

Assignment 1 (ML for TS) - MVA

Colin Coërchon colin180302@gmail.com
Chloé Habasque habasquechloe29@gmail.com

28 Octobre 2024

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 Tuesday 28th October 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: [LINK](#).

2 Convolution dictionary learning

Question 1

Consider the following Lasso regression:

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

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

Dans le problème de régression Lasso, notre objectif est de minimiser la fonction suivante :

$$\frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

Pour que $\beta = \mathbf{0}_p$ soit la solution optimale, cette valeur doit minimiser la fonction objectif. Cela signifie que si nous évaluons la fonction objectif en tout $\beta \neq \mathbf{0}_p$, elle doit être supérieure ou égale à la valeur obtenue avec $\beta = \mathbf{0}_p$.

En posant $\beta = \mathbf{0}_p$, la fonction objectif devient simplement :

$$\frac{1}{2} \|y\|_2^2$$

Pour tout $\beta \neq \mathbf{0}_p$, nous devons donc avoir :

$$\frac{1}{2} \|y\|_2^2 \leq \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

En réarrangeant les termes, cela nous donne :

$$\lambda \|\beta\|_1 \geq \frac{1}{2} (\|y\|_2^2 - \|y - X\beta\|_2^2)$$

Et donc, $\forall \beta \neq \mathbf{0}_p$, on doit avoir :

$$\lambda \geq \frac{1}{2\|\beta\|_1} (\|y\|_2^2 - \|y - X\beta\|_2^2)$$

Cela conduit à la borne suivante :

$$\lambda_{\max} = \max_{\beta \in \mathbb{R}^p \setminus \{\mathbf{0}_p\}} \left\{ \frac{1}{2\|\beta\|_1} (\|y\|_2^2 - \|y - X\beta\|_2^2) \right\}$$

Bien que cette expression pour $\lambda_{\max}$ soit correcte, elle est difficile à manipuler en pratique, car elle nécessite de connaître β et de maximiser sur toutes les valeurs possibles de $\beta \neq \mathbf{0}_p$. Une approche plus simple consiste à utiliser une **condition suffisante** pour $\lambda_{\max}$ basée sur la norme infinie $\|X^T y\|_\infty$.

" $X^T y$ " correspond aux covariances entre chaque colonne de X et y , c'est-à-dire la mesure de la corrélation entre chaque prédicteur X_j et la réponse y . En choisissant $\lambda \geq \|X^T y\|_\infty$, on garantit que l'influence de chaque X_j est dominée par la pénalité λ . Cela force chaque composante β_j à être nulle, assurant ainsi que $\beta = \mathbf{0}_p$ est la solution optimale.

On préfère prendre donc :

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

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:

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

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

1. Pour exprimer la somme des convolutions $\sum_{k=1}^K \mathbf{z}_k * \mathbf{d}_k$ sous forme matricielle $\mathbf{D}\mathbf{z}$, nous devons définir en précision cet opérateur de convolution.

- **Dimensions :**

- Le vecteur de signal $\mathbf{x} \in \mathbb{R}^n$ a $n = N$ lignes.
- Chaque atome du dictionnaire $\mathbf{d}_k \in \mathbb{R}^L$ a L éléments.
- Chaque vecteur d'activation $\mathbf{z}_k \in \mathbb{R}^{N-L+1}$.

- **Formulation de la convolution :** La i -ème composante de la convolution $(\mathbf{z}_k * \mathbf{d}_k)_i$ est donnée par :

$$(\mathbf{z}_k * \mathbf{d}_k)_i = \sum_{m=0}^{L-1} z_k[i-m] \cdot d_k[m].$$

Cette expression peut être vue comme un produit entre une matrice de convolution $\mathbf{D}_k$ et le vecteur $\mathbf{z}_k$.

