Assignment 2 (ML for TS) - MVA

Tristan Montalbetti tristan.montalbetti@etu.minesparis.psl.eu
Jules Royer jules.royer@etu.minesparis.psl.eu

November 29, 2025

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 Sunday 7$^{\text{th}}$ December 11:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname1.pdf` and
  `FirstnameLastname2_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_ValerioGuerrini.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: https://forms.gle/J1pdeHspSs9zNfWAA.

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

3 AR and MA processes

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

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

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

where $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ ($\psi = 1$) 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 (1).

Answer 2

Mean and Autocovariance of Y_t For the MA(∞) process $Y_t = \sum_{k\geq 0} \psi_k \varepsilon_{t-k}$, the square summability $\sum_k \psi_k^2 < \infty$ ensures that $\sum_k \psi_k \varepsilon_{t-k}$ converges in L^2. But L^2-convergence implies L^1-convergence, and expectation is continuous in L^1.

Thus we can swap expectation and infinite sum in both cases $\mathbb{E}[Y_t]$ and $\mathbb{E}[Y_t Y_{t-k}]$.

Mean:

$$\mathbb{E}[Y_t] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0.$$

Autocovariance:

$$\gamma_Y(k) = \mathbb{E}[Y_t Y_{t-k}] = \mathbb{E}\Bigg[\sum_{i,j\geq 0} \psi_i \psi_j \, \varepsilon_{t-i}\varepsilon_{t-k-j}\Bigg] = \sum_{i,j\geq 0} \psi_i \psi_j \, \mathbb{E}[\varepsilon_{t-i}\varepsilon_{t-k-j}].$$

Since $(\varepsilon_t)_t$ is white noise,

$$\mathbb{E}[\varepsilon_{t-i}\varepsilon_{t-k-j}] = \begin{cases} \sigma_\varepsilon^2, & i = j + k, \\ 0, & \text{otherwise}, \end{cases}$$

Consider first $k \geq 0$. Then $j = i - k \geq 0$ forces $i \geq k$, so

$$\gamma_Y(k) = \sigma_\varepsilon^2 \sum_{i=k}^{\infty} \psi_i \psi_{i-k} \underbrace{=}_{m=i-k} \sigma_\varepsilon^2 \sum_{m=0}^{\infty} \psi_{m+k}\psi_m.$$

For $k < 0$,

$$\gamma_Y(k) = \gamma_Y(-k).$$

Thus, for all $k \in \mathbb{Z}$,

$$\boxed{\gamma_Y(k) = \sigma_\varepsilon^2 \sum_{m=0}^{\infty} \psi_m \psi_{m+|k|}}$$

which depends only on k. *We already showed a constant mean, hence the process is weakly stationary.*

Power Spectrum of Y_t Take sampling frequency $f_s = 1\,\text{Hz}$. Then the power spectrum is for $f \in [-f_s/2, f_s/2]$:

$$S_Y(f) = \sum_{k\in\mathbb{Z}} \gamma_Y(k)\, e^{-2\pi i f k}.$$

For

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \mathbb{E}[\varepsilon_t] = 0,\ \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2,$$

the autocovariance is

$$\gamma_Y(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}, \qquad k \in \mathbb{Z},$$

Insert into the PSD:

$$S(f) = \sigma_\varepsilon^2 \sum_{k\in\mathbb{Z}} \left(\sum_{j=0}^{\infty} \psi_j \psi_{j+k} \right) e^{-2\pi i f k}.$$

Define the transfer function

$$\phi(z) = \sum_{j=0}^{\infty} \psi_j z^j.$$

Then

$$|\phi(e^{-2\pi i f})|^2 = \phi(e^{-2\pi i f})\,\overline{\phi(e^{-2\pi i f})} = \left(\sum_{j=0}^{\infty} \psi_j e^{-2\pi i f j} \right) \left(\sum_{l=0}^{\infty} \psi_l e^{2\pi i f l} \right) = \sum_{j,\ell}^{\infty} \psi_j \psi_\ell\, e^{-2\pi i f (j-\ell)}.$$

