Assignment 1 (ML for TS) - MVA

Hugo Pavy hugo.pavy@etu.minesparis.psl.eu
Tom Mariani tom.mariani68@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^{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 $\lambda_{\max} = \|^t Xy\|_2$

Let $\epsilon > 0$, we assume that there exists $\beta_{min} \in \mathbb{R}^p$ with $\|\beta_{min}\|_2 > 0$, s.t

$$\forall \beta \in \mathbb{R}^p, \frac{1}{2}\|y - X\beta\|_2^2 + \lambda_{max}\|\beta\|_1 \geq \frac{1}{2}\|y - X\beta_{min}\|_2^2 + \lambda_{max}\|\beta_{min}\|_1 \tag{2}$$

We want to highlight a contraction because

$$\frac{1}{2}\|y - X\beta_{min}\|_2^2 + \lambda_{max}\|\beta_{min}\|_1 > \frac{1}{2}\|y\|_2^2 \tag{3}$$

In fact by developing the left term, replacing λ_{max}, and using that $\|.\|_1 \geq \|.\|_2$

$$\frac{1}{2}\|y - X\beta_{min}\|_2^2 + \lambda_{max}\|\beta_{min}\|_1 \geq \frac{1}{2}(\|y\|_2^2 + \|X\beta_{min}\|_2^2 - 2\langle y, X\beta_{min}\rangle) + \|^t Xy\|\,\|\beta_{min}\|_2^2 \tag{4}$$

Inside the product scalar, we transpose the matrix X and use Cauchy-Schwarz,

$$\geq \frac{1}{2}(\|y\|_2^2 + \|X\beta_{min}\|_2^2 - 2\|^t Xy\|\|\beta_{min}\|) + \|^t Xy\|\,\|\beta_{min}\|_2^2 = \frac{1}{2}(\|y\|_2^2 + \|X\beta_{min}\|_2^2) \tag{5}$$

Or $\|X\beta_{min}\|_2^2 > 0$ because by assumption $\|\beta_{min}\|_2 > 0$ (we can assume $X \neq 0_{\mathbb{R}^{n\times p}}$),

We have the contradiction (3) as we wanted, $\|\beta_{min}\|_2 = 0$, and therefore $\beta_{min} = 0_{\mathbb{R}^p}$

For a larger $\lambda > \lambda_{\max}$, all inequalities hold and there is also $\beta_{min} = 0_{\mathbb{R}^p}$

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:

$$\min_{(\mathbf{d}_k)_k,(\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\|\mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k\right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{6}$$

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

By using the convolution theorem, the Parseval theorem and padding with zero in order to equal the dimension, we have that:

$$\left\|\mathbf{x}-\sum_{k=1}^{K}\mathbf{z}_k * \mathbf{d}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\|\mathbf{z}_k\|_1=\left\|\hat{\mathbf{x}}-\sum_{k=1}^{K}\hat{\mathbf{z}}_k\odot\hat{\mathbf{d}}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\|\hat{\mathbf{z}}_k\|_1 \tag{7}$$

Where $\hat{.}$ is the Fourier transform and $\odot$ is the component wise product.

With, $\beta=\begin{bmatrix}\hat{\mathbf{z}}_1\\ \vdots\\ \hat{\mathbf{z}}_K\end{bmatrix}$, we have that

$$\left\|\hat{\mathbf{x}}-\sum_{k=1}^{K}\hat{\mathbf{z}}_k\odot\hat{\mathbf{d}}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\|\hat{\mathbf{z}}_k\|_1=\left\|\hat{\mathbf{x}}-\sum_{k=1}^{K}\hat{\mathbf{z}}_k\odot\hat{\mathbf{d}}_k\right\|_2^2+\lambda\|\beta\|_1 \tag{8}$$

Let's write:

$$D=\begin{bmatrix}\hat{\mathbf{d}}_\mathbf{1}^T\\ \vdots\\ \hat{\mathbf{d}}_\mathbf{K}^T\end{bmatrix}\in\mathbb{R}^{(K*N)} \tag{9}$$

