Assignment 2 (ML for TS) - MVA

Thomas Gravier thomas.gravier@gadz.org
Richard Goudelin rgoudelin@gmail.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

Firstly:

The rate of convergence of the sample mean for i.i.d. random variables with finite variance is derived from the central limit theorem. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean θ and finite variance σ^2. The sample mean is given by:

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

According to the central limit theorem, the scaled deviation of $\hat{\theta}_n$ from θ converges in distribution to a normal random variable:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

where $\xrightarrow{D}$ denotes convergence in distribution. This indicates that the error in the estimator $\hat{\theta}_n$ decreases at the rate of $O\left(\frac{1}{\sqrt{n}}\right)$.

Secondly:

By applying the provided hint, we derive:

$$\mathbb{E}\left[(\overline{Y}_n - \mu)^2\right] = \mathbb{E}\left[\overline{Y}_n^2 - 2\overline{Y}_n\mu + \mu^2\right] = \mathbb{E}\left[\overline{Y}_n^2\right] - \mu^2 \leq \mathbb{E}\left[\overline{Y}_n^2\right].$$

Let's develop $\mathbb{E}(\overline{Y}_n^2)$:

$$\mathbb{E}(\overline{Y}_n^2) = \mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Y_iY_j\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}(Y_iY_j)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma[i-j]$$

$$\le \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\gamma[i-j]|$$

$$\le \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}2|(\gamma[|i-j|])|$$

$$\le \frac{2}{n^2}\sum_{i=1}^{n}\sum_{k=0}^{n-1}|(\gamma[k])|$$

We know that:

$$\sum_{k=0}^{\infty}|(\gamma[k])| < \infty \implies \exists C \in \mathbb{R}, \quad \sum_{k=0}^{n-1}|(\gamma[k])| \le C.$$

Thus:

$$\mathbb{E}(\overline{Y}_n^2) \le \frac{2C}{n^2}.$$

Therefore:

$$\lim_{n\to\infty}\mathbb{E}\left((\overline{Y}_n-\mu)^2\right) = 0.$$

It follows that:

$$\mathbb{E}\left[(\overline{Y}_n-\mu)^2\right] \to 0 \quad \implies \quad \overline{Y}_n \xrightarrow{L^2} \mu \quad \implies \quad \overline{Y}_n \xrightarrow{p} \mu.$$

Thus, **the sample mean is consistent.**

Given:

$$\mathbb{E}\left[(\bar{Y}_n-\mu)^2\right] \le \frac{1}{n}\sum_k|\gamma(k)|,$$

where $\bar{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t$ is the sample mean and $\gamma(k)$ is the autocovariance function, if $\sum_k|\gamma(k)| < \infty$, then:

$$\mathbb{E}\left[(\bar{Y}_n-\mu)^2\right] \le \frac{C}{n}, \quad C = \sum_k|\gamma(k)|.$$

Thus, the variance decreases as $\mathcal{O}(1/n)$ and the standard deviation as $\mathcal{O}(1/\sqrt{n})$. The sample mean therefore converges to μ with a rate of $\mathcal{O}(1/\sqrt{n})$.

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

First we derive $\mathbb{E}(Y_t)$:

$$\mathbb{E}(Y_t) = \mathbb{E}\left(\sum_{r=0}^{\infty} \psi_r \epsilon_{t-r}\right) = \sum_{r=0}^{\infty} \psi_r \mathbb{E}(\epsilon_{t-r}) = 0.$$

—

Second we derive $\mathbb{E}(Y_t Y_{t-k})$:

$$\begin{aligned}
\mathbb{E}(Y_t Y_{t-k}) &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-k-j}\right)\right] \\
&= \mathbb{E} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \epsilon_{t-i}\epsilon_{t-k-j} \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \mathbb{E}(\epsilon_{t-i}\epsilon_{t-k-j}).
\end{aligned}$$

Separate the cases:

- When $i = j + k$, the noise terms are correlated.
- When $i \neq j + k$, the noise terms are independent.

$$= \sum_{i=0}^{\infty} \psi_i \psi_{i-k} \mathbb{E}(\epsilon_{t-i}^2) + \sum_{i=0}^{\infty}\sum_{j\neq i-k} \psi_i \psi_j \mathbb{E}(\epsilon_{t-i})\mathbb{E}(\epsilon_{t-k-j}).$$

