Assignment 2 (ML for TS) - MVA

Oliver Jack olijacklu@gmail.com
Paulo Silva paulohenriquecrs@hotmail.com

December 2, 2024

1 Introduction

Objective. The goal is to better understand the properties of AR and MA processes and do signal denoising with sparse coding.

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 Monday 2nd December 11:59 PM.
- Rename your report and notebook as follows:
`FirstnameLastname1_FirstnameLastname1.pdf` and
`FirstnameLastname2_FirstnameLastname2.ipynb`.
For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: https://docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2

2 General questions

A time series $\{y_t\}_t$ is a single realisation of a random process $\{Y_t\}_t$ defined on the probability space $(\Omega, \mathcal{F}, P)$, i.e. $y_t = Y_t(w)$ for a given $w \in \Omega$. In classical statistics, several independent realizations are often needed to obtain a “good” estimate (meaning consistent) of the parameters of the process. However, thanks to a stationarity hypothesis and a "short-memory" hypothesis, it is still possible to make “good” estimates. The following question illustrates this fact.

Question 1

An estimator $\hat{\theta}_n$ is consistent if it converges in probability when the number n of samples grows to ∞ to the true value $\theta \in \mathbb{R}$ of a parameter, i.e. $\hat{\theta}_n \xrightarrow{\mathcal{D}} \theta$.

- Recall the rate of convergence of the sample mean for i.i.d. random variables with finite variance.
- Let $\{Y_t\}_{t\geq 1}$ a wide-sense stationary process such that $\sum_k |\gamma(k)| < +\infty$. Show that the sample mean $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$ is consistent and enjoys the same rate of convergence as the i.i.d. case. (Hint: bound $\mathbb{E}[(\bar{Y}_n - \mu)^2]$ with the $\gamma(k)$ and recall that convergence in L_2 implies convergence in probability.)

Answer 1

Consider a sequence of i.i.d. random variables X_n with mean μ and variance $\sigma^2 < \infty$. Let us denote the sample mean by $\overline{X_n} = \frac{X_1 + ... + X_n}{n}$. By the CLT, we have that:

$$\sqrt{n}(\overline{X_n} - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

Since the difference between X_n and μ is scaled by $\sqrt{n}$ to assure convergence, we can conclude that the unscaled difference shrinks at a rate of $O\left(\frac{1}{\sqrt{n}}\right)$. Thus, the convergence rate of the sample mean $\overline{X_n}$ is $O\left(\frac{1}{\sqrt{n}}\right)$.

Since $\{Y_t\}_{t\geq 1}$ is a wide-sense stationary process such that $\sum_k |\gamma(k)| < +\infty$, we know that:

$$\forall t, \quad \mathbb{E}[Y_t] = \mu \tag{1}$$

$$\forall t_1, t_2, \quad \mathbb{E}[Y_{t_1} Y_{t_2}] = \gamma(|t_2 - t_1|) \tag{2}$$

Let us bound $\mathbb{E}[(\bar{Y}_n - \mu)^2]$:

$$\begin{aligned}
\mathbb{E}[(\bar{Y}_n - \mu)^2] &= \mathbb{E}[\bar{Y}_n^2] - 2\mu\mathbb{E}[\bar{Y}_n] + \mu^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[Y_i Y_j] - 2\mu^2 + \mu^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(|i-j|) - \mu^2 \\
&= \frac{1}{n}\gamma(0) + \frac{1}{n^2}\sum_{k=1}^{n}2(n-k)\gamma(k) - \mu^2 \\
&\leq \frac{1}{n}\gamma(0) + \frac{1}{n^2}\sum_{k=-n,k\neq 0}^{n}(n-|k|)\gamma(|k|) \\
&\leq \frac{1}{n}|\gamma(0)| + \frac{1}{n}\sum_{k=-n,k\neq 0}^{n}|\gamma(|k|)| \\
&= \frac{1}{n}\sum_{k=-n}^{n}|\gamma(k)| \to 0 \text{ as } n \to \infty \text{ since } \sum_k |\gamma(k)| < +\infty
\end{aligned} \tag{3}$$

This implies that the sample mean $\bar{Y}_n$ converges in L_2 to μ, so converges in probability to μ. Since the variance $Var[\bar{Y}_n] = \mathbb{E}[(\bar{Y}_n - \mu)^2] = O\left(\frac{1}{n}\right)$, we have that the standard deviation of $\bar{Y}_n$ is of scale

$O\left(\frac{1}{\sqrt{n}}\right)$. Finally, we can conclude that $\bar{Y}_n$ enjoys the same convergence rate as in the i.i.d. case, namely $O\left(\frac{1}{\sqrt{n}}\right)$.

