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

Léa Khalil leakhalil@yahoo.fr
Basile Terver terverbasile@gmail.com

December 5, 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 ∞ 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 us consider iid random variables X_n with mean θ and finite variance σ^2. The sample mean is none other than a simple Monte Carlo estimator $S_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, which converges strongly to $\mathbb{E}(X) = \theta$ in virtue of the law of averages. The sample mean hence converges in probability to θ.
Let us now evaluate the rate of this convergence. Let $\epsilon \geq 0$, then the Bienaymé-Tchebychev inequality gives us:

$$\mathbb{P}(|S_n - \theta| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Let us rewrite $\eta = \frac{\sigma^2}{n\epsilon^2}$, then, with probability $1-\eta$, we have that:

$$|S_n - \theta| \leq \frac{\sigma}{\sqrt{n\eta}}$$

Finally, we get a convergence rate of $\frac{1}{\sqrt{n}}$ for the sample mean.
We could obtain a similar result by exploiting directly the central limit theorem, which states that $\sqrt{n}(S_n - \theta)$ converges in law to the normal distribution.
Let us now show that the sample mean $\bar{Y}_n$ is a consistent estimator.

$$\begin{aligned}
(\bar{Y}_n - \theta)^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=-i+1}^{n-i}(Y_i - \theta)(Y_{k+i} - \theta) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=-i+1}^{n-i} Y_i Y_{k+i} - \theta(Y_i + Y_{k+1}) + \theta^2 \\
\mathbb{E}[(\bar{Y}_n - \theta)^2] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=-i+1}^{n-i}\gamma(k) - \theta^2 \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=-i+1}^{n-i}\gamma(k) \\
&\leq \frac{1}{n}\sum_{k=-n+1}^{n-1}\gamma(k) \\
&\leq \frac{1}{n}\sum_{k=-\infty}^{+\infty}\gamma(k)
\end{aligned}$$

Indeed, we have seen that the $\gamma(k)$ family is summable. This proves that the quantity $\bar{Y}_n$ converges (L2) to θ. Again, if we use Bienaymé-Tchebychev's inequality, we can show that for any $\epsilon \geq 0$:

$$\mathbb{P}(|\bar{Y}_n - \theta| \geq \epsilon) \leq \frac{\mathbb{E}[(\bar{Y}_n - \theta)^2]}{\epsilon^2}$$

and we have just proven that :

$$\lim_{n=+\infty} \mathbb{E}[(\bar{Y}_n - \theta)^2] = 0$$

This proves that the estimator $\bar{Y}_n$ converges in probability to θ and is consistent. .

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

Let us denote $Y_t^n = \sum_{k=0}^{n} \psi_k \varepsilon_{t-k}$. We will first show that Y_t^n converges to Y_t for the L2 norm. Let $n \in \mathbb{N}$ and $N \geq n+1$.

$$Y_t^N - Y_t^n = \sum_{k=n+1}^{N} \psi_k \varepsilon_{t-k}$$

$$(Y_t^N - Y_t^n)^2 = \sum_{k=n+1}^{N} \sum_{l=n+1}^{N} \psi_k \psi_l \varepsilon_{t-k} \varepsilon_{t-l}$$

$$\mathbb{E}(Y_t^N - Y_t^n)^2 = \sum_{k=n+1}^{N} \psi_k^2 \sigma_\epsilon^2$$

Indeed, ε_t is a Gaussian white noise with 0 mean so $\gamma(k) = \sigma_\epsilon^2 \delta_0(k)$. We have also seen that the family ϕ_k is square summable, so the sum is definite when N goes to infinity. We can hence rewrite:

$$\mathbb{E}(Y_t - Y_t^n)^2 = \sum_{k=n+1}^{+\infty} \psi_k^2 \sigma_\epsilon^2$$

$$\lim_{n\to+\infty} \mathbb{E}(Y_t - Y_t^n)^2 = 0$$

Because the sum of $\phi$is convergent. We have hence proven that Y_t^n converges to Y_t in L2. However, this convergence implies the L1 convergence (indeed, the Cauchy-Schwarz inequality gives us that:

$$\lim_{n\to+\infty} \mathbb{E}\left|Y_t - Y_t^n\right| = 0$$

For all $n \in \mathbb{N}$, we have that:

$$\mathbb{E}(Y_t^n) = \sum_{k=0}^{n} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0$$