And $X=diag(D,...,D)\in\mathbb{R}^{(K.N)*(K.N)}$

Finally we have that:

$$\left\|\hat{\mathbf{x}}-\sum_{k=1}^{K}\hat{\mathbf{z}}_k\odot\hat{\mathbf{d}}_k\right\|_2^2+\lambda\sum_{k=1}^{K}\|\hat{\mathbf{z}}_k\|_1=\|\hat{\mathbf{x}}-X\beta\|_2^2+\lambda\|\beta\|_1 \tag{10}$$

Where X is the design matrix and β is the response vector.

Finally, according to question 1, there exists λ_{max} such that the sparse codes are only 0 for any $\lambda>\lambda_{max}$.

Because our sparse coding problem is a lasso regression, we have:

$$\lambda_{max}=\left\|X^Ty\right\|_2 \tag{11}$$

3 Spectral feature

Let X_n $(n=0,\dots,N-1)$ be a weakly stationary random process with zero mean and autocovariance function $\gamma(\tau):=\mathbb{E}(X_nX_{n+\tau})$. 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:

$$S(f):=\sum_{\tau=-\infty}^{+\infty}\gamma(\tau)e^{-2\pi f\tau/f_s}. \tag{12}$$

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

Let's write the autocovariance $\gamma(\tau) = \mathbb{E}(X_n X_{n+\tau})$

For $\tau \neq 0, \gamma(\tau) = \mathbb{E}(X_n X_{n+\tau}) = \mathbb{E}(X_n)\mathbb{E}(X_{n+\tau}) = 0$ because we suppose that X_n has zero mean

For $\tau = 0, \gamma(\tau) = \mathbb{E}(X_n^2) = \sigma^2$ where σ is the standard deviation of X_n

Therefore, $S(f) = \gamma(0) = \sigma^2$

Question 4

A natural estimator for the autocorrelation function is the sample autocovariance

$$\hat{\gamma}(\tau) := (1/N) \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{13}$$

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

- Show that $\hat{\gamma}(\tau)$ is a biased estimator of $\gamma(\tau)$ but asymptotically unbiased. What would be a simple way to de-bias this estimator?

Answer 4

We can compute:

$$\mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau) = \mathbb{E}[\frac{1}{N} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}] - \gamma(\tau) \tag{14}$$

By linearity of the expectation:

$$\mathbb{E}[\frac{1}{N} \sum_{n=0}^{N-\tau-1} X_n X_{n+\tau}] - \gamma(\tau) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}] - \gamma(\tau) \tag{15}$$

Thus, we have that:

$$|\mathbb{E}[\hat{\gamma}(\tau)] - \gamma(\tau)| = \frac{\tau}{N}|\gamma(\tau)| \tag{16}$$

Finally, the equation shows that we have a biased estimator but asymptotically unbiased.

In order to de-bias the estimator, we could change $\hat{\gamma}$ into:

$$\hat{\gamma}(\tau) := \frac{1}{N-\tau}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau} \tag{17}$$

Question 5

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

$$J(f) := (1/\sqrt{N})\sum_{n=0}^{N-1} X_n e^{-2\pi i f n/f_s} \tag{18}$$

The *periodogram* is the collection of values $|J(f_0)|^2, |J(f_1)|^2, \ldots, |J(f_{N/2})|^2$ where $f_k = f_s k/N$. (They can be efficiently computed using the Fast Fourier Transform.)

- Write $|J(f_k)|^2$ as a function of the sample autocovariances.
- For a frequency f, define $f^{(N)}$ the closest Fourier frequency f_k to f. Show that $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$ for $f > 0$.

