# Assignment 2 (ML for TS) - MVA 2023/2024

Yvann Le Fay yvann.lefay@ensae.fr
Antoine Schoonaert antoine.schoonaert@etu.emse.fr

December 7, 2023

## 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 Tuesday 5th 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: docs.google.com/forms/d/e/1FAIpQLSfCqMXSDU9jZJbYUMmeLCXbVeckZYNiDpPl4hRUwcJ2cBHQM

## 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 realisations 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 $\infty$ 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

Let $\{Y_t\}_{t\geq 1}$ be i.i.d random variables with mean $\mu$ and finite variance $\sigma^2$. Let $\bar{Y}_n$ be the sample mean. Tchebychev's inequality states that for any $t > 0$

$$\mathbb{P}[|\bar{Y}_n - \mu| > t] \leq \frac{\sigma^2}{nt^2}, \tag{1}$$

which goes to 0 as $n$ goes to $\infty$. Hence, the convergence rate of the sample mean estimator is $O(1/n)$.

Suppose now that $\{Y_t\}$ is a wide-sense stationary process with $\sum_k |\gamma(k)| < \infty$. We have

$$\begin{aligned}
\mathbb{E}[(\bar{Y}_n - \mu)^2] &= \frac{1}{n^2} \sum_{1\leq i,j\leq n} \mathbb{E}[(Y_i - \mu)(Y_j - \mu)] \\
&= \frac{1}{n^2} \sum_{1\leq i,j\leq n} \gamma(|i-j|) \\
&\leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{k} |\gamma(k)| \\
&= \frac{1}{n} \sum_{k} |\gamma(k)| = O(1/n).
\end{aligned} \tag{2}$$

We recover the same convergence rate by Chebychev inequality,

$$\mathbb{P}[|\bar{Y}_n - \mu| > t] \leq \frac{\mathbb{E}[(\bar{Y}_n - \mu)^2]}{t^2} = O(1/n). \tag{3}$$

## 3 AR and MA processes

**Question 2** *Infinite order moving average MA($\infty$)*

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 $\sigma_\varepsilon^2$. (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

Let $t \geq 0$, since $Y_t$ is in $L^2$, $Y_t$ is integrable. We have,

$$\mathbb{E}[Y_t] = \lim_{n\to\infty} \mathbb{E}[\sum_{k=0}^{n} \psi_k \varepsilon_{t-k}] = \lim_{n\to\infty} \sum_{k=0}^{n} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0, \tag{5}$$

where we use $\mathbb{E}[\varepsilon_k] = 0$ for any $k$. Similarly, the square-summable condition on the $\psi_k$'s implies the existence of the second-order moment of $Y_t$ and we can directly work with the infinite sums. Let $0 \leq k \leq t$, we have

$$\begin{aligned}
\mathbb{E}[Y_t Y_{t-k}] &= \mathbb{E}\left[\sum_{0\leq j,l<\infty} \psi_j \psi_l \varepsilon_{t-j} \varepsilon_{t-l-k}\right] \\
&= \sum_{0\leq j,l<\infty} \psi_j \psi_l \mathbb{E}[\varepsilon_0 \varepsilon_{j-l-k}] \\
&= \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_{l+k} \psi_l < \infty.
\end{aligned} \tag{6}$$

Since the mean and autocovariance functions of the process $Y$ is independent of $t$, $Y$ is stationnary in the wide-sense. Let $S$ be defined by $S(f) := \sum_{\tau\in\mathbb{Z}} \mathbb{E}[Y_0 Y_\tau] e^{-2i\pi f \tau}$. Starting from the right-hand side of the equality we are looking to prove, we have (dropping the $\sigma_\varepsilon^2$)

$$\begin{aligned}
|\phi(e^{-2\pi i f})|^2 &= \left(\sum_{j=0}^{\infty} e^{-2\pi i f j} \psi_j\right)\left(\sum_{l=0}^{\infty} e^{2\pi i f l} \psi_l\right) \\
&= \sum_{0\leq j,l<\infty} e^{-2\pi i f (j-l)} \psi_j \psi_l \\
&= \sum_{\tau\in\mathbb{Z}} \sum_{0\leq j,l<\infty} e^{-2\pi i f \tau} \psi_j \psi_l \mathbb{1}\{j-l=\tau\} \\
&= \sum_{\tau\in\mathbb{Z}} \sum_{l=0}^{\infty} e^{-2\pi i f \tau} \psi_{\tau+l} \psi_l,
\end{aligned} \tag{7}$$

where going from the first line to the second line is legit because the $\psi_k$'s are square summable. The obtained equation is exactly $S(f)$ up to the multiplicative constant $\sigma_\varepsilon^2$. We obtain