This means that:

$$\mathbb{E}(Y_t) = \lim_{n\to+\infty} \mathbb{E}(Y_t^n) = 0$$

Let us now compute $\mathbb{E}(Y_t Y_{t-k})$.

$$\begin{aligned}
Y_t Y_{t-k} &= \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-(j+k)} \\
&= \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i} \sum_{l=k}^{+\infty} \psi_{l-k} \varepsilon_{t-l} \\
&= \sum_{i=k}^{+\infty} \sum_{l=k}^{+\infty} \psi_i \psi_{l-k} \varepsilon_{t-i} \varepsilon_{t-l} + \sum_{i=0}^{+k-1} \sum_{l=k}^{+\infty} \psi_i \psi_{l-k} \varepsilon_{t-i} \varepsilon_{t-l} \\
\mathbb{E}[Y_t Y_{t-k}] &= \sum_{i=k}^{+\infty} \psi_i \psi_{i-k} \mathbb{E}[\varepsilon_{t-i}^2] \\
\mathbb{E}[Y_t Y_{t-k}] &= \sigma_\epsilon^2 \sum_{i=k}^{+\infty} \psi_i \psi_{i-k} = \sigma_\epsilon^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+k}
\end{aligned}$$

We can also show by reproducing the same calculation and splitting the sums in a similar way that $\gamma(k) = \gamma(-k)$, which finishes to prove that $\gamma(k)$ depends only on the absolute value of the difference of indexes between the two Y. This proves that the process Y_t is weakly stationary:

- For all t, $\mathbb{E}[Y_t] = 0$
- For all t, k, $\gamma(k) = \mathbb{E}[Y_t Y_{t-k}] = \sigma_\epsilon^2 \sum_{i=k}^{+\infty} \psi_i \psi_{i-k} = \sigma_\epsilon^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+k}$ is only a function of k
- $\gamma(-k) = \gamma(k)$
- $\mathbb{E}[Y_t + nY_{t+m}]$ is only a function of $|m - n|$

Let us now compute the spectrum of this signal, for which we take $f_s = 1$.

$$\begin{aligned}
S(f) &= \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-2ik\pi f} \\
&= \sum_{k=-\infty}^{+\infty} \sigma_\epsilon^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+|k|} e^{-2ik\pi f} \\
&= \sum_{k=1}^{+\infty} \sigma_\epsilon^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+k} (e^{2ik\pi f} + e^{-2ik\pi f}) + \sigma_\epsilon^2 \sum_{i=0}^{+\infty} \psi_i^2
\end{aligned}$$

Let us now take a look at:

$$\left|\phi(e^{-2\pi i f})\right|^2 = \sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty}\psi_k\psi_l z^{l-k}$$

$$= \sum_{l=0}^{+\infty}\sum_{k=1}^{+\infty}\psi_k\psi_{l+k}(z^k + z^{-k}) + \sum_{l=0}^{+\infty}\psi_l^2$$

With $z = e^{-2\pi i f}$, we finally conclude that:

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

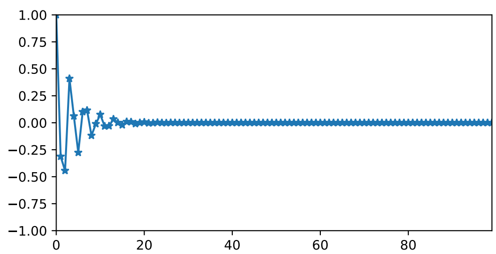
Question 3 *AR(2) process*

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

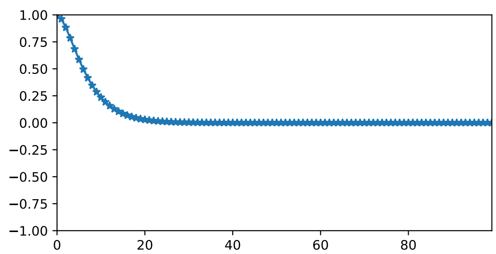
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{2}$$

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 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 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

Expressing the autocovariance

First, note that the initial condition is $Y_1 = \varepsilon_1$. Since r_1 and r_2 are outside the unit circle, the process $\{Y_t\}$ is wide-sense stationary. Therefore, we have that the mean $\mu = \mathbb{E}[Y_t] = \mathbb{E}[Y_1] = \mathbb{E}[\varepsilon_1] = 0$.

