Assignment 2 (ML for TS) - MVA

Firstname Lastname youremail1@mail.com
Firstname Lastname youremail2@mail.com

December 1, 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

For i.i.d. random variables with finite variance, the variance of the sample mean is

$$\mathrm{Var}(\bar{Y}_n) = \frac{\sigma^2}{n}$$

and the convergence rate is $1/\sqrt{n}$.

For a stationary process $\{Y_t\}$ with $\sum_k |\gamma(k)| < +\infty$,

$$\mathrm{Var}(\bar{Y}_n) = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{Cov}(Y_j, Y_k) = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \gamma(j-k)$$

which can also be written as

$$\mathrm{Var}(\bar{Y}_n) = \frac{1}{n} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)$$

If $\sum_k |\gamma(k)| < +\infty$, then as $n \to \infty$,

$$\mathrm{Var}(\bar{Y}_n) \to \frac{1}{n} \sum_{h=-\infty}^{\infty} \gamma(h)$$

so $\mathrm{Var}(\bar{Y}_n) = O(1/n)$, i.e., the convergence rate is $1/\sqrt{n}$ like in the i.i.d. case.

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 $(\psi_k)_{k\geq 0} \subset \mathbb{R}$ ($\psi_k = 1$) 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, since Y_t is a linear combination of zero mean white noises, we have

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

Now for the covariance:

$$\mathbb{E}[Y_t Y_{t-h}] = \mathbb{E}\left[\left(\sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}\right)\left(\sum_{l=0}^{\infty} \psi_l \varepsilon_{t-h-l}\right)\right] = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \psi_k \psi_l \, \mathbb{E}[\varepsilon_{t-k}\varepsilon_{t-h-l}]$$

Since $\{\varepsilon_t\}$ are uncorrelated (and independent), the only nonzero terms are when $t-k = t-h-l$, i.e., $k = h+l$. Therefore,

$$\mathbb{E}[Y_t Y_{t-h}] = \sum_{l=0}^{\infty} \psi_{h+l}\psi_l \, \mathbb{E}[\varepsilon_{t-(h+l)}^2] = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_{h+l}\psi_l$$

So the covariance at lag h is

$$\gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h}) = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \psi_{h+l}\psi_l$$

which does not depend on t: the process is weakly stationary.

To compute the power spectrum, recall that for a stationary process,

$$S(f) = \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-2\pi i f k}.$$

Now, let's expand the modulus square in $S(f) = \sigma_\varepsilon^2 |\phi(e^{-2\pi i f})|^2$:

$$\phi(e^{-2\pi i f}) = \sum_{k=0}^{\infty} \psi_k e^{-2\pi i f k}$$

so

$$|\phi(e^{-2\pi i f})|^2 = \phi(e^{-2\pi i f})\,\overline{\phi(e^{-2\pi i f})} = \left(\sum_{k=0}^{\infty} \psi_k e^{-2\pi i f k}\right)\left(\sum_{j=0}^{\infty} \psi_j e^{2\pi i f j}\right)$$

$$= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \psi_k \psi_j e^{-2\pi i f (k-j)}$$

Therefore, the power spectrum is

$$S(f) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \psi_k \psi_j \, e^{-2\pi i f (k-j)}$$

which is an explicit expansion of the complex conjugate product.

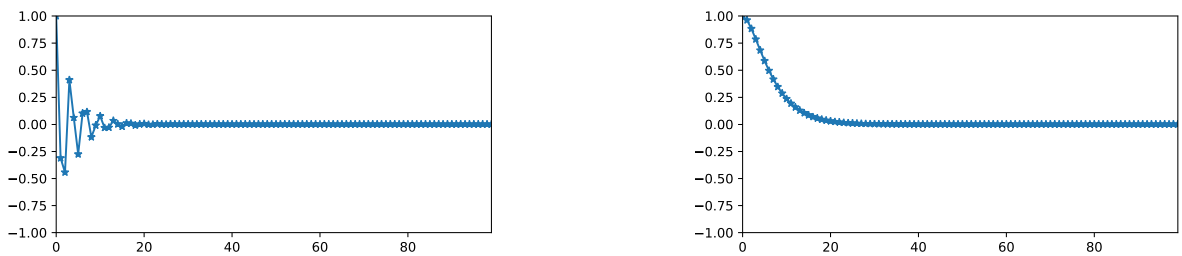
Question 3 *AR(2) process*

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

First, let us establish the stationarity and the form of the solution. Let $\nu_1 = 1/r_1$ and $\nu_2 = 1/r_2$. Since $|r_i| > 1$, we have $|\nu_i| < 1$. Using the state-space representation, let $X_t = [Y_t, Y_{t-1}]^\top$. We can write:

