0} \gamma(t)\, e^{-2i\pi f t / f_s} = \sigma^2 + 0 = \sigma^2$$

Therefore, the power spectrum is

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

Therefore, the signal has the same energy at all frequencies, which is why we call it white noise.

## 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 $\tau = 0, 1, \ldots, N-1$ 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

We consider the empirical autocovariance defined by:

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

The expectation of this empirical estimator of the autocovariance is given by:

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

Hence, since $\mathbb{E}[\hat{\gamma}(\tau)] \neq \gamma(\tau)$, $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$. However, when $N \to \infty$, we have $\frac{N-\tau}{N} \to 1$, so asymptotically the estimator becomes unbiased.

**Bias correction** A simple way to correct the bias is to multiply $\hat{\gamma}(\tau)$ by $\boxed{\dfrac{N}{N-\tau}}$. We then define

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

Thus, we obtain

$$\mathbb{E}\left[\hat{\gamma}_{\text{unbiased}}(\tau)\right] = \gamma(\tau)$$

## Question 5

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

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

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

### I. First point

Consider a random process $X_n$. Its discrete Fourier transform is defined by :

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

We aim to express $|J(f_k)|^2$ as a function of the sample autocovariances and to show that its expectation converges to the power spectral density $S(f)$ as $N \to \infty$.

We start from the definition:

$$J(f_k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n \, e^{-2i\pi \frac{k}{N} n}$$

Then :

$$|J(f_k)|^2 = J(f_k) J^*(f_k) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{p=0}^{N-1} X_n X_p \, e^{-2i\pi \frac{k}{N}(n-p)}$$

where $J^*(f_k)$ is the complex conjugate.

We reorder the double sum over the indices $(n, p)$.

**1. First case:** $p \geq n$

Let $l = p - n$, which varies from 0 to $N-1$. Hence, $n$ ranges from 0 to $N-l-1$.

We then obtain:

$$\frac{1}{N}\sum_{l=0}^{N-1}\sum_{n=0}^{N-l-1} X_n X_{n+l}\, e^{\frac{2i\pi kl}{N}} = \frac{1}{N}\sum_{l=0}^{N-1}\hat{\gamma}(l)\, e^{\frac{2i\pi kl}{N}},$$

**2. Second case:** $p < n$

Let $l = n - p$, which varies from 1 to $N-1$. Hence, $p$ ranges from 0 to $N-l-1$.

We then obtain:

$$\frac{1}{N}\sum_{l=1}^{N-1}\sum_{p=0}^{N-l-1} X_p X_{p+l}\, e^{-\frac{2i\pi kl}{N}} = \frac{1}{N}\sum_{l=1}^{N-1}\hat{\gamma}(l)\, e^{-\frac{2i\pi kl}{N}} = \frac{1}{N}\sum_{l=-(N-1)}^{-1}\hat{\gamma}(l)e^{-\frac{2i\pi kl}{N}}$$

Since $\hat{\gamma}(\tau) := \hat{\gamma}(-\tau)$ for $\tau = -(N-1), \ldots, -1$ (from Question 4).

Thus, combining both cases, we finally obtain:

$$\frac{1}{N}\sum_{l=-N+1}^{N-1}\hat{\gamma}(|l|)\, e^{-\frac{2i\pi kl}{N}}.$$

Finally we have

$$\boxed{|J(f_k)|^2 = \frac{1}{N}\sum_{l=-N+1}^{N-1}\hat{\gamma}(|l|)\, e^{-\frac{2i\pi kl}{N}}.}$$

We therefore obtain an expression of $|J(f_k)|^2$ as a sum of sample autocovariances.

**II. Second point**

We have

$$|J(f_k)|^2 = \frac{1}{N}\sum_{l=-N+1}^{N-1}\hat{\gamma}(|l|)\, e^{-\frac{2i\pi kl}{N}}$$

By taking the expectation of $|J(f_k)|^2$, we obtain (according to question 4):

$$\mathbb{E}\left[\, |J(f_k)|^2) \,\right] = \sum_{l=-N+1}^{N-1}\left(1 - \frac{|l|}{N}\right)\gamma(l)\, e^{-\frac{2i\pi kl}{N}}$$

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

$$\mathbb{E}\Big[\, |J(f^{(N)})|^2 - S(f) \,\Big] \xrightarrow[N\to\infty]{} 0$$

$$\begin{aligned}
\mathbb{E}\Big[\, |J(f^{(N)})|^2 \Big] - S(f) &= \sum_{l=-N+1}^{N-1} \left(1 - \frac{|l|}{N}\right) \gamma(l)\, e^{-2i\pi \frac{f^{(N)} l}{f_s}} \;-\; \sum_{l=-\infty}^{+\infty} \gamma(l)\, e^{-2i\pi \frac{fl}{f_s}} \\
&= \sum_{l=-N+1}^{N-1} \left(-\frac{|l|}{N}\right) \gamma(l)\, e^{-2i\pi \frac{f^{(N)} l}{f_s}} \;-\; \sum_{|l|\geq N} \gamma(l)\, e^{-2i\pi \frac{fl}{f_s}}.
\end{aligned}$$

Moreover

$$\sum_{|l|\geq N} \gamma(l)\, e^{-2\pi i \frac{f}{f_s} l} \xrightarrow[N\to+\infty]{} 0,$$

since $\gamma(l)$ is absolutely summable.