- **Construction de $\mathbf{D}$ et $\mathbf{z}$:**

- Pour chaque $\mathbf{d}_k$, nous construisons une matrice de convolution $\mathbf{D}_k$ de dimension $n \times (N-L+1)$. La matrice $\mathbf{D}_k$ est de type **Toeplitz**, construite en plaçant chaque décalage de $\mathbf{d}_k$ dans ses colonnes successives. Par exemple, si $\mathbf{d}_k = [d_k[0], d_k[1], \dots, d_k[L-1]]$, alors :

$$\mathbf{D}_k = \begin{pmatrix} d_k[0] & 0 & \dots & 0 \\ d_k[1] & d_k[0] & \ddots & 0 \\ \vdots & d_k[1] & \ddots & \vdots \\ d_k[L-1] & \vdots & \ddots & d_k[0] \\ 0 & d_k[L-1] & \dots & d_k[1] \end{pmatrix}.$$

Cette matrice applique les décalages nécessaires pour la convolution de $\mathbf{z}_k$ avec $\mathbf{d}_k$.

- La matrice complète $\mathbf{D}$ est ensuite formée en concaténant horizontalement les matrices $\mathbf{D}_k$ pour chaque $k = 1, \dots, K$, ce qui donne :

$$\mathbf{D} = (\mathbf{D}_1 \quad \mathbf{D}_2 \quad \dots \quad \mathbf{D}_K).$$

Ainsi, $\mathbf{D}$ est de dimension $n \times K(N - L + 1)$.

- Le vecteur $\mathbf{z} \in \mathbb{R}^{K(N-L+1)}$ est construit en concaténant les vecteurs d'activation $\mathbf{z}_k$ les uns à la suite des autres :

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_K \end{pmatrix}.$$

Avec cette construction, le produit $\mathbf{D}\mathbf{z}$ est de dimension $n \times 1$, ce qui permet d'écrire $\|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2$, compatible avec la dimension de $\mathbf{x}$.

Ainsi, le problème de codage parcimonieux peut être formulé comme une **régression Lasso**, où le vecteur de réponse est le signal $\mathbf{x}$ et où la matrice de conception est la matrice $\mathbf{D}$, construite à partir des convolutions des atomes du dictionnaire $\mathbf{d}_k$ avec les signaux d'activation $\mathbf{z}_k$.

2. Par analogie avec la première question, on prendra donc :

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

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:

$$S(f) := \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s}. \quad (3)$$

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

On suppose que X_n est un bruit blanc Gaussien. Par conséquent, on sait que les (X_n) sont indépendants, $\forall n = 0, \dots, N - 1$, $\mathbb{E}[X_n] = 0$ et $\mathbb{V}[X_n] = \mathbb{E}[X_n^2] = \sigma^2$.

La fonction d'autocovariance est définie par $\gamma(\tau) \triangleq \mathbb{E}[X_n X_{n+\tau}]$.

Par conséquent,

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

et

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

car X_n et $X_{n+\tau}$ sont indépendantes.

La densité spectrale de puissance est alors :

$$\begin{aligned} S(f) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-2\pi f \tau / f_s} \\ &= \gamma(0) \\ &= \sigma^2 \end{aligned}$$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \quad (4)$$

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

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

On en déduit que $\hat{\gamma}(\tau)$ est un estimateur biaisé. Mais :

$$\mathbb{E} [\hat{\gamma}(\tau)] \xrightarrow{N \rightarrow +\infty} \gamma(\tau)$$

Donc $\hat{\gamma}(\tau)$ est asymptotiquement non biaisé. On peut alors choisir

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

comme estimateur non biaisé.

Question 5

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

$$J(f) := (1/\sqrt{N}) \sum_{n=0}^{N-1} X_n e^{-2\pi i f n / f_s} \quad (5)$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \dots, |J(f_{N/2})|^2$ 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

1. Premièrement, en posant les fréquences de Fourier discrètes $f_k = \frac{f_s k}{N}$ pour $k = 0, 1, \dots, N-1$, on peut écrire :

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

On a alors :

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

Nous posons $\tau = n - m$, où τ peut varier de $-(N-1)$ à $N-1$. C'est-à-dire qu'on effectue la partition de $\llbracket 0, N-1 \rrbracket^2$ suivante :

$$\llbracket 0, N-1 \rrbracket^2 = \bigcup_{\tau=-(N-1)}^{N-1} \{(n, m) \in \mathbb{N}^2 \mid n - m = \tau\} \stackrel{\text{not.}}{=} \bigcup_{\tau=-(N-1)}^{N-1} A_\tau$$