$$X_t = \Phi X_{t-1} + \xi_t, \quad \text{with} \quad \Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \quad \xi_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}. \tag{3}$$

By iterating this equation, we obtain $X_t = \sum_{k=0}^{\infty} \Phi^k \xi_{t-k}$. Since the eigenvalues of Φ are exactly ν_1 and ν_2 (roots of $z^2 - \phi_1 z - \phi_2 = 0$), and $|\nu_i| < 1$, the spectral radius $\rho(\Phi) < 1$. Thus, the series converges in L^2, defined as a sum of independent Gaussian variables. The variance of the process converges to a finite limit $\gamma(0)$, confirming stationarity.

We now derive $\gamma(0)$ and $\gamma(1)$ algebraically using the Yule-Walker equations. Multiplying the AR(2) equation by $Y_{t-\tau}$ and taking expectations yields the recursion:

$$\gamma(\tau) = \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2), \quad \forall \tau \geq 1. \tag{4}$$

For $\tau = 0$, taking the expectation with Y_t (noting $E[\varepsilon_t Y_t] = \sigma^2$):

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2. \tag{5}$$

For $\tau = 1$:

$$\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1) \implies \gamma(1)(1-\phi_2) = \phi_1\gamma(0) \implies \gamma(1) = \frac{\phi_1}{1-\phi_2}\gamma(0). \tag{6}$$

Substituting $\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0)$ into (5):

$$\gamma(0) = \phi_1\gamma(1) + \phi_2(\phi_1\gamma(1) + \phi_2\gamma(0)) + \sigma^2 = (\phi_1^2 + \phi_2)\gamma(1) + \phi_2^2\gamma(0) + \sigma^2. \tag{7}$$

Substituting $\gamma(1)$ from (6):

$$\gamma(0)(1-\phi_2^2) - \gamma(0)\frac{\phi_1(\phi_1 + \phi_1\phi_2)}{1-\phi_2} = \sigma^2. \tag{8}$$

Factorizing and solving for $\gamma(0)$:

$$\gamma(0)\left[\frac{(1-\phi_2)(1-\phi_2^2) - \phi_1^2(1+\phi_2)}{1-\phi_2}\right] = \sigma^2 \implies \gamma(0) = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)((1-\phi_2)^2 - \phi_1^2)}. \tag{9}$$

Using Vieta's formulas, $\phi_1 = \nu_1 + \nu_2$ and $\phi_2 = -\nu_1\nu_2$. The denominator term $(1-\phi_2)^2 - \phi_1^2$ simplifies to $(1+\nu_1\nu_2)^2 - (\nu_1+\nu_2)^2 = (1-\nu_1^2)(1-\nu_2^2)$. Thus:

$$\gamma(0) = \frac{(1+\nu_1\nu_2)\sigma^2}{(1-\nu_1\nu_2)(1-\nu_1^2)(1-\nu_2^2)}. \tag{10}$$

The autocovariance satisfies the homogeneous difference equation for $\tau \geq 1$, so its solution is of the form $\gamma(\tau) = c_1\nu_1^\tau + c_2\nu_2^\tau$. We solve for c_1, c_2 using the boundary conditions at $\tau = 0, 1$:

$$\begin{pmatrix} 1 & 1 \\ \nu_1 & \nu_2 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix}. \tag{11}$$

Inverting the matrix:

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{\nu_2 - \nu_1}\begin{pmatrix} \nu_2 & -1 \\ -\nu_1 & 1 \end{pmatrix}\begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix}. \tag{12}$$