Let $0 < \epsilon < 1$. We now split the sum as follows :

$$\left| -\sum_{l=-N+1}^{N-1} \frac{|l|}{N} \gamma(l)\, e^{-2\pi i \frac{f^{(N)}}{f_s} l} \right| \;\leq\; \sum_{|l|\leq N-1} \frac{|l|}{N} |\gamma(l)| \;\leq\; \underbrace{\sum_{|l|\leq (N-1)^\epsilon} \frac{|l|}{N} |\gamma(l)|}_{(1)} \;+\; \underbrace{\sum_{(N-1)^\epsilon < |l| \leq N-1} \frac{|l|}{N} |\gamma(l)|}_{(2)}$$

Furthermore

$$(1) \leq \frac{(N-1)^\epsilon}{N} \sum_{|l|\leq N-1} |\gamma(l)| \;\xrightarrow[N\to\infty]{}\; 0. \tag{7}$$

Finally,

$$(2) \;\leq\; \sum_{(N-1)^\epsilon < |l| \leq N-1} \frac{|l|}{N} |\gamma(l)| \;<\; \sum_{|l| > (N-1)^\epsilon} |\gamma(l)|.$$

Since $\gamma(l)$ is absolutely summable, this term tends to 0 as $N \to +\infty$.

Finally, since each component of

$$\mathbb{E}\Big[\, |J(f^{(N)})|^2 \Big] - S(f)$$

tends to 0 as $N \to +\infty$ and the remaining term is bounded by 0, we conclude that

$$\mathbb{E}\Big[\, |J(f^{(N)})|^2 \Big] - S(f) \;\xrightarrow[N\to+\infty]{}\; 0.$$

Therefore,

$$\boxed{\lim_{N\to+\infty} \mathbb{E}\Big[\, |J(f^{(N)})|^2 \Big] = S(f),}$$

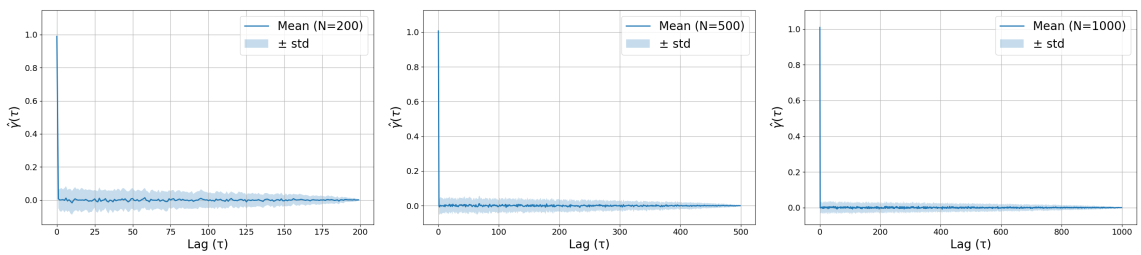
i.e., $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of the power spectral density $S(f)$ for $f > 0$.

## Question 6

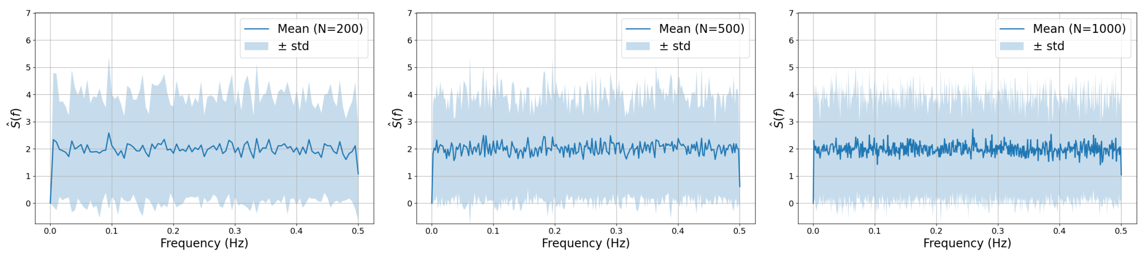
In this question, let $X_n$ ($n = 0, \dots, N-1$) be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- For $N \in \{200, 500, 1000\}$, compute the *sample autocovariances* ($\hat{\gamma}(\tau)$ vs $\tau$) 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

### I. Observations on the autocovariance plots

#### 1 - Observations within each graph :

When we look at each graph, we observe a common trend on the standard deviation interval, the larger the lag, the smaller the interval. More precisely, it converges toward 0. This can be explained by the choice of our estimator.

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

The normalization term N is constant but the number of terms in the sum is $(N - \tau - 1) - 0 + 1 = N - \tau$.

When we have a small lag, (e.g., $\tau = 1$), we sum $N - 1$ terms which have many different values (some slightly below 0, others slightly above 0), but when this sum is divided by N, we almost get a mean of the autocorrelation for this specific lag. However, for a large lag (such as $\tau = N - 1$) we sum only $N - (N - 1) = 1$ term and even if this term is not exactly zero, we divide this value by N. Hence, with this estimator, the variance decreases rapidly with the value of the lag.

**2 - Comparison between the graphs :**

When comparing the graphs, we observe two key phenomena as the number of samples $N$ increases:

First, the mean stabilizes more and more around 0. This is the true theoretical value of the autocovariance for a white noise (when $\tau > 0$). This shows that the estimator is asymptotically unbiased, as its expected value gets closer to the true value as $N$ grows.