$$S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2. \tag{8}$$

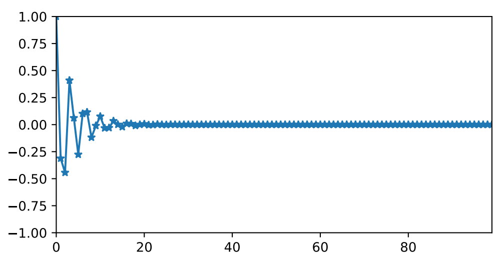
**Question 3** *AR(2) process*

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

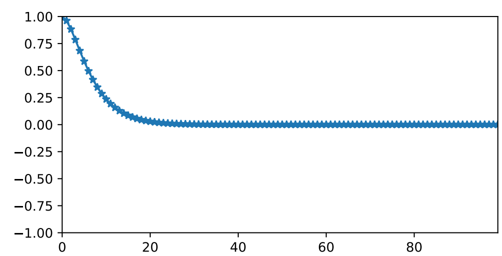
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{9}$$

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 $\phi$ 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 behaviour 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 $\phi_1$ and $\phi_2$ 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**

Assuming the process is WSS, multiplying $Y_t$ by $Y_{t+\tau}$ for any $k > 0$, and taking the expectation, we obtain for any $\tau$

$$\gamma(\tau) - \phi_1\gamma(\tau - 1) - \phi_2\gamma(\tau - 2) = 0, \tag{10}$$

i.e., $\gamma$ is the solution of the second-order recurrent linear equation with characteristic polynomial $P(z) = z^2\phi(1/z) = z^2 - \phi_1 z - \phi_2$. The roots of $P$ are the inverse of the roots of $\phi$, i.e., $1/r_1$ and $1/r_2$ and since they are distinct, we have for $|\tau| \geq 0$

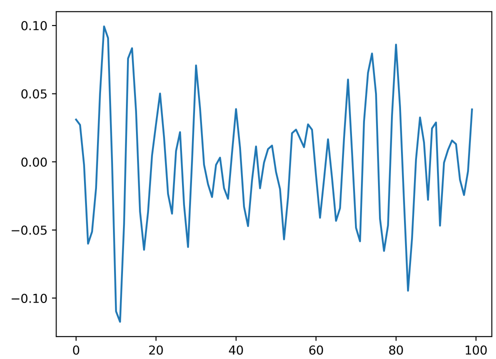
$$\gamma(\tau) = c_1/r_1^\tau + c_2/r_2^\tau, \tag{11}$$

for some constants $c_1$ and $c_2$ (determined using the initialisation laws of $Y_0$ and $Y_1$). In the case where $r_1$ has a complex component, i.e., $r_1 = re^{i\theta}$ for some $r > 0$ and $\theta \not\equiv 0[2\pi]$, we know that $r_2 = \bar{r}_1$ and (11) becomes

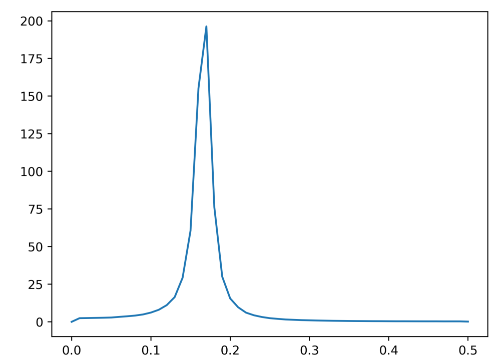
$$\gamma(\tau) = 1/r^\tau (c_1^{'} \cos(\theta\tau) + c_2^{'} \sin(\theta\tau)), \tag{12}$$

for some constants $c_1'$ and $c_2'$. Otherwise, if $r_1 \in \mathbb{R}$ then $r_2 \in \mathbb{R}$. Assuming without loss of generality $|c_1/r_1| > |c_2/r_2|$ with $|r_1| > 1$, we see that $\gamma(\tau) \sim c_1/r_1^\tau$ and $\gamma$ displays an exponential decay. We remark that if $r_1 < 0$, then $\gamma(\tau) \sim c_1/|r_1|^\tau(-1)^\tau$, hence, we observe an oscillating exponential decay.

Assuming one figure corresponds to the complex case and the other one to the real case, then the left figure corresponds to the complex root case while the right figure correspond to $r_1, r_2 \in \mathbb{R}$ with $|r_i| < 1$.