Using $\gamma(1) = \frac{\nu_1+\nu_2}{1+\nu_1\nu_2}\gamma(0)$, we calculate c_1:

$$c_1 = \frac{\nu_2\gamma(0) - \gamma(1)}{\nu_2 - \nu_1} = \frac{\gamma(0)}{\nu_2 - \nu_1}\left(\nu_2 - \frac{\nu_1+\nu_2}{1+\nu_1\nu_2}\right) = \frac{\gamma(0)}{\nu_2 - \nu_1}\frac{\nu_2 + \nu_1\nu_2^2 - \nu_1 - \nu_2}{1+\nu_1\nu_2} = \gamma(0)\frac{\nu_1\nu_2^2 - \nu_1}{(\nu_2 - \nu_1)(1+\nu_1\nu_2)}. \tag{13}$$

This simplifies to:

$$c_1 = \frac{\sigma^2\nu_1(1-\nu_2^2)}{(\nu_1 - \nu_2)(1-\nu_1\nu_2)(1-\nu_1^2)(1-\nu_2^2)} = \frac{\sigma^2\nu_1}{(\nu_1 - \nu_2)(1-\nu_1\nu_2)(1-\nu_1^2)}. \tag{14}$$

By symmetry for c_2 and expressing in terms of roots $r_i = 1/\nu_i$:

$$\gamma(\tau) = \sigma^2\sum_{i=1}^{2}\frac{r_i^{-|\tau|}}{(1-r_i^{-2})(1-r_1^{-1}r_2^{-1})(r_i^{-1} - r_{3-i}^{-1})}r_i^{-1}. \tag{15}$$

The left correlogram exhibits oscillating behavior characteristic of complex conjugate roots in the characteristic polynomial, whereas the right correlogram displays a monotonic exponential

decay typical of real roots, allowing us to identify the left as the complex case and the right as the real case.

The power spectrum density $S(f)$ is the Fourier transform of the autocovariance function:

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i2\pi f \tau}. \tag{16}$$

Using the expression derived previously:

$$\gamma(\tau) = c_1 \nu_1^{|\tau|} + c_2 \nu_2^{|\tau|}, \quad \text{where } \nu_i = 1/r_i. \tag{17}$$

We can split the sum for each component $k \in \{1, 2\}$:

$$S_k(f) = \sum_{\tau=-\infty}^{\infty} \nu_k^{|\tau|} e^{-i2\pi f \tau} = 1 + \sum_{\tau=1}^{\infty} (\nu_k e^{-i2\pi f})^\tau + \sum_{\tau=1}^{\infty} (\nu_k e^{i2\pi f})^\tau. \tag{18}$$

Since $|\nu_k| < 1$, these are convergent geometric series:

$$S_k(f) = 1 + \frac{\nu_k e^{-i2\pi f}}{1 - \nu_k e^{-i2\pi f}} + \frac{\nu_k e^{i2\pi f}}{1 - \nu_k e^{i2\pi f}} = 1 + \frac{\nu_k e^{-i\omega}}{1 - \nu_k e^{-i\omega}} + \frac{\nu_k e^{i\omega}}{1 - \nu_k e^{i\omega}}, \tag{19}$$

where $\omega = 2\pi f$. Combining terms:

$$S_k(f) = \frac{(1 - \nu_k e^{-i\omega})(1 - \nu_k e^{i\omega}) + \nu_k e^{-i\omega}(1 - \nu_k e^{i\omega}) + \nu_k e^{i\omega}(1 - \nu_k e^{-i\omega})}{|1 - \nu_k e^{-i\omega}|^2}. \tag{20}$$

The numerator simplifies to:

$$1 - \nu_k e^{i\omega} - \nu_k e^{-i\omega} + \nu_k^2 + \nu_k e^{-i\omega} - \nu_k^2 + \nu_k e^{i\omega} - \nu_k^2 = 1 - \nu_k^2. \tag{21}$$

Thus:

$$S_k(f) = \frac{1 - \nu_k^2}{|1 - \nu_k e^{-i2\pi f}|^2}. \tag{22}$$

Substituting c_k:

$$S(f) = c_1 \frac{1 - \nu_1^2}{|1 - \nu_1 z^{-1}|^2} + c_2 \frac{1 - \nu_2^2}{|1 - \nu_2 z^{-1}|^2}, \quad \text{with } z = e^{i2\pi f}. \tag{23}$$