Let us now compute the autocovariance function, for a delay $\tau \in \mathbb{N}$, we have

$$\begin{aligned}
\gamma(\tau) &= \mathbb{E}[(Y_t - \mu)(Y_{t-\tau} - \mu)] = \mathbb{E}[Y_t Y_{t-\tau}] \\
&= \mathbb{E}[(\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t) Y_{t-\tau}] \\
&= \phi_1 \mathbb{E}[Y_{t-1} Y_{t-\tau}] + \phi_2 \mathbb{E}[Y_{t-2} Y_{t-\tau}] + \mathbb{E}[\varepsilon_t Y_{t-\tau}] \\
&= \begin{cases} \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2), \text{ for } \tau \neq 0, \\ \phi_1 \gamma(\tau - 1) + \phi_2 \gamma(\tau - 2) + \sigma^2, \text{ for } \tau = 0, \end{cases}
\end{aligned}$$

where σ^2 is the variance of the white noise ε. Let us consider the following associated initial conditions

$$
\begin{aligned}
\gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2, \\
\gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(1), \\
\gamma(2) &= \phi_1\gamma(1) + \phi_2\gamma(0)
\end{aligned}
$$

Recall the known formulas on the sum and product of the roots of a second order polynomial, which yields $\phi_1 = r_1^{-1} + r_2^{-1}$ and $\phi_2 = -r_1^{-1}r_2^{-1}$. Thus, the above system becomes

$$
\begin{aligned}
\gamma(0) &= (r_1^{-1} + r_2^{-1})\gamma(1) - r_1^{-1}r_2^{-1}\gamma(2) + \sigma^2, \\
\gamma(1) &= (r_1^{-1} + r_2^{-1})\gamma(0) - r_1^{-1}r_2^{-1}\gamma(1), \\
\gamma(2) &= (r_1^{-1} + r_2^{-1})\gamma(1) - r_1^{-1}r_2^{-1}\gamma(0).
\end{aligned} \tag{3}
$$

In virtue of a result of a result [2] (Section 3.2.1), we look for solutions of the above system of the form

$$
\gamma(\tau) = \alpha_1 r_1^{-\tau} + \alpha_2 r_2^{-\tau}, \quad \tau \in \mathbb{N}, \tag{4}
$$

where α_1, α_2 are arbitrary constants. Plugging (4) into (3), we get

$$
\begin{aligned}
\alpha_1 + \alpha_2 &= (r_1^{-1} + r_2^{-1})(\alpha_1 r_1^{-1} + \alpha_2 r_2^{-1}) - r_1^{-1}r_2^{-1}(\alpha_1 r_1^{-2} + \alpha_2 r_2^{-2}) + \sigma^2, \\
\alpha_1 r_1^{-1} + \alpha_2 r_2^{-1} &= (r_1^{-1} + r_2^{-1})(\alpha_1 + \alpha_2) - r_1^{-1}r_2^{-1}(\alpha_1 r_1^{-1} + \alpha_2 r_2^{-1}), \\
(\alpha_1 r_1^{-2} + \alpha_2 r_2^{-2}) &= (r_1^{-1} + r_2^{-1})(\alpha_1 r_1^{-1} + \alpha_2 r_2^{-1}) - r_1^{-1}r_2^{-1}(\alpha_1 + \alpha_2).
\end{aligned}
$$