Second, the variance gets smaller and smaller. This is explained by the fundamental principle of the law of large numbers. The more the estimation is done on a large sample, the more the information we capture is close to the truth.

For example with $N = 200$, by pure chance, many of the $X_n X_{n+\tau}$ terms might be positive, resulting in $\hat{\gamma}(\tau) > 0$, or negative, resulting in $\hat{\gamma}(\tau) < 0$. The estimation is "noisy" and unstable. The variance (the blue area) is large.

With $N = 1000$ (Large sample): The chances of having such an imbalance collapse. The positive and negative terms will cancel each other out much more effectively. The estimate $\hat{\gamma}(\tau)$ will almost always be very close to the true value of 0. The variance (the blue area) is very small.

Finally, the combination of these two observations; the bias tending to 0 (it is asymptotically unbiased) and the variance tending to 0, proves that the sample autocovariance $\hat{\gamma}(\tau)$ is a consistent estimator of the true autocovariance $\gamma(\tau)$.

**II. Observations on the periodogram plots**

I observe that as the signal size $N$ increases, the number of frequency points increases (the resolution improves). The graph appears more unstable, but this is an illusion due to the higher number of points.

The most important observation is that, overall, the mean and the standard deviation remain constant, not only with frequency but also with the signal size $N$.

The periodogram is an asymptotically unbiased estimator of the Power Spectral Density, the mean oscillates around 2 regardless of $N$. This is the true value of the one-sided Power Spectral Density (PSD) for a white noise with variance $\sigma^2 = 1$, but the variance does not decrease with $N$. Therefore, the periodogram is not a consistent estimator of the power spectrum.

## 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 $\infty$ to the true value $\gamma(\tau)$. In this question, assume that $X$ is a wide-sense stationary *Gaussian* process.

- Show that for $\tau > 0$

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

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

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

## Answer 7

### I. Calculation of the variance

Since the hint is mentioning an expectation, we might think to calculate the variance in this way:

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

**1 - Calculation of** $(\mathbb{E}[\hat{\gamma}(\tau)])^2$

By linearity of expectation,

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

We assume that $X$ is a wide-sense stationary *Gaussian* process.

A discrete-time random process $\{X_n\}$ is wide-sense stationary, WSS if:

Its mean is constant:

$$\mathbb{E}[X_n] = \mu, \quad \forall n$$

Its covariance depends only on the time difference (the lag), not on the absolute position:

$$\mathrm{Cov}(X_n, X_m) = \mathbb{E}\big[(X_n - \mu)(X_m - \mu)\big] = \gamma(m-n),$$

i.e., it depends only on $m - n$, not on $n$ or $m$ separately.

By convention, a Gaussian White Noise is centred. Therefore,

$$\mathbb{E}[X_n] = \mu = 0, \quad \forall n$$

$$\mathrm{Cov}(X_n, X_{n+\tau}) = \mathbb{E}\big[(X_n)(X_{n+\tau})\big] = \gamma(\tau),$$

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

Hence,

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

.

**2 - Calculation of** $\mathbb{E}[\hat{\gamma}(\tau)^2]$

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

$$= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1}\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}]$$

We can use the given trick on $\mathbb{E}[Y_1 Y_2 Y_3 Y_4]$ with $Y_1 = X_n, Y_2 = X_{n+\tau}, Y_3 = X_m, Y_4 = X_{m+\tau}$ :

$$\mathbb{E}[X_n X_{n+\tau} X_m X_{m+\tau}] = \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}] + \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] + \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_{n+\tau} X_m]$$

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

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

Hence,

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

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

$$= \frac{1}{N^2}(N-\tau)(N-\tau)\gamma^2(\tau) + \frac{1}{N^2}\sum_n\sum_m\left[\gamma^2(m-n) + \gamma(m-n+\tau)\gamma(m-n-\tau)\right]$$

$$= \left(\frac{N-\tau}{N}\right)^2\gamma^2(\tau) + \frac{1}{N^2}\sum_n\sum_m\left[\gamma^2(m-n) + \gamma(m-n+\tau)\gamma(m-n-\tau)\right]$$

**3 - Calculation of Var(**$\hat{\gamma}(\tau)$**)** $= \mathbb{E}[\hat{\gamma}(\tau)^2] - (\mathbb{E}[\hat{\gamma}(\tau)])^2$

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

$$= \left[ \left( \frac{N-\tau}{N} \right)^2 \gamma^2(\tau) + \frac{1}{N^2} \sum_n \sum_m [\dots] \right] - \left( \frac{N-\tau}{N} \right)^2 \gamma^2(\tau)$$

The $\gamma^2(\tau)$ terms cancel out. We are left with:

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

Change of variable : $\sum_n \sum_m$ is a double summation over a square. We reorganize it by summing along the diagonals.

let's definine $k = m - n$ and find the range of values of k.

$k_{min} = m_{min} - n_{max} = 0 - (N-\tau-1) = -(N-\tau-1)$

$k_{max} = m_{max} - n_{min} = (N-\tau-1) - 0 = +(N-\tau-1)$

The new sum $\sum_{k=-(N-\tau-1)}^{N-\tau-1}$ iterates over each diagonal (each value of $k$). On a given diagonal $k$, the number of terms is $(N - \tau - |k|)$

Hence, we get :

$$\text{Var}(\hat{\gamma}(\tau)) = \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|) \left[ \gamma^2(k) + \gamma(k+\tau)\gamma(k-\tau) \right]$$

This is exactly the required formula by relabeling $k$ as $n$ and by moving one factor of $1/N$ from the outer $1/N^2$ to the inside of the summation.