En réorganisant les termes de la somme en fonction de τ , nous obtenons alors :

$$\begin{aligned}
 |J(f_k)|^2 &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} X_n X_m e^{\frac{-2\pi i k(n-m)}{N}} \\
 &= \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{(n,m) \in A_\tau} X_n X_{n-\tau} \right) e^{\frac{-2\pi i k \tau}{N}} \\
 &= \sum_{\tau=-(N-1)}^{N-1} \left(\frac{1}{N} \sum_{n=|\tau|}^{N-1} X_n X_{n-\tau} \right) e^{\frac{-2\pi i k \tau}{N}}
 \end{aligned}$$

On rappelle que la covariance empirique $\hat{\gamma}(\tau)$ s'écrit :

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

Ainsi, nous pouvons exprimer $|J(f_k)|^2$ en fonction de la covariance empirique comme suit :

$$|J(f_k)|^2 = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}(\tau) e^{\frac{-2\pi i k \tau}{N}}$$

2. Pour montrer que $|J(f^{(N)})|^2$ est un estimateur asymptotiquement non biaisé de $S(f)$ pour $f > 0$, nous calculons l'espérance de $|J(f^{(N)})|^2$ en utilisant la définition de $\hat{\gamma}(\tau)$ obtenue dans la première question :

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

Sous l'hypothèse de stationnarité du processus, l'espérance de $\hat{\gamma}(\tau)$ converge vers la fonction d'autocovariance réelle $\gamma(\tau)$ lorsque $N \rightarrow \infty$, i.e. :

$$\mathbb{E} [\hat{\gamma}(\tau)] \xrightarrow{N \rightarrow +\infty} \gamma(\tau) \quad (\text{cf. question 4})$$

Et de plus,

- Par définition de $f^{(N)}$ comme étant choisi comme la fréquence de Fourier la plus proche de f parmi les fréquences discrètes f_k , on a : $f^{(N)} \xrightarrow{N \rightarrow +\infty} f$.
- On a également supposé que les autocovariances sont absolument sommables : $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$.
Comme $\left| \gamma(\tau) e^{-\frac{2\pi i f \tau}{f_s}} \right| = |\gamma(\tau)|$, cette hypothèse garantit que nous pouvons **passer à la limite** dans les sommes partielles.

Ainsi, lorsque $N \rightarrow +\infty$, on a :

$$\mathbb{E} \left[|J(f^{(N)})|^2 \right] = \sum_{\tau=-(N-1)}^{N-1} \mathbb{E} [\hat{\gamma}(\tau)] e^{-\frac{2\pi i f^{(N)} \tau}{f_s}} \xrightarrow{N \rightarrow +\infty} \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) e^{-\frac{2\pi i f \tau}{f_s}}$$

Par définition, cette somme est la densité spectrale de puissance $S(f)$ pour la fréquence f . Par conséquent, on peut conclure que :

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

Cela signifie donc que $|J(f^{(N)})|^2$ est bien un **estimateur asymptotiquement non biaisé** de $S(f)$ pour $f > 0$.

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 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.

