Imperial College London

COURSEWORK

IMPERIAL COLLEGE LONDON

DEPARTMENT OF BIOENGINEERING

Advanced Signal Processing

Author:
Sorenza Bastiaens (CID: 01112757)

Date: March 29, 2018

Contents

1 Random signals and stochastic processes

1.1 Statistical estimation

1. The expected value of X is defined as $\mathrm{E}\{X\}=\int_{-\infty}^{+\infty} x f(x,n)dx$ where f is the uniform distribution of X. In the case of a uniform distribution $U(0,1)$, we have:
   $\mathrm{f(x,n)}=\begin{cases} 1 & 0\leq x\leq 1 \\ 0 & \text{otherwise} \end{cases}$ and $\mathrm{E}\{X\} = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$. Therefore the theoretical expected value is 0.5.
   Since the values generated with MATLAB are each time random, the sample mean determined with the function `mean` varies every time. The simulation was run several times.

Simulation number	1	2	3	4	5	6	7	8	9	10
Sample mean	0.491	0.488	0.499	0.502	0.523	0.487	0.496	0.498	0.481	0.505

Table 1: Sample mean for each simulation number

The maximum deviation from the theoretical mean with those values is when the sample mean is equal to 0.523. The maximum relative error is:

$$\text{Max relative error} = \frac{measuredvalue-truevalue}{truevalue} = \frac{0.523-0.5}{0.5} \approx 0.046$$

This means that the relative error with the values computed does not exceed 4.6%. Another way to assess the accuracy is with the mean square error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x-m)^2 \tag{1}$$

where x is the value generated with the model, m the theoretical mean and N the number of samples. In this case the MSE is equal to 0.0863. The equation of the sample mean shows that the accuracy of the sample mean as an estimator increases as the sample size increases.

2. The theoretical standard deviation corresponds to the square root of the variance. The variance is equal to:
   $$\sigma^2 = E\{X-E\{X\}\}^2 = \int_{-\infty}^{+\infty}(x-E\{X\})^2 f(x,n),dx \tag{2}$$
   Thus, for the uniform distribution between 0 and 1 the equation is:
   $$\sigma^2 = \int_0^1 \left(x-\frac{1}{2}\right)^2 dx = \left[\frac{1}{3}x^3-\frac{1}{2}x^2+\frac{1}{4}\right]_0^1 = \frac{1}{12} \tag{3}$$
   This means that the standard deviation is equal to $\frac{1}{2\sqrt{3}}$ which is approximatively 0.289. Just like the sample mean, the sample standard deviation varies with each computation.

Simulation number	1	2	3	4	5	6	7	8	9	10
Standard deviation	0.287	0.294	0.287	0.294	0.290	0.291	0.287	0.284	0.288	0.286

Table 2: Standard deviation for each sample number

The maximum relative error is equal to: $\frac{0.294-0.289}{0.289} \approx 0.017$ which corresponds to 1.7%.

3. The following graph represent the sample mean and standard deviation of 10 realizations of X each of a 1000 samples.

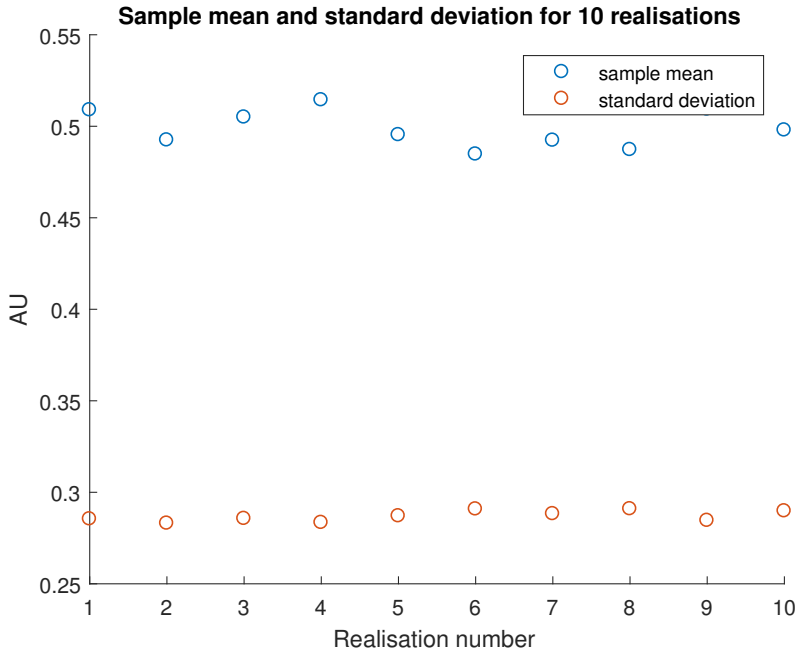


Figure 1: Plot of the sample mean and standard deviation for 10 realizations

We notice that both for the sample means and standard deviations they are close to the theoretical value. Since the bias is defined as the theoretical value minus the sampled one, the bias of those estimators is small. The maximum bias observed for sample mean is in the order of 0.01 and for standard deviation of 0.001. Indeed, the sample mean is an unbiased estimator:on average it approximates the real value. From the lecture notes *Introduction to Estimation Theory* (Page 15), we know that:

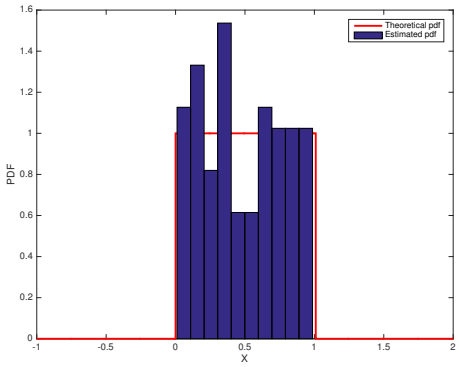
$$E\{\hat{m}\} = E\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\} = \frac{1}{N}\sum_{n=0}^{N-1} E\{x[n]\} = m \tag{4}$$

For the variance, we have:

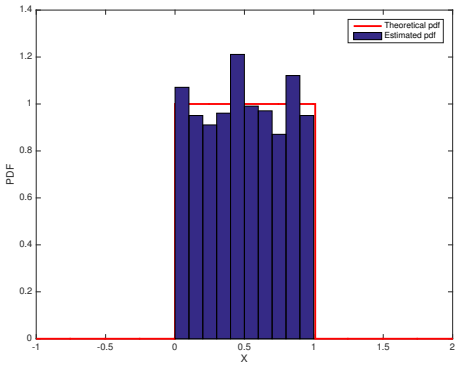
$$var\{\hat{m}\} = var\{\frac{1}{N}\sum_{n=0}^{N-1} x[n]\} = \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N} \tag{5}$$

We notice that when N tends to infinity, the variance approaches zero: it is an asymptotically unbiased estimator.

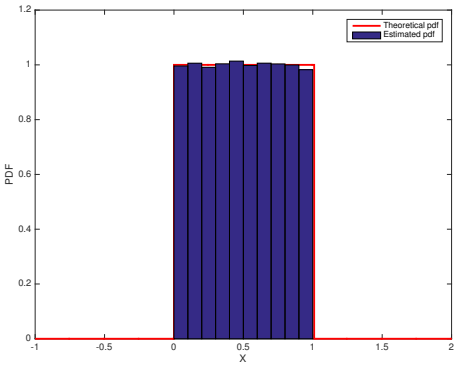
4. The probability density function (pdf) is approximated by computing the histogram and dividing the result by the total area of the histogram.
The theoretical pdf of the uniform distribution is simply a square wave ranging from 0 to 1 with amplitude 1.



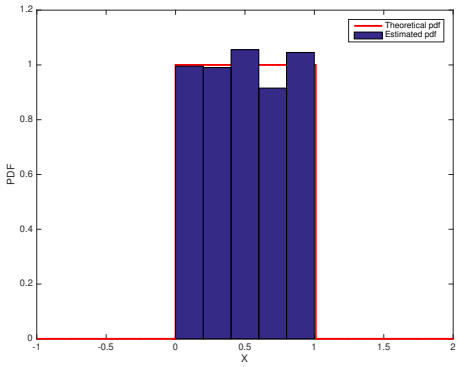
a) 100 samples and 10 bins



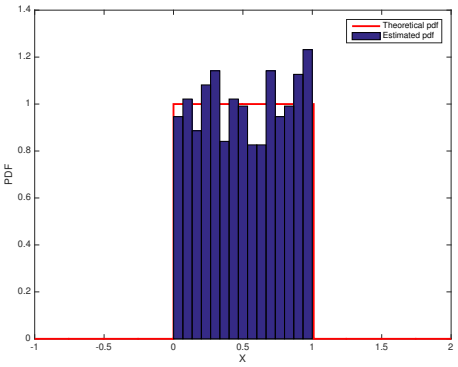
b) 10000 samples and 10 bins



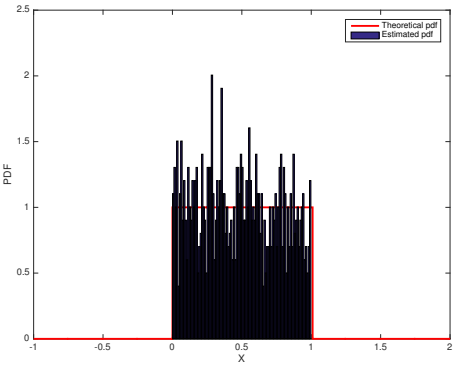
c) 100000 samples and 10 bins



a) 1000 samples and 5 bins



b) 1000 samples and 15 bins



c) 1000 samples and 100 bins

Figure 2: Theoretical and Estimated PDF with varying sample and bin numbers

From **Figure 1**, we notice that the number of sample in each realizations increases the estimated pdf approaches more the theoretical pdf as more data is accessible. Therefore, the estimated pdf converges as the number of samples generated increases.
In a uniform distribution the estimated pdf has to approximate a straight line, no curvature. Therefore the less number of bins the better to have the highest ratio of samples per bin an average over more samples.

5. In this case the theoretical mean is 0 and the standard deviation is 1 as the randn function normally distributes random variables. In this case, one of the sample mean computes is -0.0024. The MSE is equal to 0.9867. One of the sample standard deviation is 0.9722. The relative error is approximatively equal to 3%

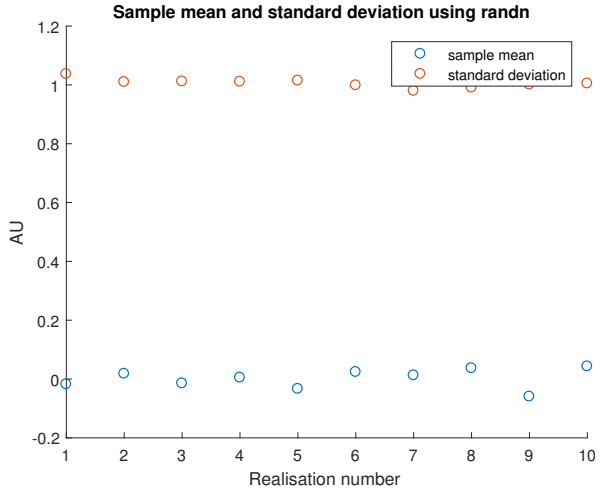
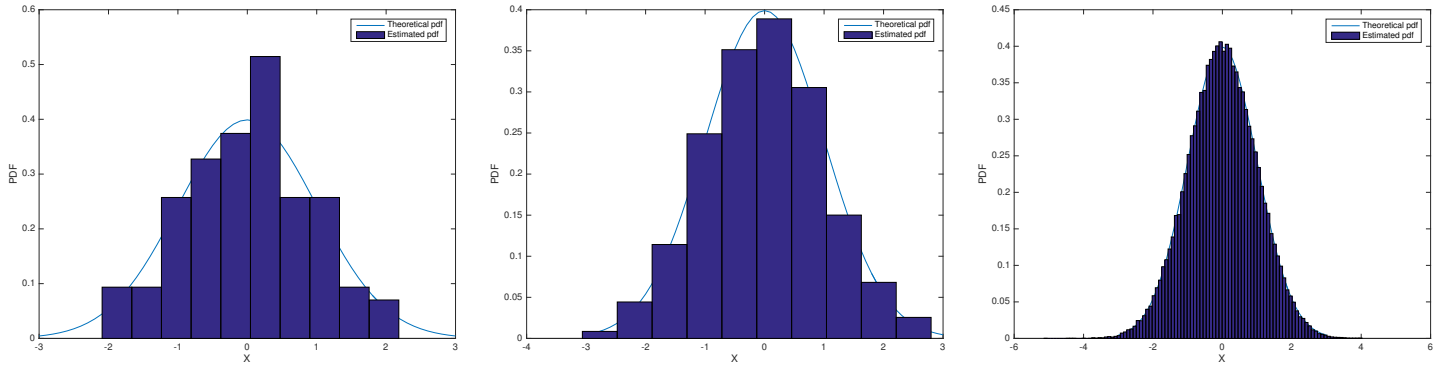
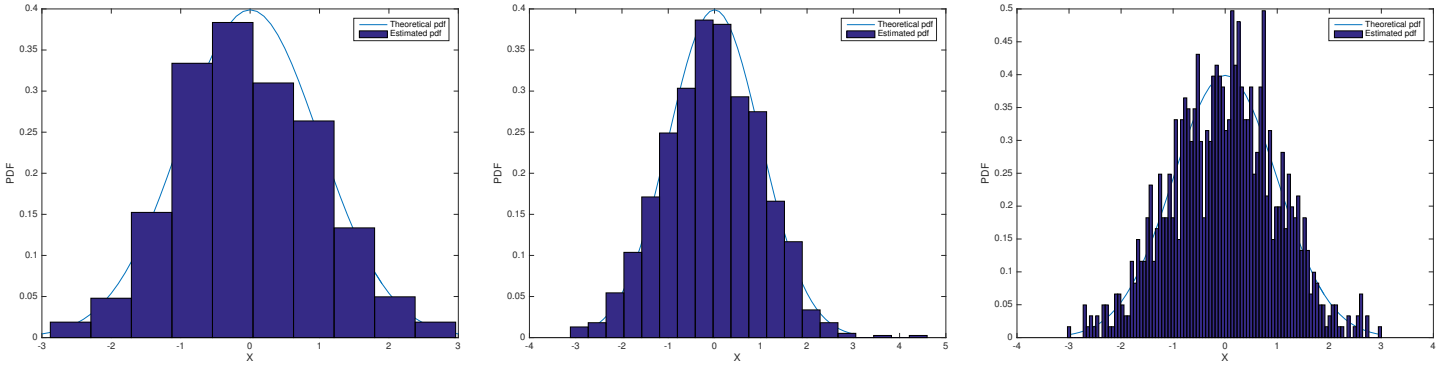


Figure 3: Plot of the sample mean and standard deviation of the `randn` function

Just as previously, since the sample mean and standard deviation are unbiased and asymptotically unbiased estimator respectively, they are in average close to the theoretical value.



a) Sample number=100 b) Sample number=1000 c) Sample number=10000



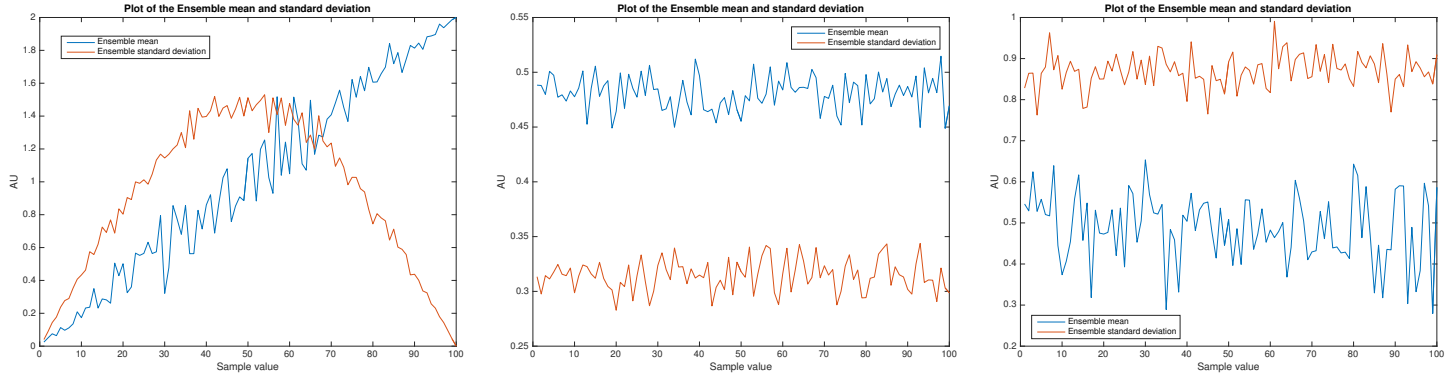
a) Bin number=10 b) Bin number=20 c) Bin number=100

Figure 4: Theoretical and Estimated PDF for `randn` with different bin numbers and samples (Top figures: a and b 10 bins and c 100 bins; Below figures: 1000 samples

Figure 2 permits to visualize the fact that increasing the number of samples enables a better pdf approximation. **Figure 3** shows that the number of bins needs to be adequately chosen. For example, 10 bins does not approximate enough the curve. However, with a 100 bins for 1000 samples, the estimated pdf seem noisy. A 100 bins is a good choice for 10000 samples as seen on **Figure 2 c)**. To approximate a Gaussian pdf with a histogram, there needs to be a sufficient number of bins to approximate the curvature. However a too high number of bins gives a spiked shaped curve as it does not average over enough samples. Therefore it is important to have an appropriate ratio of samples per bins.

1.2 Stochastic process

1. In this part, three different types of stochastic processes are analysed.



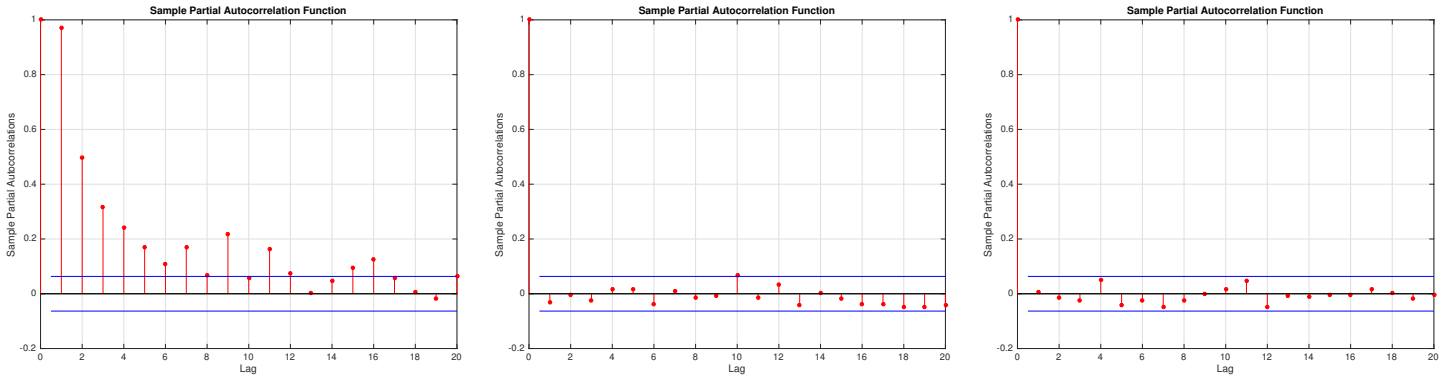
a) Stochastic process 1 b) Stochastic process 2 c) Stochastic process 3

Figure 5: Ensemble mean and standard deviation of the three different processes where M=100 and N=100

For practical applications, the wide-sense stationarity (WSS)condition is considered. The process is WSS when

- **Mean** : $E\{x[m]\} = E\{x[m+n]\}$
- **Covariance** and c(m,n) = c(m-n,0 = c(m-n)

By examining the ensemble mean on the first graph on **Figure 4**, the first process is not stationary as an increasing trend of the mean is observed over time. The other two processes seem stationary as the ensemble mean and variance are centered around a certain value. The third process has a wider range of ensemble mean values. It presents more variations than the second process. To determine if they are actually WSS, the partial autocorrelation can be examined:



a) Process 1 b) Process 2 c) Process 3

Figure 6: Partial autocorrelation of the three different processes

For process 1, the partial autocorrelation presents a slow decay. The process is non stationary which correlates with **Figure 5 a)**. The other two processes have a sharp decreases after the first sample. The values then stay in the confidence interval which means that they are both stationary.