By linear combination and substitution, we solve this system with only 2 unknowns (α_1, α_2), and get

$$
\gamma(\tau) = \frac{\sigma^2 r_1^2 r_2^2}{(r_1 r_2 - 1)(r_2 - r_1)}[(r_1^2 - 1)^{-1} r_1^{1-\tau} - (r_2^2 - 1)^{-1} r_2^{1-\tau}]. \tag{5}
$$

Identifying the correlogram depending on the roots of the characteristic polynomial

From [2] again, we get that, if $r_1 = re^{i\theta}$ and $r_2 = re^{-i\theta}, 0 < \theta < \pi$, we can write (5) as

$$
\gamma(\tau) = \frac{\sigma^2 r^{4-\tau}\sin(\tau\theta + \psi)}{(r^2 - 1)(r^4 - 2r^2\cos(2\theta) + 1)\sin(\theta)}. \tag{6}
$$

where

$$
\tan(\psi) = \frac{r^2 + 1}{r^2 - 1}\tan(\theta) \tag{7}
$$

and $\cos\psi$ has the same sign as $\cos\theta$. Therefore in this case of complex roots, γ has the form of a damped sinusoid with damping factor r^{-1} and period $2\pi/\theta$. If the roots are close to the unit circle, then r is close to 1, the damping is slow, and we obtain a nearly sinusoidal autocovariance function.

Since an $(AR(p)$ process is wide-sense stationary, we compute the autocorrelation function ρ of such a process from its autocovariance function γ as

$$
\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}. \tag{8}
$$

Thus, the correlogram is just a rescaling of the plot of the autocovariance function.

Therefore, since the correlogram of the second AR(2) is not a sinusoid, we know it has real roots. Thus, the first AR(2) has to be the one with complex roots.

Expressing the power spectral density

The power spectral density is defined for wide-sense stationary processes as the Fourier transform of the autocovariance function. Thus, the power spectral density of the $AR(2)$ process Y with sampling frequency $f_s = 1$ Hz is

$$\begin{aligned} S(f) &= \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-2ik\pi f} \\ &= \frac{\sigma^2 r_1^2 r_2^2}{(r_1 r_2 - 1)(r_2 - r_1)} \sum_{k=-\infty}^{+\infty} [(r_1^2-1)^{-1} r_1^{1-|k|} - (r_2^2-1)^{-1} r_2^{1-|k|}] e^{-2ik\pi f} \\ &= \frac{\sigma^2 r_1^2 r_2^2}{(r_1 r_2 - 1)(r_2 - r_1)} \left[(r_1^2-1)^{-1} r_1 \sum_{k=-\infty}^{+\infty} r_1^{-|k|} e^{-2i\pi fk} - (r_2^2-1)^{-1} r_2 \sum_{k=-\infty}^{+\infty} r_2^{-|k|} e^{-2i\pi fk} \right]. \end{aligned} \tag{9}$$

Let us focus on the first term, where the geometric series equals

$$\begin{aligned} \sum_{k=-\infty}^{+\infty} r_1^{-|k|} e^{-2i\pi fk} &= 1 + \sum_{k=1}^{+\infty} \frac{[e^{-2i\pi fk} + e^{2i\pi fk}]}{r_1^k} = 1 + \frac{e^{-2i\pi f}}{r_1(1-\frac{e^{-2i\pi f}}{r_1})} + \frac{e^{2i\pi f}}{r_1(1-\frac{e^{2i\pi f}}{r_1})} \\ &= 1 + \frac{e^{-2i\pi f}}{r_1 - e^{-2i\pi f}} + \frac{e^{2i\pi f}}{r_1 - e^{2i\pi f}} = 1 + \frac{e^{-2i\pi f} r_1 - 1 + e^{2i\pi f} r_1 - 1}{r_1^2 - r_1 e^{-2i\pi f} - e^{2i\pi f} r_1 + 1}. \end{aligned} \tag{10}$$

Similarly, we get

$$\sum_{k=-\infty}^{+\infty} r_2^{-|k|} e^{-2i\pi fk} = 1 + \frac{e^{-2i\pi f} r_2 - 1 + e^{2i\pi f} r_2 - 1}{r_2^2 - r_2 e^{-2i\pi f} - e^{2i\pi f} r_2 + 1}. \tag{11}$$