3 AR and MA processes

Question 2 *Infinite order moving average MA(∞)*

Let $\{Y_t\}_{t\geq 0}$ be a random process defined by

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \tag{4}$$

where $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ ($\psi_0 = 1$) are square summable, i.e. $\sum_k \psi_k^2 < \infty$ and $\{\varepsilon_t\}_t$ is a zero mean white noise of variance σ_ε^2. (Here, the infinite sum of random variables is the limit in L_2 of the partial sums.)

- Derive $\mathbb{E}(Y_t)$ and $\mathbb{E}(Y_t Y_{t-k})$. Is this process weakly stationary?
- Show that the power spectrum of $\{Y_t\}_t$ is $S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2$ where $\phi(z) = \sum_j \psi_j z^j$. (Assume a sampling frequency of 1 Hz.)

The process $\{Y_t\}_t$ is a moving average of infinite order. Wold's theorem states that any weakly stationary process can be written as the sum of the deterministic process and a stochastic process which has the form (4).

Answer 2

Since $\mathbb{E}[\epsilon_t] = 0, \forall t$, we have that:

$$\mathbb{E}[Y_t] = \mathbb{E}\left[\sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}\right] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\epsilon_{t-k}] = 0 \tag{5}$$

Furthermore, we have that:

$$\mathbb{E}[\epsilon_t \epsilon_{t'}] = \begin{cases} \sigma_\epsilon^2 & \text{if } t = t', \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Thus:

$$\begin{aligned}
\mathbb{E}[Y_t Y_{t-k}] &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-k-j}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \epsilon_{t-i} \epsilon_{t-k-j}\right] \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}[\epsilon_{t-i} \epsilon_{t-k-j}] \\
&= \sum_{j=0}^{\infty} \psi_{j+k} \psi_j \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_{j+k} \psi_j
\end{aligned} \tag{7}$$

Since the autocovariance $\mathbb{E}[Y_t Y_{t-k}]$ only depends on the lag k and not on the specific time t, we can conclude that the process $\{Y_t\}_{t\geq 0}$ is wide-sense stationary.

Based on the previous result, we can write the autocovariance as $\gamma(k) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$, $\forall k \in \mathbb{Z}$. Assume that the sample frequency is 1 Hz. Let us first compute $\left|\sum_{j=0}^{N} \psi_j e^{-2\pi i f j}\right|^2$, $\forall N \in \mathbb{N}$:

$$
\begin{aligned}
\left|\phi\left(e^{-2\pi i f}\right)\right|^2 &= \left|\sum_{j=0}^{N} \psi_j e^{-2\pi i f j}\right|^2 \\
&= \left(\sum_{j=0}^{N} \psi_j e^{-2\pi i f j}\right)\left(\sum_{\ell=0}^{N} \psi_\ell e^{2\pi i f \ell}\right) \\
&= \sum_{j=0}^{N}\sum_{\ell=0}^{N} \psi_j \psi_\ell e^{-2\pi i f (j-\ell)} \\
&= \sum_{\tau=-N+1}^{N-1} \sum_{n=0}^{N-|\tau|-1} \psi_n \psi_{n+|\tau|} e^{-2\pi i f \tau}
\end{aligned}
\tag{8}
$$

using the same change of variables as in Assignment 1. As $N \to \infty$, we obtain:

$$
\left|\phi\left(e^{-2\pi i f}\right)\right|^2 = \sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} e^{-2\pi i f k} \tag{9}
$$

Next, let us compute the power spectrum:

$$
\begin{aligned}
S(f) &= \sum_{k=-\infty}^{\infty} \gamma(k) e^{-2\pi i f k} \\
&= \sigma_\epsilon^2 \sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} e^{-2\pi i f k} \\
&= \sigma_\epsilon^2 \left|\phi\left(e^{-2\pi i f}\right)\right|^2
\end{aligned}
\tag{10}
$$

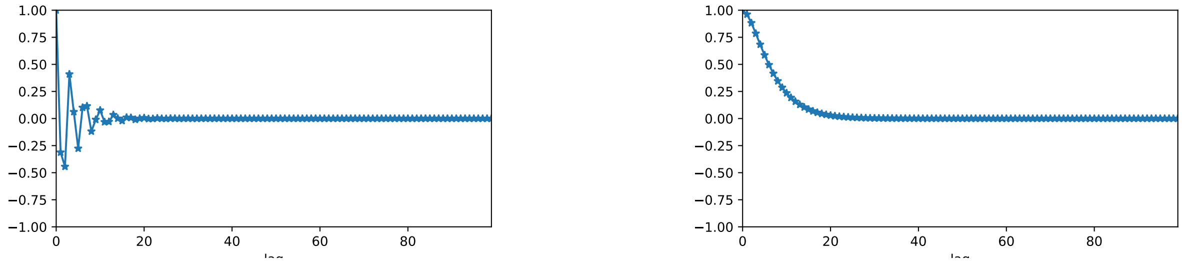
Question 3 *AR(2) process*

Let $\{Y_t\}_{t\geq 1}$ be an AR(2) process, i.e.

$$
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{11}
$$

with $\phi_1, \phi_2 \in \mathbb{R}$. The associated characteristic polynomial is $\phi(z) := 1 - \phi_1 z - \phi_2 z^2$. Assume that ϕ has two distinct roots (possibly complex) r_1 and r_2 such that $|r_i| > 1$. Properties on the roots of this polynomial drive the behavior of this process.