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

**II. Conclusion on the Consistency of $\hat{\gamma}(\tau)$**

To show that the estimator $\hat{\gamma}(\tau)$ is consistent, we must show that it converges in probability to the true value $\gamma(\tau)$ as $N \to \infty$:

$$\hat{\gamma}(\tau) \xrightarrow{\mathcal{P}} \gamma(\tau)$$

A sufficient condition for convergence in probability is convergence in mean square ($L^2$). This means we need to show that the quadratic risk (R) of the estimator tends to zero:

$$\lim_{N \to \infty} \mathbb{E} \left[ \left( \hat{\gamma}(\tau) - \gamma(\tau) \right)^2 \right] = 0$$

The quadratic risk can be decomposed into its bias and variance components:

$$R(\hat{\gamma}(\tau)) = (\text{Biais}(\hat{\gamma}(\tau)))^2 + \text{Var}(\hat{\gamma}(\tau))$$

where $\text{Biais}(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau)$.

The bias is :

$$\text{Biais}(\hat{\gamma}(\tau)) = \mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau) = \left(1 - \frac{\tau}{N}\right)\gamma(\tau) - \gamma(\tau) = -\frac{\tau}{N}\gamma(\tau)$$

As $N \to \infty$, the bias tends to zero:

$$\lim_{N\to\infty} \text{Biais}(\hat{\gamma}(\tau)) = \lim_{N\to\infty} -\frac{\tau}{N}\gamma(\tau) = 0$$

Consequently, the squared bias also tends to zero. The estimator is asymptotically unbiased.

From the first part of the exercise, we have the expression for the variance:

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

We must show that $\lim_{N\to\infty} \text{Var}(\hat{\gamma}(\tau)) = 0$. As $N \to \infty$, the summation range expands to cover all integers $(-\infty, \infty)$. For any fixed $n$, the term $\left(1 - \frac{\tau+|n|}{N}\right)$ approaches 1.

The expression is of the form $\frac{1}{N} \times \text{Sum}(N)$. For the limit to be zero, the sum $\text{Sum}(N)$ must not grow faster than $N$. The standard condition to ensure this is that the process has a property related to ergodicity, which implies that its autocovariance function is square-summable.

First term: $\sum_{n=-\infty}^{\infty} \gamma^2(n)$ is finite by our assumption, as this is explicitly given in the problem statement at the beginning of Part 3.

Second term: For $\sum_{n=-\infty}^{\infty} \gamma(n-\tau)\gamma(n+\tau)$, we can apply the Cauchy-Schwarz inequality with a change of variable ($k = n - \tau$ and $m = n + \tau$) and prove that it is also finite

Hence, the infinite series converges to a finite constant $C$:

$$\lim_{N\to\infty} \sum_{n=-(N-\tau-1)}^{N-\tau-1} \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right] = \sum_{n=-\infty}^{\infty} \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right] = C < \infty$$

Given this standard assumption, we can evaluate the limit of the variance:

$$\lim_{N\to\infty} \text{Var}(\hat{\gamma}(\tau)) = \lim_{N\to\infty} \frac{1}{N} \times C = 0$$

We have shown that both terms of the quadratic risk tend to zero:

$\lim_{N\to\infty} (\text{Biais}(\hat{\gamma}(\tau)))^2 = 0$

$\lim_{N\to\infty} \text{Var}(\hat{\gamma}(\tau)) = 0$

Therefore, the quadratic risk tends to zero:

$$\lim_{N\to\infty} \mathrm{R}(\hat{\gamma}(\tau)) = 0$$

> Since convergence in mean square ($L^2$) implies convergence in probability, we have successfully shown that $\hat{\gamma}(\tau)$ is a consistent estimator for $\gamma(\tau)$.

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 Gaussian white noise, but this holds for more general stationary processes.

## Question 8

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

- Derive the mean and variance of $A(f)$ and $B(f)$ for $f = f_0, f_1, \dots, f_{N/2}$ 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

**I. Derive the mean and variance of $A(f)$ and $B(f)$.**

**1. Mean Calculation**

$A(f_k)$ and $B(f_k)$ are weighted sums of random variables $X_n$. We use the linearity of expectation.$X_n$ is a white noise, which implies it is zero-mean:

$$\mathbb{E}[X_n] = 0 \text{ for all } n$$

Mean of $A(f_k)$:

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

$$\boxed{\mathbb{E}[A(f_k)] = \mathbf{0}}$$

Mean of $B(f_k)$:

$$\mathbb{E}[B(f_k)] = \mathbb{E}\left[-\sum_{n=0}^{N-1} X_n \sin(2\pi kn/N)\right]$$
$$= -\sum_{n=0}^{N-1} \mathbb{E}[X_n] \sin(2\pi kn/N)$$

$$\boxed{\mathbb{E}[B(f_k)] = \mathbf{0}}$$

Both $A(f_k)$ and $B(f_k)$ have zero mean for all frequencies $f_k$.

