Imperial College London

COURSEWORK

IMPERIAL COLLEGE LONDON

ELECTRICAL AND ELECTRONIC ENGINEERING

Advanced Signal Processing

Author: Zhongxuan Li

CID: 01050018

Date: March 29, 2018

Contents

Chapter 1

Random signals and stochastic process

1.1 Statistical estimation

Q1&2. A stochastic process is an infinite collection of random variables, in another word, it is a generalisation of random variables. Here we have a 1000-sample vector **X**, which can be considered as one 1000-sample realisation of a uniform distributed stationary stochastic process $X_n \sim \mathcal{U}(0,1)$. By the definition of uniform distribution, theoretical mean $m = 0.5$ and variance $\sigma^2 = \frac{1}{12}(1-0)^2 = 0.0833$, thus theoretical standard deviation $\sigma = 0.289$.
An analytical proof is given in equation 1.1 and 1.2. By estimation theory, the sample mean is an unbiased estimator of theoretical mean, and the sample variance tends to zero as N increases.

$$E(\hat{M}) = E(\frac{1}{N}\sum_{n=1}^{N} X_n) = \frac{1}{N}\sum_{n=1}^{N} E(X_n) = m \tag{1.1}$$

$$Var(\hat{M}) = var(\frac{1}{N}\sum_{n=1}^{N} X_n) = \frac{1}{N^2}\sum_{n=1}^{N} var(X_n) = \sigma^2/N \tag{1.2}$$

After running the code in MATLAB, we have obtained sample mean = 0.5091, which is within in 2% variance of theoretical mean. Sample standard deviation is 0.2825, which is within 3% variance of the theoretical value we discussed above. The two estimators will converge to theoretical values as data size increases.

Q3. Now we generate an ensemble of ten 1000-sample realisations of **X**. Sample mean and variance of each realisation is recorded in figure 1.1.

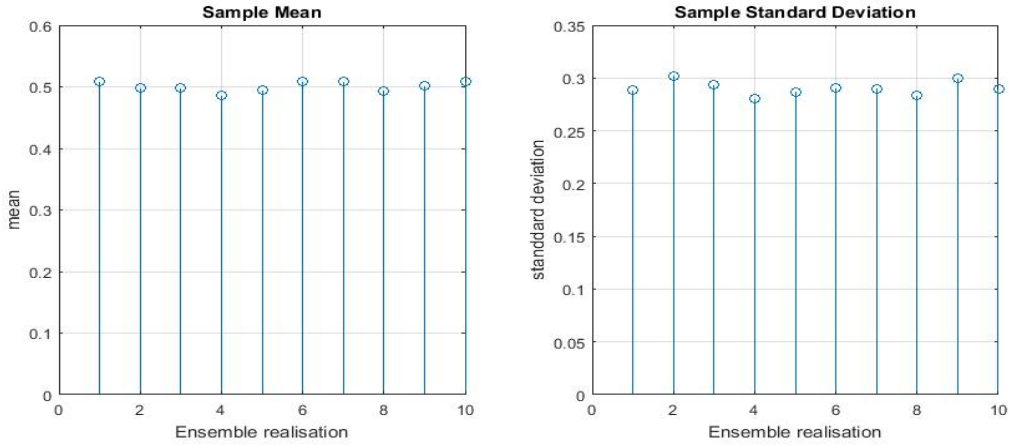

Figure 1.1: Ten ensemble mean and variance of uniform distribution

Sample mean clusters about the theoretical mean with a maximum bias of 0.015, while sample standard deviation clusters about the theoretical value with a maximum bias of 0.014. Therefore, we conclude that both estimators have no bias as they are very clustered and only deviates slightly from their respective theoretical values.

Q4. Plot probability density function of uniform distribution in 1.1.1 and 1.1.2 with reference lines :

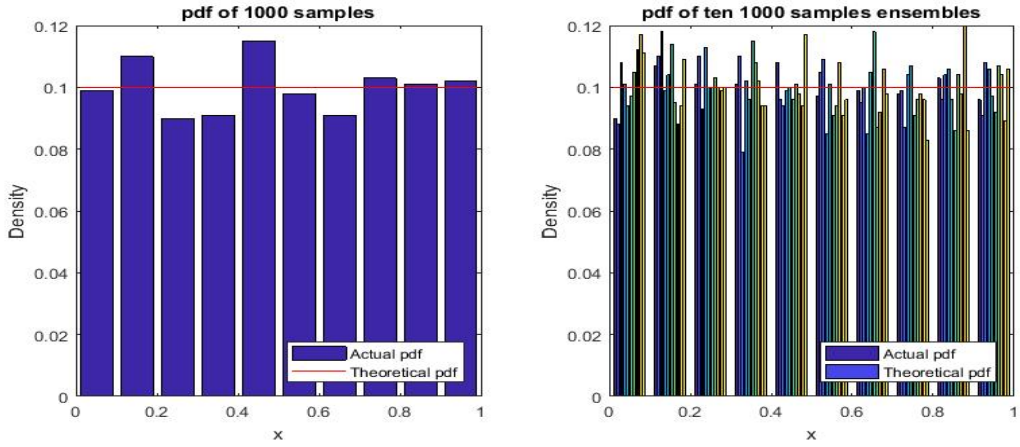

Figure 1.2: Probability density function of uniform distribution

From where we observe pdf plot converges to 0.1 as number of realisation increases. A larger number of bins will provide a more intuitionistic view of how pdf plot converges to theoretical line.

Q5. Now we consider Gaussian distribution $X \sim \mathcal{N}(0,1)$. In single 1000-sample ensemble, $\hat{m} = 0.0300$ and $\hat{\sigma} = 0.9709$. In a ten 100-sample population, the sample mean clusters around 0 and standard deviation clusters around 1 as shown in figure 1.3.