- Express the autocovariance coefficients $\gamma(\tau)$ using the roots r_1 and r_2.
- Figure 1 shows the correlograms of two different AR(2) processes. Can you tell which one has complex roots and which one has real roots?
- Express the power spectrum $S(f)$ (assume the sampling frequency is 1 Hz) using $\phi(\cdot)$.
- Choose ϕ_1 and ϕ_2 such that the characteristic polynomial has two complex conjugate roots of norm $r = 1.05$ and phase $\theta = 2\pi/6$. Simulate the process $\{Y_t\}_t$ (with $n = 2000$) and display the signal and the periodogram (use a smooth estimator) on Figure 2. What do you observe?



Correlogram of the first AR(2)

Correlogram of the second AR(2)

Figure 1: Two AR(2) processes

Answer 3

Express the autocovariance coefficients $\gamma(\tau)$ using the roots r_1 and r_2

Given the AR(2) process:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \tag{12}$$

the characteristic polynomial is:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2. \tag{13}$$

Assume the roots r_1 and r_2 of the characteristic polynomial satisfy $|r_i| > 1$.

The autocovariance function $\gamma(\tau)$ can be expressed in terms of the roots r_1 and r_2. Since r_1 and r_2 are the roots of the characteristic polynomial, the general solution for the AR(2) process is:

$$\gamma(\tau) = A r_1^{\tau} + B r_2^{\tau}, \tag{14}$$

where A and B are constants determined by the initial conditions.

Using the Yule-Walker equations, we have:

$$\begin{aligned}
\gamma(0) &= \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma_\varepsilon^2, \\
\gamma(1) &= \phi_1 \gamma(0) + \phi_2 \gamma(1), \\
\gamma(\tau) &= \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2) \quad \text{for} \quad \tau \geq 2.
\end{aligned} \tag{15}$$

By solving these equations, we obtain the values of A and B, thus fully expressing $\gamma(\tau)$ in terms of the roots r_1 and r_2.

Which AR(2) process has complex roots and which one has real roots?

To determine whether the AR(2) processes have complex or real roots, we visually analyze the correlograms present in the plots of Figure 1:

- **First AR(2) Process (Left)**: The correlogram shows oscillatory behavior. This pattern indicates that the process has complex roots, as complex roots lead to oscillatory behavior in the autocorrelation function.
- **Second AR(2) Process (Right)**: The correlogram exhibits a smooth exponential decay. This pattern indicates that the process has real roots, as real roots lead to non-oscillatory, exponentially decaying autocorrelation functions.

Express the power spectrum $S(f)$ using $\phi(\cdot)$

Given the AR(2) process:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

the associated characteristic polynomial is:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2$$

The power spectrum $S(f)$ of the AR(2) process can be expressed using the characteristic polynomial $\phi(\cdot)$. The general form for the power spectrum of an AR process is:

$$S(f) = \sigma_\varepsilon^2 \left| \frac{1}{\phi(e^{-2\pi i f})} \right|^2$$

For the AR(2) process, substituting the characteristic polynomial $\phi(z)$, we get:

$$\phi(e^{-2\pi i f}) = 1 - \phi_1 e^{-2\pi i f} - \phi_2 e^{-4\pi i f}$$

Therefore, the power spectrum $S(f)$ is:

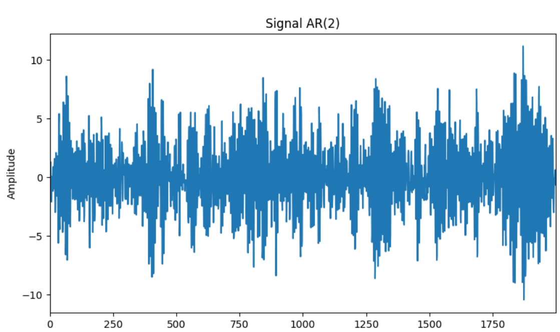
$$S(f) = \sigma_\varepsilon^2 \left| \frac{1}{1 - \phi_1 e^{-2\pi i f} - \phi_2 e^{-4\pi i f}} \right|^2$$

We can also express it using the roots r_1 and r_2 of the characteristic polynomial:

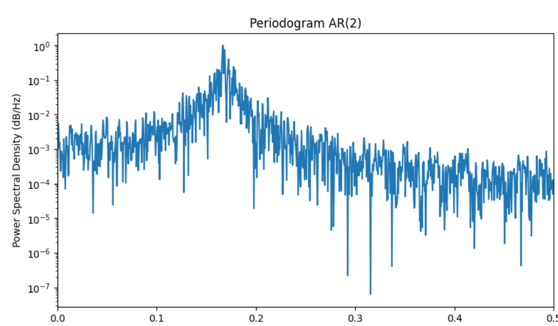
$$S(f) = \frac{\sigma_\varepsilon^2 r_1^2 r_2^2}{(r_1^2 - 1)(r_2 - r_1)} \left[\frac{r_1}{r_1^2 - e^{-2\pi i f} r_1} - \frac{r_2}{r_2^2 - e^{-2\pi i f} r_2} \right]$$

where $\phi(z) = (z - r_1)(z - r_2)$.

Simulate the process $\{Y_t\}_t$ and display the signal and the periodogram



Signal



Periodogram

Figure 2: AR(2) process

The simulated AR(2) time series plot displays clear oscillations, characteristic of an AR(2) process with complex conjugate roots. The periodogram reveals a dominant frequency peak around 0.2 Hz, confirming the oscillatory nature observed in the time domain. This peak, highlighted using a logarithmic scale, reflects the influence of the process parameters on the spectral properties of the AR(2) signal.

4 Sparse coding

The modulated discrete cosine transform (MDCT) is a signal transformation often used in sound processing applications (for instance, to encode an MP3 file). A MDCT atom $\phi_{L,k}$ is defined for a length 2L and a frequency localisation k ($k = 0, \dots, L-1$) by

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{16}$$

where w_L is a modulating window given by

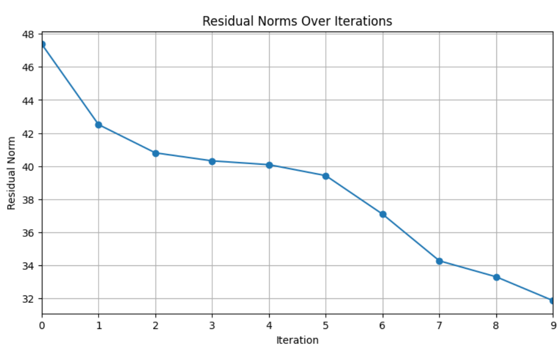
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{17}$$

Question 4 *Sparse coding with OMP*

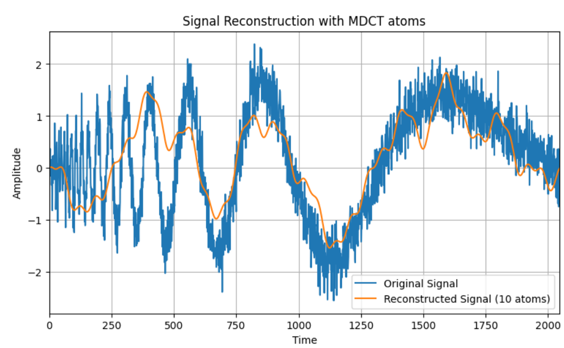
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- For the sparse coding, implement the Orthogonal Matching Pursuit (OMP). (Use convolutions to compute the correlation coefficients.)
- Display the norm of the successive residuals and the reconstructed signal with 10 atoms.

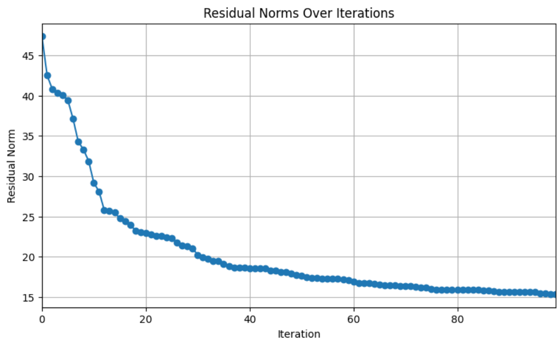
Answer 4



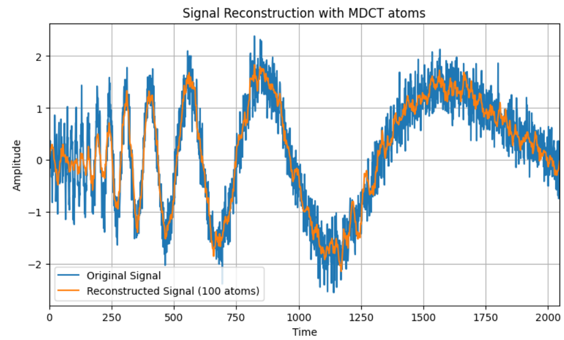
Norms of the successive residuals (10 iterations)



Reconstruction with 10 atoms



Norms of the successive residuals (100 iterations)



Reconstruction with 100 atoms

Figure 3: Question 4