Set $k = j - \ell$ and sum over all pairs (j, ℓ):

$$\left|\phi(e^{-2\pi i f})\right|^2 = \sum_{k=-\infty}^{\infty} \left(\sum_{m=0}^{\infty} \psi_{m+k} \psi_m \right) e^{-2\pi i f k},$$

where we use the convention $\psi_{m+k} = 0$ if $m + k < 0$. But for $k \geq 0$, we have the autocovariance relation:

$$\sum_{m=0}^{\infty} \psi_{m+k} \psi_m = \frac{1}{\sigma_\varepsilon^2} \gamma_Y(k)$$

and for $k < 0$ the same by symmetry of $\gamma_Y(k)$. So, substituting $\frac{1}{\sigma_\varepsilon^2}\gamma_Y(|k|)$ for the inner sum:

$$\left|\phi(e^{-2\pi i f})\right|^2 = \frac{1}{\sigma_\varepsilon^2} \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-2\pi i f k} = \frac{1}{\sigma_\varepsilon^2} S_Y(f)$$

That is:

$$\boxed{S_Y(f) = \sigma_\varepsilon^2 \left|\phi(e^{-2\pi i f})\right|^2, \quad \phi(z) = \sum_{m=0}^{\infty} \psi_m z^m,\ f \in \left[-\tfrac{1}{2}, \tfrac{1}{2}\right].}$$

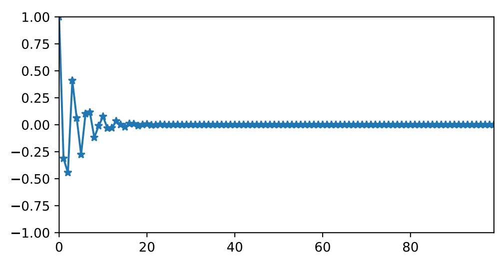
Question 3 *AR(2) process*

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

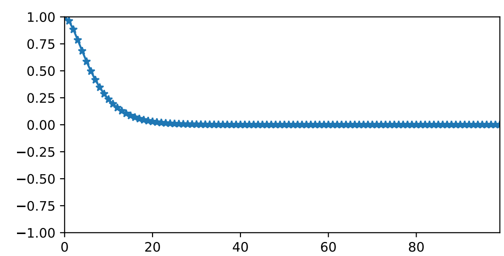
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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?



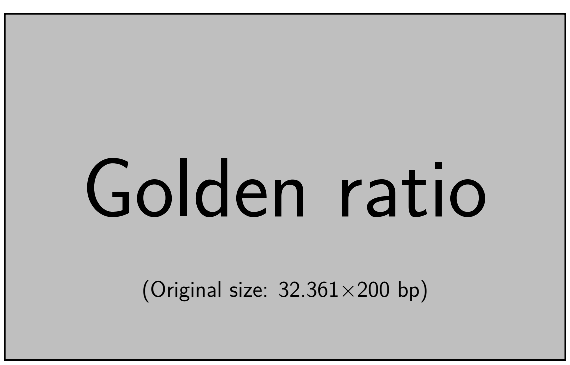
Correlogram of the first AR(2)



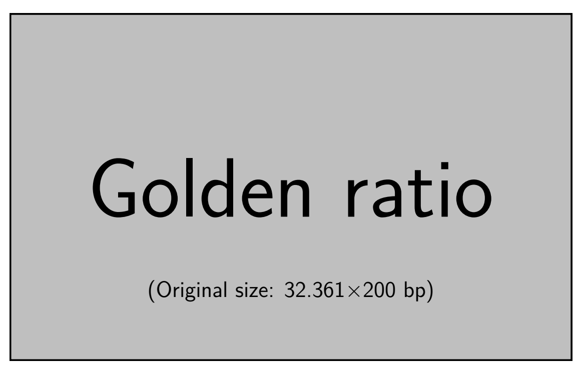
Correlogram of the second AR(2)

Figure 1: Two AR(2) processes

Answer 3



Signal



Periodogram

Figure 2: AR(2) process

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

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{3}$$

where w_L is a modulating window given by

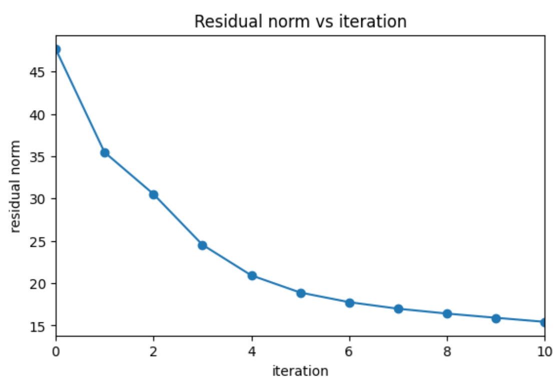
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