**2. Variance Calculation**

To calculate the variance of a sum of independent variables, we use the property:

$\text{Var}\left(\sum_n a_n X_n\right) = \sum_n a_n^2 \text{Var}(X_n)$ Where the $a_n$ (the cos and sin terms) are constants and The $X_n$ variables are independent (since it's Gaussian white noise, uncorrelated implies independent).

We know that $\text{Var}(X_n) = \sigma^2$ for all $n$.

Hence, variance of $A(f_k)$:

$$\text{Var}(A(f_k)) = \sigma^2 \sum_{n=0}^{N-1} \cos^2(2\pi kn/N)$$

Variance of $B(f_k)$:

$$\text{Var}(B(f_k)) = \sigma^2 \sum_{n=0}^{N-1} \sin^2(2\pi kn/N)$$

- Case 1: Frequencies where $0 < k < N/2$

  Let's define

  $S_c = \sum_{n=0}^{N-1} \cos^2(2\pi kn/N)$ and $S_s = \sum_{n=0}^{N-1} \sin^2(2\pi kn/N)$.

  We use two identities:

$$S_c + S_s = \sum_{n=0}^{N-1} (\cos^2(2\pi kn/N) + \sin^2(2\pi kn/N)) = \sum_{n=0}^{N-1} 1 = N$$

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

$$S_c - S_s = \text{Re}\left(\sum_{n=0}^{N-1} e^{i\frac{2\pi(2k)n}{N}}\right)$$

$$\sum_{n=0}^{N-1} e^{i\frac{2\pi(2k)n}{N}} = \sum_{n=0}^{N-1} \left(e^{i\frac{2\pi(2k)}{N}}\right)^n$$

This is a geometric series of the form $\sum_{n=0}^{N-1} r^n$.

The ratio is $r = e^{i\frac{2\pi(2k)}{N}}$.

We are in the case $0 < k < N/2$, which implies $0 < 2k < N$. The ratio $r = e^{i\frac{2\pi(2k)}{N}}$ can only equal 1 if its exponent $\frac{2\pi(2k)}{N}$ is a multiple of $2\pi$. This would require $2k$ to be a multiple of $N$.However, since $2k$ is strictly between 0 and $N$, it cannot be a multiple of $N$. Therefore, $r \neq 1$.

Since, $r \neq 1$, and we can use the sum formula: $\sum_{n=0}^{N-1} r^n = \frac{1-r^N}{1-r}$.

$$r^N = \left(e^{i\frac{2\pi(2k)}{N}}\right)^N = e^{i\frac{2\pi(2k)N}{N}} = e^{i2\pi(2k)}$$

$e^{i2\pi(2k)}$ is equal to $\cos(2\pi(2k)) + i\sin(2\pi(2k))$.

$$\cos(2\pi(2k)) = 1 \quad \text{and} \quad \sin(2\pi(2k)) = 0$$

So, $r^N = 1$. Now we substitute this back into the sum formula:

$$\sum_{n=0}^{N-1} r^n = \frac{1-r^N}{1-r} = \frac{1-1}{1-r} = \frac{0}{1-r} = 0$$

We have proven that the complex sum $\sum_{n=0}^{N-1} r^n$ is 0.

$$S_c - S_s = \text{Re}(0) = 0$$

Therefore:

$S_c - S_s = 0 \implies S_c = S_s$

Combining with $S_c + S_c = N$, we have $S_c + S_c = N \implies \mathbf{S_c = N/2}$ and $\mathbf{S_s = N/2}$.

$\text{Var}(A(f_k)) = \sigma^2(N/2)$

$\text{Var}(B(f_k)) = \sigma^2(N/2)$

- Case $2 : k = 0$

  $S_c = \sum_{n=0}^{N-1} \cos^2(0) = \sum_{n=0}^{N-1} 1^2 = N$

  $S_s = \sum_{n=0}^{N-1} \sin^2(0) = \sum_{n=0}^{N-1} 0^2 = 0$

  $\text{Var}(A(f_0)) = \sigma^2 N$

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

- Case 3 : $k = N/2$

  $S_c = \sum_{n=0}^{N-1} \cos^2(2\pi(N/2)n/N) = \sum_{n=0}^{N-1} \cos^2(\pi n) = \sum_{n=0}^{N-1} 1 = N$

  $S_s = \sum_{n=0}^{N-1} \sin^2(2\pi(N/2)n/N) = \sum_{n=0}^{N-1} \sin^2(\pi n) = 0$

  $\text{Var}(A(f_{N/2})) = \sigma^2 N$

  $\text{Var}(B(f_{N/2})) = 0$

In summary, we have :

$$\boxed{\begin{array}{l} \text{Var}(A(f_k)) = \begin{cases} \sigma^2 N & \text{if } k = 0 \text{ or } k = N/2 \\ \sigma^2 \frac{N}{2} & \text{if } 0 < k < N/2 \end{cases} \\ \\ \text{Var}(B(f_k)) = \begin{cases} 0 & \text{if } k = 0 \text{ or } k = N/2 \\ \sigma^2 \frac{N}{2} & \text{if } 0 < k < N/2 \end{cases} \end{array}}$$

**II. the distribution of the periodogram values.**

$X_n$ is a Gaussian white noise. The terms $A(f_k)$ and $B(f_k)$ are linear combinations of the Gaussian variables $X_n$. Any linear combination of Gaussian variables is also Gaussian. Therefore, $\mathbf{A}(\mathbf{f_k})$ and $\mathbf{B}(\mathbf{f_k})$ are Gaussian variables.