2. A process is ergodic when the ensemble average corresponds to the time average. TIME averages and standard deviations where determined for 4 realizations.

Realization	1	2	3	4
Mean P1	10.0165	10.0173	9.9815	10.0554
Mean P2	0.0631	0.6227	0.3269	0.0579
Mean P3	0.4668	0.5004	0.4869	0.5274

Table 3: Table of the time averages of the three processes for 4 different realizations

Realization	1	2	3	4
Standard deviation P1	5.8363	5.8493	5.8238	5.8960
Standard deviation P2	0.0614	0.2981	0.0526	0.0956
Standard deviation P3	0.8689	0.8219	0.9499	0.8569

Table 4: Table of time standard deviations of the three processes for 4 realizations

For process 1 and 2 the ensemble mean seems to be around 0.5. The second process is not ergodic since some mean values deviate considerably from 0.5 (such as 0.0631). The third process is ergodic as all the mean values are close to the ensemble mean and only vary due to the introduction of some noise. The first process is not ergodic, since the ensemble mean varies over time and the time average is between 9 and 10. Furthermore, that value would vary with different sample numbers.

3. The following equations describe each processes mathematically:

- **Function process 1** :

$$P1[n] = 5sin\left(\frac{\pi n}{N}\right)x[n] + 0.02n \tag{6}$$

with $x[n]$ a realization of a uniform random variable $X \sim U$(-0.5,0.5)

- **Function process 2** :

$$P2[n] = x[n]u[n] + u[n] \tag{7}$$

with $x[n]$ a realization of a uniform random variable $X \sim U$(-0.5,0.5) and $u[n]$ of a uniform random variable $X \sim U$(0,1)

- **Function process 3** :

$$P3[n] = 3x[n] + 0.5 \tag{8}$$

with $x[n]$ a realization of a uniform random variable $X \sim U$(-0.5,0.5).

The expected values of the three processes are:

- **Mean P1** :

$$E\{P1\} = 5sin(\frac{\pi n}{N}) * Ex[n] + 0.02n = 5sin(\frac{\pi n}{N}) * 0 + 0.02n = 0.02n \tag{9}$$

since the mean of U(-0.5,0.5) is 0.

- **Mean P2** :

$$E\{P2\} = E\{x[n]u[n]\} + E\{u[n]\} = E\{(u[n] - 0.5)u[n]\} + E\{u[n]\} = E\{u[n]^2\} - E\{0.5u[n]\} + E\{u[n]\} \tag{10}$$

To determine what $E\{x[n]^2\}$ for $X \sim U$(a,b) we have:
$E\{X^2\} = \int_a^b x^2 f(x,n)dx$ where f(x,n)is the probability density function of a uniform distribution.
$E\{X^2\} = \int_a^b \frac{x^2}{b-a}dx = \frac{1}{3(b-a)}[x^3]_a^b = \frac{b^3-a^3}{3(b-a)}$
For $X \sim U$(0,1), $E\{X^2\} = \frac{1}{3}$
Therefore we have:

$$E\{P2\} = \frac{1}{3} - \frac{1}{2}\frac{1}{4} + \frac{1}{4} = \frac{7}{12} \approx 0.583 \tag{11}$$

- **Mean P3** :

$$E\{P3\} = 3E\{x[n]\} + 0.5 = 3 * 0 + 0.5 = 0.5 \tag{12}$$

The theoretical variance, which corresponds to $Var\{X\} = E\{$(X - $E\{X\})^2\} = E\{X^2\}$ - $(E\{X\})^2$, for each process are:

- **Variance P1** :

$$\begin{aligned} Var\{P1\} &= E\{(5sin(\frac{\pi n}{N})x[n] + 0.02n)^2\} - (0.02n)^2 \\ Var\{P1\} &= E\{(25sin^2(\frac{\pi n}{N})x^2[n]\} + E\{(0.02n)^2\} - (0.02n)^2 = 25sin^2(\frac{\pi n}{N})E\{x^2[n]\} \end{aligned} \tag{13}$$

From previous calculation we know that$E\{x^2[n]\}$ is equal to $\frac{1}{12}$. Thus we have:

$$Var\{P1\} = \frac{25}{12}sin^2(\frac{\pi n}{N}) \tag{14}$$

- **Variance P2** :

$$Var\{P2\} = E\{(x[n]u[n] + u[n])^2\} - (0.58)^2 = E\{x[n]^2u[n]^2 + u[n]^2\} - (0.58)^2 \tag{15}$$

If we replace $x[n]^2$ with (u[n] - 0.5)^2, the equation becomes:

$$Var\{P2\} = E\{u[n])^4 - u[n]^3 + 0.5u[n]^3 + u[n]^2\} - (0.58)^2 \tag{16}$$

We have $E\{u^4[n]\} = \int_0^1 x^4 dx = \frac{1}{5}$ $E\{u^3[n]\} = \int_0^1 x^3 dx = \frac{1}{4}$ and $E\{u^2[n]\} = \int_0^1 x^2 dx = \frac{1}{3}$
We now have:

$$Var\{P2\} = \frac{1}{5} - \frac{1}{4} + \frac{1}{4}\frac{1}{2} + \frac{1}{3} - (0.58)^2 \approx 0.07 \tag{17}$$

- **Variance P3** :

$$Var\{P3\} = E\{(3x[n] + 0.5)^2\} - (0.5)^2 = E\{9x[n]^2\} + E\{3x[n]^2\} + (0.5)^2 - (0.5)^2 = 9E\{x[n]^2\}. \tag{18}$$

Since the expected value of U(-0.5,0.5) is 0, we have in this case $Var\{X\} = E\{X^2\}$. The variance of a uniform distribution is $\frac{1}{12}$(upper limit - lower limit). Therefore the variance of process 3 is: $Var\{\text{P3}\} = \frac{9}{12} = \frac{3}{4}$

The ensemble average is computed by calculating the mean at a specific sample (time) across all the realizations (vertical). The time average is the mean value of one realization over all the samples (horizontal).
Compared to the theoretical values, only process 3 has time averaged and standard deviation that correspond which corresponds to the earlier results showing that it is in fact ergodic.
Process 2 has only one standard deviation close to the theoretical value and the mean does not always correspond to the theoretical value: the process is not ergodic.
Process 1 is not ergodic since the mean and the standard deviation will change with the number of samples.

1.3 Estimation of probability distribution

1. The code to estimate the pdf with the function `hist` can be found in the attached m-file with the different matlab codes under the function named `pdf_1`. The following plot is a test of the function:

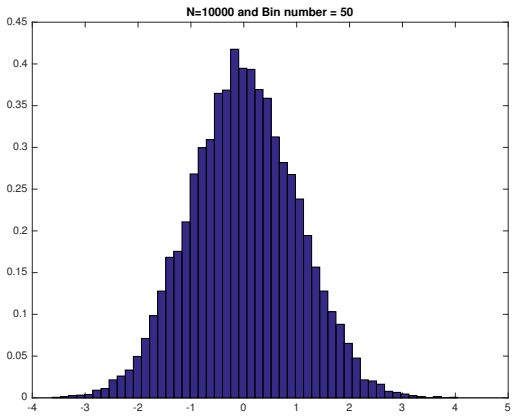


Figure 7: Plot of the estimated pdf of a stationary process with Gaussian pdf

The curve does follow a Gaussian pdf, which makes the code a valid estimation of the pdf.

2. The only process that is stationary and ergodic is process 3: P3[n] = 3x[n] + 0.5 with x[n] a realization of a uniform random variable X~U(-0.5,0.5). The multiplication by 3, makes the uniform distribution 3 times wider and the amplitude is then $\frac{1}{3}$. The addition by 0.5 shifts the distribution to the right by that amount. Therefore, the theoretical pdf is a uniform distribution of amplitude $\frac{1}{3}$ ranging from -1 to 2.

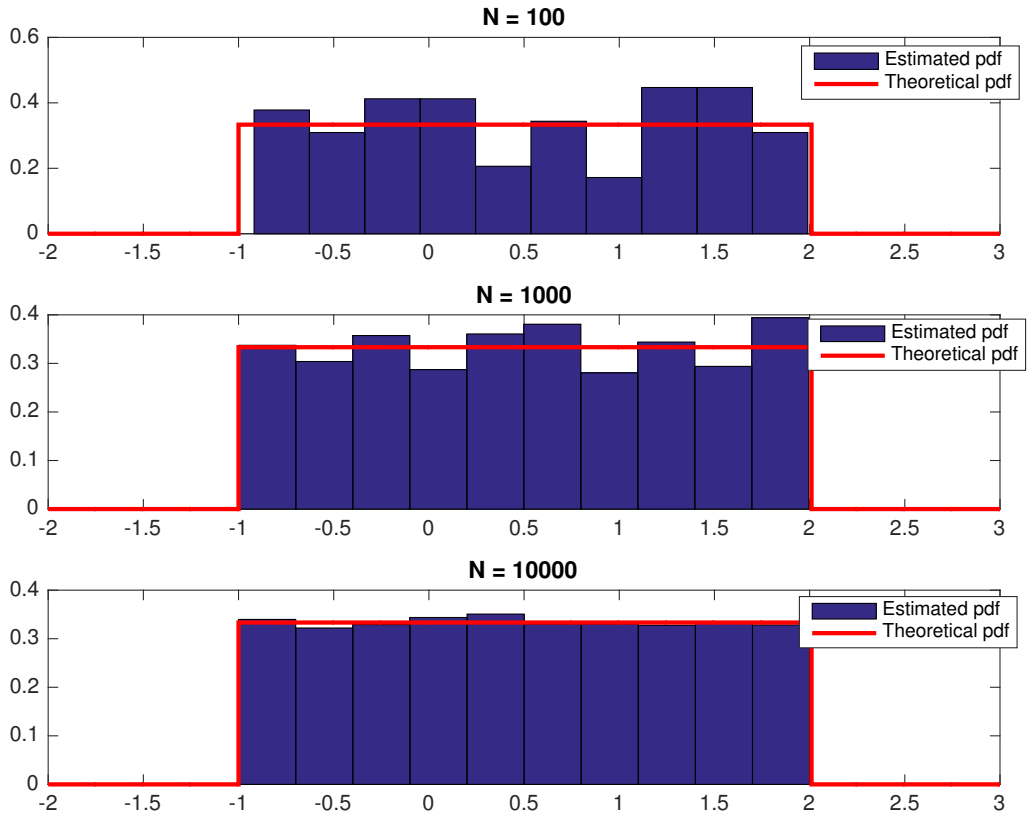


Figure 8: Plot of estimated densities and theoretical pdf for different sample values

As data length increases, the estimated pdf converges towards the theoretical pdf.

3. It is not possible to estimate the pdf of a non stationary process with the function pdf computed on Matlab because in a non stationary process the mean value changes over time and in the estimated pdf it tries to center around the pdf around a mean.
When the mean changes from 0 to 1 after sample point N = 500 for a 1000-sample-long signal, the problem that arises is that instead of presenting two peaks it will only present on peak at 0.5 which does not correspond to the theoretical non stationary pdf. In order to compute it, need to do short sliding window data: compute the pdf from 0 to 500 samples and then from 500 to 1000 and add them together.

2 Linear stochastic modelling

2.1 ACF of uncorrelated and correlated sequences

The unbiased estimate of the autocorrelation function is given by:

$$\hat{R}_X(\tau) = \frac{1}{N-|\tau|} \sum_{n=0}^{N-|\tau|-1} x[n]x[n+\tau], \qquad \tau = -N+1,...,N-1 \tag{19}$$

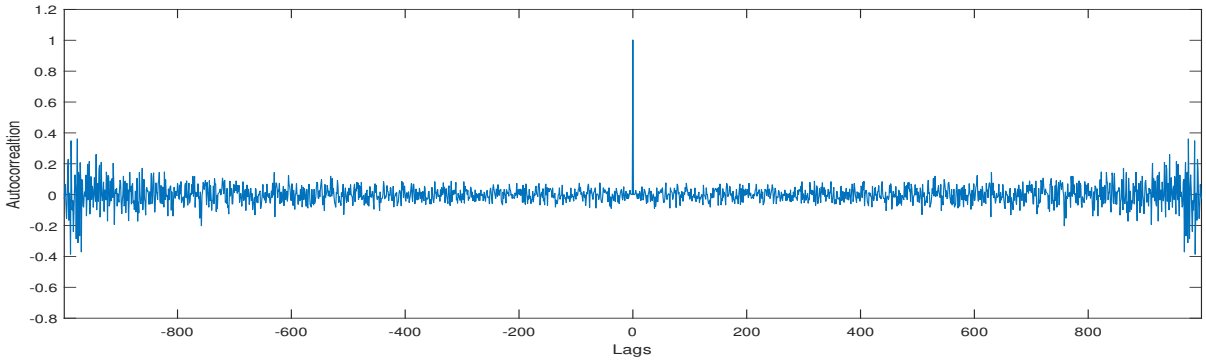


Figure 9: Autocorrelation of WGN

1. The maximum autocorrelation is observed at 0 and is equal to 1. The signal is most self-similar when there is no lag. This is expected since the signal is made of white gaussian noise and the samples are uncorrelated. Therefore, for τ different than 0, the autocorrelation should approximate 0. Indeed, if we had an infinite number of samples, the autocorrelation function would correspond to the discrete Dirac function. This is shown mathematically as the autocorrelation function is also defined as:

$$R_X(\tau) = E\{x[n]\}E\{x[n+\tau]\} \tag{20}$$

Since the mean is zero, for τ different then zero it should theoretically always be 0. If τ equals 0 the equation becomes:

$$R_X(0) = E\{x^2[n]\} = \sigma^2 \tag{21}$$

The variance is equal to 1 for white gaussian noise. Combining the two equations, the result is a Dirac function. Since a finite number of values are used, variations are observed. They increase as $|\tau|$ increases.
The symmetry of the autocorrelation function is explained by the property of the autocorrelation being an even function. If p = n $+\tau$, we have:

$$R_X(\tau) = E\{x[n]\}E\{x[n+\tau]\} = E\{x[p-\tau]\}E\{x[p]\} = E\{x[n-\tau]\}E\{x[n]\} = R_X(-\tau) \tag{22}$$

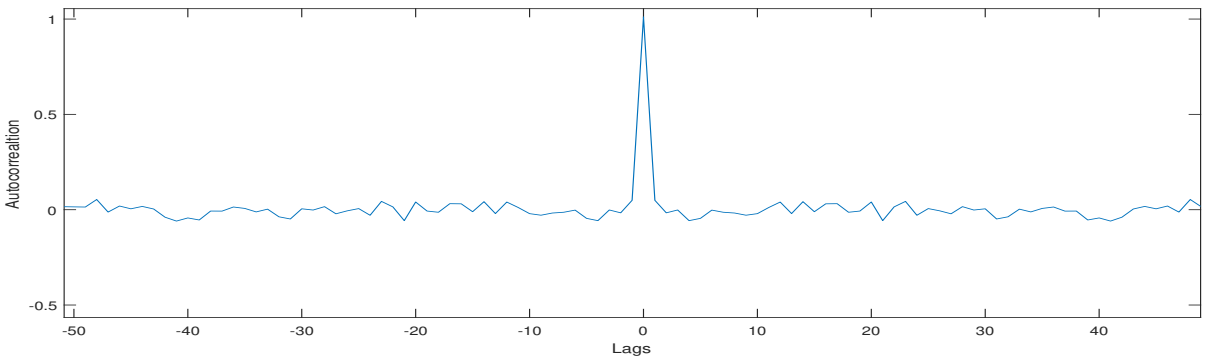
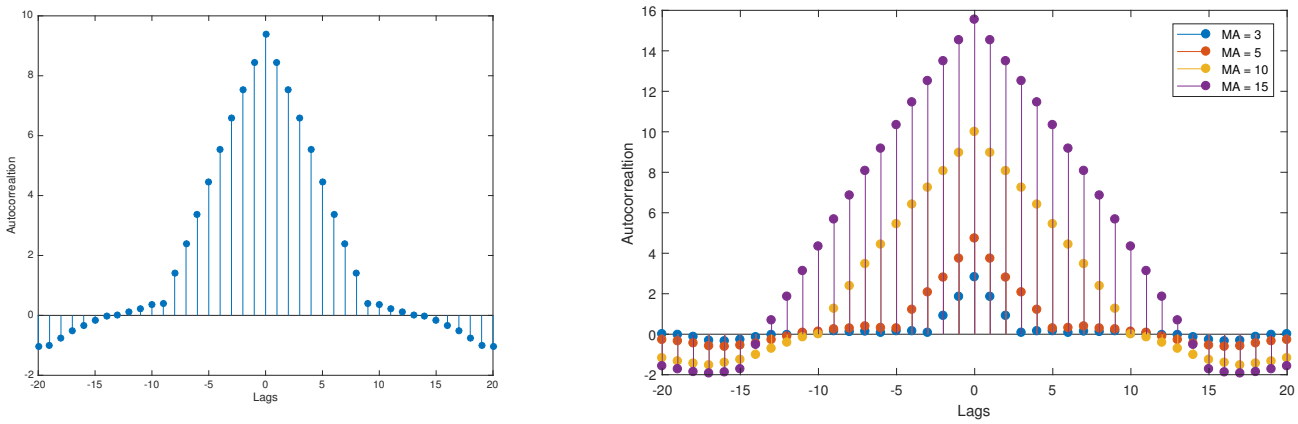


Figure 10: Zoom around 0 lag of the autocorrelation

2. For values ofτ close to 0, the autocorrelation oscillates around 0: the variations are small. For large values of τ, as we can see in Figure 7, it the values still oscillate around 0 but with a greater variability. The maximum autocorrelation observed is close to 0.4 (0.413 in this case).

3. This effect can be explained with the equation of the unbiased estimate of the autocorrelation (1). As τ increases, the value $N-|\tau|$ decreases and $\hat{R}_X(\tau)$ is then averaged over less values. More error is

therefore introduced, and consequently more variability is observed in the autocorrelation. For large values of τ the autocorrelation estimate function estimates are statistically unreliable. The bound of $|\tau|$ depends on the number of sample values: the more sample you have, the greater the bound for $|\tau|$ as it will become unreliable at a further lag value. An empirical bound that could be suggested is $|\tau| < N \times \frac{2}{5}$.



(a) Coefficient of MA equal to 9

(b) ACF with different MA coefficients

Figure 11: ACF of WGN filtered with moving average filter of different coefficients

4. From Figure 9a, we can see that the autocorrelation of the filtered WGN corresponds to a triangular pulse of length 18 centered around 0. This indicates that the filtering introduces correlation between successive samples in the range of the order of the MA filter ($|\tau| < 9$). As the filter order increases, the more correlation will be introduced. Furthermore, the shape is similar to the ACF of the impulse response of the filter.
   When τ is equal to 0, the unbiased estimate autocorrelation function correspond to:

$$\hat{R}_X(0) = \frac{1}{N}\sum_{n=0}^{N-1}(x[n])^2 \tag{23}$$

   This is equivalent to the sum of the mean squared. Therefore from $\hat{R}_X(0)$ the local mean can be estimated.