Answer 5

We have that:

$$\tag{19}$$

$$\begin{aligned}
|J(f_k)|^2 &= J(f_k).J(f_k)^* &(20)\\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{N-1} X_n X_j e^{\frac{2ik(n-j)\pi}{N}} &(21)
\end{aligned}$$

This sum is the same as creating an array with N-1 raws and N-1 columns where the column j and raw n would be $X_i X_j e^{\frac{2i(n-j)\pi}{N}}$

The sum we are computing is the sum of all the elements of this matrix. We can see that $N\hat{\gamma}(\tau)e^{\frac{2i(n-j)\pi}{N}}$ where $\tau = n - j$ will be the sum of the elements of the diagonals going from the coordinate (n, 0) to the coordinate (N, N-n). Same when we are going from (n, 0) up until (N-n, N) following the diagonal.

With $\gamma(-\tau) = \gamma(\tau)$, at the end our sum is equal to:

$$\begin{aligned}
|J(f_k)|^2 &= \frac{1}{N}\sum_{h=1-N}^{N-1} N\hat{\gamma}(h)e^{\frac{2ikh\pi}{N}} &(22)\\
&= \sum_{h=1-N}^{N-1} \hat{\gamma}(h)e^{\frac{2ikh\pi}{N}} &(23)
\end{aligned}$$

Then, we have $f^{(N)} = argmin_{f_k}(|f - f_k|r)$

We have $f^{(N)} \to f$, $\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty$ and because $\hat{\gamma}(\tau)$ is asymptotically unbiased, we have:

$$|J(f^{(N)})|^2 \to S(f) \tag{24}$$

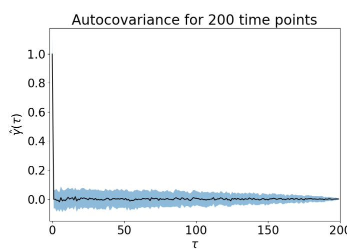
Thus $|J(f^{(N)})|^2$ is an asymptotically unbiased estimator of $S(f)$

Question 6

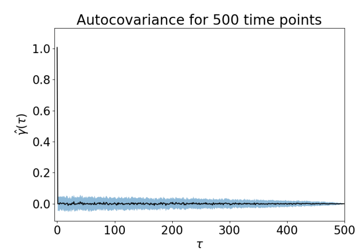
In this question, let X_n $(n = 0, \ldots, N-1)$ be a Gaussian white noise with variance $\sigma^2 = 1$ and set the sampling frequency to $f_s = 1$ Hz

- 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?

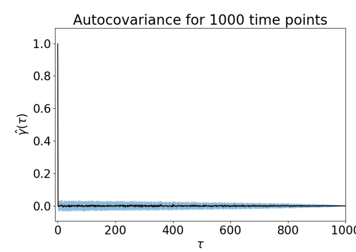
Add your plots to Figure 1.



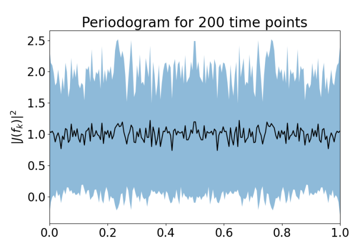
Autocovariance ($N = 200$)



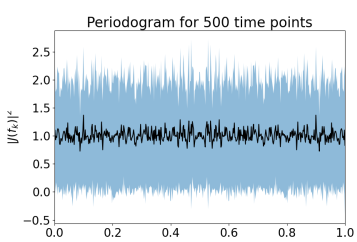
Autocovariance ($N = 500$)



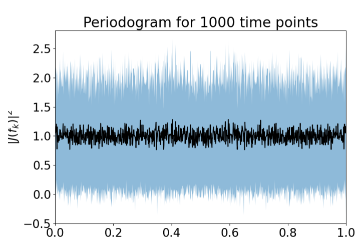
Autocovariance ($N = 1000$)



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

Answer 6

- The first thing to notice is that we are close to the theoretical solution of Q1, i.e the standard deviation for $\tau = 0$, which is the case here $(\sigma = 1)$ and 0 elsewhere. The second thing to notice is that the standard deviation decreases with the number of timepoints (or the sampling frequency depending on how we see it) which is a good sign to approach $\gamma(\tau)$ with $\hat{\gamma}(\tau)$

- The periodogram seems quite consistent with the number of timepoints, average value of 1 and standard deviation of 1 also.