Using the properties of white noise ($\mathbb{E}(\epsilon_t^2) = \sigma_\epsilon^2$ and $\mathbb{E}(\epsilon_t) = 0$:

$$= \sum_{i=k}^{\infty} \psi_i \psi_{i-k} \sigma_\epsilon^2 + 0$$

$$= \sigma_\epsilon^2 \sum_{i=k}^{\infty} \psi_i \psi_{i-k}.$$

This value do not depend on t, this imply that the autocovariance depends only on lag

—

Finally to show that a process is weakly stationary we need to show that the variance is constant $\mathrm{Var}(Y_t)$:

$$\mathrm{Var}(Y_t) = \mathbb{E}(Y_t^2) - [\mathbb{E}(Y_t)]^2$$

Since $\mathbb{E}(Y_t) = 0$:

$$\mathrm{Var}(Y_t) = \mathbb{E}(Y_t^2)$$

$$= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}\right)^2\right]$$

$$= \mathbb{E}\left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \epsilon_{t-i} \epsilon_{t-j}\right]$$

$$= \sum_{i=0}^{\infty} \psi_i^2 \sigma_\epsilon^2$$

$$= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2.$$

and we know that $\sum_k \psi_k^2 < \infty$.

To conclude we got a constant variance, autocorrelation that depend only on lag and an average that is egual to 0 thus this process is weakly stationnary.

Now we can compute

$$\left| \sum_{j=0}^{N} \psi_j e^{-2i\pi f j} \right|^2, \quad \text{for } N \in \mathbb{N}:$$

$$\left| \sum_{j=0}^{N} \psi_j e^{-2i\pi f j} \right|^2 = \left(\sum_{j=0}^{N} \psi_j e^{-2i\pi f j} \right) \left(\sum_{l=0}^{N} \psi_l e^{2i\pi f l} \right)$$

$$= \sum_{j=0}^{N} \sum_{l=0}^{N} \psi_j \psi_l e^{-2i\pi f(j-l)}$$

$$= \sum_{\tau=-N+1}^{N-1} \sum_{n=0}^{N-\tau-1} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau} \quad \text{(e.g Assignment 1).}$$

We make $N \to \infty$, and we finally get:

$$\left| \phi\left(e^{-2i\pi f}\right) \right|^2 = \sum_{\tau=-\infty}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau}.$$

Let's compute the power spectrum, let f:

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-2i\pi f \tau}, \quad \text{with } f_s = 1 \text{ Hz.}$$

$$S(f) = \sigma_\epsilon^2 \sum_{\tau=-\infty}^{\infty} \sum_{n=0}^{\infty} \psi_n \psi_{n+\tau} e^{-2i\pi f \tau}.$$

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

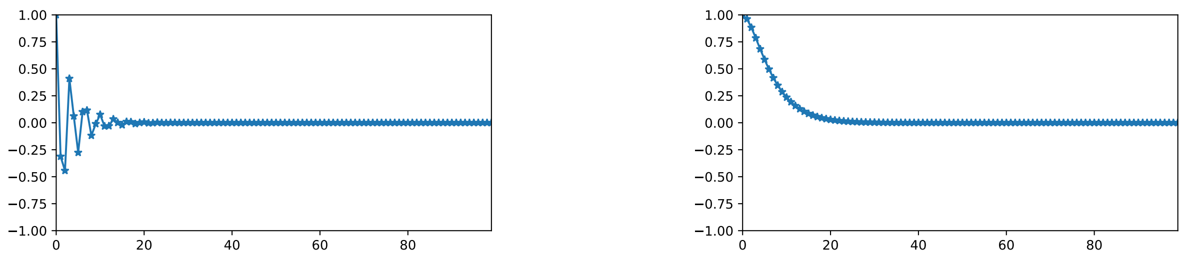
Question 3 *AR(2) process*

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

1. **Expression of Autocovariance Coefficients $\gamma(\tau)$ Using Roots r_1 and r_2:**

   To express $Y(z)$ using r_1 and r_2, we do the same logic for the parameter estimation of an AR model in the course:

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

$$\Leftrightarrow \quad Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t$$

$$\Rightarrow \quad Y_t Y_{t-\tau} - \phi_1 Y_{t-1} Y_{t-\tau} - \phi_2 Y_{t-2} Y_{t-\tau} = \epsilon_t Y_{t-\tau}$$

$$\Rightarrow \quad \mathbb{E}(Y_t Y_{t-\tau}) - \phi_1 \mathbb{E}(Y_{t-1} Y_{t-\tau}) - \phi_2 \mathbb{E}(Y_{t-2} Y_{t-\tau}) = \mathbb{E}(\epsilon_t Y_{t-\tau})$$

$$= \mathbb{E}(\epsilon_t) \cdot \mathbb{E}(Y_{t-\tau}) = 0 \quad \text{(by independence)}$$

$$\Rightarrow \quad \gamma(\tau) = \phi_1 \gamma(\tau-1) + \phi_2 \gamma(\tau-2)$$

So here it is a recurrent series of order 2, with the following polynomial characteristic:

$$z^2 - \phi_1 z - \phi_2 = P(z).$$

So we can see that

$$P(z) = z^2 \Phi(1/z)$$

we deduce that the roots of P are the inverse of the roots of Φ

- **Cas 1 :** $\Phi(z)$ admet deux racines réelles distinctes $\frac{1}{r_1}$ et $\frac{1}{r_2}$ $(r_1 \neq r_2)$

$$\exists (\lambda, \mu) \in \mathbb{R}^2 \text{ tel que } \forall \tau \in \mathbb{N}, \quad \gamma(\tau) = \lambda \left(\frac{1}{r_1}\right)^{\tau} + \mu \left(\frac{1}{r_2}\right)^{\tau}.$$

- **Cas 2 :** $\Phi(z)$ admet une racine double $\frac{1}{r_1} = \frac{1}{r_2} = \frac{1}{r}$

$$\exists (\lambda, \mu) \in \mathbb{R}^2 \text{ tel que } \forall \tau \in \mathbb{N}, \quad \gamma(\tau) = \lambda \left(\frac{1}{r}\right)^{\tau} + \mu \tau \left(\frac{1}{r}\right)^{\tau}.$$

- **Cas 3 :** $\Phi(z)$ admet deux racines complexes conjuguées $\frac{1}{r_1} = \frac{1}{R} e^{-i\theta}$ et $\frac{1}{r_2} = \frac{1}{R} e^{i\theta}$

$$\exists (\lambda, \mu) \in \mathbb{R}^2 \text{ tel que } \forall \tau \in \mathbb{N}, \quad \gamma(\tau) = \left(\frac{1}{R}\right)^{\tau} \big(\lambda \cos(\tau\theta) + \mu \sin(\tau\theta)\big).$$

2. **Process with complex roots:**

   Based on the observations above, we conclude that $r_1, r_2 \in \mathbb{C}$ correspond to an oscillatory system whose autocorrelation fades to 0 over lag (in fact $|r_i| > 1$) . Thus, the correlogram on the left indicates the presence of complex roots. On the other hand, $r_1, r_2 \in \mathbb{R}$ correspond to a second-order system with monotonic decay to 0. Therefore, the correlogram on the right is associated with real roots.

3. **Expression of the Power Spectrum $S(f)$:**

   Let's introduce the lag operator L such that $LY_t = Y_{t-1}$.
   So by applying $\phi(L)$ over Y_t we have:

   $\phi(L)Y_t = (1 - \phi_1 L - \phi_2 L^2)Y_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t \quad (\, Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t)$

   $r1$ and $r2$ being the roots of $\phi(z)$ we have:

$$-\phi_2 (r_1 - L)(r_2 - L) Y_t = -\phi_2 \left(1 - \frac{L}{r_1}\right)\left(1 - \frac{L}{r_2}\right) Y_t = \epsilon_t$$

equivalent to:

$$Y_t = \frac{1}{-\phi_2 \left(1 - \frac{L}{r_1}\right)\left(1 - \frac{L}{r_2}\right)} \epsilon_t$$