5. From equation (2) and (3), we know that $R_X(\tau)$ is 0 expect for $R_X(0)$, when it is equal to σ^2 as the input is uncorrelated. Thus, $R_Y(\tau)$ is simply equal to $R_h(\tau)$ times the variance. In our case that variance is equal to 1, and therefore we have:

$$R_Y(\tau) = R_h(\tau) \tag{24}$$

   The autocorrelation function of the filtered version of the process represents the autocorrelation function of the impulse response of the filter times the white noise variance.

2.2 Cross-correlation function

1. An unbiased estimate of the cross correlation function is:

$$\hat{R}_{XY}(\tau) = \frac{1}{N-|\tau|}\sum_{n=0}^{N-|\tau|-1} x[n]y[n+\tau], \qquad \tau = -N+1,\ldots,N-1 \tag{25}$$

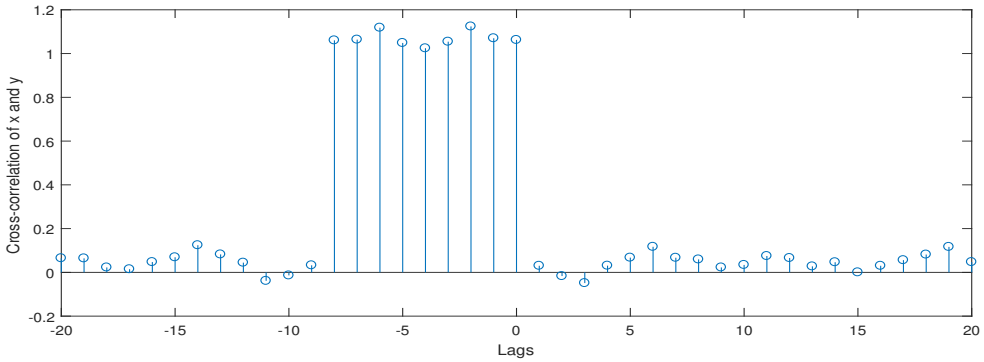


Figure 12: Cross-correlation function of WGN and filtered WGN

The shape of the cross-correlation is a rectangular pulse which corresponds to the impulse response of an moving average filter.
With the same reasoning as $R_Y(\tau)$ previously, for an uncorrelated stochastic process, $R_{XY}(\tau)$ will be equal to $h(\tau)$ times the variance of the signal. The variance is equal to 1, and therefore we have:

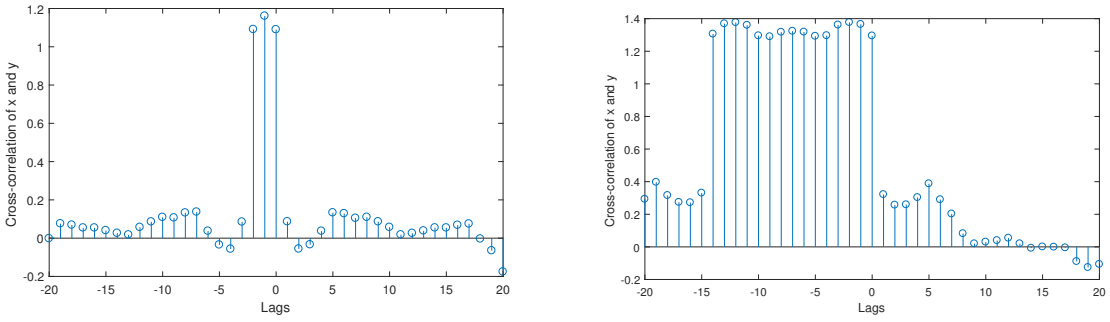
$$R_{XY}(\tau) = h(\tau) \tag{26}$$

This correlates with the result of **Figure 10**.

2. If the input is white gaussian noise, the impulse response can be estimated from the cross-correlation of the input and the filtered input, enabling the identification of the system. Indeed, if we have a white noise zero mean variance as input, then we have:

$$h(\tau) = \frac{R_{XY}(\tau)}{\sigma_x^2} \tag{27}$$

where $h\tau$) is the impulse response, $R_{XY}(\tau)$ the cross correlation between input and filtered input.
The impulse response of an MA filter is a rectangular pulse which is what is seen in Figure 11. The order of the filter determines the width of the rectangle and thus enables to determine the coefficient of the filter from the width of the impulse response..



(a) MA coefficient = 3 **(b)** MA coefficient = 15

Figure 13: CCF of WGN and filtered WGN with different filter coefficient for the MA filter

2.3 Autoregressive modelling

1. The aim is to determine for what values of a1 and a2 the AR(2) process is stable, thus when *x[n]* converges. To evaluate this, the function `isstable` was used.

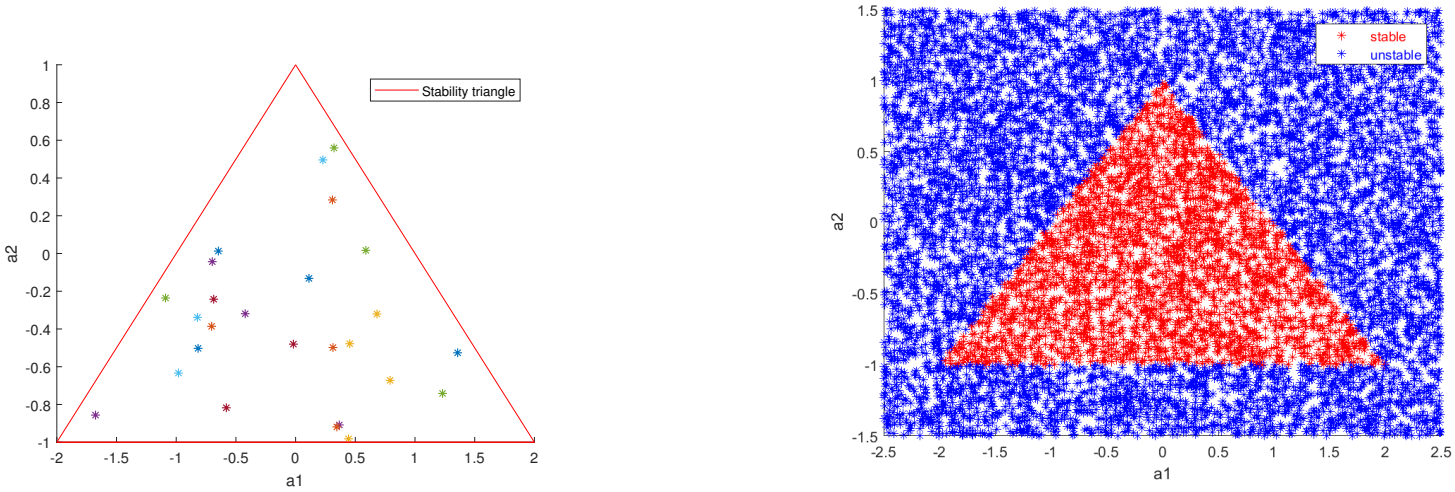


Figure 14: Plot of the pairs of a1 and a2 for which the process is stable (left figure) or stable and unstable (right figure)

The values of the pairs (a1,a2) which preserve the stability of this process seem to be constrained in the triangle as shown in **Figure 12(a)**. This corresponds to the stability triangle. If more values of a1 and a2 are computed, this triangle is clearly visible as it can be seen in **Figure 12(b)**.
The conditions for an AR(2) process to be stable are:

$$\begin{cases} a_1 + a_2 < 1 \\ a_2 - a_1 < 1 \\ -1 < a_2 < 1 \end{cases} \tag{28}$$

These conditions are determined by the correlation coefficient $\rho(1)$ and the variance of an AR(2) process. Indeed, from the Yule-equations for p=2, we know that:

$$\rho(1) = \frac{a_1}{1 - a_2} \tag{29}$$

Since $\rho(1) < \rho(0) = 1$ (as $\rho(0)$ is the normalized Autocorrelation function at 0) we have the following condition:

$$\frac{a_1}{1 - a_2} < 1 \Leftrightarrow a_1 + a_2 < 1 \tag{30}$$

which corresponds to condition 1.
Furthermore, the variance, which is derived from the general AR(2) process equation, is equal to:

$$\sigma_x^2 = \left(\frac{1 - a_2}{1 + a_2}\right)\frac{\sigma_w^2}{(1 - a_2)^2 - a_1^2} = \frac{(1 - a_2)\sigma_w^2}{(1 + a_2)(1 - a_2 - a_1)(1 - a_2 + a_1)} \tag{31}$$

The variance is always positive. Therefore, the numerator and the denominator need to be of the same sign.

$$\text{If} \quad 1 - a_2 > 0 \quad \text{then} \quad (1 + a_2)(1 - a_2 - a_1)(1 - a_2 + a_1) > 0 \tag{32}$$

This happens if in the denominator either all the factors are positive or 2 of them are negative and one positive. Due to the fact that $1 > a_2$ and condition 1, the only possibility for the denominator to be positive is if all the factors are positive. Therefore we have:

$$\begin{aligned} 1 + a_2 > 0 &\Leftrightarrow -1 < a_2 \\ 1 - a_2 + a_1 > 0 &\Leftrightarrow 1 > a_2 - a_1 \\ 1 - a_2 - a_1 > 0 &\Leftrightarrow 1 > a_2 + a_1 \end{aligned} \tag{33}$$

which correspond to our three conditions. If $1 - a_2 < 0$, there is no possible solution for the denominator to be negative and therefore a_2 is always smaller than 1 and also greater than -1.
These three conditions explain the triangular shape of the convergence region of an AR(2) process.

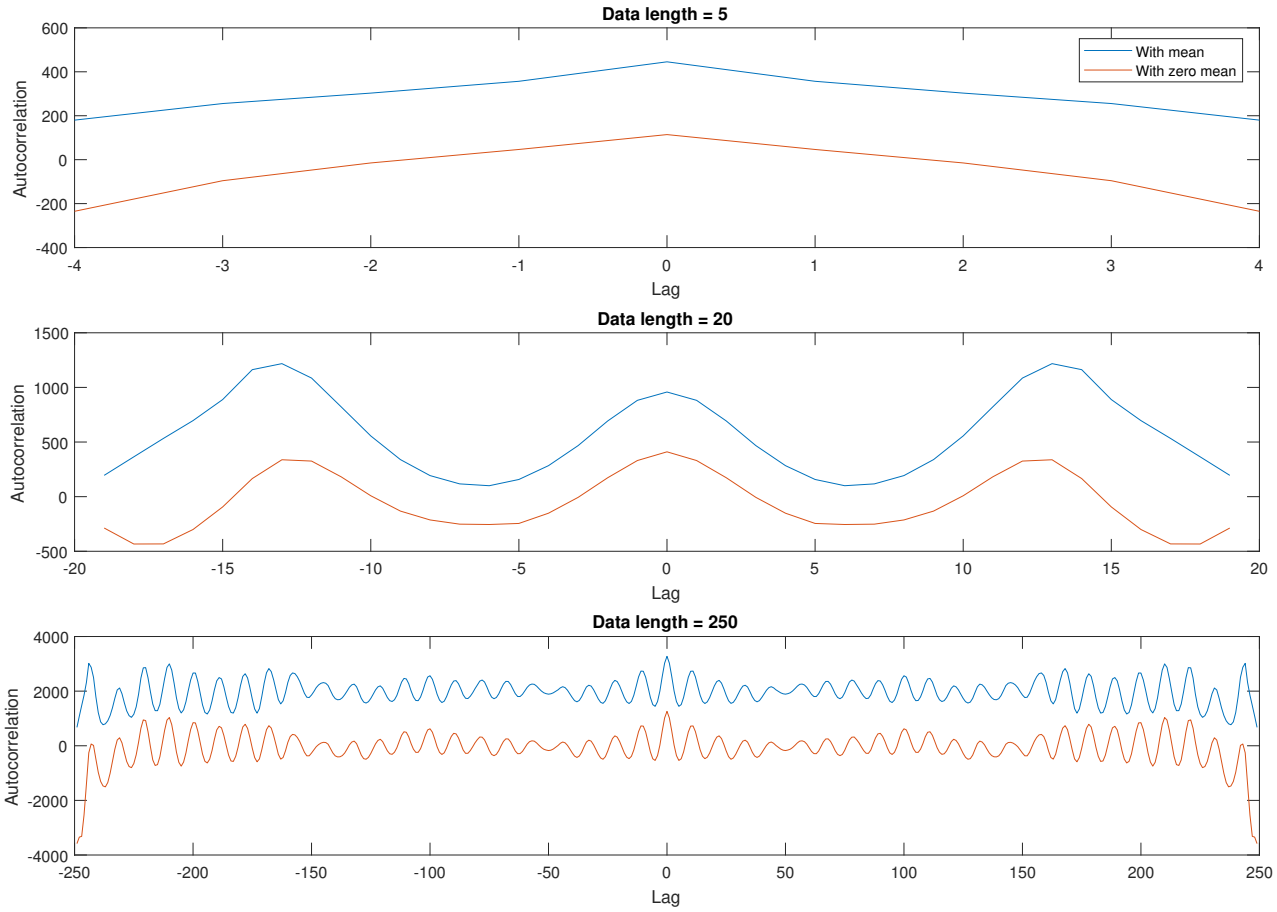


Figure 15: Plot of the Autocorrelation functions of the Sunspot time series with mean and with zero mean for different data lengths

2. From Figure 14, it is deduced that a certain number of data is necessary to gain sufficient information about the autocorrelation. Indeed, with a data length of 5, no conclusion about the autocorrelation can be taken. For a data length of 20, even though the number of data is still insufficient, an oscillatory behaviour seems to take place. No other information can be deduced from this autocorrelation. For data length 250, the autocorrelation is more accurate and for values of lag between 0 and 50 a decaying sinusoidal is observed suggesting an AR process. The ACF is periodic which means that the data is periodic and has a cyclic component. For larger values of lag, more variability is observed and the autocorrelation values cannot be trusted. This is expected due to the unbiased estimate of the autocorrelation that is averaged over less values as the lag increases as seen previously (2.1.1). The mean of the series is introducing a bias and corresponds to a DC offset. This is a deterministic component that should be removed in order to analyze only the stochastic component of the series. This is achieved by bringing the mean of the data series to 0. By comparing the autocorrelation with and without the offset, we notice that they are in fact of similar shape and only differ by an offset that increases with the sample number. The zero mean series is always centered around 0.

3. The coefficients of sunspot time series for different model order are estimated with the function

`aryule` which uses the Yule-Walker method. For model order 2, the coefficients are given on MATLAB as the following: a = [1 -a1 -a2];

p	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1	0.9295									
2	1.4740	-0.5857								
3	1.5492	-0.7750	0.1284							
4	1.5167	-0.5788	-0.2638	0.2532						
5	1.4773	-0.5377	-0.1739	0.0174	0.1555					
6	1.4373	-0.5422	-0.1291	0.1558	-0.2248	0.2574				
7	1.3669	-0.4807	-0.1718	0.1912	-0.0764	-0.1359	0.2736			
8	1.3016	-0.4483	-0.1535	0.1456	-0.0355	-0.0212	-0.0523	0.2384		
9	1.2615	-0.4395	-0.1500	0.1515	-0.0599	0.0046	0.0230	0.0197	0.1680	
10	1.2573	-0.4400	-0.1506	0.1514	-0.0584	0.0007	0.0268	0.0308	0.1362	0.0252

Table 5: Table of coefficients of the sunspot time series for different model order

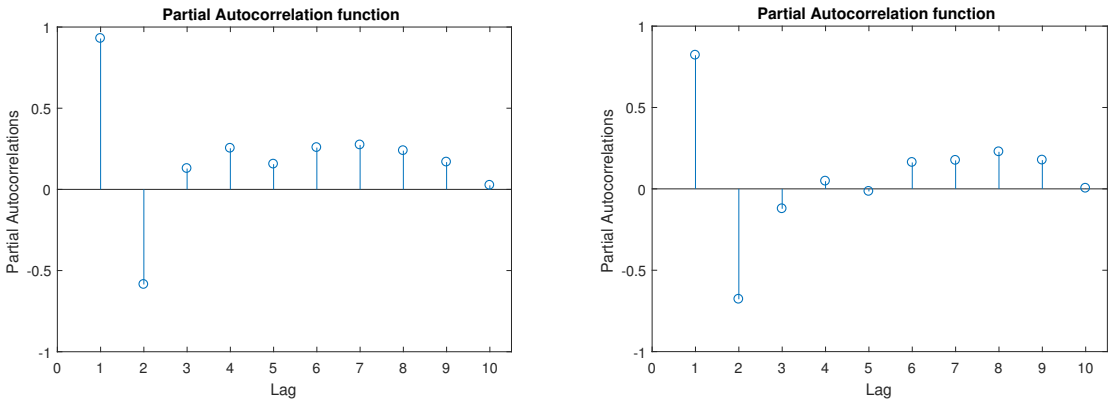
From the coefficients obtained with the Yule-Walker equations, the most likely model order for the sunspot time series is model order p=2. Indeed after model order 2, the last coefficient of each model is significantly smaller which is an indicator that these coefficients are unnecessary.

p	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1	0.8212									
2	1.3783	-0.6783								
3	1.2953	-0.50970	-0.1223							
4	1.3011	-0.4856	-0.1836	0.0473						
5	1.3018	-0.4885	-0.1912	0.0676	-0.0156					
6	1.3044	-0.4994	-0.1602	0.1469	-0.2269	0.1623				
7	1.2760	-0.4597	-0.1859	0.1749	-0.1394	-0.0661	0.1751			
8	1.2361	-0.4447	-0.1541	0.1351	-0.0971	0.0385	-0.1154	0.2276		
9	1.1959	-0.4243	-0.1609	0.1523	-0.1210	0.0658	-0.0368	0.0093	0.1766	
10	1.1952	-0.4243	-0.1608	0.1520	-0.1205	0.0652	-0.0362	0.0109	0.1721	0.0038

Table 6: Table of coefficients of the standardised sunspot time series for different model order

Compared to the non standardised sunspot time series, the coefficients of the standardised series seem to have a greater difference between the last coefficient of model order 2 and 3. The choice of the model order is clearer and more easily defined. This is visible in **Figure 16** on the next page with the reflection coefficients. They correspond to the partial autocorrelation coefficients scaled by –1. In the case of the standardised series, after model order 2, the coefficients are closer to 0 and more inside the confidence interval than the original series. This supports the fact that the sunspot time series should be modelled with an AR(2) process

4. The optimal model order can be assessed with different criterion which take into account the computational complexity and model accuracy.
The model order at which the minimum value of the minimum description length criterion (MDL) and the Akaike information criterion (AIC) is found correspond to the optimal model order. Both equations have a term called the loss function(E_p). It is typically the cumulative squared error. Here, it is the estimated variance of the white noise input. Those criterions are presented in **Figure 17** on the next page.