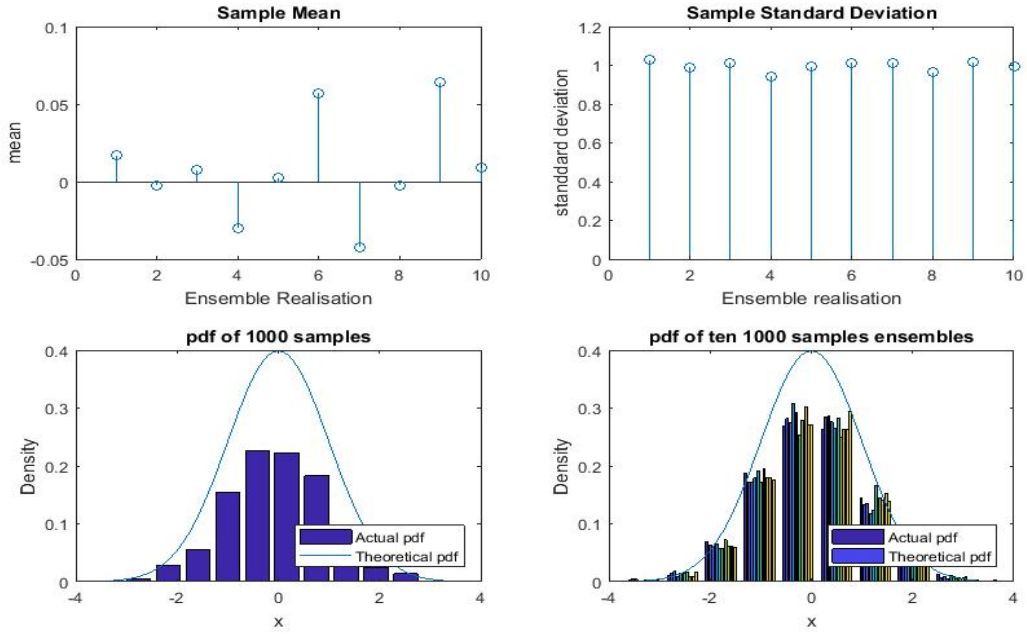

Figure 1.3: Sample mean, standard deviation and pdf of Gaussian distribution

The sample pdf basically follows the shape of theoretical pdf. Because Gaussian process is a stochastic process, such that every finite collection of those random variables has a multivariate normal distribution.

1.2 Stochastic process

A very special class of stochastic processes, called stationary processes, is based on the assumption that the process is in a particular state of statistical equilibrium. A stochastic process is said to be strictly stationary if its properties are unaffected by a change of time origin, that is, if the joint probability distribution associated with m observations $z_{t1}, z_{t2}, ..., z_{tm}$, made at any set of times $t_1, t_2, ..., t_m$, is the same as that associated with m observations $z_{t1+k}, z_{t2+k}, ..., z_{tm+k}$, made at times $t_{1+k}, t_{2+k}, ..., t_{m+k}$. Thus, for a discrete process to be strictly stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward or backward by any integer amount k [1]. Because pdf $p(z)$ is the same over time.
Hence stochastic process has a constant mean:

$$\mu = E(z_t) = \int_{-\infty}^{+\infty} zp(z)dz \tag{1.3}$$

and a constant variance:

$$\sigma^2 = E[(z_t - \mu)^2] = \int_{-\infty}^{+\infty} (z-\mu)^2 p(z)dz \tag{1.4}$$

In addition, the mean μ of the stochastic process can be estimated by the sample mean of the time series, and the variance σ^2 of the stochastic process can be estimated by the sample variance.

Q1. Here we have three stochastic processes and we what to comment on each of their stationarity. We use M = 100 members of the ensemble, each of length N =100; Plot the ensemble mean and standard deviation for each process as a funciton of N (time):

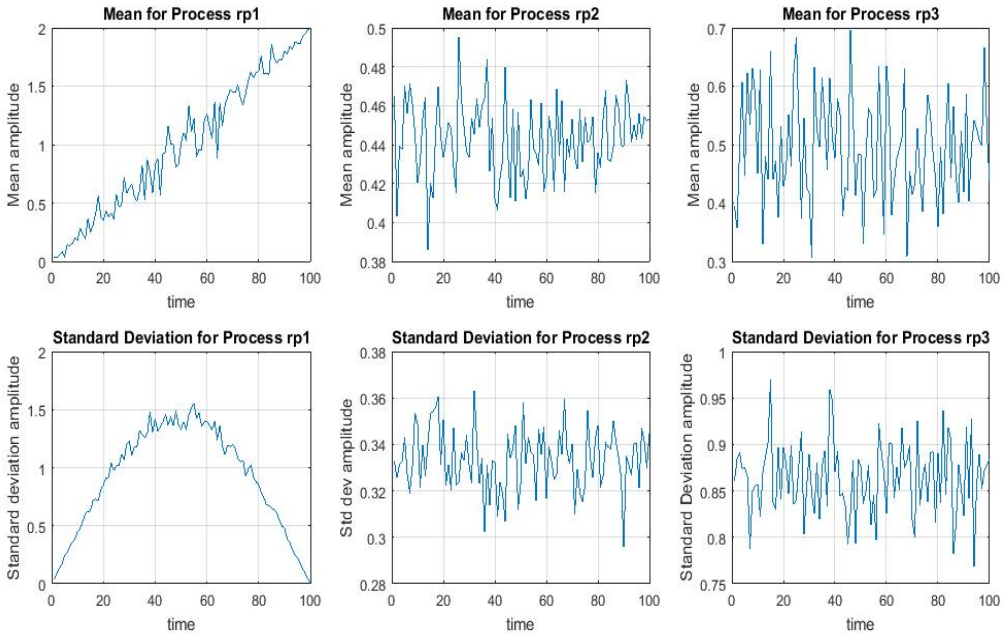

Figure 1.4: Mean and standard deviation for 3 ensembled processes

1. Process rp1 is not a stationary process because its mean and stand deviation change over time.

2. Process rp2 and rp3 are stationary proccesses because we observe their mean and standard deviation fluctuate around a certain level.

The property that the time averages are equal to the ensemble averages is called ergodicity, in particular first-order ergodic. Instead of calculating the ensemble averages, we calculating the time average:

$$\lim_{T\to\infty} E(\mu_x)_T = \mu_x(t) \tag{1.5}$$

In a similar sense a random process X(t) is said to be ergodic in the ACF, i.e, second-order ergodic(wide sense ergodicity) if covariance is equal to the time autocorrelation:

$$\lim_{T\to\infty} E[R_{xx}(\tau)_T] = R_{xx}(\tau) \tag{1.6}$$

Q2. Now let M = 4 and N =1000 for each process, plot the mean and standard deviation for each realisation:

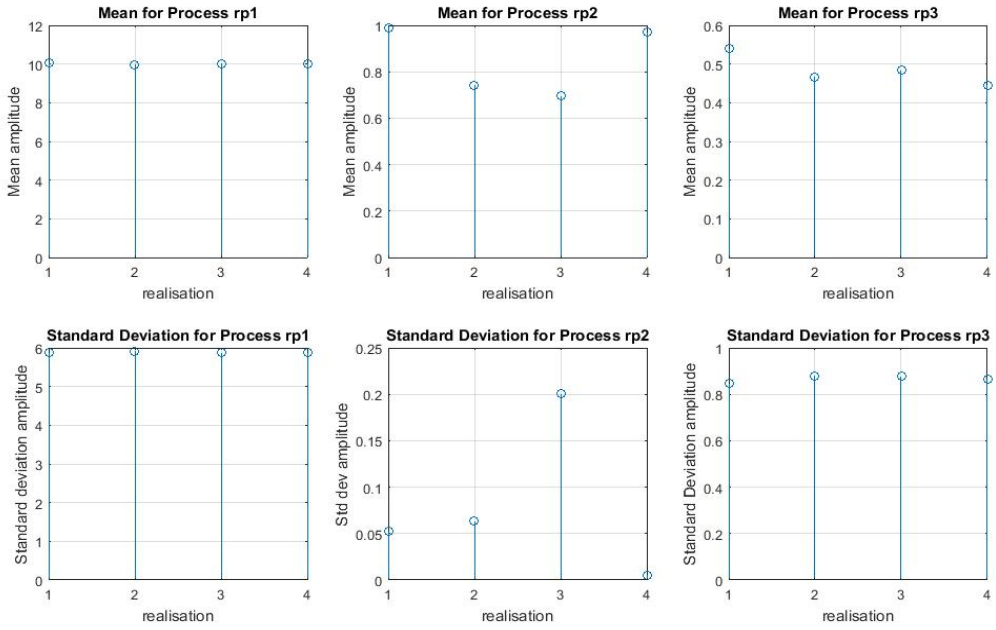

Figure 1.5: Mean and standard deviation for each realisation

1. Process 1 has a time average of 10 ± 0.05 and an increasing ensemble average. Thus process 1 is not an ergodic process.
2. Process 2 has a non-converging time average while a fluctuating ensemble average around 0.45. It can be deducted that process 2 is not a ergodic process either.
3. Process 3 has both time average and ensemble average converge at 0.5. Thus it is ergodic.

Q3. We derive the theoretical mean for each process: First, define a random process X uniformly distributed in rang 0 1:

$$X \sim \mathcal{U}(0,1)$$

1. For Process 1, we only care about Ac. In the case of section two, it is a 4 by 1000 matrix of double valued elements.

$$v = \begin{pmatrix} 0.02 & 0.04 & ... & 0.02 \times N \\ 0.02 & 0.04 & ... & 0.02 \times N \\ 0.02 & 0.04 & ... & 0.02 \times N \\ 0.02 & 0.04 & ... & 0.02 \times N \end{pmatrix}$$

   Calculating the mean need averaging over 4 realisations. Theoretical average should be a linear incrementing function with gradient 0.02 and variance is just a second order polynomial by $Var(X) = \mathbb{E}[(X - \mu)^2]$

   As the Matrix Mc with sine terms, we just regard it as noise with zero mean. So, process1 is intrinsically a linear growing matrix plus a zero mean noise term.

   Furthermore if we increase the number of ensembles(M), The noise averages to zero and we will observe a more linear shaped mean and clear 2nd order polynomial shaped σ.

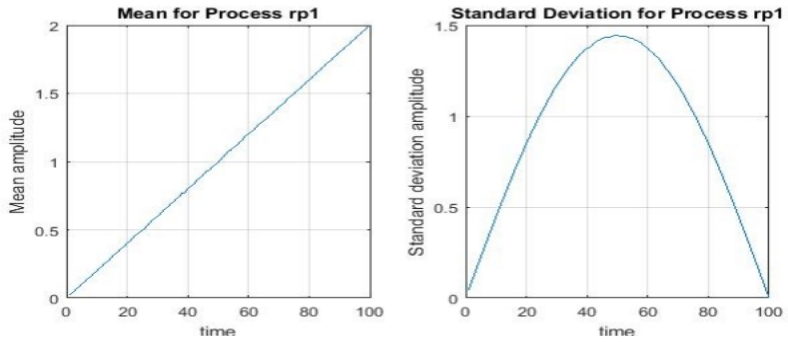

Figure 1.6: Plot process 1 over 1e5 realisations

Theoretical mean and standard deviation are coincident with the plot in 1.3.1.

2. Process 2 is a 4 by 1000 matrix X composed of uniform distributed elements, subtracted 0.5 from each element, then dot product with a 4 by 1000 matrix Mr, at last plus a M by N matrix Ar, which has equal elements in the each individual row.

$$Mr = Ar = \begin{pmatrix} Ar_1 & \dots & Ar_1 \\ Ar_2 & \dots & Ar_2 \\ Ar_3 & \dots & Ar_3 \\ Ar_4 & \dots & Ar_4 \end{pmatrix}$$

where $Ar = X \sim \mathcal{U}(0,1)$

Calculating theoretical mean only need to consider Matrix Ar, because $\mathbb{E}[rand(M,N) - 0.5]$=0. Therefore, theoretical mean of the whole process:

$$E(v) = (0.5 - 0.5) \times 0.5 + 0.5 = 0.5$$

and variance

$$\begin{aligned} \sigma^2 &= E[(X-0.5)^2] \cdot E(X)^2 - E^2(X-0.5) \cdot E^2(Y) + Var(X) \\ &= E(x^2 - x + 0.25) \cdot E(X^2) - E^2(X-0.5) \cdot E^2(Y) + Var(X) \\ &= (E(X^2) - E(X) + 0.25) \cdot \frac{1}{3} + \frac{1}{12} \\ &= (\frac{1}{3} - 0.5 = 0.25) \times \frac{1}{3} + \frac{1}{12} \\ &= \frac{1}{9}, \quad \sigma \approx \frac{1}{3} \end{aligned}$$

Given variance of two independent variable is:

$$Var(XY) = E(X^2) \times E(Y^2) - E^2(X) \times E^2(Y)$$

$$Var(X+Y) = Var(X) + Var(Y)$$

and expectation of uniform distribution squared:

$$E(X^2) = \int_0^1 X^2 dx = \frac{1}{3}$$

Conclusion: The time average does not converge to theoretical average either the standard deviation. We confirmed that process 2 is not ergodic.

3. Process 3 overall is a 4 by 1000 matrix:

$$v = \begin{pmatrix} Vr_{11} - 0.5 & ... & Vr_{1n} - 0.5 \\ Vr_{21} - 0.5 & ... & Vr_{2n} - 0.5 \\ Vr_{31} - 0.5 & ... & Vr_{3n} - 0.5 \\ Vr_{41} - 0.5 & ... & Vr_{4n} - 0.5 \end{pmatrix} \times 3 + 0.5$$

where $Vr = X \sim \mathcal{U}(0,1)$
Obtain theoretical mean by the linearity property of expectation,

$$\mathbb{E}(v) = \mathbb{E}[(X - 0.5) \times 3 + 0.5] = 0.5$$

and variance

$$Var(v) = 1/12 \times 9 = 0.75$$

Hence standard deviation

$$\sigma = \sqrt{0.75} \approx 0.866$$

Conclusion: In process 3, The time average converges to theoretical average, standard deviation also converges to the theoretical standard deviation. Hitherto, We have confirmed that process 3 is the only ergodic process among the three. Overall only process 3 is a stationary ergodic process.