On the other hand, we have $\phi(z) = -\phi_2(z-r_1)(z-r_2)$ so

$$|\phi(e^{-2i\pi f})|^2 = \phi_2^2 (e^{-2i\pi f} - r_1)(e^{-2i\pi f} - r_2)(e^{2i\pi f} - r_1)(e^{2i\pi f} - r_2) \tag{12}$$

which, using $\phi_2 = -r_1^{-1} r_2^{-1}$, yields

$$|\phi(e^{-2i\pi f})|^2 = r_1^{-2} r_2^{-2} (e^{-2i\pi f} - r_1)(e^{-2i\pi f} - r_2)(e^{2i\pi f} - r_1)(e^{2i\pi f} - r_2) \tag{13}$$

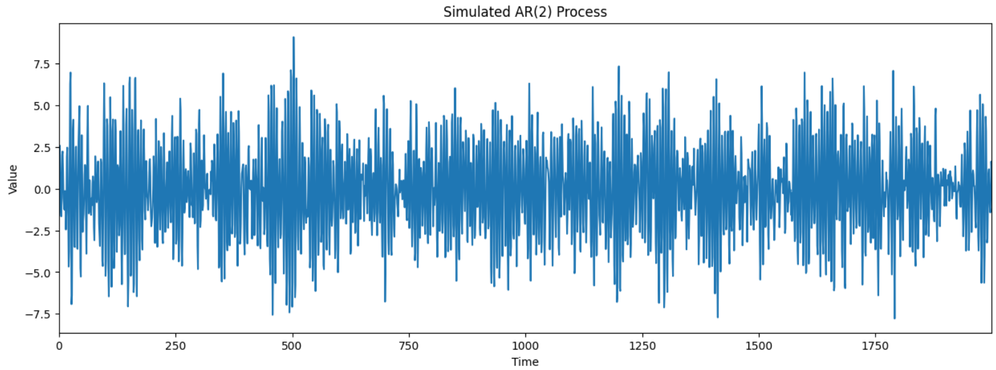
and therefore

$$\begin{aligned} \frac{\sigma^2}{|\phi(e^{-2\pi if})|^2} &= \sigma^2 r_1^2 r_2^2 \left[(e^{-2i\pi f} - r_1)(e^{-2i\pi f} - r_2)(e^{2i\pi f} - r_1)(e^{2i\pi f} - r_2) \right]^{-1} \\ &= \sigma^2 r_1^2 r_2^2 \times \\ & \Big[1 - (r_2 + r_1) e^{-2i\pi f} + r_1 r_2 e^{-4i\pi f} - r_2 e^{2i\pi f} + r_2^2 + r_1 r_2 - r_2^2 r_1 e^{-2i\pi f} + \\ & -r_1 e^{2i\pi f} + r_1 r_2 + r_1^2 - r_1^2 r_2 e^{-2i\pi f} + r_1 r_2 e^{4i\pi f} - r_1 r_2^2 e^{2i\pi f} - r_1^2 r_2 e^{2i\pi f} + r_1^2 r_2^2 \Big]^{-1} \end{aligned} \tag{14}$$

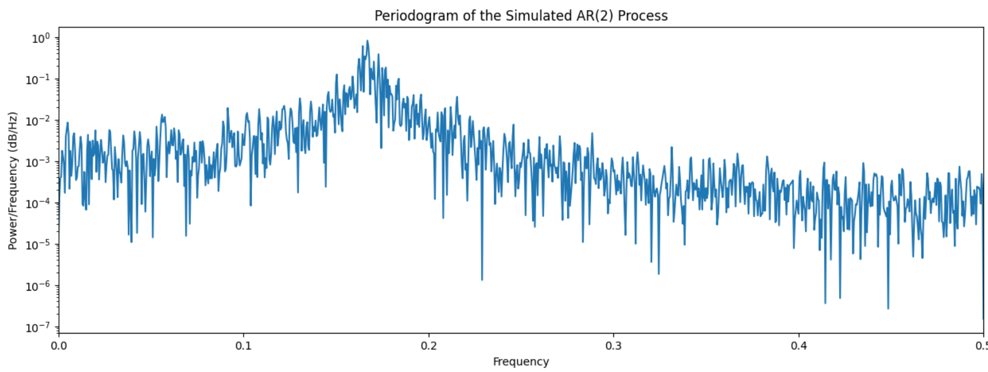
Plugging (10) and (11) in (9), and subsquently putting at the same denominator, the result equals (14). Thus, we conclude