(a) Sunspot time series **(b)** Standardised Sunspot time series

Figure 16: Partial Autocorrelation function of the standardised and non standardised sunspot time series using the inverse of the reflection coefficients

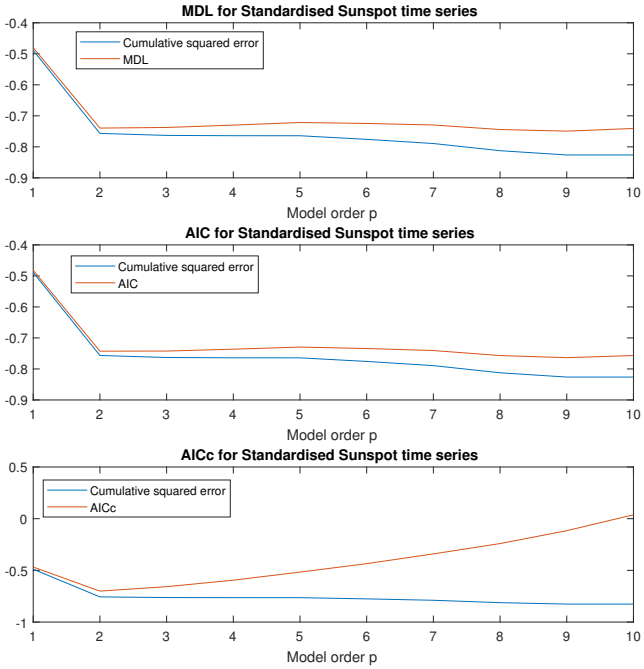


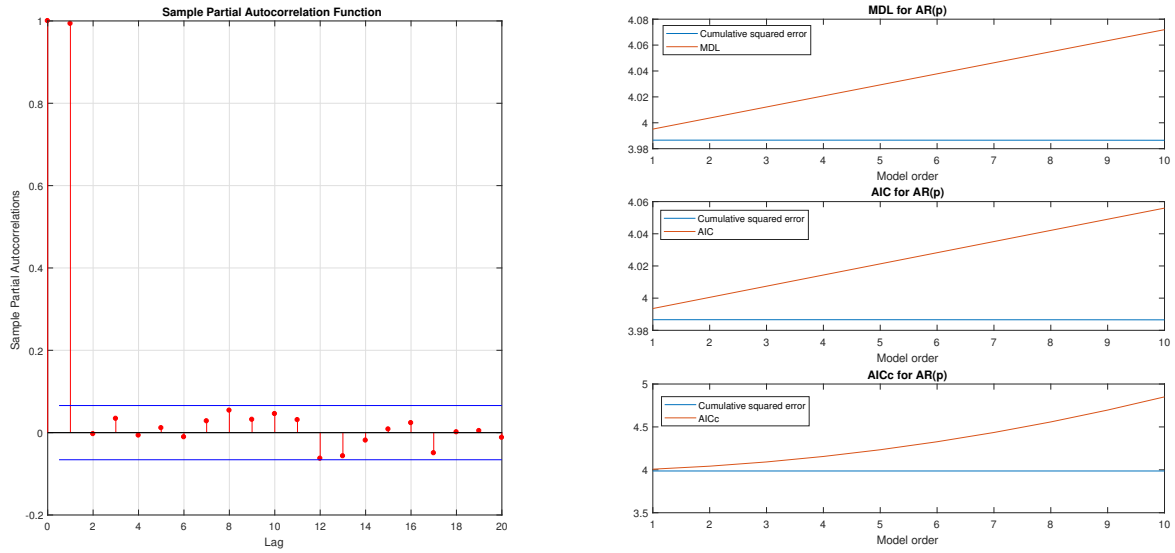
Figure 17: Plot of the three criterions with the cumulative squared error

5. The AR model are used to predict the outcome of the sunspot time series m steps ahead. The AR model were tested to predict the outcome 1,2,5 and 10 steps ahead (the code used to implement those graphs can be found in the MATLAB code).
   It was noticed that as the value of m increases, the less accurate the signal is. Indeed, if try to predict 1 step ahead, the shape corresponds to the original data only delayed by one. Furthermore, the sunspot time series, presents a cyclical data, which means that greater prediction horizons will yield predicted values that correspond less to the actual value.
   Comparing the results of different AR model, which where 1,2 and 10 it was noticed that the higher order model than 1 were more accurate. This is expected as the model order defines the number of previous data values it remembers. The higher the model order, the more data values it keeps in memory. However, not much differences were observed between the predicted values of AR(2) and AR(10) due to the fact that the coefficients were considerably smaller after a2. In addition, a higher

order model implies more computational and memory complexity.
If the model is undermodelled, it will predict poorly as it will miss important features of the model. However, overmodelling may remove some of the residual variation and will then also be a less good predictor than a lower model order.

2.4 Cramer-Rao Lower Bound

1. a) To determine if an AR(1) model is sufficient to describe the daily returns of the NASDAQ financial Index, the partial autocorrelation function and the three different criterions were found.



(a) Partial Autocorrelation **(b)** MDL, AIC and AICc criterion

Figure 18: Elements to determine optimal model order for the daily returns of the NASDAQ financial Index

With the partial autocorrelation function, after model order 1, the coefficients are very small and are in the confidence interval region. For all three criterions, the minimum value is found at model order 1. The two methods correlate which suggests that a AR(1) is sufficient to model the daily returns of the index.

b) The Fisher information matrix $[\mathbf{I}(\theta)]_{ij}$ is essential to compute the Cramer-Rao Lower Bound (CRLB) which determines the lower bound on the variance of any unbiased estimator. Indeed the CRLB is of the form:

$$var(\hat{\theta}_i) \geqslant [I^{-1}(\theta)]_{ii}, \quad [I(\theta)]_{ij}] = -E\frac{\partial^2 p(x;\theta)}{\partial\theta_i \partial\theta_j} \tag{34}$$

In order to calculate $[\mathbf{I}(\theta)]_{22}$, we determine the derivative of $\ln[\hat{P}_X(f;\theta)]$ over θ_2 which corresponds to σ_2. Therefore, we have:

$$\frac{\partial ln[\hat{P}_X(f;\theta)]}{\partial\sigma^2} = \frac{\partial(ln[\sigma^2] - ln[1-\sum_{m=1}^{p} a_m e^{-j2\pi f m}] - ln[1-\sum_{m=1}^{p} a_m e^{j2\pi f m}])}{\partial\sigma^2} = \frac{1}{\sigma^2} \tag{35}$$

We have now:

$$[I(\theta)]_{22} = \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial ln[\hat{P}_X(f;\theta)]}{\partial\sigma^2}\frac{\partial ln[\hat{P}_X(f;\theta)]}{\partial\sigma^2}df = \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{\partial ln[\hat{P}_X(f;\theta)]}{\partial\sigma^2}\right)^2 df = \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\frac{1}{\sigma^2}\right)^2 df$$
$$[I(\theta)]_{22} = \frac{N}{2\sigma^4} \tag{36}$$

Therefore, the Fisher matrix information is:

$$I(\theta) = \begin{bmatrix} \frac{Nr_{xx}(0)}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix} \tag{37}$$

c) From equation 15, the equations we have are:

$$\begin{aligned} var(\hat{\theta_2}) \geqslant [I^{-1}(\theta)]_{22} \Leftrightarrow var(\hat{\sigma}^2) \geqslant \frac{2\sigma^4}{N} \\ var(\hat{\theta_1}) \geqslant [I^{-1}(\theta)]_{11} \Leftrightarrow var(\hat{a1}) \geqslant \frac{\sigma^2}{Nr_{xx}(0)} \end{aligned} \tag{38}$$

For an AR(1) process, the autocorrelation function at zero lag is equal to the variance of the signal.

$$r_{xx}(0) = \sigma_x^2 = \frac{\sigma_w^2}{1-\rho_1 a_1} \tag{39}$$

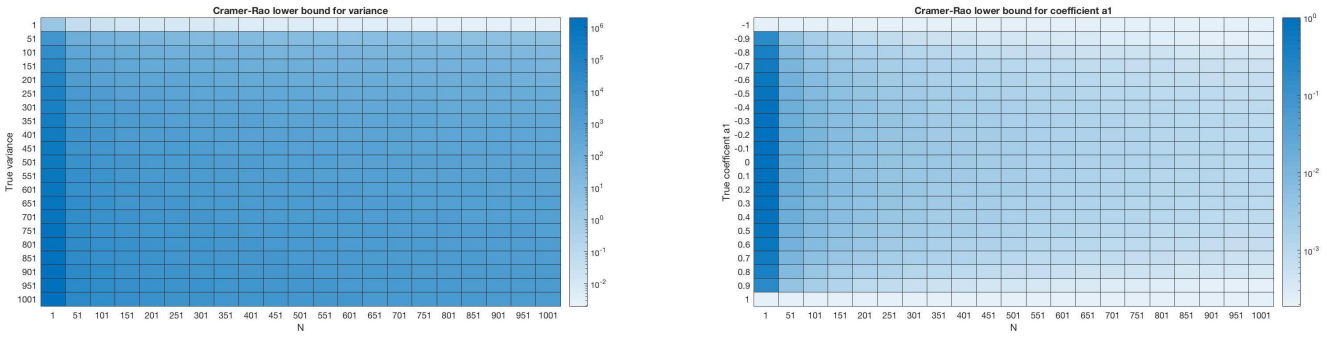
where σ_w^2 is the variance of the driving noise (σ^2) The correlation coefficient ρ_1 is equal to a_1 times ρ_0. The last term is also equal to 1 as it is the normalized autocorrelation at 0 lag. If ρ_1 is replace in equation (20) we have:

$$r_{xx}(0) = \frac{\sigma_w^2}{1-a_1^2} \tag{40}$$

Implementing this result in equation (19) for $var(\hat{a_1})$ gives:

$$var(\hat{a_1}) \geqslant \frac{\sigma^2}{N}\frac{1-a_1^2}{\sigma^2} = \frac{1}{N}(1-a_1^2) \tag{41}$$

i)



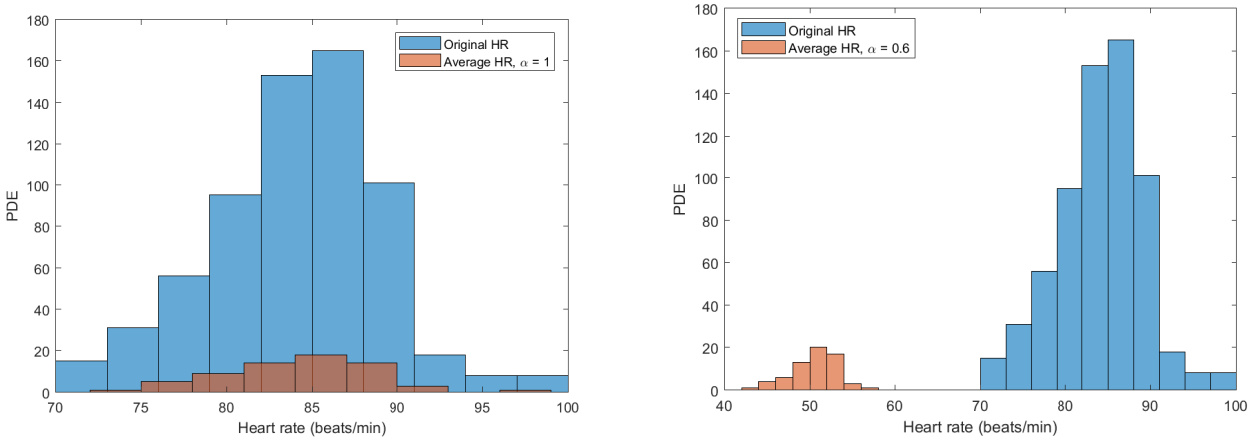
(a) CRLB of $\hat{\sigma}^2$ **(b)** CRLB of $\hat{a}_1$

Figure 19: Plot of the CRLB for the variance and coeffficient a1 against the number of data points.

Note: The plots in **Figure 19** have a log scale for the CRLB value.
We notice that the CRLB in both cases is high for small values of N and decreases as N increases. In the case of CRLB a1, as N increases, the CRLB decreases when a1 diverges from 0. For CRLB of the variance, the minimum values are observed for high variance and high N.

2.5 Real world signals: ECG from iAmp experiment

a) The following plot corresponds to the PSD of the original and the averaged heart rates.



(a) $\alpha = 1$ **(b)** $\alpha = 0.6$

Figure 20: Probability density function of the original and the averaged heart rates for 2 different α values

b) The amplitude of the PDE's of the average heart rates are smaller than the original heart rate. This is due to the fact that each heart rate are equal to the average of 10 heart rates. Therefore, since there are less heart rates in total the overall frequency will be lower.
Changing the value of α shifts the mean of the heart rate. Indeed, the α acts as a weighting and each heart rate are weighted by the same value. If α is less than 1 the mean is shifted to the left and if α is greater than 1 it will be shifted to the right. The constant does not affect the amplitude of the PDE.

c) The Autocorrelation function of the RRI trials is **Figure 21**. For a Moving Average process (MA), the autocorrelation presents a triangular shape extending over twice the number of coefficients and is then approximatively equal to 1. In the case of a Autoregressive process (AR), the autocorrelation function is a relatively slow decay. From the plot of the autocorrelation of the three RRI trials, in general, a pseudo-periodic decaying sinusoidal is observed. In an MA process, a triangular shape is observed and the ACF function approximates 0 over time. Therefore, we can infer that the RRI data is more likely an AR process.

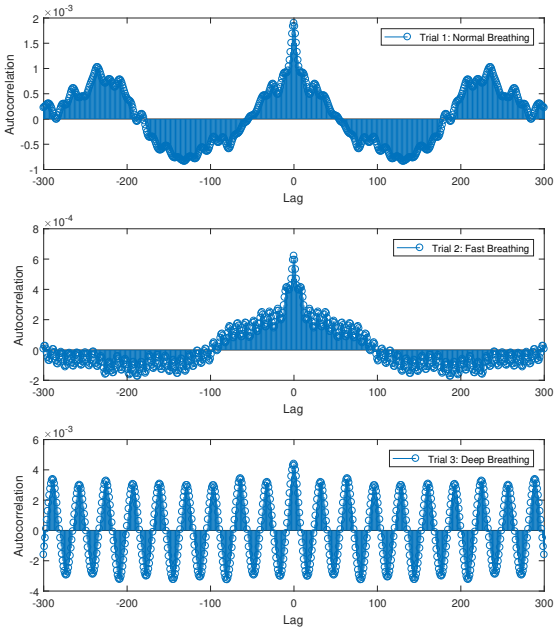
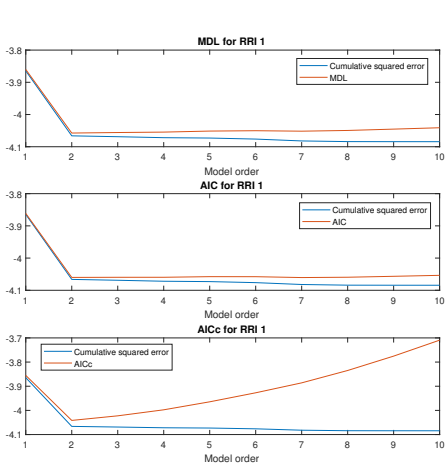
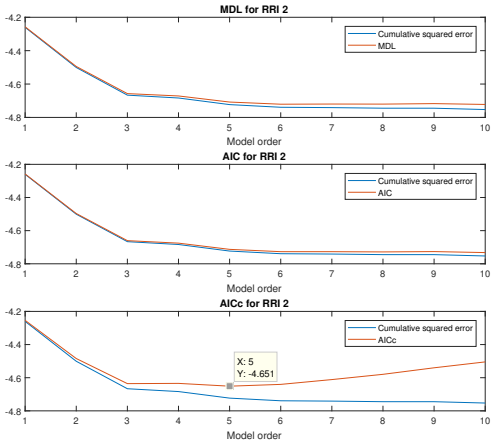


Figure 21: Plot of the Autocorrelation function of the three RRI trials

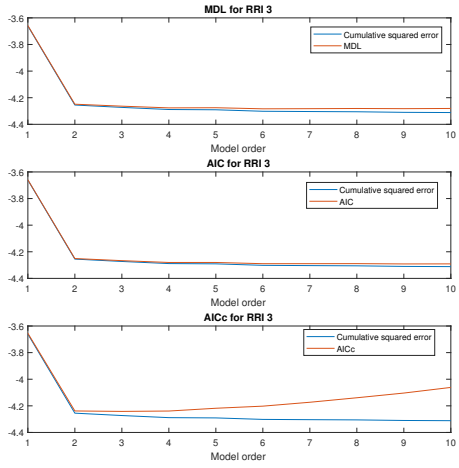
d) To determine if the three RRI signal can be modelled by an AR(p) process, the partial autocorrelation, the MDL, the AIC and the AICc are computed.



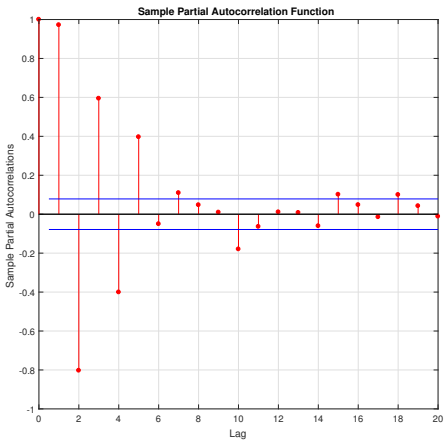
a) Criterion for RRI 1



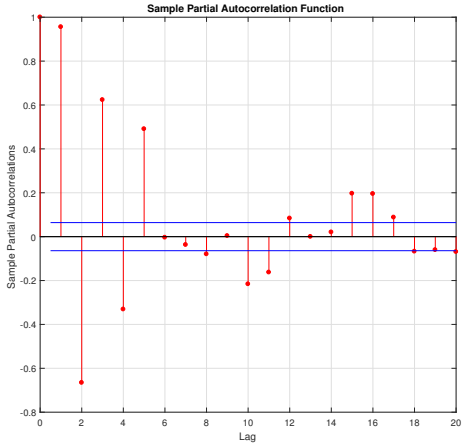
b) Criterion for RRI 2



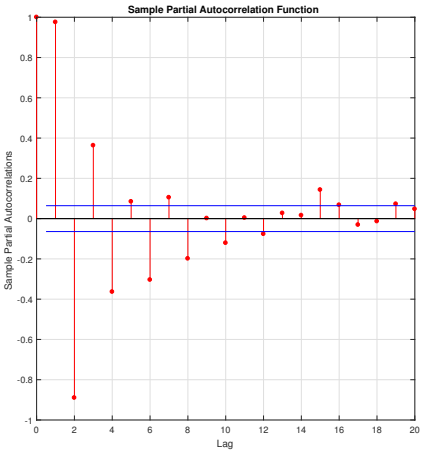
c) Criterion for RRI 3



a) PACF for RRI 1



b) PACF for RRI 2



c) PACF for RRI 3

Figure 22: Analysis of optimal AR model for the three different RRI trials

From the criterion, we can say that the optimal model order for RRI trial 1 and trial 3 would be order 2 and order 3 for RRI trial 2.
However, looking at the partial autocorrelations of the different trials, RRI 1 the values are in the confidence interval from order 5, for RRI 2 as well. For RRI 3, even though, the values are not in the confidence interval a significant decrease is observed after model order 2. Therefore, for RRI 3, an AR process of model order 2 could model the data but it might not be an accurate representation.

3 Spectral estimation and modelling

The periodogram enabling the estimation of the power spectral density (PSD) (which refers to the distribution of the power of an ergodic stochastic process across the frequency components) is defined as:

$$\hat{P}_X(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi f\frac{n}{N}}\right|^2 \qquad f = 0, 1/N, ..., (N-1)/N \tag{42}$$

where N are the discrete frequencies at which the PSD is estimated.