1.3 Estimation of probability distributions

Q1&2. Guassian distribution is known as a stationary ergodic process, thus it has constant pdf over time.

Use the Matlab code to approximate pdf of a Guassian process for $N \subset 100, 1000, 10000$.

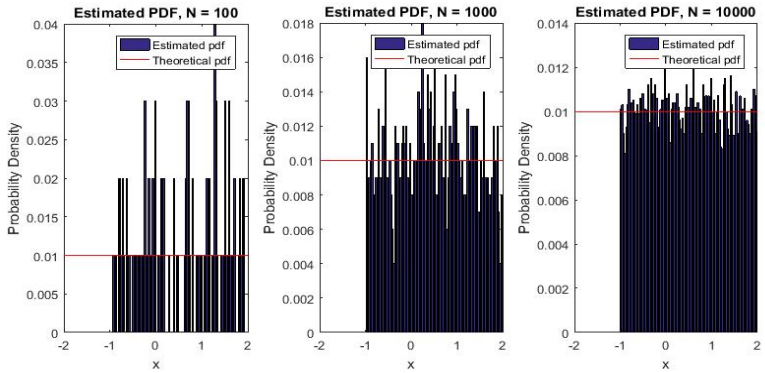

Figure 1.7: Gaussian pdf

Because 100 bins is defined in the *pdf.m* function script, probability of each bin of this discrete Gaussian distribution is 0.01. As number of samples N increase, the pdf plot converges to the 0.01 reference line.

Q3. Probability density function of a non-stationary process changes over time, so it cannot be estimated using the *pdf.m* function directly. Use process 1 to examine the argument:

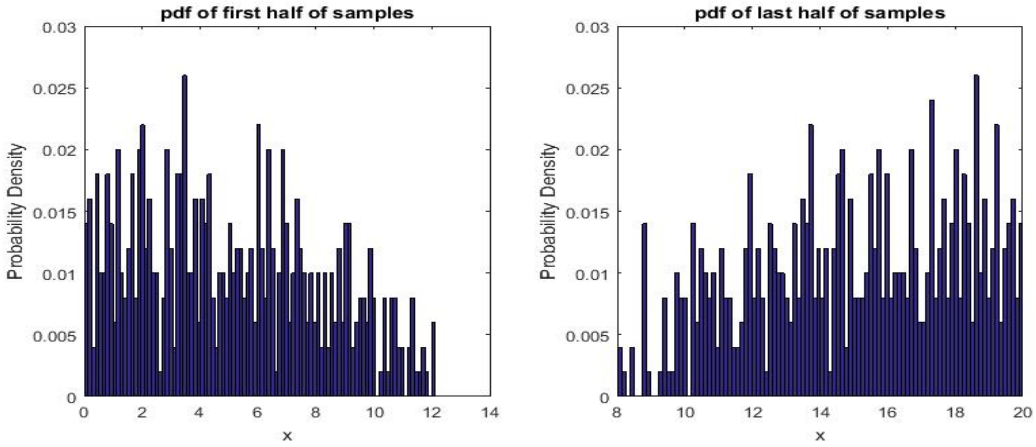

Figure 1.8: Process 1 pdf first and last 500 samples

pdf of first 500 samples does not provide a veracious overview of pdf of last half samples, while both half samples' pdf plot fail to estimate pdf of 1000 samples(figure 1.8) correctly.

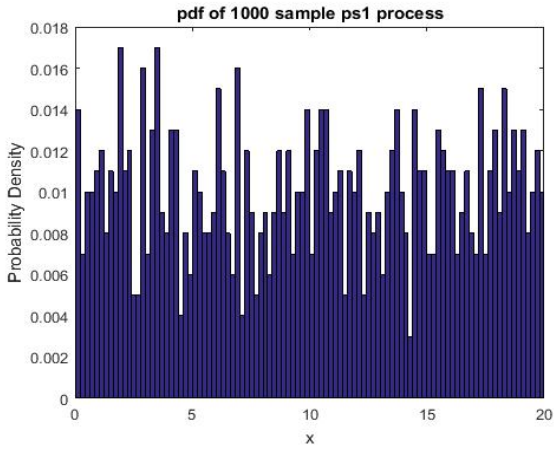

Figure 1.9: Process 1 pdf

Assume the ensemble mean increases linearly from 0 to 1 over first 500 samples, we deduce that differentiating the data samples will give us a flat mean, in another word, a stationary process.
In order to differentiate the data, the difference between consecutive observations is computed. Mathematically, this is shown as

$$y'_t = y_t - y_{t-1} \tag{1.7}$$

Differentiating removes the changes in the level of a time series, eliminating trend and seasonality and consequently stabilizing the pdf of the time series. Now plot pdf of y'_t over first 500 samples and last 500 samples respectively:

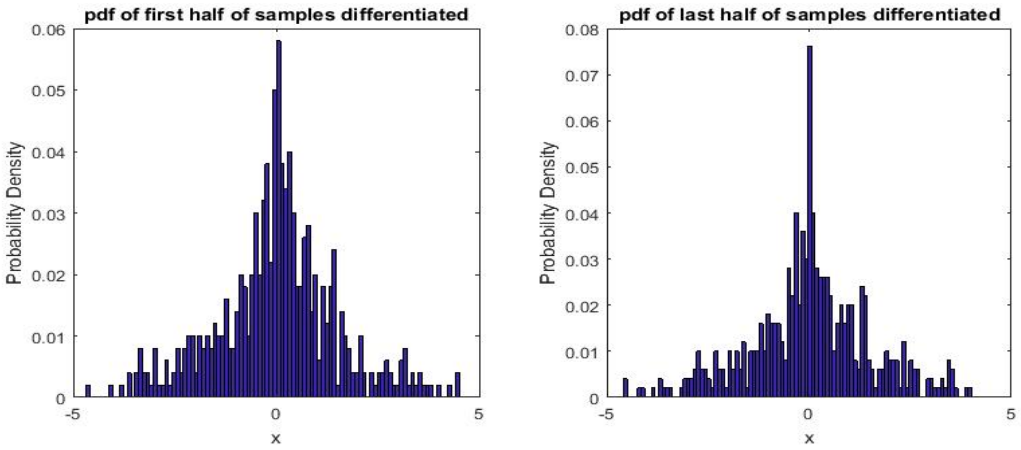

Figure 1.10: Process 1 pdf after differentiation

pdf of first 500 samples gives a true and fair prediction of last 500 samples' pdf. Hence we are able to predict the pdf based on y'_t.