Recall from the derivation of $\gamma(0)$:

$$c_1(1 - \nu_1^2) = \frac{\sigma^2 \nu_1}{(\nu_1 - \nu_2)(1 - \nu_1 \nu_2)}, \quad c_2(1 - \nu_2^2) = \frac{\sigma^2 \nu_2}{(\nu_2 - \nu_1)(1 - \nu_1 \nu_2)}. \tag{24}$$

So:

$$S(f) = \frac{\sigma^2}{1 - \nu_1 \nu_2} \left(\frac{\nu_1}{(\nu_1 - \nu_2)|1 - \nu_1 z^{-1}|^2} + \frac{\nu_2}{(\nu_2 - \nu_1)|1 - \nu_2 z^{-1}|^2} \right). \tag{25}$$

Finding a common denominator $D = |1 - \nu_1 z^{-1}|^2 |1 - \nu_2 z^{-1}|^2 = |(1 - \nu_1 z^{-1})(1 - \nu_2 z^{-1})|^2 = |1 - (\nu_1 + \nu_2) z^{-1} + \nu_1 \nu_2 z^{-2}|^2$. This matches $|\phi(e^{-i2\pi f})|^2$ since $\phi_1 = \nu_1 + \nu_2$ and $\phi_2 = -\nu_1 \nu_2$.

The numerator term is:

$$N = \frac{\sigma^2}{(\nu_1 - \nu_2)(1 - \nu_1 \nu_2)} \left[\nu_1 (1 - \nu_2 z)(1 - \nu_2 z^{-1}) - \nu_2 (1 - \nu_1 z)(1 - \nu_1 z^{-1}) \right]. \tag{26}$$

Expanding the bracket:

$$\nu_1(1-\nu_2(z+z^{-1})+\nu_2^2)-\nu_2(1-\nu_1(z+z^{-1})+\nu_1^2) \tag{27}$$
$$=\nu_1-\nu_1\nu_2(z+z^{-1})+\nu_1\nu_2^2-\nu_2+\nu_1\nu_2(z+z^{-1})-\nu_1^2\nu_2 \tag{28}$$
$$=(\nu_1-\nu_2)+\nu_1\nu_2(\nu_2-\nu_1)=(\nu_1-\nu_2)(1-\nu_1\nu_2). \tag{29}$$

Thus, the entire numerator N simplifies to just σ^2.

$$S(f)=\frac{\sigma^2}{|\phi(e^{-i2\pi f})|^2}=\frac{\sigma^2}{|1-\phi_1 e^{-i2\pi f}-\phi_2 e^{-i4\pi f}|^2}. \tag{30}$$

To analyze the peak of the spectrum, let us factorize the characteristic polynomial as $\phi(z)=(1-z/r_1)(1-z/r_2)$. Assuming complex conjugate roots $r_{1,2}=\rho e^{\pm i\theta}$ with $\rho>1$, the power spectrum is given by:

$$S(f)=\frac{\sigma^2}{|1-\rho^{-1}e^{i(2\pi f-\theta)}|^2|1-\rho^{-1}e^{i(2\pi f+\theta)}|^2}. \tag{31}$$

T For complex conjugate roots $r_{1,2}=re^{\pm i\theta}$ with $r=1.05$ and $\theta=2\pi/6$, we have $\nu_{1,2}=r^{-1}e^{\pm i\theta}$ where $r^{-1}\approx 0.952$. The power spectrum is:

$$S(f)=\frac{\sigma^2}{|1-\phi_1 e^{-i2\pi f}-\phi_2 e^{-i4\pi f}|^2}=\frac{\sigma^2}{|(1-\nu_1 e^{-i2\pi f})(1-\nu_2 e^{-i2\pi f})|^2}. \tag{32}$$

For frequency $f=f_0:=\theta/(2\pi)=1/6$ Hz, we have $e^{-i2\pi f_0}=e^{-i\theta}$. Thus:

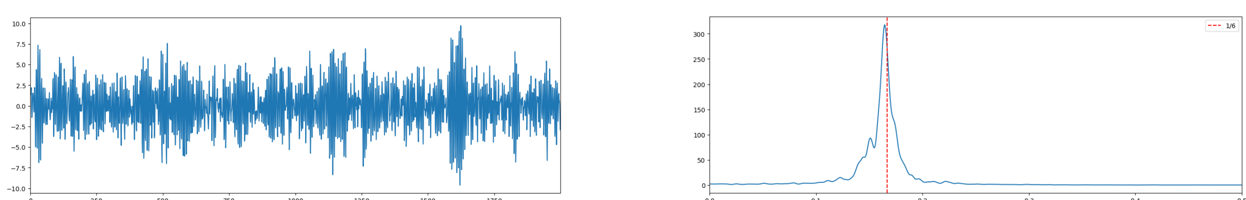
$$|1-\nu_1 e^{-i\theta}|^2=|1-r^{-1}e^{i\theta}e^{-i\theta}|^2=|1-r^{-1}|^2=(1-r^{-1})^2. \tag{33}$$

Similarly, $|1-\nu_2 e^{-i\theta}|^2=|1-r^{-1}e^{-i\theta}e^{-i\theta}|^2=|1-r^{-1}e^{-i2\theta}|^2$. Since $r\approx 1$, the denominator $|1-r^{-1}|^2=(r-1)^2/r^2\approx(r-1)^2$ is very small. For $r=1.05$, we have $(r-1)^2=0.05^2=0.0025$, yielding:

$$S(f_0)\approx\frac{\sigma^2}{(r-1)^2\cdot|1-r^{-1}e^{-i2\theta}|^2}\approx\frac{\sigma^2}{0.0025\cdot C}\sim 400\sigma^2/C, \tag{34}$$

where $C=|1-r^{-1}e^{-i2\theta}|^2\approx O(1)$. This produces a sharp spectral peak at frequency $f=\theta/(2\pi)=1/6$ Hz, corresponding to the imaginary part (phase) of the complex roots, with amplitude scaling as $(r-1)^{-2}$, explaining the observed peak around magnitude ~ 350 in the periodogram.

We can see that the process is stationnary non diverging and the fourier spectrum shows a peak at the frequency of the roots.



Signal Spectrum

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

$$\forall u = 0, \dots, 2L-1, \quad \phi_{L,k}[u] = w_L[u]\sqrt{\frac{2}{L}}\cos[\frac{\pi}{L}\left(u + \frac{L+1}{2}\right)(k + \frac{1}{2})] \tag{35}$$

where w_L is a modulating window given by

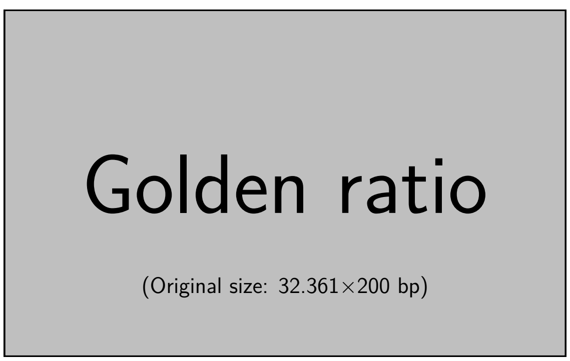
$$w_L[u] = \sin\left[\frac{\pi}{2L}\left(u + \frac{1}{2}\right)\right]. \tag{36}$$

Question 4 *Sparse coding with OMP*

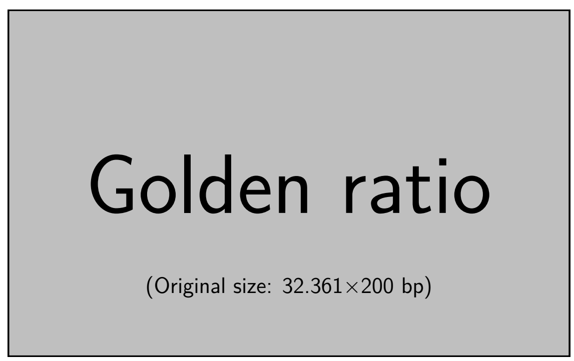
For the signal provided in the notebook, learn a sparse representation with MDCT atoms. The dictionary is defined as the concatenation of all shifted MDCDT atoms for scales L in $[32, 64, 128, 256, 512, 1024]$.

- 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