$$S(f) = \frac{\sigma^2}{|\phi(e^{-2\pi if})|^2} \tag{15}$$

Plotting the periodogram in the case of complex roots



Signal



Periodogram

Figure 2: AR(2) process

For the last part of the question, we simulate an $AR(2)$ process and plot its periodogram. One can see the plots in 2. To have complex conjugate roots equal to $r_1 = re^{-i\pi/3}, r_2 = re^{i\pi/3}$, we used $\phi_1 = r_1^{-1} + r_2^{-1} = \frac{2}{r}\cos(\pi/3) = r^{-1}$ and $\phi_2 = -r_1^{-1}r_2^{-1} = -r^{-2}$.

One can see a peak in the periodogram. We printed the frequency corresponding to this peak and obtained `The frequency associated to the peak on the spectrum is 0.167`. This is expected because we have complex conjugate roots of ϕ and, as said above in this answer, this corresponds to an autocovariance function being a damped sinusoid of frequency $\frac{\theta}{2\pi}$. Here, $\theta = \pi/3$ so the frequency of the sinusoid is $\frac{\pi}{6\pi} = \frac{1}{6} = 0.1666\ldots$. Since the periodogram is the Fourier transform of the autocovariance function, it should have a peak at the fundamental frequency of the autocovariance function, which is precisely $\frac{1}{6}$. This is indeed the case.

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

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

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

Answer 4

Contrary to the previous assignment, we are now looking for sparse coding that is not convolutional, and with l_0-regularization (where we constrain the activation vectors to have K_0 non-null entries). In our case, this yields

$$\min_{(\mathbf{z}_k)_k, \|\mathbf{z}\|_0 = K_0} \left\| \mathbf{x} - \sum_{L \in \mathcal{L}} \sum_{k=0}^{L-1} \mathbf{z}_{L,k}\phi_{L,k} \right\|_2^2 , \tag{18}$$

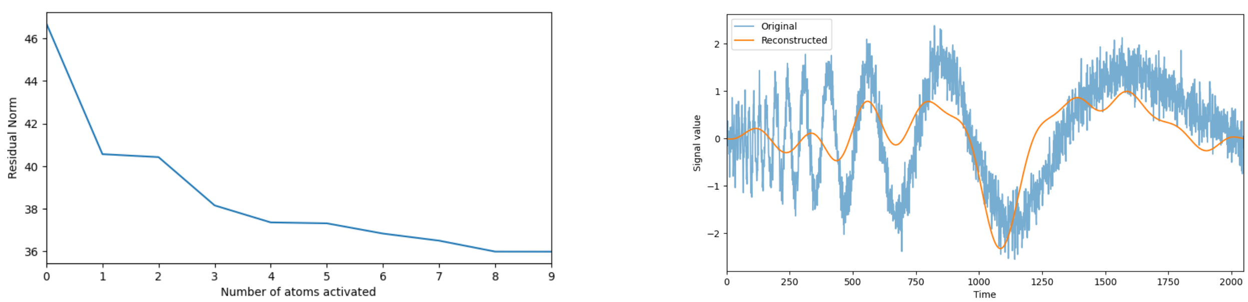
where $\mathbf{z}_{L,k} \in \mathbb{R}$ are activations signals, with $\mathcal{L} = \{32, 64, 128, 256, 512, 1024\}$. Thus, the size of the learnable activation vector $\mathbf{z}$ is $MDCT(\mathcal{L})$ with $MDCT(\mathcal{L}) = \sum_{L \in \mathcal{L}} \sum_{k=0}^{L-1} 1 = 1008$. Denote $K = MDCT(\mathcal{L})$ in what follows.