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 $\tau > 0$

$$\operatorname{var}(\hat{\gamma}(\tau)) = (1/N) \sum_{n=-(N-\tau-1)}^{n=N-\tau-1} \left(1 - \frac{\tau + |n|}{N}\right) \left[\gamma^2(n) + \gamma(n-\tau)\gamma(n+\tau)\right]. \tag{25}$$

(Hint: if $\{Y_1, Y_2, Y_3, Y_4\}$ are four centered jointly Gaussian variables, then $\mathbb{E}[Y_1Y_2Y_3Y_4] = \mathbb{E}[Y_1Y_2]\mathbb{E}[Y_3Y_4] + \mathbb{E}[Y_1Y_3]\mathbb{E}[Y_2Y_4] + \mathbb{E}[Y_1Y_4]\mathbb{E}[Y_2Y_3]$.)

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

Answer 7

- Let's compute the variance of $\hat{\gamma}(\tau)$

$$\operatorname{var}(\hat{\gamma}(\tau)) = \mathbb{E}(\hat{\gamma}(\tau)^2) - \mathbb{E}(\hat{\gamma}(\tau))^2 \tag{26}$$

$$= \mathbb{E}((\frac{1}{N}\sum_{n=0}^{N-\tau-1} X_n X_{n+\tau})^2) - (\frac{N+\tau}{N}\gamma(\tau))^2 \text{ , using definitions and Q4} \tag{27}$$

$$= \frac{1}{N^2}\sum_{n,m=0}^{N-\tau-1} \mathbb{E}(X_n X_{n+\tau} X_m X_{m+\tau}) - (\frac{N+\tau}{N}\gamma(\tau))^2 \text{ , using linearity of } \mathbb{E} \tag{28}$$

$$= \frac{1}{N^2}\sum_{n,m=0}^{N-\tau-1} \mathbb{E}[X_n X_{n+\tau}]\mathbb{E}[X_m X_{m+\tau}] + \mathbb{E}[X_n X_m]\mathbb{E}[X_{n+\tau} X_{m+\tau}] \tag{29}$$

$$+ \mathbb{E}[X_n X_{m+\tau}]\mathbb{E}[X_m X_{n+\tau}] - (\frac{N+\tau}{N}\gamma(\tau))^2 \text{ using the given hint} \tag{30}$$

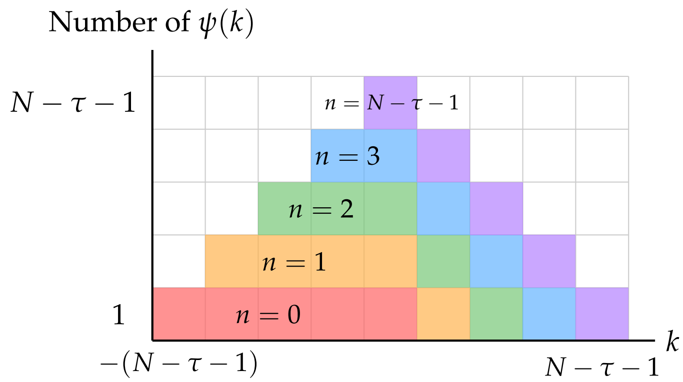
We can use the fact that X_n is weakly stationary and therefore $\forall m, n, \mathbb{E}(X_n, X_{n+\tau}) = \mathbb{E}(X_m, X_{m+\tau})$ and write this based on γ

$$\operatorname{var}(\hat{\gamma}(\tau)) = \frac{1}{N^2}\sum_{n,m=0}^{N-\tau-1} \gamma(n)^2 + \gamma(n-m)^2 + \gamma(n-k-\tau)\gamma(n-m+\tau) - (\frac{N+\tau}{N}\gamma(\tau))^2 \tag{31}$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{m=0}^{N-\tau-1} \gamma(n-m)^2 + \gamma(n-m-\tau)\gamma(n-m+\tau) \tag{32}$$

$$= \frac{1}{N^2}\sum_{n=0}^{N-\tau-1}\sum_{k=n}^{n-(N-\tau-1)} \underbrace{\gamma(k)^2 + \gamma(k-\tau)\gamma(k+\tau)}_{\psi(k)} \text{ with change of variable } k = n-m \tag{33}$$