From the previous exercise, we know their means are 0.

$A(f_k) \sim \mathcal{N}(0, \text{Var}(A(f_k)))$

$B(f_k) \sim \mathcal{N}(0, \text{Var}(B(f_k)))$

The Periodogram Formula is $|J(f_k)|^2 = \left| \frac{1}{\sqrt{N}}(A(f_k) + iB(f_k)) \right|^2$.

$$|J(f_k)|^2 = \frac{1}{N} \left( A(f_k)^2 + B(f_k)^2 \right)$$

The distribution of the periodogram is therefore the distribution of a sum of squared Gaussian variables.

Let's recall the definition of a chi-squared distribution :

If $Z_1, \ldots, Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$X = \sum_{i=1}^{k} Z_i^2$$

is distributed according to the chi-squared distribution with $k$ degrees of freedom. This is usually denoted as

$$X \sim \chi^2(k) \text{ or } X \sim \chi_k^2.$$

- Case 1: The General Frequencies ($0 < k < N/2$)

  Let $Z_A = \frac{A(f_k)}{\sqrt{\text{Var}(A(f_k))}} = \frac{A(f_k)}{\sqrt{\sigma^2 N/2}}$. Then $Z_A \sim \mathcal{N}(0, 1)$.

Let $Z_B = \frac{B(f_k)}{\sqrt{\mathrm{Var}(B(f_k))}} = \frac{B(f_k)}{\sqrt{\sigma^2 N/2}}$. Then $Z_B \sim \mathcal{N}(0,1)$.

Now, rewrite $A(f_k)^2$ and $B(f_k)^2$ in terms of $Z_A$ and $Z_B$:

$A(f_k)^2 = (\sigma^2 N/2) \cdot Z_A^2 B(f_k)^2 = (\sigma^2 N/2) \cdot Z_B^2$

Substitute these into the periodogram formula:

$$|J(f_k)|^2 = \frac{1}{N}\left(A(f_k)^2 + B(f_k)^2\right)$$

$$|J(f_k)|^2 = \frac{1}{N}\left((\sigma^2 N/2)Z_A^2 + (\sigma^2 N/2)Z_B^2\right)$$

$$|J(f_k)|^2 = \frac{\sigma^2 N}{2N}\left(Z_A^2 + Z_B^2\right)$$

$$|J(f_k)|^2 = \frac{\sigma^2}{2}\left(Z_A^2 + Z_B^2\right)$$

Hence, $|J(f_k)|^2 \sim \frac{\sigma^2}{2}\chi_2^2$

- Case 2: The Boundary Frequencies ($k = 0$ and $k = N/2$)

  For these two special frequencies, we found:$\mathrm{Var}(A(f_k)) = \sigma^2 N$ and $\mathrm{Var}(B(f_k)) = 0$

  Since $\mathbb{E}[B(f_k)] = 0$ and $\mathrm{Var}(B(f_k)) = 0$, $B(f_k)$ is not a random variable; it is deterministically 0.

  The periodogram formula simplifies:

$$|J(f_k)|^2 = \frac{A(f_k)^2}{N}$$

  Let's standardize $A(f_k)$:Let $Z_A = \frac{A(f_k)}{\sqrt{\mathrm{Var}(A(f_k))}} = \frac{A(f_k)}{\sqrt{\sigma^2 N}}$. Then $Z_A \sim \mathcal{N}(0,1)$.

  Rewrite $A(f_k)^2 = (\sigma^2 N) \cdot Z_A^2$

  Then

$$|J(f_k)|^2 = \frac{(\sigma^2 N) \cdot Z_A^2}{N}$$
$$|J(f_k)|^2 = \sigma^2 Z_A^2$$
$$|J(f_k)|^2 \sim \sigma^2 \chi_1^2$$

The periodogram values $|J(f_k)|^2$ follow a scaled Chi-squared distribution.

$$\boxed{|J(f_k)|^2 \sim \begin{cases} \sigma^2\chi_1^2 & \text{if } k = 0 \text{ or } k = N/2 \\ \\ \frac{\sigma^2}{2}\chi_2^2 & \text{if } 0 < k < N/2 \end{cases}}$$

**III. Variance of the $|J(f_k)|^2$ and conclusion on the consistency of the periodogram.**

**1. Calculation of the variance of the $|J(f_k)|^2$**

The variance of a chi-squared distribution is $\text{Var}(\chi_k^2) = 2k$.

- Case 1: General Frequencies ($0 < k < N/2$) $\text{Var}(|J(f_k)|^2) = \text{Var}\left(\frac{\sigma^2}{2}\chi_2^2\right) = \left(\frac{\sigma^2}{2}\right)^2 \text{Var}(\chi_2^2) = \frac{(\sigma^2)^2}{4} \cdot (4) = (\sigma^2)^2$
- Case 2: Boundary Frequencies ($k = 0$ and $k = N/2$) $\text{Var}(|J(f_k)|^2) = \text{Var}(\sigma^2\chi_1^2) = (\sigma^2)^2\text{Var}(\chi_1^2) = 2(\sigma^2)^2$

Therefore,

$$\boxed{\text{Var}(|J(f_k)|^2) = \begin{cases} 2(\sigma^2)^2 & \text{if } k = 0 \text{ or } k = N/2 \\ (\sigma^2)^2 & \text{if } 0 < k < N/2 \end{cases}}$$

**2. Conclusion: the periodogram is not consistent**

Notice that in both cases, the variance does not depend on $N$. Therefore, the limit is:

$$\lim_{N\to\infty} \text{Var}(|J(f_k)|^2) = \lim_{N\to\infty} (\sigma^2)^2 = (\sigma^2)^2 \neq 0$$

Since the variance does not tend to zero, we can conclude that the periodogram $|J(f_k)|^2$ is not a consistent estimator of the power spectrum. This confirms our observation from the plots (question 6).