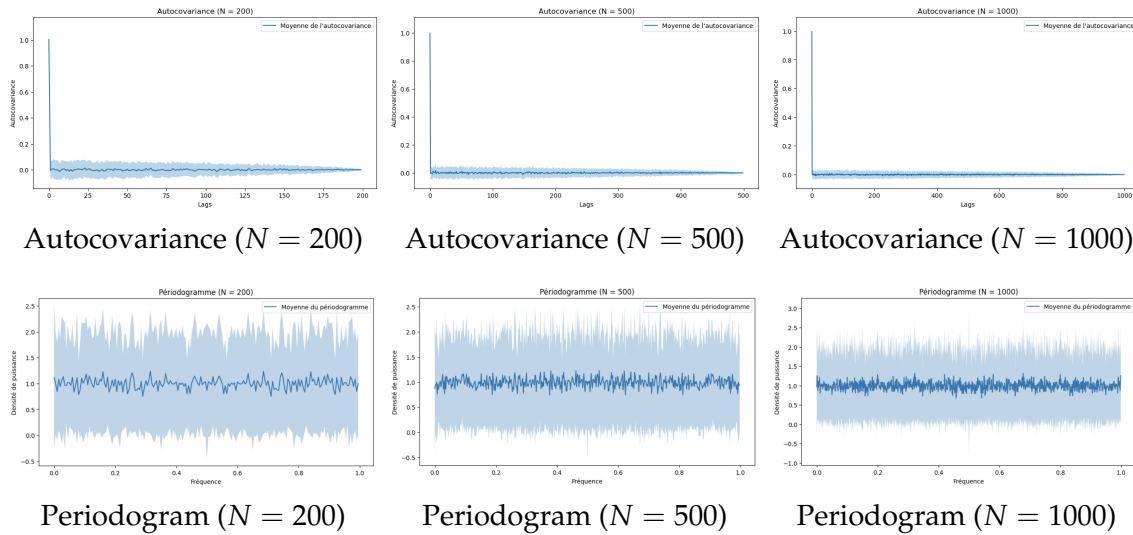


Figure 1: Autocovariances and periodograms of a Gaussian white noise (see Question 6).

Answer 6

- **Autocovariances :**

Pour $\tau = 0$, l'autocovariance vaut bien $\sigma^2 = 1$ et pour les autres valeurs de τ , la valeur moyenne se rapproche de 0, comme les valeurs de $\gamma(\tau)$. Cela illustre le fait que $\hat{\gamma}(\tau)$ est un estimateur de $\gamma(\tau)$ pour un bruit blanc gaussien.

De plus, on observe qu'en augmentant N , les variances diminuent. Ce qui semblerait indiquer que $\hat{\gamma}(\tau)$ serait un estimateur consistant de $\gamma(\tau)$.

- **Périodogrammes :**

On observe que la moyenne des périodogrammes est d'environ 1, ce qui correspond à la valeur de $S(f) = \sigma^2 = 1$ pour un bruit blanc gaussien. Mais contrairement à l'autocovariance, la variance des périodogrammes ne diminue pas en augmentant N . L'estimateur ne semble pas être consistant.

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 $\tau > 0$

$$\text{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) [\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)]. \quad (6)$$

(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

1. Nous savons que pour $\tau > 0$, l'estimateur de la covariance empirique $\hat{\gamma}(\tau)$ s'écrit :

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

La variance de $\hat{\gamma}(\tau)$ est donnée par :

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

Il va donc être nécessaire de calculer $\mathbb{E}[\hat{\gamma}(\tau)^2]$. On peut déjà remarquer en premier lieu que :

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

Et on peut donc développer notre calcul d'espérance :

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

En appliquant la propriété des processus gaussiens mentionnée dans l'indice de l'exercice, on peut exprimer cette espérance comme suit :

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

En utilisant la stationnarité de notre processus, et la définition de la fonction d'autocovariance, on obtient :

$$\begin{aligned} \mathbb{E}[X_n X_m] &= \gamma(n - m), & \mathbb{E}[X_{n+\tau} X_{m+\tau}] &= \gamma(n - m), \\ \mathbb{E}[X_n X_{m+\tau}] &= \gamma(n - m - \tau), & \text{et} & \quad \mathbb{E}[X_{n+\tau} X_m] = \gamma(n - m + \tau). \end{aligned}$$

Cela nous donne :

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

En substituant cette expression dans la somme de $\mathbb{E}[\hat{\gamma}(\tau)^2]$, nous obtenons :

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

Pour développer cette somme, on effectue le changement de variables $h = n - m$. Pour être plus précis, la sommation double s'effectue dans l'ensemble $\llbracket 0, N - \tau - 1 \rrbracket^2$, et nous le partitionnons de cette manière :

$$\llbracket 0, N - \tau - 1 \rrbracket^2 = \bigcup_{h=-(N-\tau-1)}^{N-\tau-1} \{(n, m) \in \mathbb{N}^2 \mid n - m = h\} \stackrel{\text{not.}}{=} \bigcup_{h=-(N-\tau-1)}^{N-\tau-1} A_h$$

On obtient alors :