In the signal given in the notebook, we have a number of samples $N = 2048$ and we are asked to use OMP with a sparsity constraint of $K_0 = 10$.

For OMP algorithm we take inspiration from the pseudo-code in [1] to produce our own Algorithm 1. We implement this pseudo-code and obtain Figure 3. We test what the OMP algorithm outputs for 50 atoms and display it in 4.

Algorithm 1: OMP for MDCT

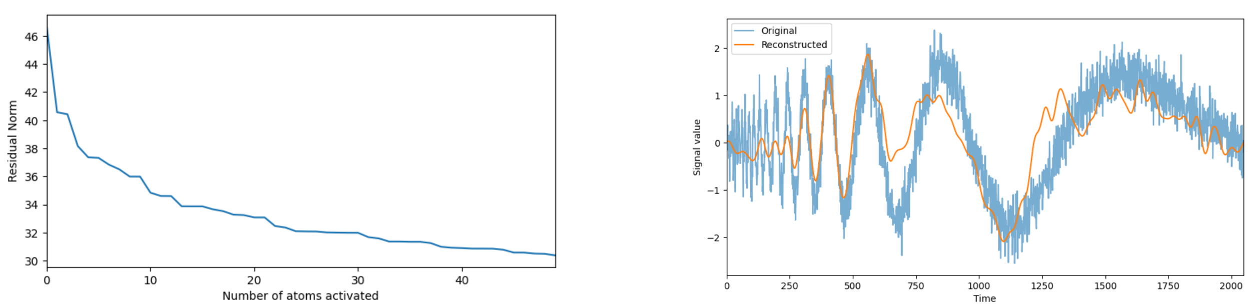
Data: Time series $\mathbf{x} \in \mathbb{R}^N$
Dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$, recall $K = MDCT(\mathcal{L})$
Number of non-null coefficients K_0
Initial support (index set) $\mathcal{S} = \emptyset$
Result: Activation vector $\mathbf{z} \in \mathbb{R}^K$ with K_0 non-null entries
Initialisation
$\mathbf{z} \leftarrow \mathbf{0}_K$; `/* Initial solution */`
$\mathbf{r} \leftarrow \mathbf{x}$; `/* Initial residual */`
$it \leftarrow 1$; `/* Iteration */`
while $it \leq K_0$ **do**
  $\forall j \in \{1, \ldots, K\}, t_j \leftarrow \dfrac{(\mathbf{D}_j^\top \mathbf{r}^{(it-1)})^2}{\|\mathbf{D}_j\|^2}$; `/* Compute correlations */`
  $idx \leftarrow \arg\max_{k \notin \mathcal{S}} t_k$; `/* Find an atom in` $\mathbf{D}$ `with maximum correlations */`
  $\mathcal{S} \leftarrow \mathcal{S} \cup idx$; `/* Support update */`
  $\mathbf{z}(\mathcal{S}) \leftarrow \mathbf{D}(\mathcal{S})^+ \mathbf{x} = \arg\min_{\mathbf{z}} \|\mathbf{x} - \mathbf{D}(\mathcal{S})\mathbf{z}\|_2^2$; `/* Sparse coefficient update */`
  $\mathbf{r}^{(it)} = \mathbf{x} - \mathbf{D}(\mathcal{S})\mathbf{z}(\mathcal{S})$; `/* Residual update */`
  $it \leftarrow it + 1$;
end



Norms of the successive residuals

Reconstruction with 10 atoms

Figure 3: Question 4



Norms of the successive residuals

Reconstruction with 50 atoms

Figure 4: Question 4

References

[1] Ilkay Oksuz. Chapter 5 - dictionary learning for medical image synthesis. In Ninon Burgos and David Svoboda, editors, *Biomedical Image Synthesis and Simulation*, The MICCAI Society book Series, pages 79–89. Academic Press, 2022.

[2] Richard A. Davis Peter J. Brockwell. *Introduction to Time Series and Forecasting, Third Edition*. Springer, 2016.