Chapter 2

Linear stochastic modelling

2.1 ACF of uncorrelated sequences

Q1&2. Compute an unbiased estimate of the signal autocorrelation with lag $|\tau| < 999$ and $|\tau| < 50$ respectively.

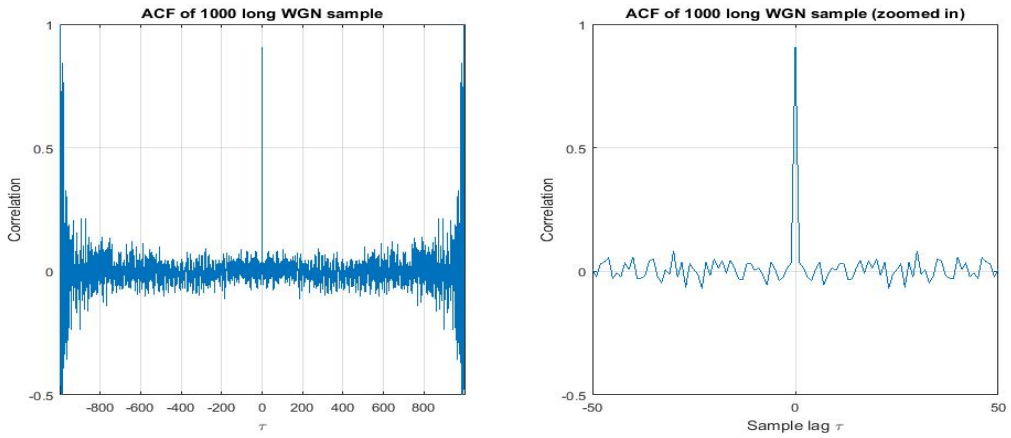

Figure 2.1: Autocorrelation of 1000 samples of WGN

From discrete time view, white noise is a wide sense stationary signal whose samples are regarded as a sequence of serially uncorrelated random variables with zero mean and finite variance [5]. We know that

$$R_{xx}(\tau) = \mathbb{E}[x(t)x(t-\tau)] \tag{2.1}$$

by independence,

$$R_{xx}(\tau) = \begin{cases} \sigma^2, & \tau = 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

At $\tau = 0$ we get the variance which is also the power of x(t). Furthermore, autocorrelation function based on a lags between samples, hence the fact it is an even function can be stated as $R(\tau) = R(-\tau)$. We can therefore deduct that the autocorrelation function of white noise must be a delta function.
We observe a better shaped delta function on the zoomed autocorrelation plot (LHS figure 2.1). Whereas with a larger $|\tau|$, ACF is not strictly zero beyond certain level (RHS plot).

Q3. This is because a smaller $|\tau|$ makes $N-|\tau|$ in equation 2.3 relatively larger and $\widehat{R}_x(\tau)$ has more samples to sum up with. Ergodicity applies and hence $\widehat{R}_x(\tau)$ can converge to zero.

$$\widehat{R}_x(\tau) = \frac{1}{N-|\tau|} \sum_{n=0}^{N-|\tau|-1} x[n]x[n+\tau] \tag{2.3}$$

We conclude that a large autocorrelation lag $|\tau|$ harms the accuracy of ACF estimates according to equation (2.1) and is therefore statistically unreliable. A suggested empirical bound will be $|\tau| < 400$.

Q4. Filter a 1000-sample WGN by moving average filter with unit coefficients of order 9. Since a moving average filter is a low-pass FIR filter, this explains why the curve in autocorrelation is smoothed out.

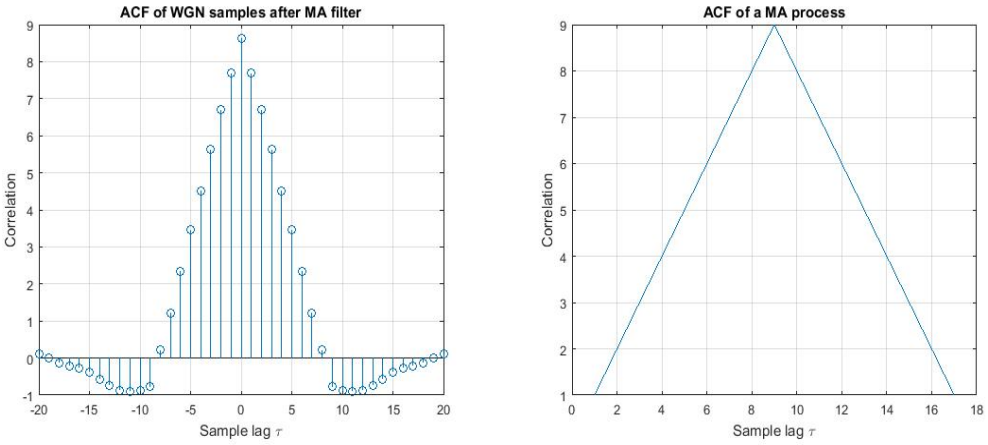

Figure 2.2: Autocorrelation of a MA filter

With higher filter order, the result ACF will be smoothed out more, as the MA filter removes the trend within a wider range. Eventually, $R_y(\tau)$ converges to triangle shaped as shown in RHS figure 2.2.

Q5. The driver sample autocorrelation approaches an impulse scaled by the white-noise variance

$$R_Y(\tau) = R_X(\tau) * R_h(\tau) \tag{2.4}$$

For finite number of observed samples of a filtered white noise process, we may say that the sample autocorrelation of filtered white noise is given by the autocorrelation of the filter's impulse response convolved with the sample autocorrelation of the driving white-noise sequence.
If X_n is an uncorrelated process(e.g. white noise) with autocorrelation function $R_x(\tau) = \delta(\tau)$. By the shifting property of convolution:

$$\begin{aligned} R_Y(\tau) &= R_h(\tau) * \delta(\tau) \\ &= R_h(0) \\ &= \lim_{n\to\infty} \frac{1}{N} \sum_{N=0}^{N-1} h[n]^2 \\ &= \sigma_h^2 \end{aligned}$$

In this case, $R_Y(\tau)$ is variance of process h_n.