**IV. Explaination of the erratic behavior of the periodogram in Question 6 by looking at the covariance between the $|J(f_k)|^2$.**

Because the Gaussian white noise process consists of independent variables in the time domain (a property we observed with the autocovariance), a fundamental result of applying the Discrete Fourier Transform is that the resulting frequency components ($|J(f_k)|^2$ and $|J(f_l)|^2$) are also independent. This means:

$$\text{Cov}\left(|J(f_k)|^2, |J(f_l)|^2\right) = \mathbb{E}\left[|J(f_k)|^2|J(f_l)|^2\right] - \mathbb{E}\left[|J(f_k)|^2\right]\mathbb{E}\left[|J(f_l)|^2\right]$$

$$\text{Cov}\left(|J(f_k)|^2, |J(f_l)|^2\right) = \left(\mathbb{E}\left[|J(f_k)|^2\right]\mathbb{E}\left[|J(f_l)|^2\right]\right) - \left(\mathbb{E}\left[|J(f_k)|^2\right]\mathbb{E}\left[|J(f_l)|^2\right]\right) = \mathbf{0} \quad \text{for } k \neq l$$

The estimate we get at frequency $f_k$ gives us no information about the estimate we will get at the very next frequency, $f_{k+1}$. Hence, the plot of the mean curve jumps unpredictably from one point to the next. Moreover, the constant variance at each frequency $f_k$ explains the large vertical spread (the wide blue band), meaning each estimate can be far from the mean. Because the adjacent estimates are independent, the jumps between them can be large. And finally as N increases, the resolution in frequence increases and we observes more jumps in the periodogram.

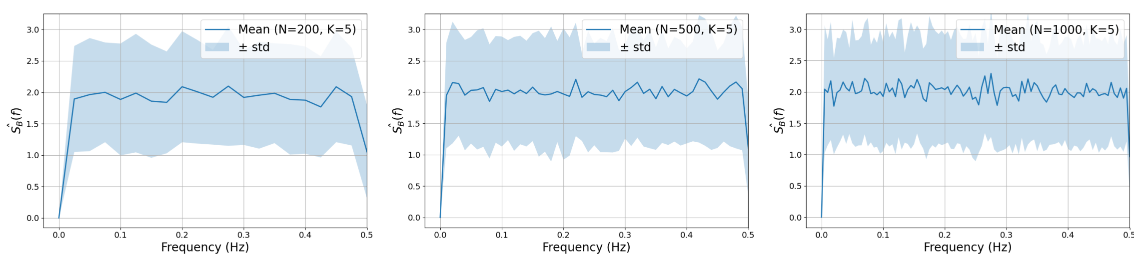
### Question 9

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 into $K$ 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).

Both the raw periodogram (Question 6) and Bartlett's estimator (Question 9) show the mean centered around 2, which is the correct theoretical value for the one-sided Power Spectral Density $S(f)$. The main difference is in the variance.

The raw periodogram plots from Question 6 show a very large $\pm$ standard deviation band, confirming their high variance. This is consistent with the theory that the variance of the raw periodogram is approximately the square of its mean: $\text{Var}(\hat{S}(f)) \approx (S(f))^2 \approx (2.0)^2 = 4.0$.

Indeed from Question 5 we have $\mathbb{E}[|J(f_k)|^2] \approx S(f)$ (The periodogram is an asymptotically unbiased estimator for the power spectrum $S(f)$) and from Question 8 $|J(f_k)|^2 \sim \frac{\sigma^2}{2}\chi^2_2$ and $\text{Var}(|J(f_k)|^2) = (\sigma^2)^2$ (for $0 < k < N/2$).

Let's calculate the mean (Expectation) from this distribution:

$$\mathbb{E}[|J(f_k)|^2] = \mathbb{E}\left[\frac{\sigma^2}{2}\chi_2^2\right]$$

Using the property $\mathbb{E}[\chi_k^2] = k$, for $k = 2$:

$$\mathbb{E}[|J(f_k)|^2] = \frac{\sigma^2}{2}\mathbb{E}[\chi_2^2] = \frac{\sigma^2}{2}(2) = \sigma^2$$

Hence

$$\mathrm{Var}(|J(f_k)|^2) = (\sigma^2)^2 \implies \mathrm{Var}(|J(f_k)|^2) = (S(f))^2$$

Bartlett's estimator (Question 9) shows a drastically reduced variance. This is because its variance is the raw variance divided by $K$: $\mathrm{Var}(\hat{S}_B(f)) \approx \mathrm{Var}(\hat{S}(f))/K = 4.0/5 = 0.8$.