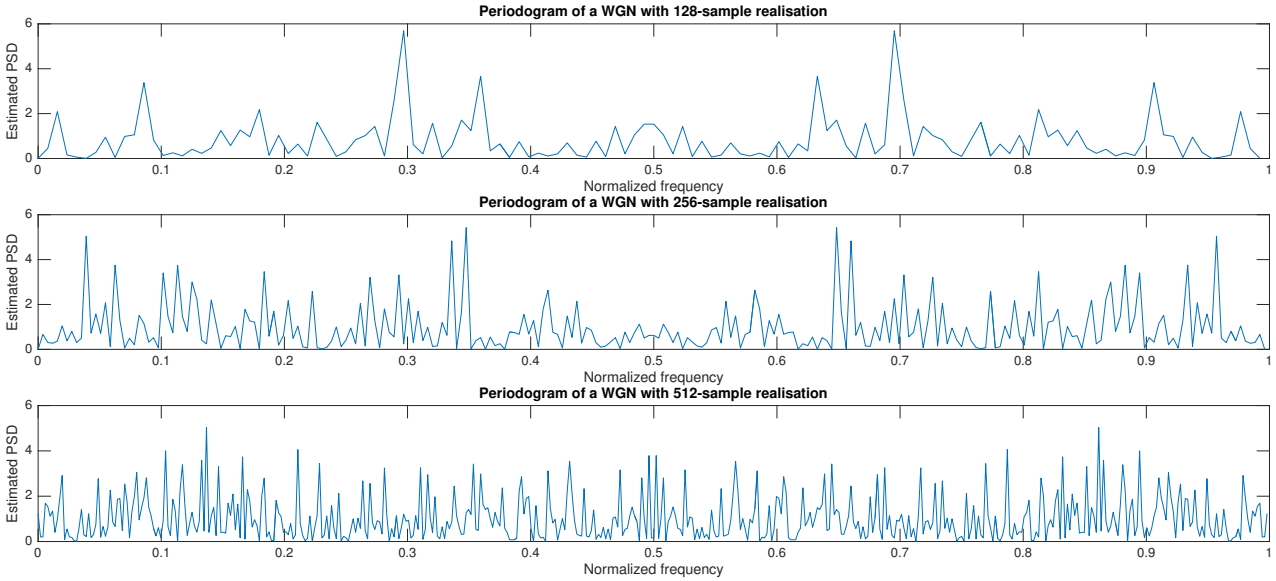


Figure 23: Estimated PSD of WGN for different sample lengths using the periodogram

The PSDs are symmetric at around the y axis at the frequency value of 0.5. Theoretically, the PSD of White Gaussian noise (WGN) is equal to the variance of the process, thus 1 for all frequencies. However, multiple errors are observed which are explained by the finite number of samples and frequency range. This causes the power to not be equally distributed. Increasing the number of samples does not change the variance of the PSD. The values of the three estimated PSDs seem to be in the same range since the variation does not vary much between the three results. The periodogram is an asymptotically unbiased estimator as there is a bias and the mean converges to the PSD as the number of sample increases.

3.1 Averaged periodogram estimates

1. To improve the estimated PSDs, a filter can be added to perform frequency smoothing as to decrease the variance. A zero-phase filter with impulse response [0.2 0.2 0.2 0.2 0.2] is applied to each PSD.

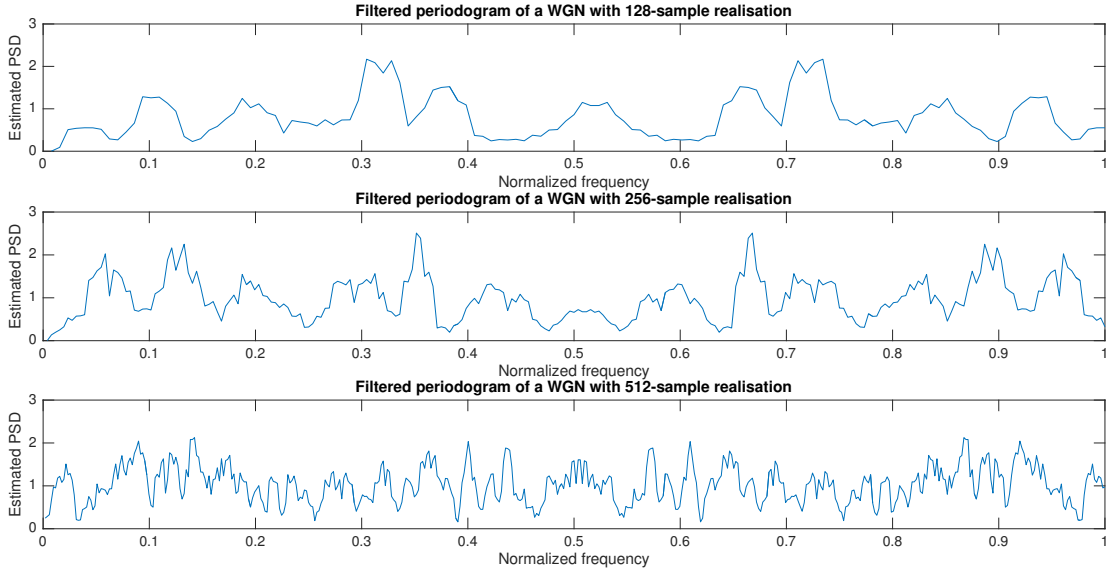


Figure 24: Filtered periodogram of WGN for different sample lengths

Compared to the previous figure, the variability of the PSD has considerably decreased. The maximum values of PSD observed are in the range [2 3] instead of [5 6]. The PSD now approaches more the theoretical PDS which is equal to 1.

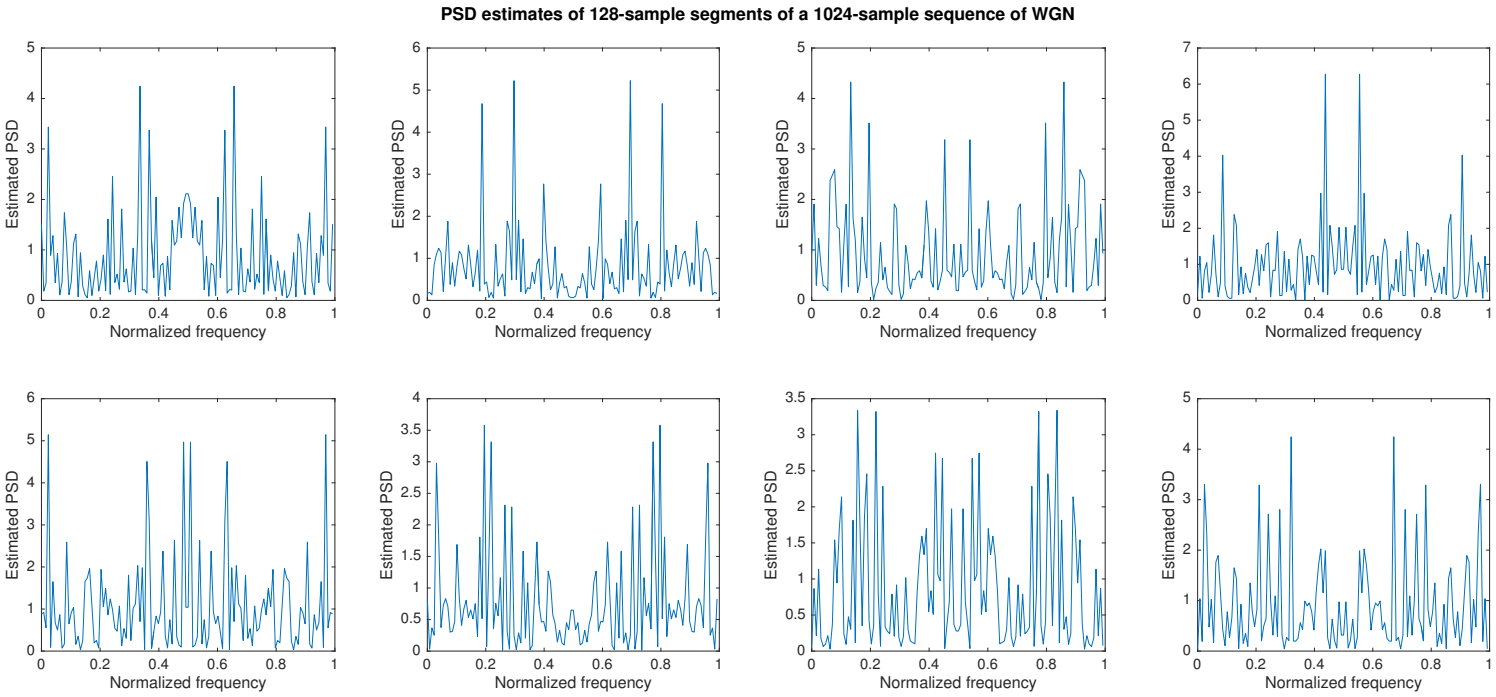


Figure 25: Periodogram of eight non-overlapping 128-sample segments of 1024-sample sequence of WGN

2. The eight PSD's derived are all similar in variability and resemble the periodogram obtained in figure 20.

3. To obtain a more accurate PSD, the eight individual PSD's are averaged to yield a new estimator called the averaged periodogram as the segments are independent of each other. This averaged periodogram can yield a better estimate if it is filtered as well. This is computed in **Figure 26** on the next page

	Mean	Standard deviation
128-sample segment	0.8845	0.8806
Averaged periodogram	0.9011	0.3222
Filtered average periodogram	0.8897	0.1414

Table 7: Table of Mean and Stadard deviation of three different periodograms

With the averaged periodogram, the variability is similar to the result found with the filtered periodogram with a zero-phase FIR filter. The standard deviation, shown in table 5 highlights the fact that the averaged periodogram has less variability than the periodogram of a 128-sample segment. If the averaged periodogram is filtered as well, the periodogram approximates the theoretical one even more since the standard deviation is smaller (0.1414).

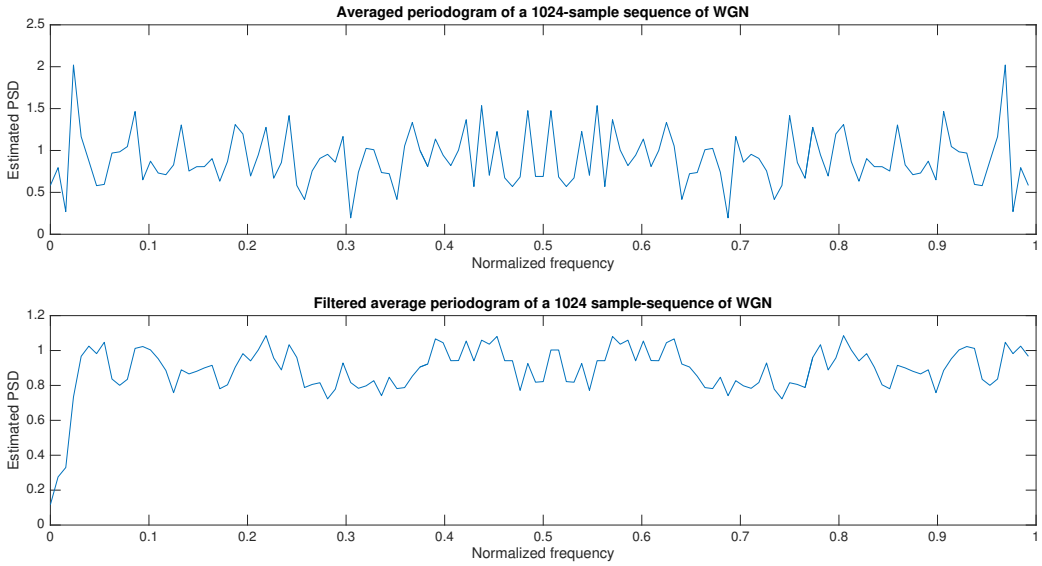


Figure 26: Plot of the Averaged periodogram of the 8 segments and the Filtered averaged periodogram

3.2 Spectrum of autoregressive processes

1. The power spectrum of the infinite impulse response (IIR) filter with coefficients b=1 and a=[1 0.9] is given by the following equation:

$$P_Y(f) = \frac{1}{|\,1 + 0.9e^{-j2\pi f}\,|^2} \tag{43}$$

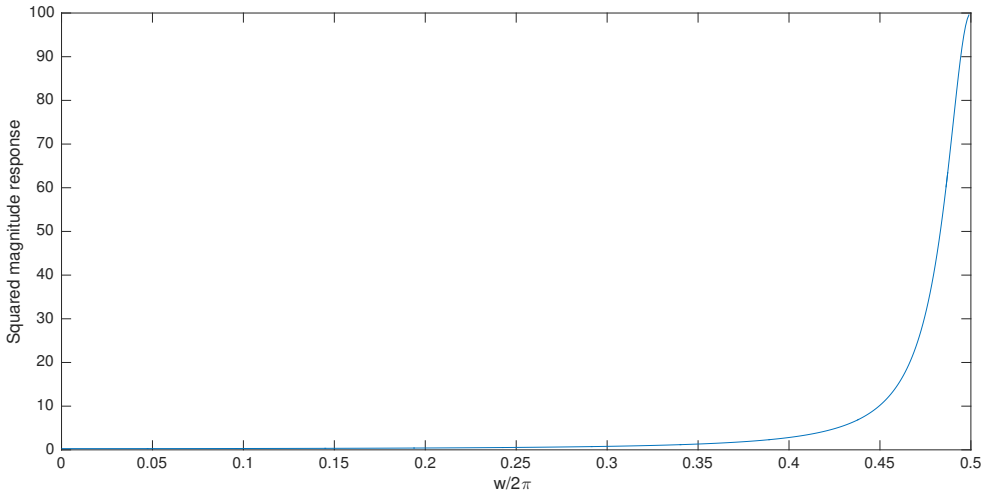
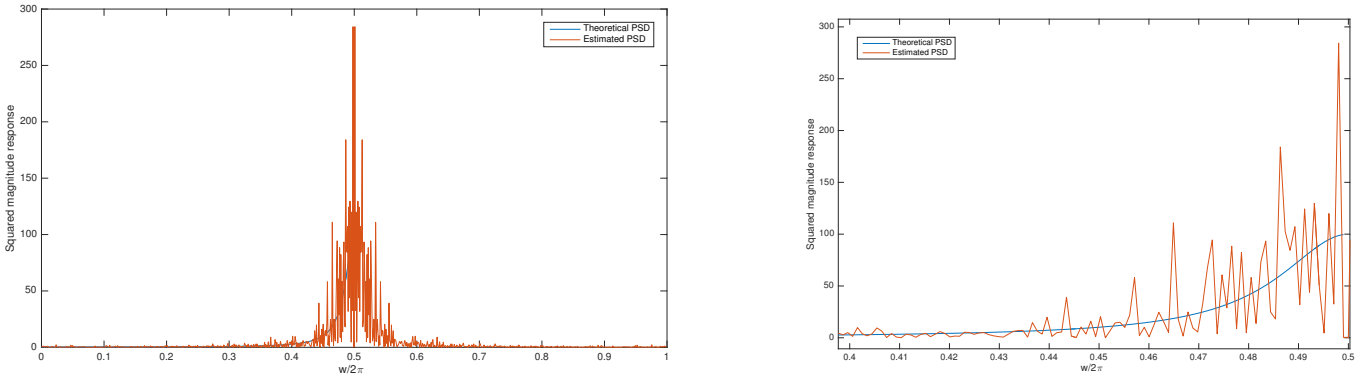


Figure 27: Exact PSD of the considered IIR filter

From the shape of the PSD, we can deduce that the IIR filter act as a high pass filter. This is expected from the coefficients as 0.9 is smaller than 1, a minimum is expected at f=0.
The cut off frequency is the frequency at the half power point. This signifies that it is the frequency where the power transmitted is half the maximum power transmitted in the pass band. In this case, it is the frequency at which the squared magnitude response is equal to 50, which is approximatively $f_c = 0.4834$.



(a) Theoretical and estimated PSD

(b) Zoom on 0.4 and 0.5 frequency of the PSD's

Figure 28: Plot of the estimated PSD with the periodogram and the exact PSD generated with `freqz`

2. From **Figure 28**, we have the theoretical and estimated PSD Due to the definition of the Fourier transform, the periodogram is symmetric around 0.5Hz. Only the left hand side is useful in this case to compare with the theoretical result.
   Before 0.42Hz, the periodogram follows the theory. After that more variability is observed for the periodogram as random spikes appear as we get closer to 0.5Hz. This is due to the fact that the periodogram is not an exact estimate of the PSD since it has a limited number of values.

3. These spikes, or ripples observed are due to the underlying rectangular window. Indeed, the window used determines the bias of the estimator and it introduces ripples in the frequency domain. To solve this problem a Hanning or Hamming window can be used to lessen those ripples and reduce the variance.

4. The model based PSD is generated by finding the variance and the coefficients with the autocorrelation. The results found are: $\hat{a_1} = 0.9028$ and $\hat{\sigma_X^2} = 1.0824$. The PSD is then computed with the `freqz` function of MATLAB.

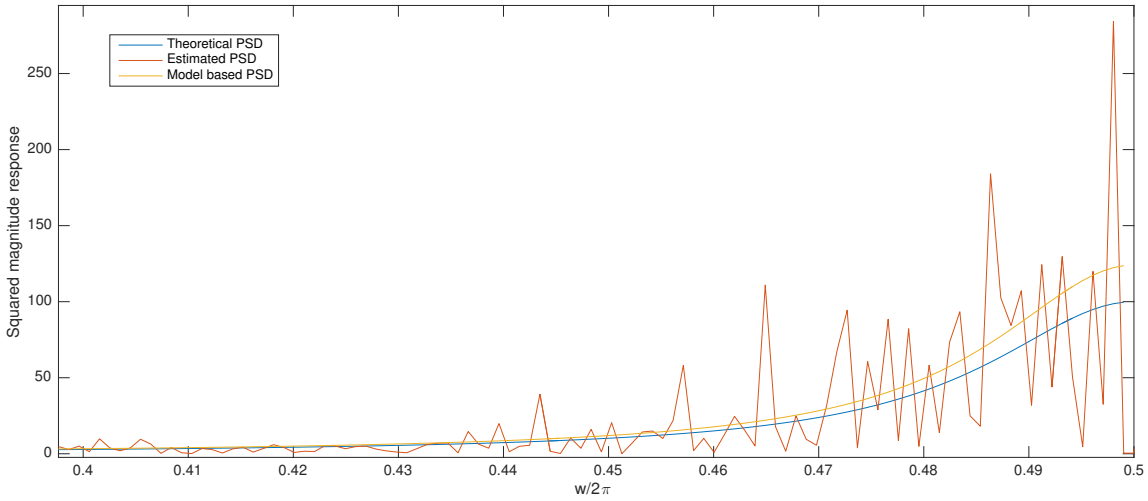
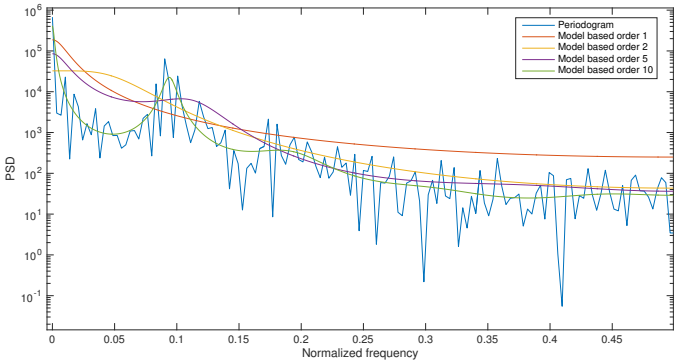


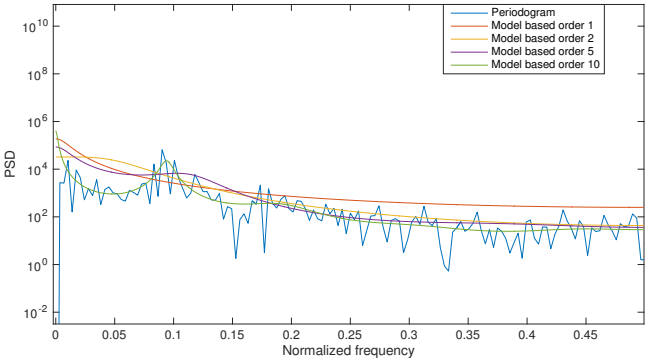
Figure 29: PSD's computed with three different methods: theoretical, estimated with periodogram and model based

The Model based result is closely related to the theoretical one due to the fact that it has implied an infinitely number of samples. The difference observed is due to the inexactitude of finding the

coefficients and variance with the autocorrelation function. The periodogram estimate has a finite number of samples, which implies an underlying windowing which introduces variability and spikes. The model based result is thus more reliable and accurate than the periodogram.