2.2 Cross-correlation function

Q1. Plot Cross-correlation function for sequence **x** and **y** generated in Part 2.2. When a white-noise sequence is filtered, successive samples generally become correlated. Due to 9th order MA filter, consecutive 9 samples (i.e.,$x[n], x[n-1], x[n-2]....x[n-8]$) in process **x** are correlated to each other. The plot has 9 peaks from lag -8 to 0 consequently.

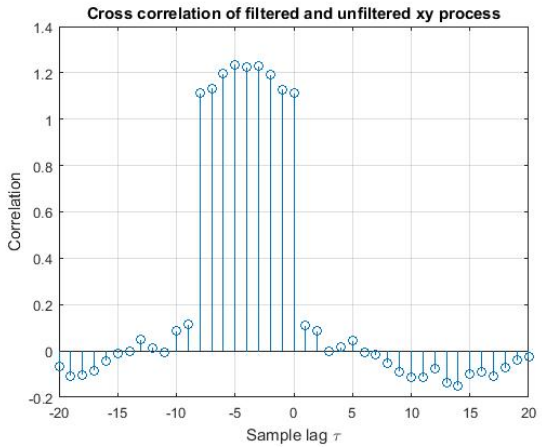

Figure 2.3: Cross-correlation of sequence X and Y

The cross-correlation function R_{XY} between the input and output of a filter is given by

$$R_{XY}(\tau) = h(\tau) * R_X(\tau) \tag{2.5}$$

If X_t is an uncorrelated stochastic process, $R_X(\tau)$ is a Dirac shaped impulse function $\delta(\tau)$,

$$R_{XY}(\tau) = h(\tau) * R_X(\tau) = h(0)$$

which is impulse response at time zero.

Q2. Assume a system having finite impulse response(e.g. MA filter in part 2.2). The estimator of cross-correlation between the input x(t) and output y(t) of the system for an infinite time series, is given by:

$$\widehat{R}_{XY}(\tau) = \sum_{n=0}^{N-1} h[\tau] R_x[t-\tau] with i = 0...N-1 \tag{2.6}$$

Deploying equation(2.3) in am matrix form:

$$\begin{pmatrix} r_x(0) & r_x(-1) & ... & r_x(1-N) \\ r_x(1) & r_x(0) & ... & ... \\ ... & r_x(1) & r_x(0) & r_x(-1) \\ r_x(N-1) & ... & r_x(1) & r_x(0) \end{pmatrix} \cdot \begin{pmatrix} h(0) \\ h(1) \\ \cdots \\ h(N-1) \end{pmatrix} = \begin{pmatrix} r_{XY}(0) \\ r_{XY}(1) \\ \cdots \\ r_{XY}(N-1) \end{pmatrix}$$

Thus, the impulse response can be calculated as:

$$h = R_X^{-1} \cdot R_{XY} \tag{2.7}$$

Equation 2.7 has the same form as optimum Wener filter solution. The key in system identification is how to choose the probing signal $X[n]$. Assume $X[n]$ is an uncorrelated sequence,that is $r_x[k] = 0$ for $k \neq 0$. Hence $[R_x]$ is a diagonal matrix $r_x(0) \cdot \mathbf{I}$.
And the impulse response is calculated by

$$h(k) = \frac{r_{xy}(k)}{r_x(0)} \tag{2.8}$$

This result is coincident with the MVU estimator of a FIR system, which is "the MVU estimator is the ration of the input-output cross-correlation to the input autocorrelation" [2]. This is obvious because the MA filter we deal with is in fact a FIR filter. With higher filter order, cross-correlation function will become more flat and will have more pulses cluster above 1.

2.3 Autoregressive modelling

Q1. Plot (a_1, a_2) which preserves the stability of this AR(2) process:

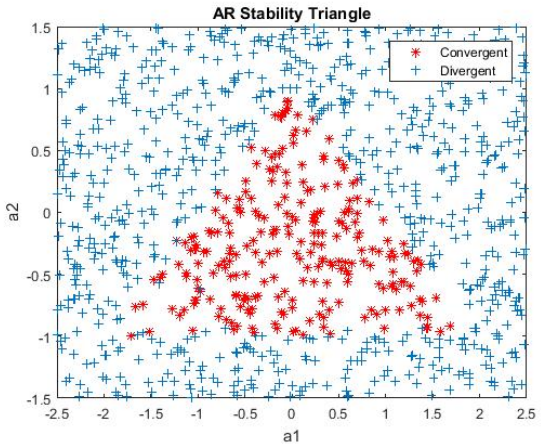

Figure 2.4: AR stability triangle

A theoretical justification as follows, an AR(2) process has the form

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + w[n] \tag{2.9}$$

Perform Laplace transform

$$X(z) = (a_1 z^{-}1 + a_2 z^{-2})X(z) + W(z) \tag{2.10}$$

Transfer function

$$H(z) = \frac{X(z)}{W(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} \tag{2.11}$$

For the system to be stable, all poles must lie outside the unit circle, that is roots of polynomial $1 - a_1 z^{-1} - a_2 z^{-2} = 0$ lies **inside** the unit circle, which is equivalent to roots of $z^2 - a_1 z - a_2 = 0$ lie **outside** the unit circle. Hence we have,

$$-1 < \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2} < 1 \tag{2.12}$$

implies that,

$$a_1 + a_2 < 1; \quad a_2 - a_1 < 1$$

Moreover, if z is complex, then

$$|\frac{a_1}{2} \pm i\frac{\sqrt{a_1^2 + 4a_2}}{2}| = \frac{a_1^2}{4} - \frac{a_1^2 + 4a_2}{4} = -a_2 \leq 1 \tag{2.13}$$

$$-1 < a_2 < 1$$

Hence stability requires a triangle shaped restriction on a_1 and a_2 as shown in figure 2.4.

Q2. Plot sunspot series autocorrelation function for N = 5, 20 ,250.