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\tau)^2] &= \frac{1}{N^2} \sum_{n=0}^{N-\tau-1} \sum_{m=0}^{N-\tau-1} \gamma^2(n-m) + \gamma(n-m-\tau)\gamma(n-m+\tau) + \gamma^2(\tau) \\
&= \frac{1}{N^2} \sum_{h=-(N-\tau-1)}^{N-\tau-1} \sum_{(n,m) \in A_h} \gamma^2(h) + \gamma(h-\tau)\gamma(h+\tau) + \gamma^2(\tau) \\
&= \frac{1}{N^2} \sum_{h=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|h|) [\gamma^2(h) + \gamma(h-\tau)\gamma(h+\tau) + \gamma^2(\tau)] \\
&\quad (|A_h| = N-\tau-|h|) \\
&= \frac{1}{N} \sum_{h=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|h|}{N}\right) [\gamma^2(h) + \gamma(h-\tau)\gamma(h+\tau) + \gamma^2(\tau)] \\
&= \frac{1}{N} \sum_{h=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|h|}{N}\right) [\gamma^2(h) + \gamma(h-\tau)\gamma(h+\tau)] + \gamma^2(\tau) \\
&\quad (\text{car } \gamma^2(\tau) \text{ ne dépend pas de } h \text{ et la somme vaut } 1)
\end{aligned}$$

En utilisant cette expression, on en déduit que :

$$\begin{aligned}
\text{var}(\hat{\gamma}(\tau)) &= \mathbb{E}[\hat{\gamma}(\tau)^2] - \gamma^2(\tau), \\
\Rightarrow \quad \text{Var}(\hat{\gamma}(\tau)) &= \frac{1}{N} \sum_{n=-(N-\tau-1)}^{N-\tau-1} \left(1 - \frac{\tau+|n|}{N}\right) [\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)]
\end{aligned}$$

2. Nous venons d'établir que :

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

Lorsque N devient grand, le facteur $\frac{1}{N}$ devant la somme implique que la variance $\text{Var}(\hat{\gamma}(\tau))$ pourrait tendre vers 0, à condition que la somme à l'intérieur soit **bornée**. Nous allons maintenant montrer que c'est bien le cas.

(a) **Sommabilité absolue de $\gamma(\tau)$** : Puisque $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$, il existe une constante C_1 telle que :

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| \leq C_1.$$

(b) **Sommabilité quadratique de $\gamma(\tau)$** : De même, l'hypothèse $\sum_{\tau \in \mathbb{Z}} \gamma^2(\tau) < \infty$ implique qu'il existe une constante C_2 telle que :

$$\sum_{h \in \mathbb{Z}} \gamma^2(h) \leq C_2.$$

En utilisant ces deux hypothèses, nous pouvons majorer chaque terme de la somme dans l'expression de $\text{Var}(\hat{\gamma}(\tau))$:

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

Puisque $\left(1 - \frac{\tau + |h|}{N}\right) \leq 1$ pour $h \in \llbracket -(N-\tau-1), N-\tau-1 \rrbracket$, il vient que :

$$\text{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N} \sum_{h=-(N-\tau-1)}^{N-\tau-1} (\gamma^2(h) + |\gamma(h-\tau)\gamma(h+\tau)|)$$

Maintenant, en étendant cette somme à $\mathbb{Z}$ (grâce à la décroissance de $\gamma(\tau)$ et le fait que les termes restants sont négligeables pour N grand), nous obtenons :

$$\text{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N} \left(\sum_{h \in \mathbb{Z}} \gamma^2(h) + \sum_{h \in \mathbb{Z}} |\gamma(h-\tau)\gamma(h+\tau)| \right)$$

En utilisant la sommabilité quadratique de $\gamma(\tau)$, la première somme est bornée par C_2 . Pour la deuxième somme, on utilise l'inégalité de Cauchy-Schwarz :

$$\sum_{h \in \mathbb{Z}} |\gamma(h-\tau)\gamma(h+\tau)| \leq \sqrt{\sum_{h \in \mathbb{Z}} \gamma^2(h-\tau)} \sqrt{\sum_{h \in \mathbb{Z}} \gamma^2(h+\tau)} \leq \sqrt{C_2} \sqrt{C_2} = C_2$$