(a) Original Sunspot series



(b) Mean-centered Sunspot series

Figure 30: Plot of the periodogram and model based PSD's for different orders of the original and mean-centered sunspot time series

5. A log scale was chose to present the data in order to see better the periodogram. With a normal scale, the values of the periodogram ar all 0, exept at the 0.1 normalized frequency peak. Undermodelling or overmodelling changes the frequency at which the maximum power peak occurs. With the periodogram, that peak seems to be at 0.1Hz, which corresponds to the result we get with a model order of 10. For order model 1 and 2 that peak seems to occur earlier. Furthermore, the highest peak of each PSD decreases with increasing model orders. The model order also affects the spread of power. For model order 10, there is a signifcant peak at 0.1Hz and the rest is close to 0 (in a non log scale). For the other model order that lobe is wider. With the mean-centered data, less variations seems to be observed with the periodogram. The results are similar to the original series.

3.3 The Least Squares Estimation (LSE) of AR Coefficients

With the LSE method, the aim is to minimize the squared difference between the observed data and the assumed model of noiseless data. This implies finding the minimum of the cost function:

$$J(\theta) = \sum_{n=0}^{N-1} e^2[n] \tag{44}$$

where e is the difference between the observed signal and the signal model.

1. For an AR(p) process, the LS cost function to find the unknown coefficients is of the form:

$$J = \sum_{k=1}^{M} \left[\hat{r}_{xx}[k] - \sum_{i=1}^{p} a_i \hat{r}_{xx}[k-i] \right]^2, \quad for \quad M \geqslant p \tag{45}$$

To simplify this equation, the cost function can be rewritten in matrix form.
If

$$\mathbf{x} = \begin{bmatrix} \hat{r}_{xx}[1] \\ \hat{r}_{xx}[2] \\ \vdots \\ \hat{r}_{xx}[M] \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} \hat{r}_{xx}[0] & \hat{r}_{xx}[-1] & \cdots & \hat{r}_{xx}[1-p] \\ \hat{r}_{xx}[1] & \hat{r}_{xx}[0] & \cdots & \hat{r}_{xx}[2-p] \\ \vdots & \vdots & & \vdots \\ \hat{r}_{xx}[M-1] & \hat{r}_{xx}[M-2] & \cdots & \hat{r}_{xx}[M-p] \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{46}$$

then

$$\sum_{i=1}^{p} a_i \hat{r}_{xx}[k-i] = \mathbf{Ha} \quad and \quad J = \sum_{k=1}^{M} \left[\hat{r}_{xx}[k] - \sum_{i=1}^{p} a_i \hat{r}_{xx}[k-i] \right]^2 = (\mathbf{x} - \mathbf{Ha})^T (\mathbf{x} - \mathbf{Ha}) \tag{47}$$

This can be rewritten as:

$$\begin{aligned} J &= (\mathbf{x} - \mathbf{Ha})^T (\mathbf{x} - \mathbf{Ha}) \\ J &= \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{Ha} + \mathbf{a}^T\mathbf{H}^T\mathbf{Ha} \end{aligned} \tag{48}$$

To find the minimum, the gradient is set to 0 and we have:

$$\begin{aligned} 0 &= -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{Ha} \\ \mathbf{H}^T\mathbf{Ha} &= \mathbf{H}^T\mathbf{x} \end{aligned} \tag{49}$$

The Least Squares estimates for the unknown coefficents **a** is then:

$$\hat{\mathbf{a}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \tag{50}$$

We notice that if p=M, the least square estimation corresponds to the Yule -walker estimates.

2. With the Autocorrelation function, the stochastic components of the signal are analysed. Since the matrix **H** is composed of the autocorrelation of x for different lag, the observation matrix **H** is a stochastic matrix.

3.4 Spectrogram for time-frequency analysis: dial tone pad

1. From the first plot of the figure above, we can see that the tone lasts 5.25s with idle time in between. By comparing tone 2 and 0, it is clear that the output waveform is different between the two, which enables to specify each of time as each number will have a specific waveform. The last plot of the figure highlights the fact that the tone is indeed the addition of two sinusoids with specific frequencies. Each number has two specific frequencies.
   The sampling rate 32768 is appropriate as it needs to be at least 2 times bigger than the Nyquist rate. Generally, the closest power of 2 to that value is taken.

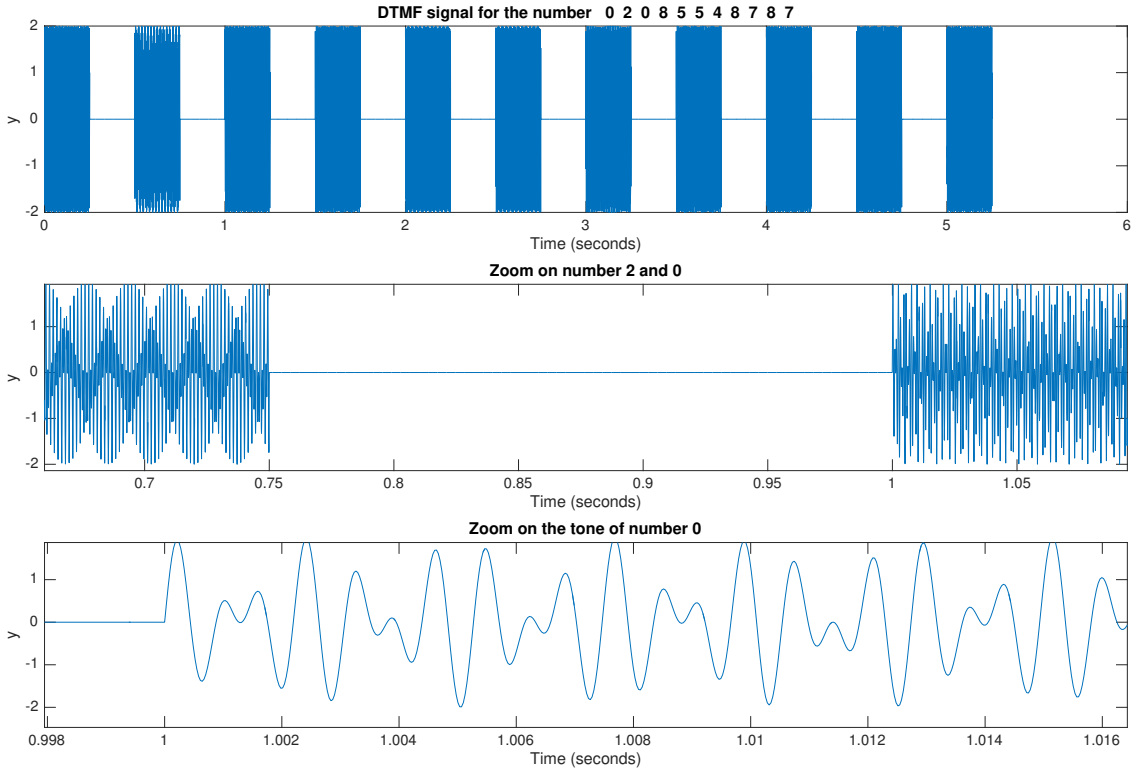
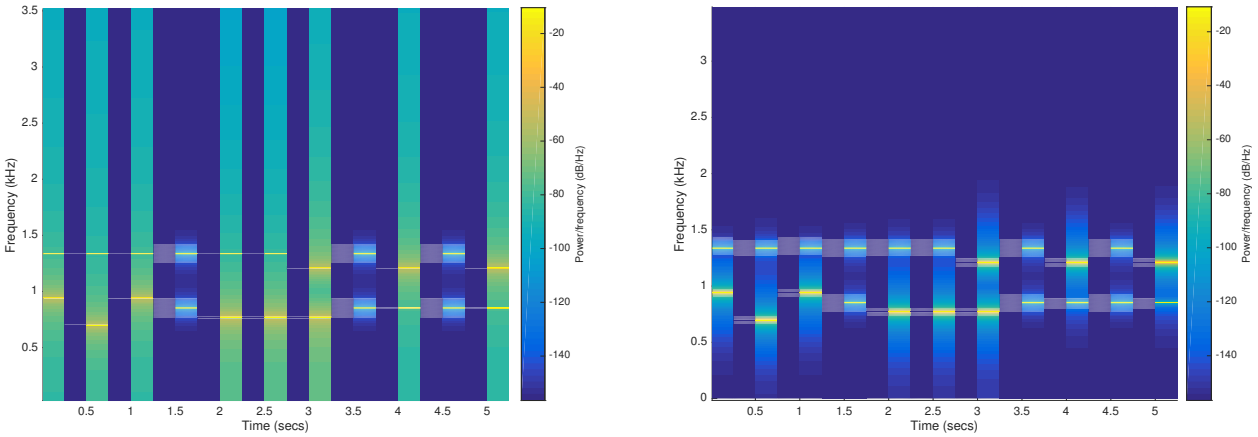


Figure 31: DTMF of a Landline number with different scales

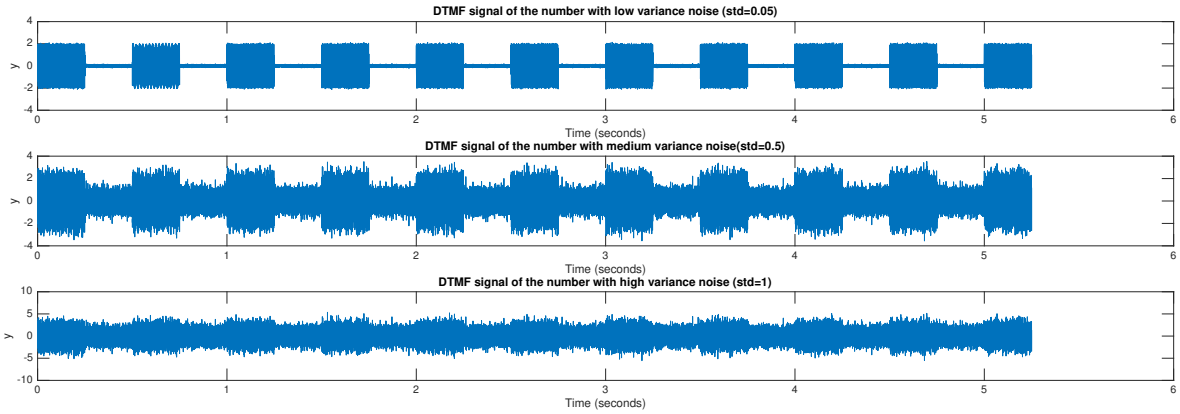
2. To compute the FFT for non-overlapping segments, the segment length and number of DFT points used in the function `spectogram`is equal to 8192 which corresponds to 0.25 times the sampling rate. Indeed, in that case each segment of 0.25s is analysed which is corresponds to the time of a touch-tone information or idle time. The spectrogram are presented on the next page with **Figure 32**
   With the Hamming window, the difference between idle time and touch-tone information is clear as in idle time, the bands are completely dark blue. With the Hanning window the frequency at which the power is maximum is highlighted as only these parts are not dark blue.
   For each number, two yellow lines are observed corresponding to the pairs of frequency characterizing the key.

3. The maximum power frequency is observed for each pair characterizing the keys as mentioned previously. Therefore, by looking at which frequency this occurs, we can determine what number it is. For example, for the first key press, there is a maximum power at around 932HZ and 1336HZ. If we look at the table, this corresponds to 0. If we do this for each key press, the number can be reconstructed. Windowing spreads the power, which decreases the selectivity. In this case, the frequency are different enough that even tough there is a small spread of power, we are able to identify which key it is. If the frequencies where closer together, this would pose a problem and another windowing should be used.

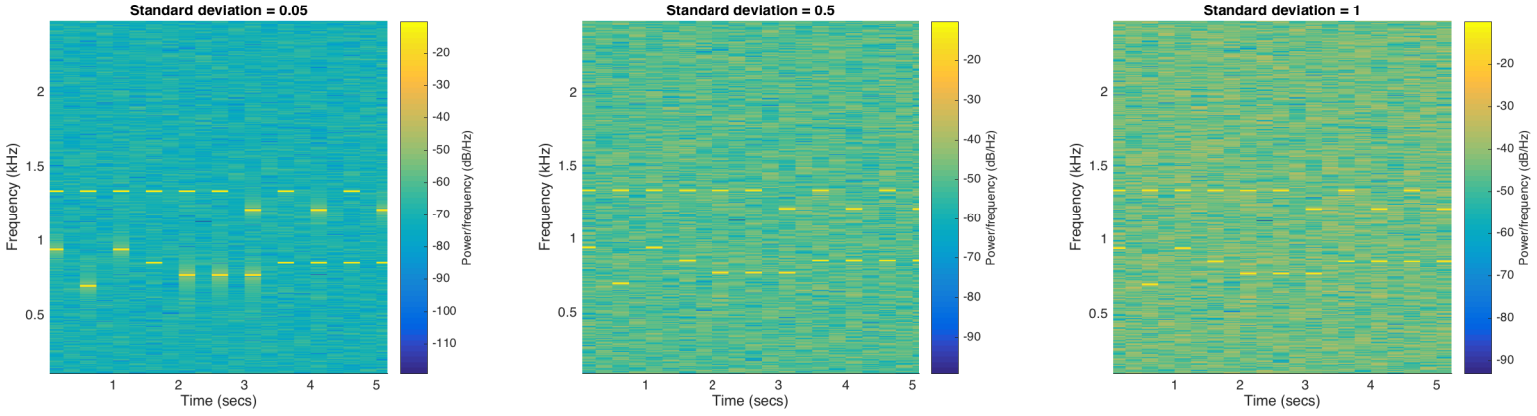


(a) Spectogram using the default Hamming window **(b)** Spectogram using Hanning window

Figure 32: Spectogram of the DTMT of the number 020 8554 8787]



(a) DTMF of signal with added noise of different variance



(b) Spectogram of the signals with noise of different variance

Figure 33: DTMF and spectogram of the original signal with noise of different variance

4. As the variance of the noise increases, the idle time and the time corresponding to a single key become indistinguishable from each other in the time domain. The tone is hidden in the noise. This is shown in **Figure 33a)**
   In the frequency domain (spectogram) the background is all the same: we are unable to distinguish

the idle and single key time. As the variance increases, the background tends to more yellow colors, closer to the maximum power. However, the yellow lines are still visible although more spread out and blurry and thus less selective. In our case, it is still possible to perform key identification as the frequencies of characterization are far from each other. Otherwise, the noise would have to be filtered in order to get more accurate and reliable key identifications.

3.5 Real world signals: Respiratory sinus arrhythmia from RR-Intervals

1. The periodogram of the RRI derived from the three different trials were computed:

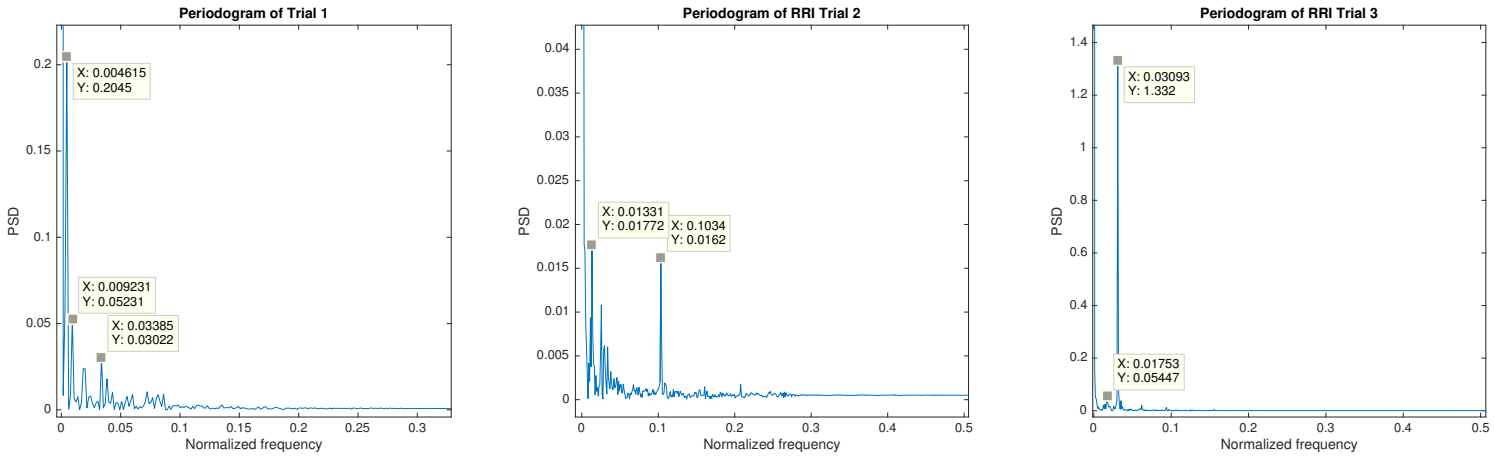


Figure 34: Periodogram of the three different RRI: Normal breathing, Fast breathing and Deep breathing

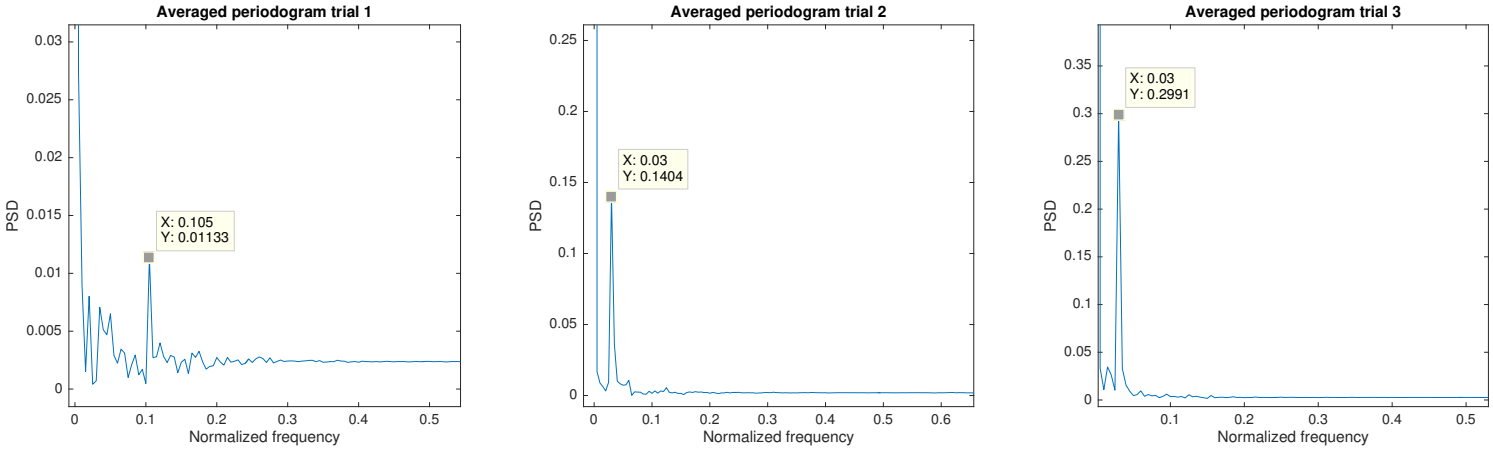


Figure 35: Averaged Periodogram of the three different RRI: Normal breathing, Fast breathing and Deep breathing

2. Due to the constrained fast breathing a peak at a higher frequency is observed for trial 2 (0.1Hz). It is expected that in trial 3, deep breathing, a peak would be observed at a lower frequency than in trial 2, which is the case as a peak occurs at 0.03Hz. The maximum peak highlighted on each plot, thus correspond to frequencies of respiration.
   Some small harmonics are observed with the periodogram, but it is also mixed with noise. Therefore, it is quite hard to identify all of them. The noise might be due to noise during the recording of the ECG.
   With the averaged periodogram, the harmonics are not visible, only the fundamental. For trial 3,

the maximum peak are approximatively the same. For trial 1, only one of the peaks appear in the averaged periodogram. For trial 2, the second peak at the higher rate indicating the fast breathing is not recorded. This suggest that the periodogram is more reliable in this case, as some important peaks seem to be missing in the averaged periodogram.