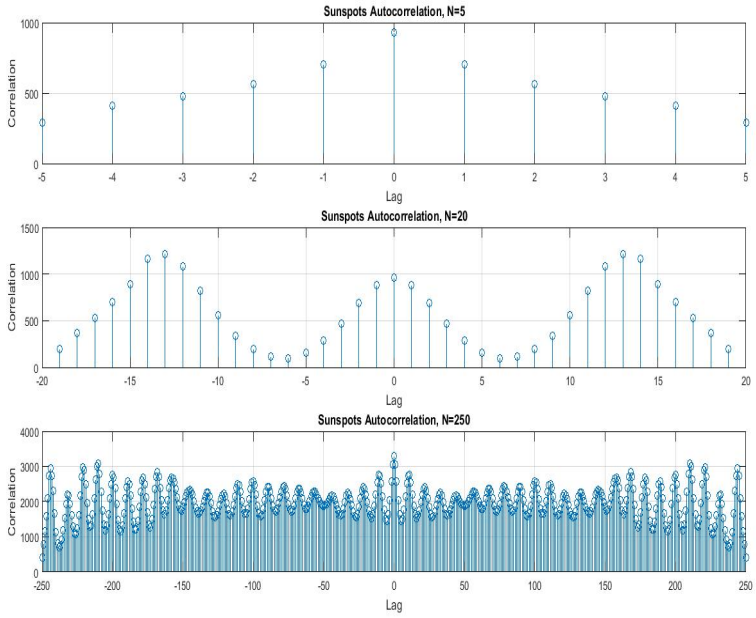

Figure 2.5: Sunspot ACF for N = 5, 20, 250

As N gets larger, a clear pseudo-periodic trend in the autocorrelation function is observed. Generally, ACF at N=5 is a condensed version of ACF at N=250. On the ACF plot with zero mean (figure 2.6), a even clear pseudo-periodic trend is observed.

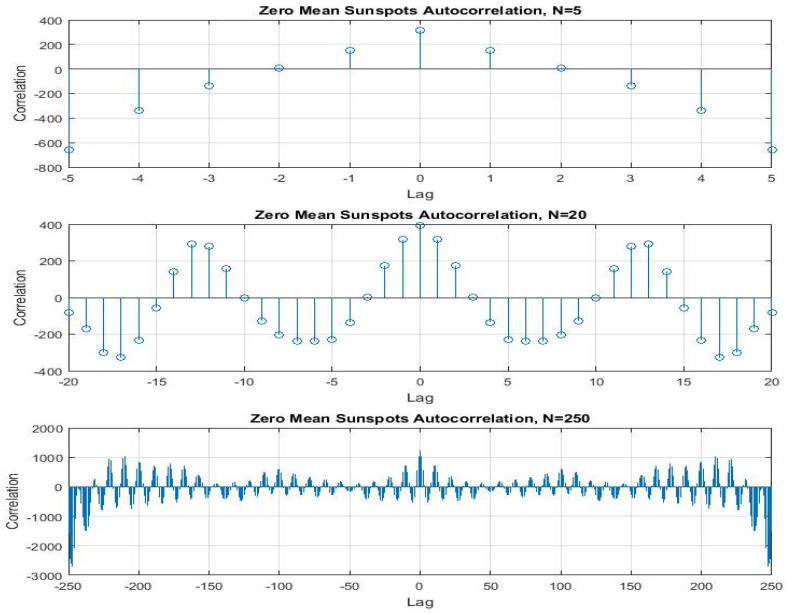

Figure 2.6: Zero mean sunspot ACF for N = 5, 20, 250

The non-zero mean of the series acts as an offset in the autocorrelation function, The larger N is, to larger extent is offset effect is, making the ACF trend invisible. Therefore, normalization on the data set is essential for correct ACF plot, as we do not want modeling the DC offset deterministic component bu the stochastic component(AR process driven by white noise). Wold's theorem states that any purely non-deterministic covariance0-stationary process can be well approximated by an ARMA process. Our focus will be modeling the stochastic component of the signal rather than the deterministic part.

Q3. Find partial correlations up to model order 10.

No.	1	2	3	4	5	6	7	8	9	10
$AR(1)$	-0.9295									
$AR(2)$	-1.4740	0.5857								
$AR(3)$	-1.5492	0.7750	-0.1284							
$AR(4)$	-1.5167	0.5788	0.2638	-0.2532						
$AR(5)$	-1.4773	0.5377	0.1739	-0.0174	-0.1555					
$AR(6)$	-1.4373	0.5422	0.1291	-0.1558	0.2248	-0.2574				
$AR(7)$	-1.3669	0.4807	0.1718	-0.1912	0.0764	0.1359	-0.2726			
$AR(8)$	-1.3016	0.4483	0.1535	-0.1456	0.0355	0.0212	0.0523	-0.2384		
$AR(9)$	-1.2615	0.4395	0.1500	-0.1515	-0.0599	0.0046	-0.0230	-0.0197	0.1680	
$AR(10)$	-1.2573	0.4400	0.1506	-0.1514	0.0584	-0.0007	-0.0268	-0.0308	-0.1362	-0.0252

Table 2.1: AR coefficients up to model order 10

After lag 2, the partial correlation coefficients becomes significantly small, therefore the most likely model order of the sunspot series would be AR(2).
Find partial correlations of normalised data up to model oder 10.

No.	1	2	3	4	5	6	7	8	9	10
$AR(1)$	-0.8212									
$AR(2)$	-1.3783	0.6783								
$AR(3)$	-1.2953	0.5097	0.1223							
$AR(4)$	-1.3011	0.4856	0.1836	-0.0473						
$AR(5)$	-1.3018	0.4885	0.1912	-0.0676	-0.0156					
$AR(6)$	-1.3044	0.4994	0.1602	-0.1469	0.2269	-0.1623				
$AR(7)$	-1.2760	0.4597	0.1859	-0.1749	0.1394	0.0661	-0.1751			
$AR(8)$	-1.2361	0.4447	0.1541	-0.1351	0.0971	-0.0385	0.1154	-0.2276		
$AR(9)$	-1.1959	0.4243	0.1609	-0.1523	0.1210	-0.0658	0.0368	-0.0093	-0.1766	
$AR(10)$	-1.1952	0.4243	0.1608	-0.1520	0.1205	-0.0652	0.0362	-0.0109	-0.1721	-0.0038

Table 2.2: AR coefficients up to model order 10 normalised

After the data is normalised, partial ACF a_{kk}converges faster within the confidence level. The difference between the first two coefficients and the others is exaggerated. As Wold's theorem implies, any covariance-stationary time series can be decomposed into deterministic part and stochastic part. And we aim on modeling the stochastic part, therefore, normalisation is required as we do not want modeling the DC offset deterministic component bu the stochastic component(AR process driven by white noise).

Q4. Plot MLD and AIC for different AR(p) model.