Ainsi, on obtient :

$$\text{Var}(\hat{\gamma}(\tau)) \leq \frac{1}{N} (C_2 + C_2) = \frac{2C_2}{N} \xrightarrow{N \rightarrow +\infty} 0$$

Nous avons donc montré que :

$$\boxed{\text{Var}(\hat{\gamma}(\tau)) \xrightarrow{N \rightarrow +\infty} 0}$$

De plus,

- D'après de le théorème de **Bienaymé-Tchebychev**,

$$\mathbb{P}(|\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq \epsilon^2) \leq \frac{\text{Var}(\hat{\gamma}(\tau))}{\epsilon^2}$$

- Et le fait que $\mathbb{E}[\hat{\gamma}(\tau)] \xrightarrow{N \rightarrow +\infty} \gamma(\tau)$ (d'après la question 4).

Nous arrivons donc au résultat final :

$$\begin{aligned} & \mathbb{P}(|\hat{\gamma}(\tau) - \mathbb{E}[\hat{\gamma}(\tau)]| \geq \epsilon^2) \xrightarrow{N \rightarrow +\infty} 0 \\ \implies & \mathbb{P}(|\hat{\gamma}(\tau) - \gamma(\tau)| \geq \epsilon^2) \xrightarrow{N \rightarrow +\infty} 0 \end{aligned}$$

Cela permet de prouver que $\hat{\gamma}(\tau)$ converge en probabilité vers $\gamma(\tau)$:

$$\boxed{\hat{\gamma}(\tau) \xrightarrow{\mathbb{P}} \gamma(\tau)}$$

Cela prouve la **consistance** de l'estimateur $\hat{\gamma}(\tau)$.

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 σ^2) and let $A(f) := \sum_{n=0}^{N-1} X_n \cos(-2\pi f n / f_s)$ and $B(f) := \sum_{n=0}^{N-1} X_n \sin(-2\pi f n / f_s)$. Observe that $J(f) = (1/N)(A(f) + iB(f))$.

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

Answer 8

1. Moyenne et Variance de $A(f)$ et $B(f)$

Puisque X est un bruit blanc Gaussien centré, chaque X_n est une variable aléatoire centrée, donc :

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

et de même,

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

Pour la variance, utilisons l'indépendance des X_n et la propriété $\mathbb{E}[X_n^2] = \sigma^2$:

$$\text{Var}(A(f_k)) = \mathbb{E}[A(f_k)^2] = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \mathbb{E}[X_n X_m] \cos\left(-2\pi \frac{kn}{N}\right) \cos\left(-2\pi \frac{km}{N}\right)$$

Comme X est un bruit blanc, on a $\mathbb{E}[X_n X_m] = 0$ pour $n \neq m$ et $\mathbb{E}[X_n^2] = \sigma^2$, ce qui donne :

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

De même, on obtient :

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

2. Nous n'avons pas eu le temps de traiter la suite de la question.

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

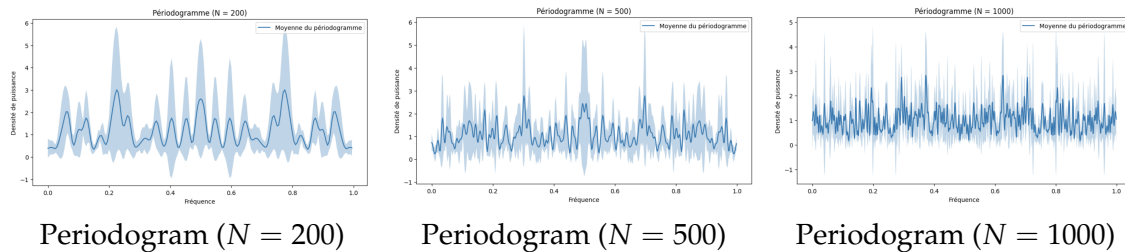


Figure 2: Bartlett's periodograms of a Gaussian white noise (see Question 9).

Les résultats obtenus à l'issue de la question 6 montraient un vrai problème dans la variance de l'estimateur $|J(f^{(N)})|^2$ pour estimer la densité spectrale de puissance $S(f)$.