4 Optimal filtering -fixed and adaptive

4.1 Wiener filter

The Signal-to-Noise ration (SNR) in dB for output signal z[n], which is the output filter y[n] with added noise, is equal to 20.16dB.

1. The optimal coefficients of the Wiener filter are given by the following equation:

$$\mathbf{w}_{opt} = \mathbf{R}xx^{-1}\mathbf{p}_{xx} \tag{51}$$

Thus $\mathbf{R}_{xx}$ and $\mathbf{p}_{xx}$ need to be calculated using the autocorrelation of the input x[n] and the correlation between z[n] and x[n]

$$\mathbf{R}_{xx} = \begin{bmatrix} 0.9984 & -0.0434 & -0.0212 & 0.0016 & 0.0215 \\ -0.0434 & 0.9984 & -0.0434 & -0.0212 & 0.0016 \\ -0.0212 & -0.0434 & 0.9984 & -0.0434 & -0.0212 \\ 0.0016 & -0.0212 & -0.0434 & 0.9984 & -0.0434 \\ 0.0215 & 0.0016 & -0.0212 & -0.0434 & 0.9984 \end{bmatrix} \quad \mathbf{p}_{xx} = \begin{bmatrix} 0.2088 \\ 0.4310 \\ 0.6733 \\ 0.4324 \\ 0.2051 \\ -0.0121 \\ -0.0104 \\ -0.0335 \\ -0.0513 \end{bmatrix} \quad \mathbf{w}_{opt} = \begin{bmatrix} 0.9735 \\ 2.0000 \\ 2.9973 \\ 1.99542 \\ 0.9891 \end{bmatrix} \tag{52}$$

Note: The obtain the correct coefficients, $\mathbf{w}_{opt}$ was multiplied by the scaling factor that was applied on the unknown system output to have a variance unity.
The coefficients are similar to those of the unknown system as the coefficients b are [1,2,3,2,1]

2. In the following table, the variance correspond to the variance of the additive noise.

Variance	SNR (dB)
0.1	10.4702
0.5	4.4207
1	3.1928
2	1.7178
5	0.7276
10	0.4117

$\mathbf{w}_{opt}(0.1)$	$\mathbf{w}_{opt}(0.5)$	$\mathbf{w}_{opt}(1)$	$\mathbf{w}_{opt}(2)$	$\mathbf{w}_{opt}(5)$	$\mathbf{w}_{opt}(10)$
0.9688	1.0216	1.1390	1.0656	1.2456	0.9089
0.9688	1.0216	1.1390	1.0656	1.2456	0.9089
1.9860	2.0793	1.8824	2.2325	1.8219	1.8987
2.9886	2.7650	3.0882	2.7992	2.8904	3.0187
1.9707	2.0470	2.0310	1.9529	1.9686	2.0842
1.0157	1.0372	0.9305	1.1296	0.6512	1.4750

We notice that as the variance of the noise increases, the SNR decreases which means that the level of the background noise is higher and we are less able to discern the signal. It also seems that the coefficients are then less accurate as the variance increases. However, the error is relatively low, which means that even though the SNR is small, the wiener filter is still able to compute values close to the theoretical coefficients.
If we increase N_w we are assuming that there is more than 5 coefficients. In this case, these additional values should approximate 0 as the theoretical filter only has 5 coefficients. To test this N_w equal to 6, 8 and 10 were tested (plus 2 coefficients)

$\mathbf{w}_{opt}(N_w = 6)$	$\mathbf{w}_{opt}(N_w = 8)$	$\mathbf{w}_{opt}(N_w = 10)$
0.9863	0.9861	0.9731
1.9997	1.9994	1.9993
2.9976	2.9979	2.9983
1.9953	1.9961	1.9967
0.9887	0.9884	0.9892
-0.0034	-0.0037	-0.0033
-0.0139	-0.0143	-0.0146
	-0.0165	-0.0164
	0.01695	0.0174
		0.0038
		0.0163

As expected, the additional coefficients are close to 0. Increasing the value of N_w does not affect the coefficients much. Therefore, it is better to avoid using higher values of N_w than necessary as it will add more complexity and calculations.

3. Many additions and multiplications are performed to determined the optimal wiener solution:

   - $O(N_w + 1)$ operations for the inverse autocorrelation matrix
   - $N_w(N_w + 1)$ additions between $\mathbf{p}_{zx}$ and the inverse autocorrelation. The number of multiplications would then be $(N_w + 1)^2$
   - $(N_w + 1)^3$ for the number of multiplication to find the autocorrelation matrix. $N_w(N_w + 1)^2$ for the number of additions.
   - $N_w(N_w + 1)$ for the number of additions for the cross correlation matrix. The number of multiplications is $(N_w + 1)^2$

These number of operations are valid if assume that the number of data points x[n] and w[n] are equal to $(N_w + 1)$. As we can see this adds up in the end to many computational work and thus, the lower N_w, the less calculations there are.

4.2 The least mean square (LMS) algorithm

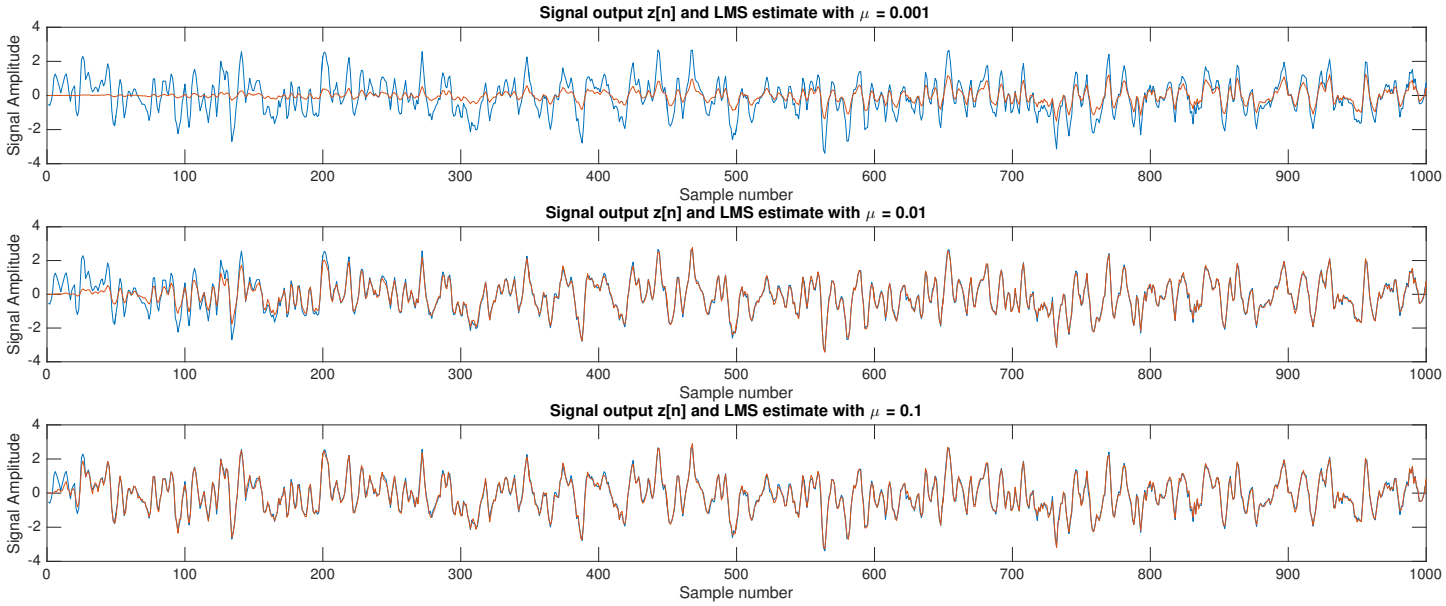
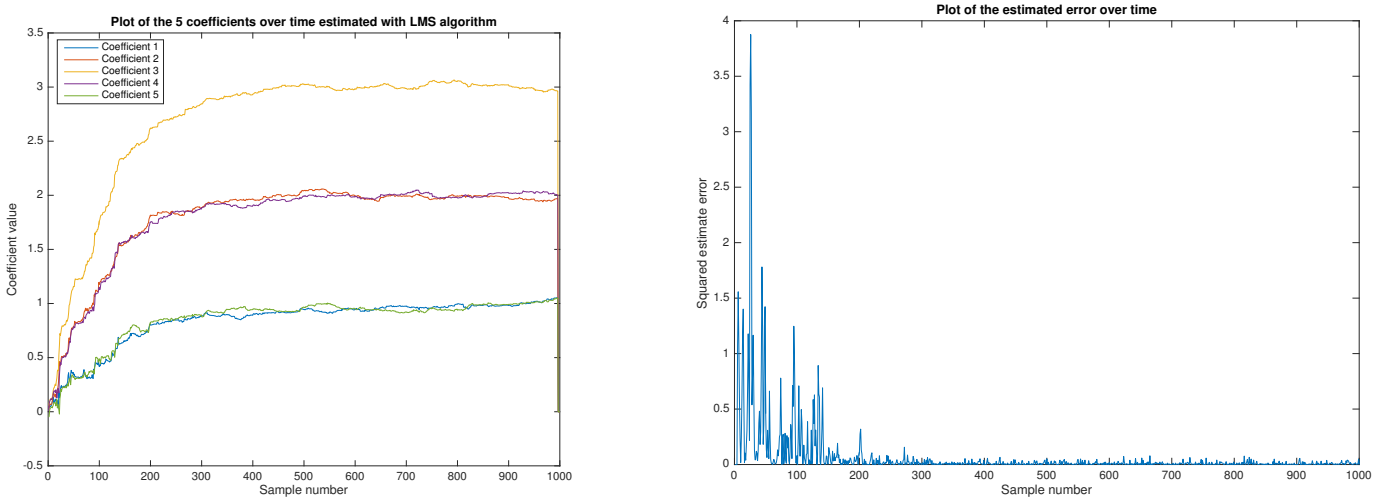


Figure 36: Plot of the output *z[n]* and the LMS estimate for different values of adaptation gain

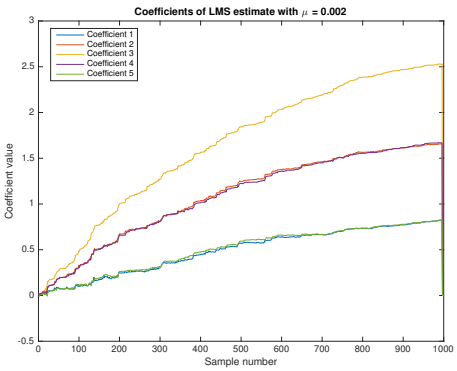
1. From the figure above, we can see that as the adaptation gain increases, the estimate is more accurate more quickly: it approximates the theoretical output quicker if $\mu = 0.1$ than $\mu = 0.001$. However, after enough time, the approximation will be more accurate if the adaptation gain is smaller. Indeed, the minimum value is reached in smaller step, which enables a greater precision than a greater step. The value of the adaptation gain is thus important and varies depending on the application. A ideal μ can be estimated.



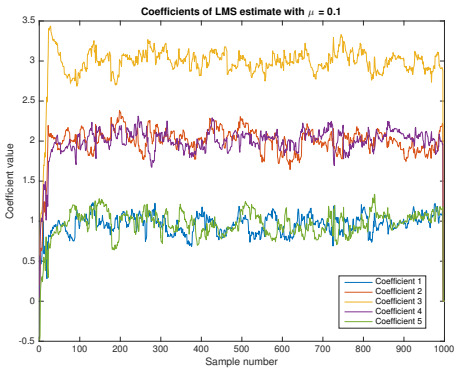
(a) Coefficients estimated with LMS algorithm **(b)** Squared error over time

Figure 37: Plot of the results of the coefficients and the error with an adaptation gain of 0.01

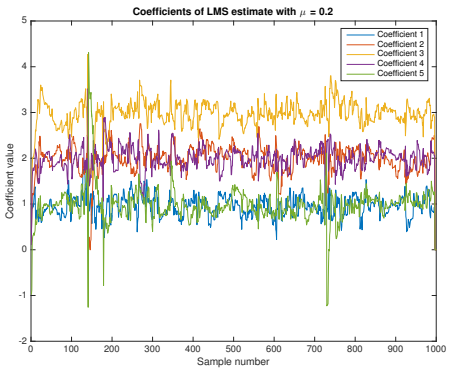
2. Over time, the LMS estimate becomes more and more accurate and the error considerably decreases.



a) $\mu = 0.002$



b) $\mu = 0.1$



c) $\mu = 0.2$

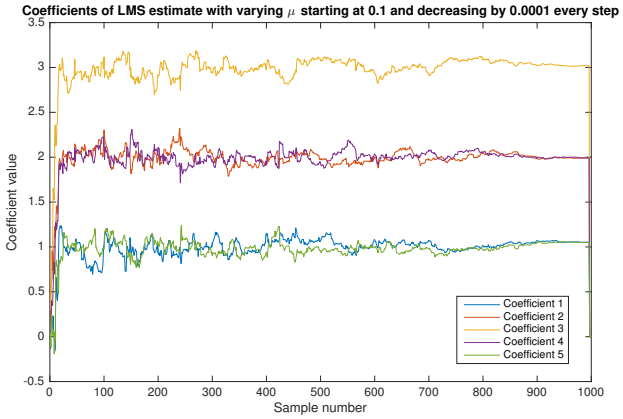
Figure 38: Plot of the five coefficients determined with LMS algorithm with different adaptation gain values

From the figure above, we can infer that with a too small value of adaptation gain, the correct coefficients may never be attained and never approximate the system. However, if the adaptation gain is too high, there is a higher variability in the value of the coefficients and thus do not get an accurate result. The system becomes unstable. The aim is too find a balance between the time taken too reach the correct results and a small variability in the coefficient.
In our case, the best result is found with 0.01 adaptation gain.

3. For the LMS algorithm, the number of additions is estimated at $N^3 N_w$ and the number of multiplications at $3N(N_w + 1)$

4.3 Gear shifting

1. In order to see the effect of a time varying μ, the value was changed every step by 0.0001 starting with 0.1.



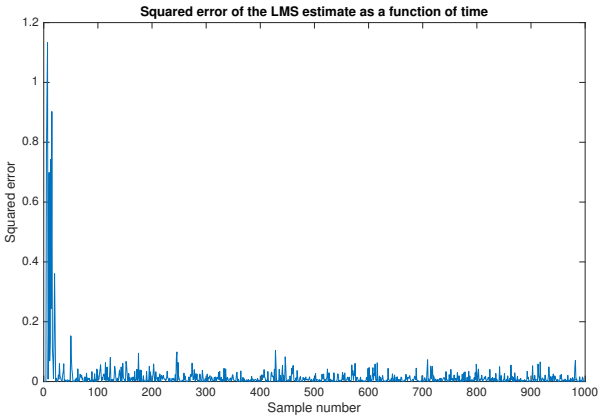


Figure 39: Plot of the coefficients over time as the adaptation gain varies with the corresponding squared error

Compared to the constant μ, the decreasing varying μ enables a fast approximation of the coefficient that then becomes more accurate over time since a smaller μ has a better tracking behaviour. This means that the squared error decays faster than if we started with a constant smaller μ and the variability of the squared error is then relatively smaller than if we started with higher constant μ value.

4.4 Identification of AR processes

1. If the adaptive LMS algorithm is correctly implemented, a_1 and a_2 should respectively converge to -0.9 and -0.2 if the AR model is defined with the coefficients $\mathbf{a} = [1\ 0.9\ 0.2]$.

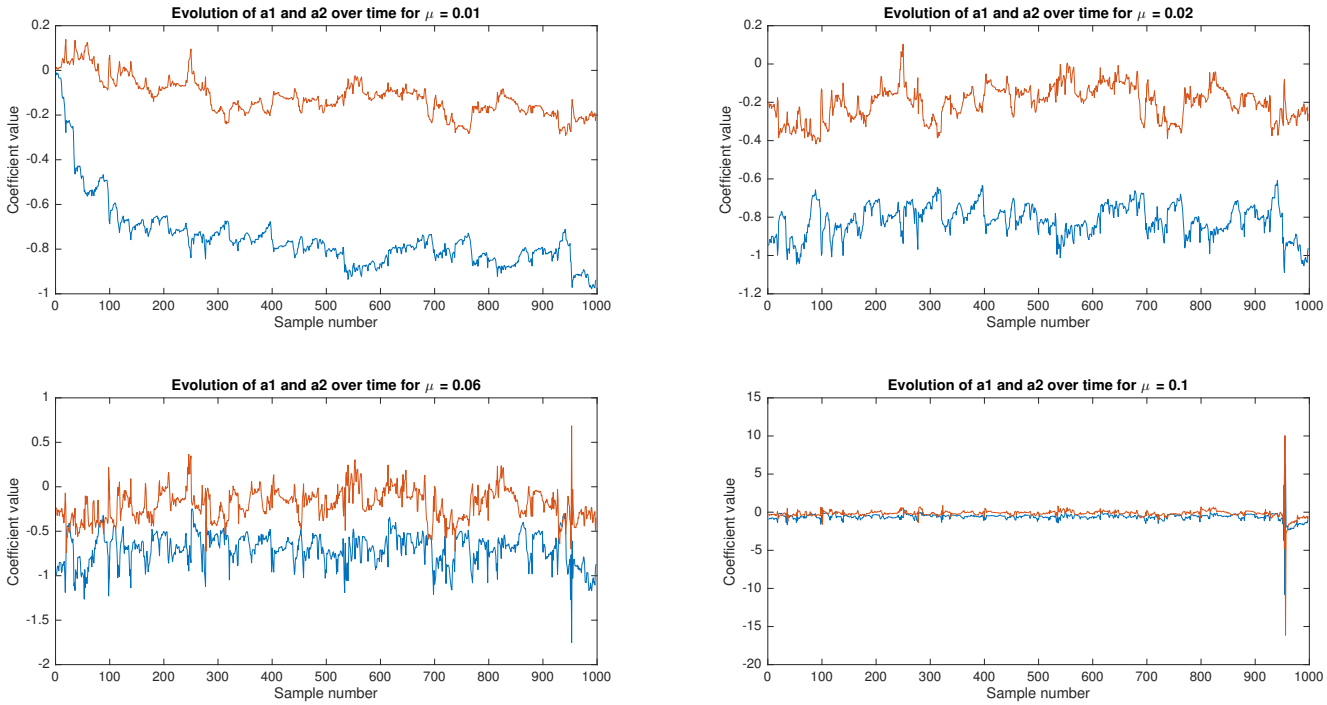


Figure 40: Graphs of the evolution of a_1 and a_2 over time determined with the synthesis and analysis structure of an AR model (a_1 blue line and a_2 red line

From the figure above, we can see that the lower the value of μ, the longer it takes to reach the theoretical value. However, as μ increases, more oscillations are observed, indicating a variability and a less reliable result. Indeed for $\mu = 0.06$, a lot of noise is observed. However, by taking the mean we get a reasonable approximation of the coefficients.At $\mu = 0.1$, the results are unreliable and the model seems to break down. Therefore, a lower adaptation gain has a better tracking behaviour.