As $\psi(k)$ is independant we can remark that the sum can be rewritten (the graph under explains how the sum should be seen)



$$\operatorname{var}(\hat{\gamma}(\tau)) = \frac{1}{N^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} (N-\tau-|k|)\, \psi(k) \tag{34}$$

$$= \frac{1}{N} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \left(1-\frac{\tau+|k|}{N}\right) \left(\gamma^2(k)+\gamma(k-\tau)\gamma(k+\tau)\right) \tag{35}$$

- Now we want to show that, for $\epsilon > 0$,

$$\mathbb{P}(|\hat{\gamma}_N(\tau)-\gamma(\tau)| \geq \epsilon) \xrightarrow{N\rightarrow\infty} 0$$

First, let's show that for a N large enough we have

$$\mathbb{P}(|\hat{\gamma}_N(\tau)-\gamma(\tau)| \geq \epsilon) \leq \mathbb{P}(|\hat{\gamma}_N(\tau)-\frac{N+\tau}{N}\gamma(\tau)| \geq \epsilon) \ ...$$

Then, let's use Tchebychev inequality for r.v $\hat{\gamma}(\tau)$

$$\mathbb{P}(|\hat{\gamma}_N(\tau)-\frac{N+\tau}{N}\gamma(\tau)| \geq \epsilon) \leq \frac{\operatorname{var}(\hat{\gamma}(\tau))}{\epsilon^2} \tag{36}$$

$$\leq \frac{1}{N\epsilon^2} \sum_{k=-(N-\tau-1)}^{N-\tau-1} \underbrace{\left(1-\frac{\tau+|k|}{N}\right)}_{\leq 1} \Big(\underbrace{\gamma^2(k)}_{\text{summable family}} + \gamma(k-\tau)\gamma(k+\tau)\Big) \tag{37}$$

$$\leq \frac{1}{N\epsilon^2} \left(C + \sum_{k=-(N-\tau-1)}^{N-\tau-1} \gamma(k-\tau)\gamma(k+\tau) \right) \tag{38}$$

$$\leq \frac{1}{N\epsilon^2} \left(C + \underbrace{\sum_{k=-N}^{N} \gamma^2(k-\tau)}_{\text{summable family}} \underbrace{\sum_{k=-N}^{N} \gamma^2(k+\tau))}_{\text{summable family}} \right) \text{ using Cauchy-Schwarz} \tag{39}$$

$$\leq \frac{\tilde{C}}{N\epsilon^2} \xrightarrow{N\rightarrow\infty} 0 \tag{40}$$

Therefore $\hat{\gamma}(\tau) \xrightarrow{proba} \gamma(\tau)$ hence $\hat{\gamma}(\tau)$ 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 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

The cos and the sin are just scale factors, thus the mean is unchanged for A,B we have:

$$\mathbb{E}(A(f)) = \mathbb{E}(B(f)) = 0 \tag{41}$$

Then:

$$\begin{aligned}
\mathrm{Var}(A(f_k)) &= \mathbb{E}\left[A(f_k)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{n=0}^{N-1} X_n \cos\left(\frac{2\pi kn}{N}\right)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)\right] \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}[X_n X_m]\mathbb{E}\left[\cos\left(\frac{2\pi kn}{N}\right)\cos\left(\frac{2\pi km}{N}\right)\right] \quad \text{(hint)} \\
&= \sum_{n=0}^{N-1} \mathbb{E}[X_n^2]\mathbb{E}\left[\cos^2\left(\frac{2\pi kn}{N}\right)\right] \quad (\mathbb{E}[X_n X_m] \neq 0 \text{ for } n = m) \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos^2\left(\frac{2\pi kn}{N}\right) \\
&= \frac{\sigma^2}{2} \sum_{n=0}^{N-1}\left(1 + \cos\left(\frac{4\pi kn}{N}\right)\right) \\
&= \frac{\sigma^2}{2}\left(N + \mathrm{Re}\left(\sum_{n=0}^{N-1} e^{\frac{4i\pi kn}{N}}\right)\right) \\
&= \frac{\sigma^2 N}{2}
\end{aligned}$$

In the same way, we have:

$$Var(B(f_k)) = \frac{\sigma^2 N}{2} \tag{42}$$

Finally we have found that A and B are just the same Gaussians.

–

Then, we have $J(f) = \frac{1}{N}(A(f) + iB(f))$, thus:

$$|J(f_k)|^2 = \frac{1}{N^2}(A(f_k)^2 + B(f_k)^2) \tag{43}$$

Let's compute:

$$\begin{aligned}
\text{cov}(A(f_k), B(f_k)) &= \mathbb{E}\left[A(f_k)B(f_k)\right] && (44)\\
&= \mathbb{E}\left[-\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X_n X_m \cos\left(\frac{2\pi kn}{N}\right)\sin\left(\frac{2\pi km}{N}\right)\right] && (45)\\
&= -\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathbb{E}\left[X_n X_m\right] \mathbb{E}\left[\cos\left(\frac{2\pi kn}{N}\right)\sin\left(\frac{2\pi km}{N}\right)\right] \quad \text{(hint)} && (46)\\
&= -\sum_{n=0}^{N-1} \mathbb{E}\left[X_n^2\right]\cos\left(\frac{2\pi kn}{N}\right)\sin\left(\frac{2\pi kn}{N}\right) \quad (\mathbb{E}[X_n X_m] \neq 0 \text{ for } n = m) && (47)\\
&= -\frac{\sigma^2}{2}\sum_{n=0}^{N-1}\sin\left(\frac{4\pi kn}{N}\right) && (48)\\
&= -\frac{\sigma^2}{2}\text{Im}\left(\sum_{n=0}^{N-1} e^{\frac{4i\pi kn}{N}}\right) && (49)\\
&= 0 && (50)
\end{aligned}$$

Since $A(f_k)$ and $B(f_k)$ are independent Gaussian random variables, $|J(f_k)|^2$ is a chi squared distribution of rank 2.

–

We know that if $X \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}[X^4] = 3\sigma^4$. We then have $\text{Var}(X^2) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = 2\sigma^4$.

$$\begin{aligned}
\text{Var}(|J(f_k)|^2) &= \frac{1}{N^2}\left(\text{Var}(A(f_k)^2) + \text{Var}(B(f_k)^2)\right) \quad \text{as } A^2 \text{ and } B^2 \text{ are i.i.d.} && (51)\\
&= \frac{1}{N^2}\left(2\left(\sigma\sqrt{\frac{N}{2}}\right)^4 + 2\left(\sigma\sqrt{\frac{N}{2}}\right)^4\right) \quad \text{based on the previous remark} && (52)\\
&= \sigma^4 && (53)
\end{aligned}$$

Furthermore, we can compute the expectation of $|J(f_k)|^2$:

$$\mathbb{E}[|J(f_k)|^2] = \frac{1}{N}\left(\mathbb{E}[A(f_k)^2] + \mathbb{E}[B(f_k)^2]\right) \tag{54}$$

$$= \frac{1}{N}\left(\mathrm{Var}(A(f_k)) + \mathrm{Var}(B(f_k))\right) \tag{55}$$

$$= \sigma^2 \tag{56}$$

The distribution of $|J(f_k)|^2$ doesn't depend on N. This means that $\mathbb{E}(|J(f_k)|^2)$ and $\mathbb{V}(|J(f_k)|^2)$ are not changing with N, i.e. $|J(f_k)|^2$ is not consistent.