Or, ici, en utilisant la procédure de Bartlett, on observe que la variance de nos estimations diminue en augmentant N , cette méthode apporte donc une solution à **l'absence de consistance de l'estimateur** $|J(f^{(N)})|^2$. De plus, la moyenne des périodogrammes est d'environ 1, ce qui correspond à la valeur de $S(f)$.

Ainsi, en utilisant la procédure de Bartlett, $|J(f^{(N)})|^2$ devient un bon estimateur de $S(f)$.

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

En combinant l'algorithme de classification KNN avec la distance Dynamic Time Warping (DTW) et en utilisant la méthode de validation croisée K-fold pour déterminer le nombre optimal de voisins, nous avons obtenu un **F-score de 0,46** en prenant en compte les **5 voisins** les plus proches pour classer les pas.

Le F-score de 0,46 indique que le modèle de classification KNN utilisant la distance DTW n'arrive pas à classer correctement les pas en *healthy/not healthy*. Cela peut en partie être dû au fait que, dans l'ensemble de test, 73 des 88 pas sont des pas de classe *not healthy*. Cette classe est donc surreprésentée dans le test.

Question 11

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

Answer 11

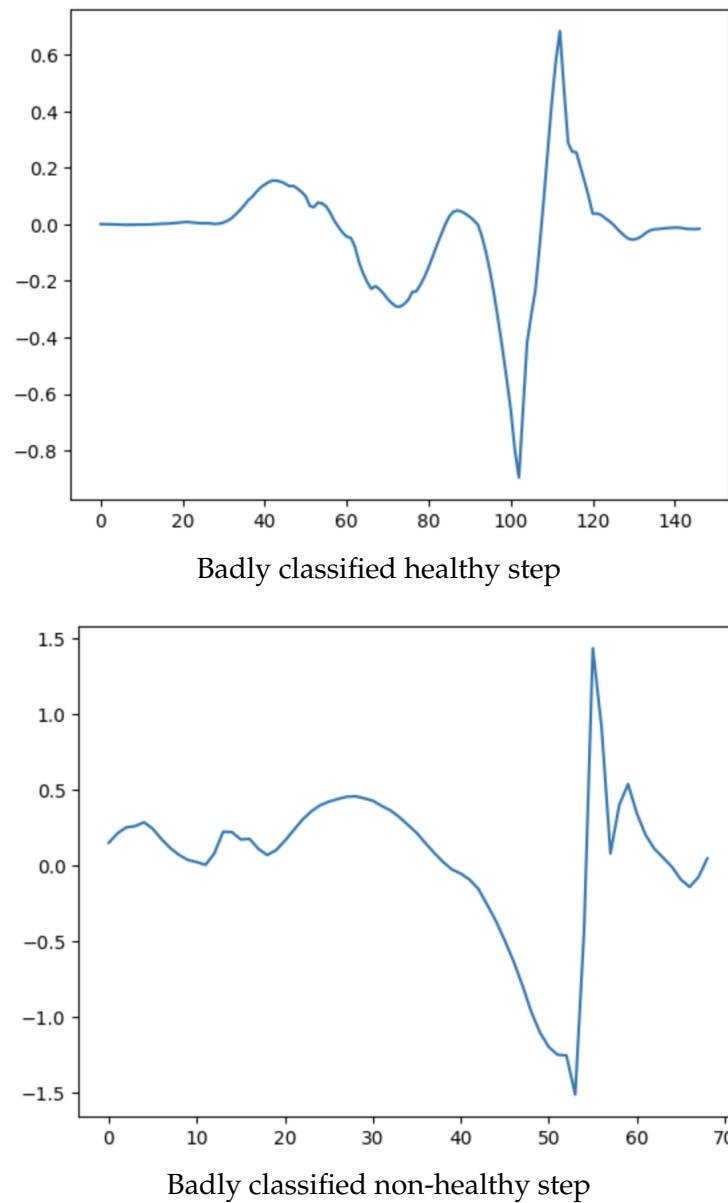


Figure 3: Examples of badly classified steps (see Question 11).