4.5 Speech recognition

1. Five different letters were recorded to compute the following graphs. The analysis part of the AR model structure was used to assess how the adaptation gain and order of predictor affects the predicted data.

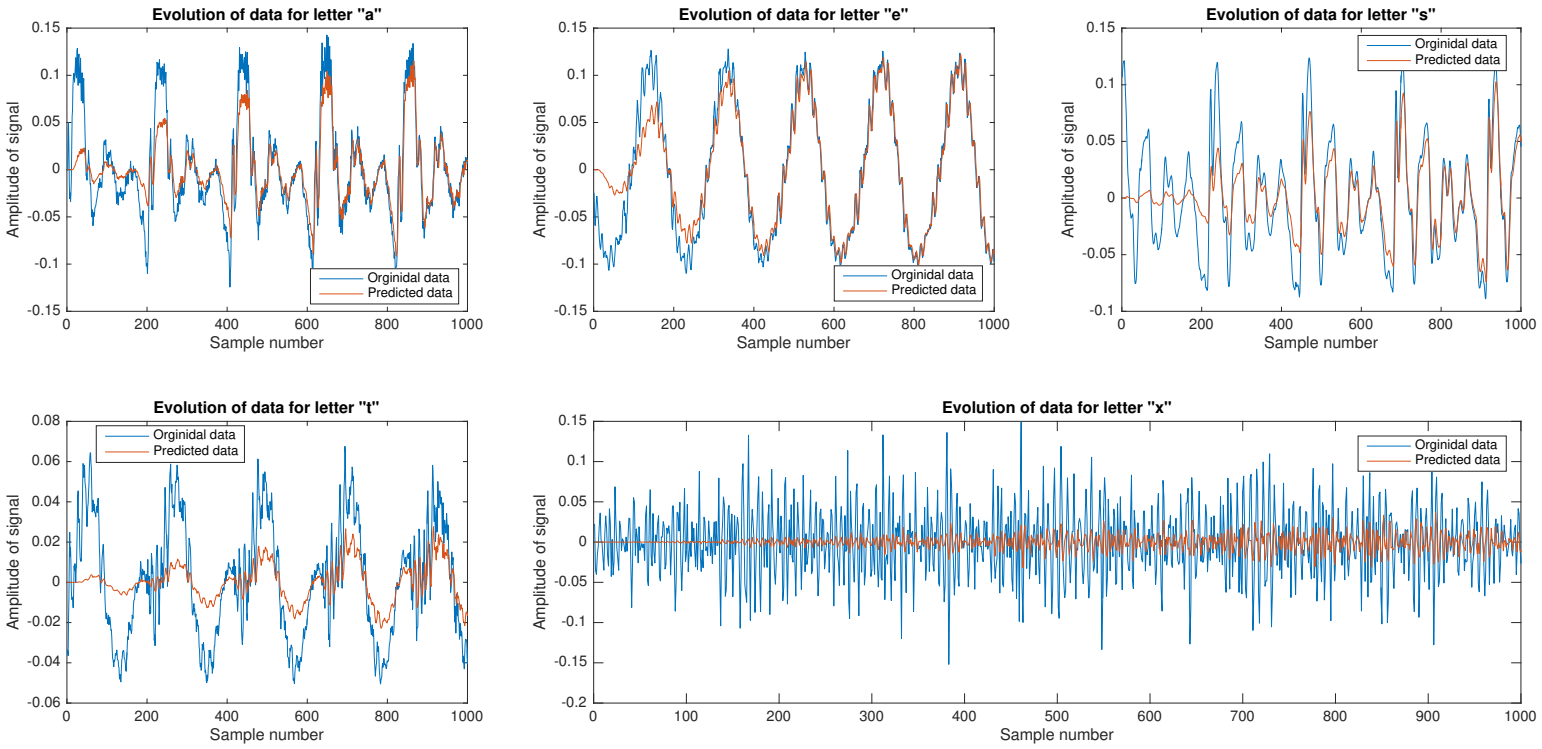


Figure 41: Plot of the actual and predicted data for different letters with μ=0.5 and order 20

From the results, we can see that the less spikes there are the better the prediction. Indeed, for ”x” even at the end, the prediction does not correspond to the actual data. In order to resolve this a higher predictor order could be chosen, to increase the accuracy at the expense of complexity, or a higher adaptation gain. In our case, the order predictor is 10, and it seems to be enough for the other letters. For ”e”, a sinusoidal pattern is observed and thus the prediction corresponds early on to the actual data. To assess the adaptation gain, ”t” was chosen to be observed (as the general trend would be the same).

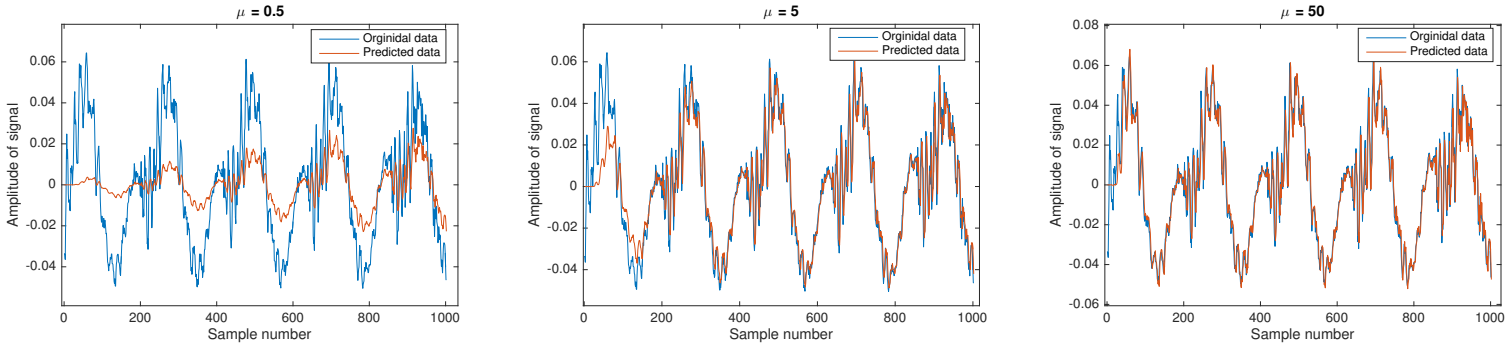


Figure 42: Plot of the evolution of data ”t” with different μ for the predicted output

From the figure below, we can say that a too small adaptation gain will take too long to correctly predict the output. It converges more rapidly to the original data when the adaptation gain is higher. However, as seen previously, a high gain introduces oscillations and inacurracies over time. This is less visible in this case as a cyclical pattern is observed. There is again a trade-off between the time to get an accurate prediction and the inacurracies introduced. Another way, an heuristic approach, would be to look at the difference between the original and the predicted for different filter length for a letter. From this comparison, it is possible to determmine an optimal fitler length.

2. An optimal filter length can be determined analytically with the different criterion such as Minimum description length (MDL) and Akaike information criterion (AIC) encountered previously. The aim is to have the shortest filter length that still yields a suitable estimate of the data.

3. With a smaller frequency samples, less data are taken in an interval of time. Therefore, the prediction gain is expected to be of lower amplitude, as more errors are likely to be introduced.

4.6 Dealing with computational complexity: sign algorithms

In order to reduce calculations and simplify the LMS algorithm, there is the sign LMS algorithm. Either only the sign of the error or the regressor or both are taken

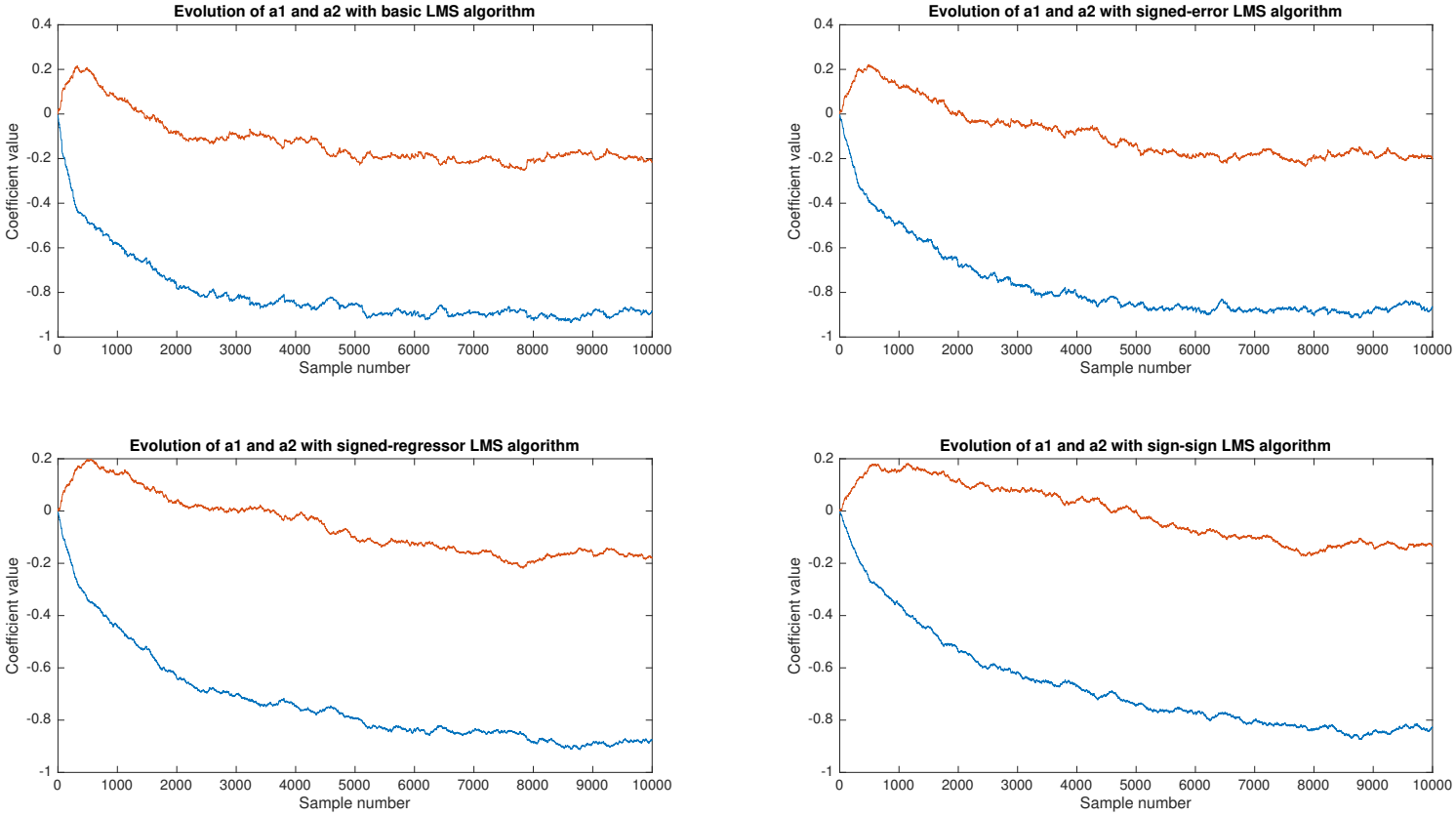


Figure 43: Evolution of coefficients a1 and a2 over time for the basic LMS algorithm and the three different sign LMS algorithms

Comparing the sign LMS algorithm performance to the basic one, we can see that the basic one converges more quickly to the desired coefficient. The slowest one is the sign-sign algorithm. This is expected as with the sign, the previous result is less precise and accurate and thus it will take more time to get the to the expected coefficient value. Even tough it takes longer, all of the sign algorithm converge to the expected coefficients and are as accurate that the basic algorithm. The only difference observed is the time to reach the value.
We are know going to assess how it predicts the speech data:

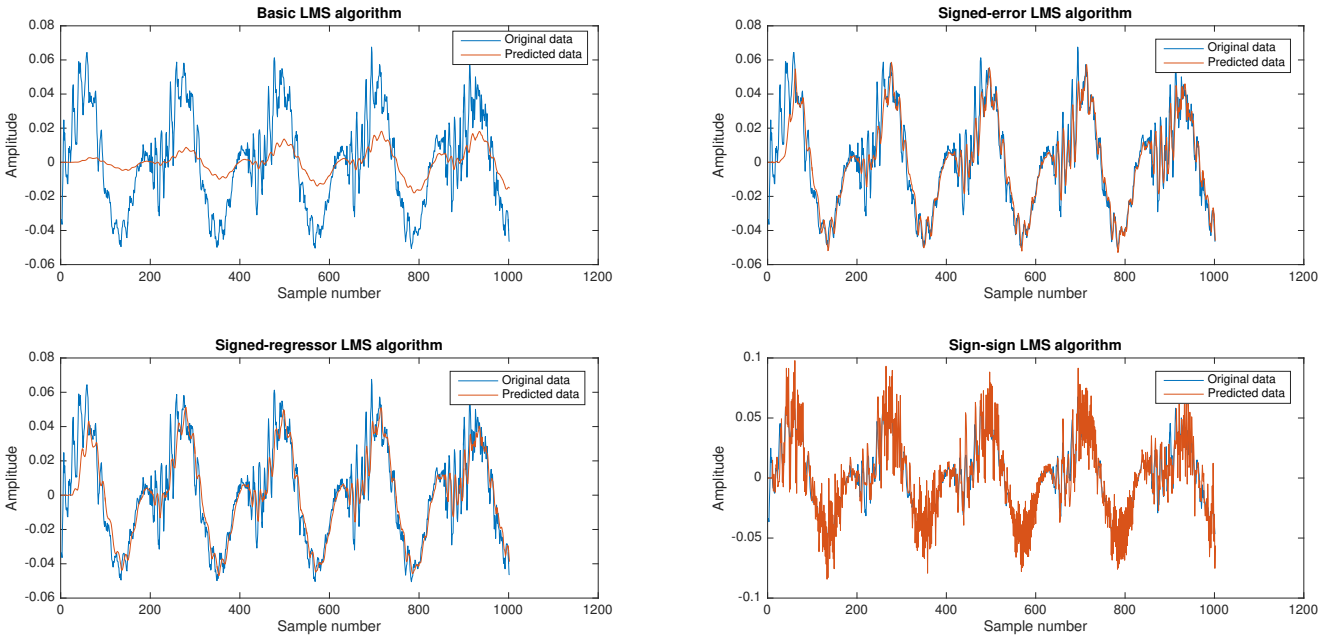


Figure 44: Original and predicted data computed with the basic LMS algorithm and three sign LMS algorithm with $\mu = 0.1$ and order 10

From **Figure 44**, we can see that the signed-error and signed-regressor or more accurate than the basic algorithm. This is due to the fact that the adaptation gain is insufficient for the basic LMS algorithm to attain the correct shape. If a higher adaptation gain was chose, the basic LMS algorithm might be more accurate. However, here the signed-error and regressor are better estimator and converge more quickly to the correct shape. For the sign-sign LMS algorithm, there seems to be a lot of errors and oscillations. This indicates that the sign is insufficient to gain an adequate shape. If the value of μ was smaller, the sign-sign LMS algorithm would be more appropriate. Thus at higher adaptation gain, the basic algorithm is a better estimator and at low values, the sign algorithms are more accurate.

5 MLE for the Frequency of a Signal (Optional)

1. The maximum-likelihood estimate (MLE) of the frequency f_0 can be found by minimizing the following equation:

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - Acos(2\pi f_0 n + \phi))^2 \tag{53}$$

To facilitate the calculations, the equation can be rewritten by using the cosine trigonometric identity: $cos(\alpha + \beta) = cos\alpha cos\beta + sin\alpha sin\beta$. Therefore, we have:

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - Acos(2\pi f_0 n)cos\phi + Asin(2\pi f_0 n)sin\phi)^2 \tag{54}$$

Now, to simplify the equation, some elements are defined:

$$\alpha_1 = Acos\phi \quad \alpha_2 = -Asin\phi \quad c = \begin{bmatrix} 1 \\ cos(2\pi f_0) \\ \vdots \\ cos(2\pi f_0(N-1)) \end{bmatrix} \quad s = \begin{bmatrix} 0 \\ sin(2\pi f_0) \\ \vdots \\ sin(2\pi f_0(N-1)) \end{bmatrix} \tag{55}$$

The equation becomes then:

$$J(\theta) = (x - \alpha_1 c - \alpha_2 s)^T (x - \alpha_1 c - \alpha_2 s) \tag{56}$$

If we define:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ cos(2\pi f_0) & sin(2\pi f_0) \\ \vdots & \vdots \\ cos(2\pi f_0(N-1) & sin(2\pi f_0(N-1)) \end{bmatrix} = [\mathbf{c}, \mathbf{s}] \quad and \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \tag{57}$$

we get the following form:

$$J'(\alpha_1, \alpha_2, f_0) = (\mathbf{x} - \mathbf{H}\alpha)^T (\mathbf{x} - \mathbf{H}\alpha) \tag{58}$$

2. The minimizing solution is found with the same principle as in 3.3 and thus, the minimizing solution is:

$$\hat{\alpha} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \alpha \tag{59}$$

3. Using the following equations:

$$\mathbf{H}^T = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{s}^T \end{bmatrix} \quad and \quad \mathbf{x}^T \mathbf{H} = (\mathbf{H}^T \mathbf{x})^T \tag{60}$$

We have:

$$\mathbf{x}^T \mathbf{H} = \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \quad \mathbf{H}^T \mathbf{H} = \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix} \quad \mathbf{H}^T \mathbf{x} = \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix} \tag{61}$$

This means that maximizing $\mathbf{x}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})\mathbf{H}^T \mathbf{x}$ corresponds to maximizing:

$$\begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix} \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix} \tag{62}$$

4. f_0 cannot be close to 0 or $\frac{1}{2}$ because otherwise the assumption that $\cos(2\pi f_0)$ and $\sin(2\pi f_0)$ approximates 0, does not hold which is the assumption made.

$$J = \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \frac{2}{N} & 0 \\ \frac{2}{N} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix} = \frac{2}{N} [(\mathbf{c}^T \mathbf{x})^2 + (s^T \mathbf{x})^2] \tag{63}$$

Replacing $\mathbf{c}^T \mathbf{x}$ by the sum of $x[n]$ and $\cos(2\pi f_0 \mathrm{n})$ squared and as well for $\mathbf{s}^T \mathbf{x}$ but replace cos by sin and ten transforming them in exponential form we get the periodogram.