–

We have:

$$\mathbb{E}[A(f_k)^2 A(f_l)^2] = \mathbb{E}\left[\sum_{n,m,s,t=0}^{N-1} X_n X_m X_s X_t \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ls}{N}\cos\frac{2\pi lt}{N}\right]$$

$$= \sum_{n,m,s,t=0}^{N-1} \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ls}{N}\cos\frac{2\pi lt}{N}\left(\mathbb{E}[X_n X_m X_s X_t]\right)$$

$$= \sum_{n,m,s,t=0}^{N-1} \cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ls}{N}\cos\frac{2\pi lt}{N}\left(\mathbb{E}[X_n X_m]\mathbb{E}[X_s X_t] + \mathbb{E}[X_n X_s]\mathbb{E}[X_m X_t] + \mathbb{E}[X_n X_t]\mathbb{E}[X_m X_s]\right)$$

The first term of the parenthesis leads to study cases where $n = m$ and $s = t$:

$$\sum_{n,s}\cos^2\left(\frac{2\pi kn}{N}\right)\cos^2\left(\frac{2\pi ls}{N}\right)\mathbb{E}[X_n^2]\mathbb{E}[X_s^2] = \sigma^4\sum_{n,s}\cos^2\left(\frac{2\pi kn}{N}\right)\cos^2\left(\frac{2\pi ls}{N}\right) = \frac{\sigma^4 N^2}{4}$$

The second term of the parenthesis leads to study cases where $n = s$ and $m = t$:

$$\sum_{n,m}\cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ln}{N}\cos\frac{2\pi lm}{N}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2]$$

$$= \sigma^4\sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}\sum_m \cos\frac{2\pi km}{N}\cos\frac{2\pi lm}{N}$$

$$= \frac{\sigma^4}{2}\sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}\sum_m\left[\cos\frac{2\pi(l-k)m}{N} + \cos\frac{2\pi(l+k)m}{N}\right]$$

$$= \frac{\sigma^4 N}{2}\sum_n \cos\frac{2\pi kn}{N}\cos\frac{2\pi ln}{N}\delta_{lk} = \frac{\sigma^4 N^2}{4}\delta_{lk}$$

The third and last term of the parenthesis leads to study cases where $n = t$ and $m = s$:

$$\sum_{n,m}\cos\frac{2\pi kn}{N}\cos\frac{2\pi km}{N}\cos\frac{2\pi ln}{N}\cos\frac{2\pi lt}{N}\mathbb{E}[X_n^2]\mathbb{E}[X_m^2] = \frac{\sigma^4 N^2\delta_{l,k}}{4} \quad \text{(same method as 2nd term)}$$

Finally, we have

$$\mathbb{E}[A(f_k)^2A(f_l)^2] = \frac{\sigma^4N^2}{4}(1+2\delta_{l,k})$$

The same method gives us:

$$\mathbb{E}[B(f_k)^2B(f_l)^2] = \frac{\sigma^4N^2}{4}(1+2\delta_{l,k})$$

We can compute the covariance of $|J(f_k)|^2$ using what we did above:

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

$$= \frac{1}{N^2}\mathbb{E}[A(f_k)^2A(f_l)^2 + B(f_k)^2B(f_l)^2 + A(f_k)^2B(f_l)^2 + B(f_k)^2A(f_l)^2] - \sigma^4$$

$$= \frac{1}{N^2}\left(\frac{\sigma^4N^2}{4}(1+2\delta_{l,k}) + \frac{\sigma^4N^2}{4}(1+2\delta_{l,k}) + \frac{\sigma^4N^2}{4} + \frac{\sigma^4N^2}{4}\right) - \sigma^4$$

$$\text{i.e.}\quad \text{cov}(|J(f_k)|^2, |J(f_l)|^2) = \sigma^4(1+\delta_{l,k}) - \sigma^4$$

In particular, for $l \neq k$, $\text{cov}(|J(f_k)|^2, |J(f_l)|^2) = 0$. As there is no correlation between the $|J(f_k)|^2$, it explains the randomness in the periodogram. No pattern can be extracted from white noise.