Signal: mean for 1000 sample paths




Periodogram: mean for 1000 sample paths

Figure 2: AR(2) process

Let $t \geq 2$, we have

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \\ &= \phi_1 \varepsilon_{t-1} + (\phi_1^2 + \phi_2) Y_{t-2} + \phi_1 \phi_2 Y_{t-3} + \varepsilon_t \\ &= \dots \\ &= \sum_{j=0}^{t} \varepsilon_{t-j} a_j, \end{aligned} \tag{13}$$

where we define $a_j$ by

$$a_j = \sum_{l,k \geq 0, l+2k=j} \phi_1^l \phi_2^k. \tag{14}$$

We observe that $a_j$ is the $j$-th coefficient of the series $\sum_{j\geq 0}(\phi_1 z + \phi_2 z^2)^j = \frac{1}{1-\phi_1 z - \phi_2 z^2} = \frac{1}{\phi(z)}$. Thus, using the previous equality for the power-spectrum of a moving-average process, we have

$$S(f) = \sigma_\varepsilon^2 |\tilde{\phi}(e^{-2i\pi f})|^2, \tag{15}$$

where $\tilde{\phi}(z) = \frac{1}{\phi(z)}$, i.e.,

$$S(f) = \frac{\sigma_\varepsilon^2}{|\phi(e^{-2i\pi f})|^2} \tag{16}$$

We have, $z^2 P(1/z) = \phi(z)$, with $P(z) = (z - 1/r_1)(z - 1/r_2)$, hence,

$$\frac{1}{r_1} + \frac{1}{r_2} = \phi_1 \quad -\frac{1}{r_1 r_2} = \phi_2. \tag{17}$$

We find $\phi_1 \approx 0.952$ and $\phi_2 \approx -0.907$.

## 4 Sparse coding

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

$$\forall u = 0, \ldots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k+\frac{1}{2})] \tag{18}$$

where $w_L$ is a modulating window given by

$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{19}$$

**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 correlations coefficients.)
- Display the norm of the successive residuals and the reconstructed signal with 10 atoms.

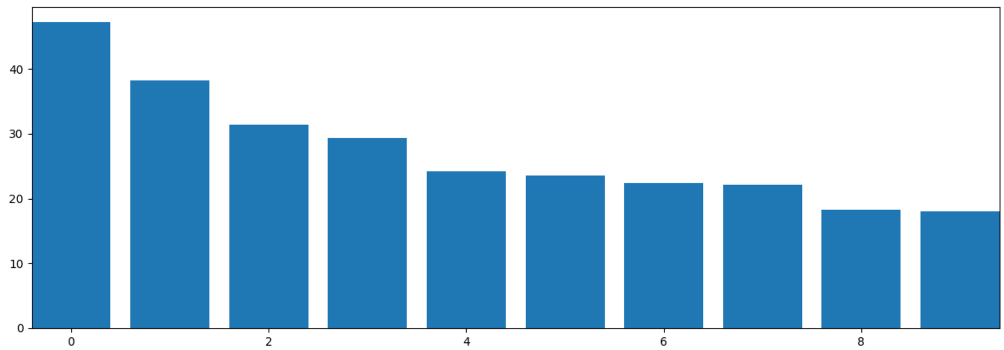
### Answer 4

The aim of this exercise is to re-create a signal from atoms defined by 18.

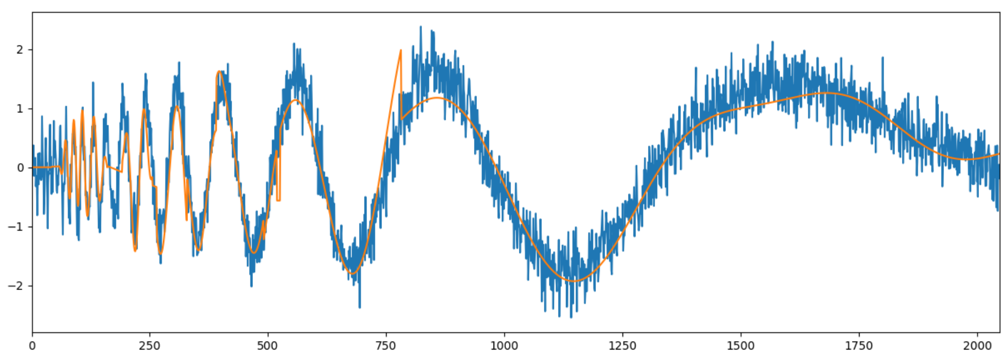
Firstly, we implement the creation of the dictionary with normalisation.

Secondly, we implement the Orthogonal Matching Pursuit using convolution to calculate the correlation coefficients. At each stage, we calculate the convolution between the residual signal and each atom, so the correlation coefficient corresponds to the maximum of the convolution result. The new atom to be added is the one with the highest coefficient. In doing this, the only atoms taken were those of length 2048 because, since they are longer, the convolution has a greater maximum. So we decided to divide each coefficient obtained by a value proportional to the size of the atom. The value corresponds to the average (over atoms of the same length) of the maximum values of the convolution. It is by doing this that we obtain the best reconstruction.

Figure 3 shows our results. We are aware that there is still room for improvement. We can see a peak at around 750, due to the end of an atom. This appears when we "force" the acceptance of small frequencies by dividing the coefficients as explained above. The visual aspect is very bad at this point but thanks to this the final norm of the residue decreased from 21 to 18.




Norms of the successive residuals




Reconstruction with 10 atoms

Figure 3: Question 4