We can remark a geometric sum, with $|L/ri| < 1$ cause $|ri| > 1$ that

$$\frac{1}{1 - \frac{L}{r_i}} = \sum_{k=0}^{\infty} \frac{L^k}{r_i^k}$$

Which led to:

$$Y_t = \frac{\epsilon_t}{-\phi_2} \sum_{k=0}^{\infty} \frac{L^k}{r_1^k} \sum_{l=0}^{\infty} \frac{L^l}{r_2^l}$$

$$Y_t = \frac{\epsilon_t}{-\phi_2} \sum_{k=0}^{\infty} \frac{L^k}{r_1^k} \sum_{l=0}^{\infty} \frac{L^l}{r_2^l}$$

$$Y_t = \frac{\epsilon_t}{-\phi_2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{L^{k+l}}{r_1^k r_2^l}$$

$$Y_t = \frac{1}{-\phi_2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{\epsilon_{t-(k+l)}}{r_1^k r_2^l}$$

$$Y_t = \frac{1}{-\phi_2} \sum_{j=0}^{\infty} \sum_{k=0}^{j} \frac{\epsilon_{t-j}}{r_1^k r_2^{j-k}}$$

$$Y_t = \frac{1}{-\phi_2} \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \quad \text{where } c_j = \sum_{k=0}^{j} \frac{1}{r_1^k r_2^{j-k}}.$$

Finally here we recognize that a AR(2) model is a MA(∞)] model thus from question two we deduce that it's power spectrum is :

$$S(f) = \frac{1}{\phi_2^2} \sigma_\epsilon^2 \left| \phi\left(e^{-2i\pi f}\right) \right|^2.$$

4. **Simulation and Periodogram:**

First, let's find the values of ϕ_1 and ϕ_2: Knowing $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$, we have the following system of equations:

$$\begin{cases} 1 - \phi_1 r_1 - \phi_2 r_1^2 = 0, \\ 1 - \phi_1 r_2 - \phi_2 r_2^2 = 0. \end{cases}$$

Solving the system:

$$\phi_1 = \frac{1 - \phi_2 r_1^2}{r_1},$$

$$1 - \frac{1 - \phi_2 r_1^2}{r_1} r_2 - \phi_2 r_2^2 = 0,$$

$$1 - \frac{r_2}{r_1} + \phi_2 r_2 (r_1 - r_2) = 0,$$

$$\phi_2 = -\frac{1}{r_1 r_2}.$$

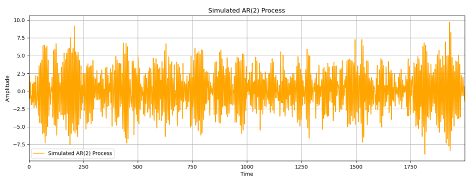
Substituting $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}$:

$$\phi_2 = -\frac{1}{(re^{i\theta})(re^{-i\theta})} = -\frac{1}{r^2}.$$

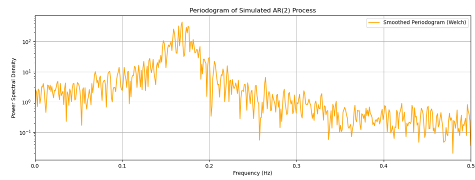
$$\phi_1 = \frac{r_1 + r_2}{r_1 r_2} = \frac{re^{i\theta} + re^{-i\theta}}{r^2} = \frac{2r\cos\theta}{r^2} = \frac{2\cos\theta}{r}.$$

Final expressions:

$$\phi_1 = \frac{2\cos\theta}{r}, \quad \phi_2 = -\frac{1}{r^2}.$$



Signal



Periodogram

Figure 2: AR(2) process

A peak is observed near the frequency $f = 0.17$, indicating a dominant frequency component in the time series. This observation aligns with $\theta = \frac{2\pi}{6}$, as it corresponds to $2\pi f = \theta$, leading to $f = \frac{1}{6}$.

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

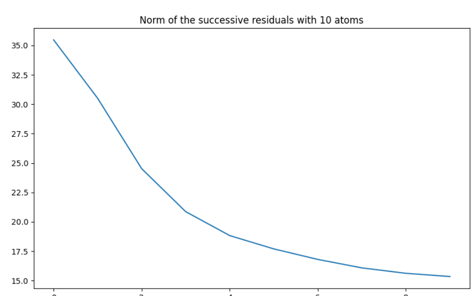
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{4}$$

Question 4 *Sparse coding with OMP*

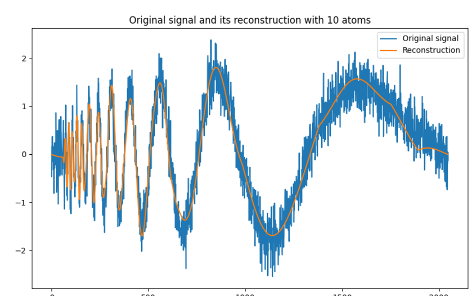
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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