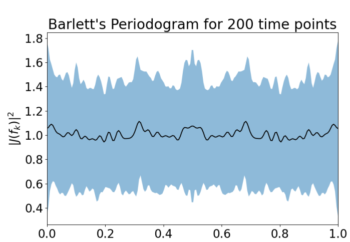
Question 9

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.

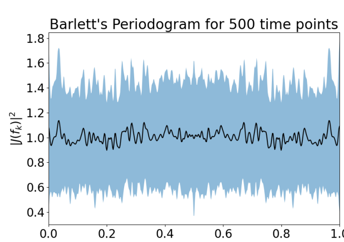
- Rerun the experiment of Question 6, but replace the periodogram by Barlett's estimate (set $K = 5$). What do you observe?

Add your plots to Figure 2.

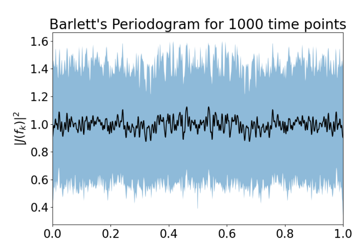
Answer 9



Periodogram ($N = 200$)



Periodogram ($N = 500$)



Periodogram ($N = 1000$)

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

These graphs show a better standard deviation compared to Q6. This is coherent with the estimation of Barlett, indeed, we choose $K = 5$ and the standard deviation went from 1 to $1/\sqrt{5} \approx 0.4$

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 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

We can see that the optimal number of neighbors is 5 with a validation f1 score of 0.77. However the associated f1-score with the test dataset is 0.46.

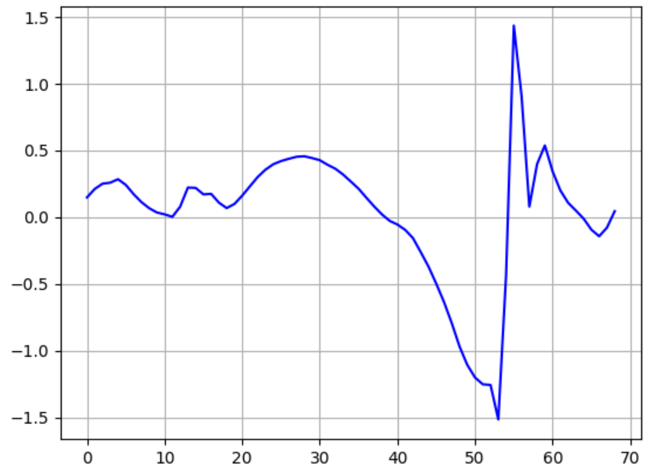
Our model is not generalizing really well on unseen data.

Maybe we could try to combine more data from the accelerometer in order to have a better result.

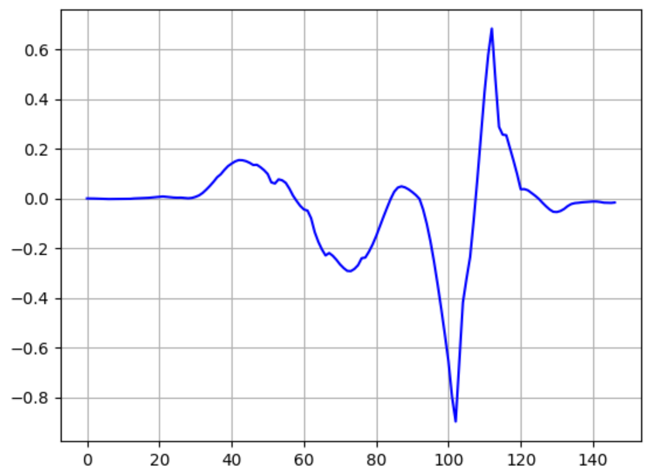
Question 11

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).