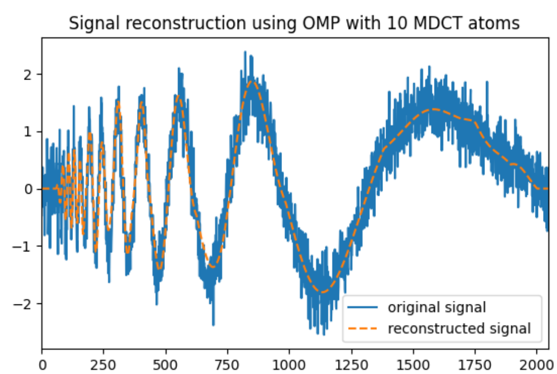
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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



Reconstruction with 10 atoms

Figure 3: Question 4

As shown in Figure 3, the residual norms decrease fast at the beginning, up to iteration 4, then more slowly. Because the residual norm $\|r^{(k)}\|^2$ is the amount of signal that is not explained by the selected atoms at iteration k, this means the first selected atoms capture a large amount of signal (maybe the enveloppe or the low frequencies). From iteration 7, the curve flattens, meaning the algorithm now does not capture much more signal.

Looking at the reconstruction with 10 atoms on the right of Figure 3, we see the reconstruction captures the large-scale oscillations of the signal, but misses the high-frequency fluctuations and the noise.

A Complementary material

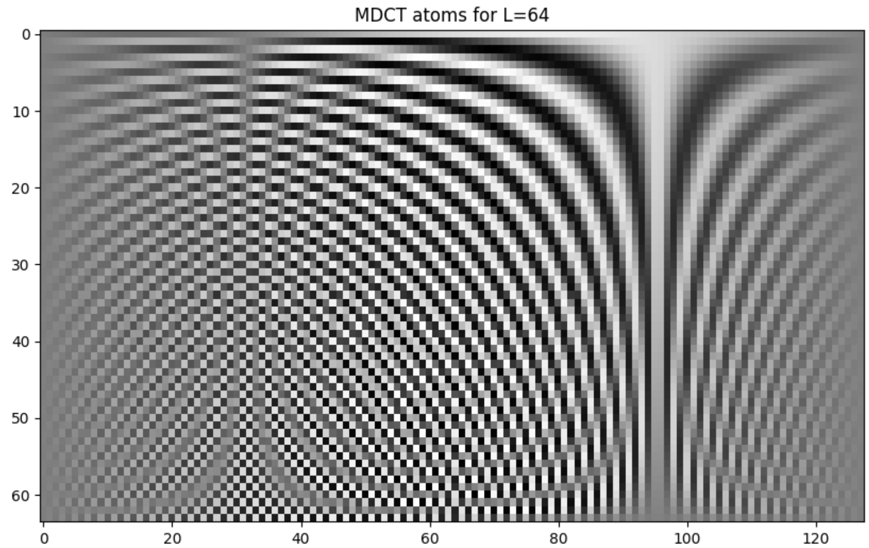


Figure 4: MDCT atoms for L=64

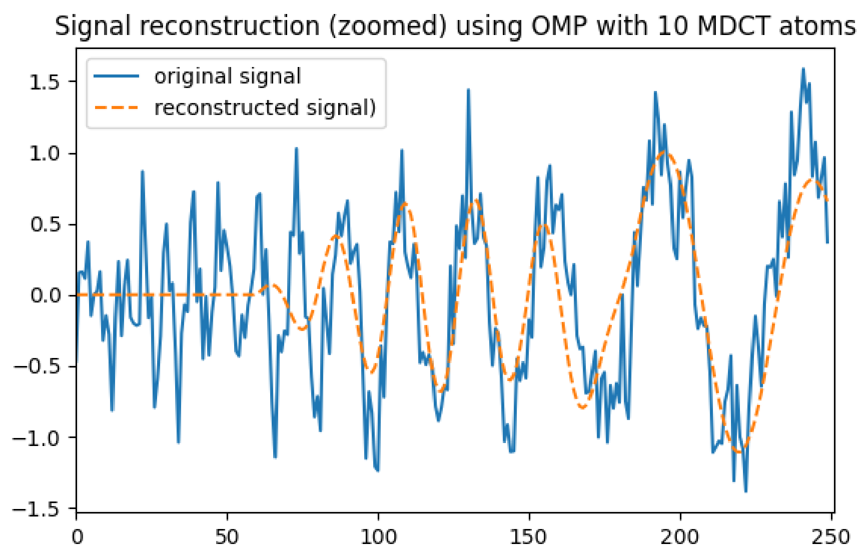


Figure 5: Reconstruction (zoomed) with 10 atoms