Indeed $\hat{S}_B(f) = \frac{1}{K}\sum \hat{S}_i(f)$ :

$$\mathrm{Var}(\hat{S}_B(f)) = \mathrm{Var}\left(\frac{1}{K}\sum \hat{S}_i(f)\right) = \frac{1}{K^2}\cdot\left(\sum_{i=1}^{K}\mathrm{Var}(\hat{S}_i(f))\right)$$

Each segment has the same variance (let's say $\mathrm{Var}(S\hat{(f)})$), hence :

$$\mathrm{Var}(\hat{S}_B(f)) = \frac{1}{K^2}\cdot\left(K\cdot\mathrm{Var}(S\hat{(f)})\right) = \frac{\mathrm{Var}(S\hat{(f)})}{K}$$

Our plots confirm these values (a variance of $\approx 4.0$ for the periodogram vs. $\approx 0.8$ for Bartlett's method), which proves that averaging the $K = 5$ sections works and makes the estimator much more stable.

This also clearly illustrates the method's trade-off. The variance, Var $\approx S(f)^2/K$, only depends on $K$. In our experiment, we fixed $K = 5$, which is why the variance remained constant ($\approx 0.8$) even when $N$ increased. By keeping $K$ constant, we used the extra data from increasing $N$ only to improve the frequency resolution (since the segment length $M = N/K$ increased).

This is the core of the trade-off: to make the variance tend to 0, we must increase $K$, but as $K$ increases, the segment size $M = N/K$ gets smaller, which degrades the frequency resolution ($\Delta f = 1/M$).

# 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 falls. 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 and 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 Dynamic Time Warping (DTW) distance was combined with a k-nearest neighbors (k-NN) classifier to distinguish healthy from non-healthy footsteps. Using a 5-fold cross-validation, the optimal number of neighbors was found to be k = 5, with a mean F1-score of 0.78. When applied to the test set, the model achieved an overall F1-score of 0.51, with good detection of non-healthy steps (F1 = 0.51) but poor performance on healthy ones (F1 = 0.13). The overall accuracy reached 0.38, confirming a bad generalization ability.

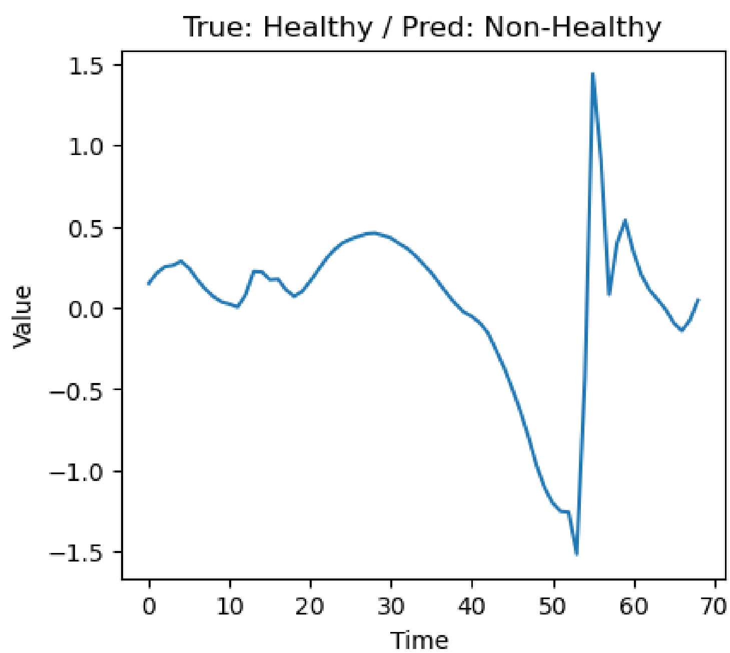
**Possible explanations for the poor generalization**

- **Overfitting** : the classifier overfits the dataset.
- **Representativeness** : the training dataset may not fully capture the diversity of walking patterns across subjects, leading to poor generalization.
- **Model simplicity** : the training dataset may not fully capture the diversity of walking patterns across subjects, leading to poor generalization.

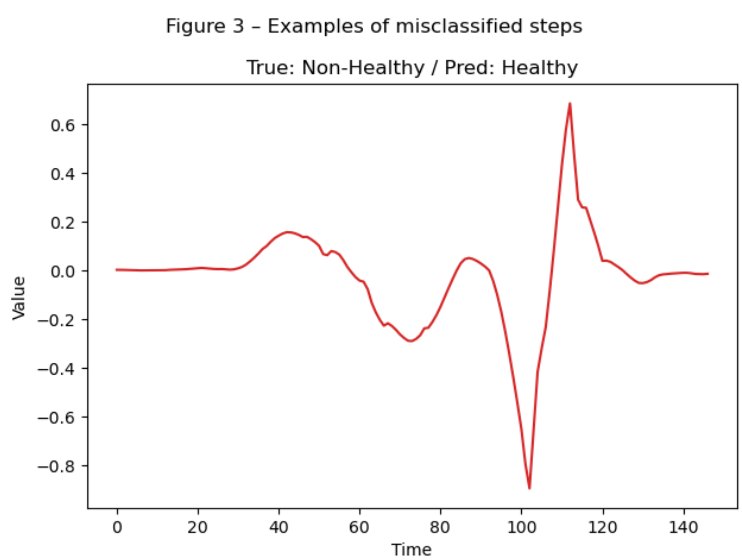
### Question 11

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).