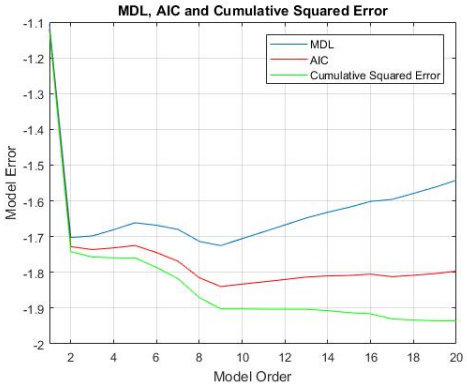

Figure 2.7: Penalty term and error

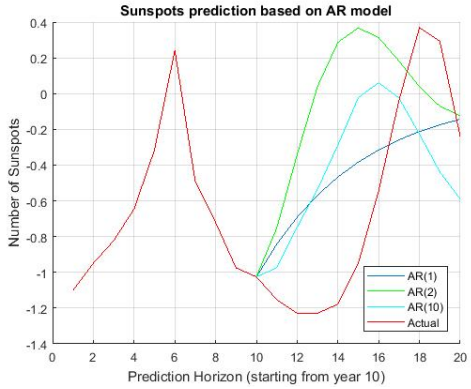

Figure 2.8: Sunspot prediction

MDL(minimum description length) and AIC(Akaike information criterion) is composed of a loss function($log(E_p)$) correlates to model accuracy and a penalty term($\frac{plogN}{N}$ or $\frac{2p}{N}$) regards of computational complexity.
From figure 2.7 the prediction error increases after model order 2. Correct model order for the standardised data is therefore order 2.

On figure 2.8, use data prior than year 10 to predict sunspots numbers using different AR models and compared them with actual data. It is obvious that AR(2) gives an most accurate estimation of original data as well as low computational complexity. Although AR(10) gives a moderately close estimation, a model order higher than the optimal one tries to fit some residual deterministic part in the signal rather than stochastic part we want to model. This may result in a relative lower model error(interpolation) but will not give a reasonable prediction hence high prediction error(extrapolation).

2.4 Cramer-Rao Lower Bound

Q(a). Find partial correlations of AR(10) with confidence intervals: $\pm 1.96\sqrt{288}$. Also plot AIC and MDL for the NASDAQ series:

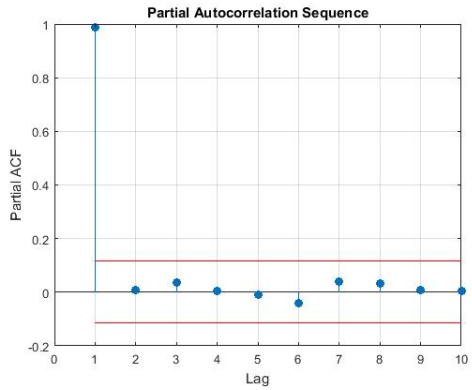

Figure 2.9: Partial ACF for NASDAQ data]

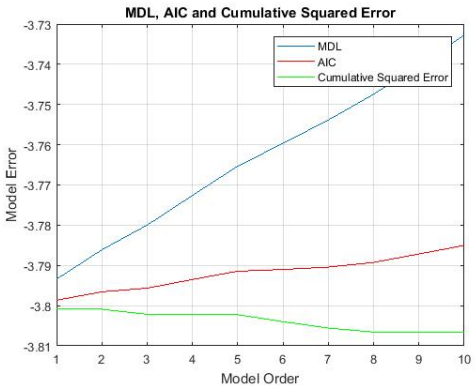

Figure 2.10: penalty terms for NASDAQ data

Partial autocorrelation sequence drops within in the confidence level after lag 1. In AIC and MDL criterion, curve monotonically increases since moder order 1. We conclude that AR(1) model is sufficient to describe daily returns of the index. This conclusion makes sense because, stock price is greatly related to yesterday's closing price while even remote price seems irrelevant.

Q(b). Given fisher information matrix of AR(1) model:

$$[\mathbf{I}(\theta)]_{ij} = \frac{N}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln[\hat{P}_X(f;\theta)]}{\partial \theta_i}\frac{\partial \ln[\hat{P}_X(f;\theta)]}{\partial \theta_j} df \tag{2.14}$$

where $\hat{P}_X(f;\theta) = \frac{\hat{\sigma}^2}{|1-\hat{a}e^{-j2\pi f}|^2}$ and the parameter vector $\theta = [a_1 \quad \sigma^2]^T$.
Partial derivative with respect to θ_2:

$$\frac{\partial \ln[\hat{P}_X(f;\theta)]}{\partial \theta_2} = \frac{\partial \ln[\hat{\sigma}^2]}{\partial \hat{\sigma}^2} = \frac{1}{\sigma^2} \tag{2.15}$$

Therefore

$$[\mathbf{I}(\theta)]_{22} = \frac{N}{2} \times \frac{1}{\sigma^4}\int_{-\frac{1}{2}}^{\frac{1}{2}} df = \frac{N}{2\sigma^4} \tag{2.16}$$

Similarly

$$[\mathbf{I}(\theta)]_{11} = \frac{N r_{xx}(0)}{\sigma^2} \quad [\mathbf{I}(\theta)]_{12} = [\mathbf{I}(\theta)]_{21} = 0 \tag{2.17}$$

Q(c). CRLB theorem states that the minimum variance unbiased (MVU) estimator $\hat{\theta}$ has variance

$$var(\hat{\theta}) \geq \frac{1}{\mathbf{I}(\theta)} \tag{2.18}$$

Also note AR(1) model has sequence variance

$$\sigma_x^2 = r_{xx}(0) = \frac{\sigma^2}{1-a_1^2}$$

Hence result from the above equations

$$var(\hat{\sigma}^2) \geq [\mathbf{I}^{-1}(\theta)]_{22} = \frac{2\sigma^4}{N} \tag{2.19}$$

$$var(\hat{a}) \geq [\mathbf{I}^{-1}(\theta)]_{11} = \frac{\sigma^2}{N r_{xx}(0)} = \frac{\sigma^2(1-a_1^2)}{N\sigma^2} = \frac{(1-a_1^2)}{N} \tag{2.20}$$

Plot CRLB for both $\hat{\sigma}^2$ and $\hat{a}$

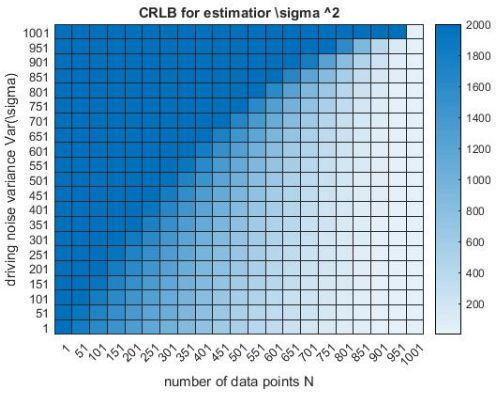

Figure 2.11: CRLB for $\hat{\sigma}^2$ in [1:50:1001]

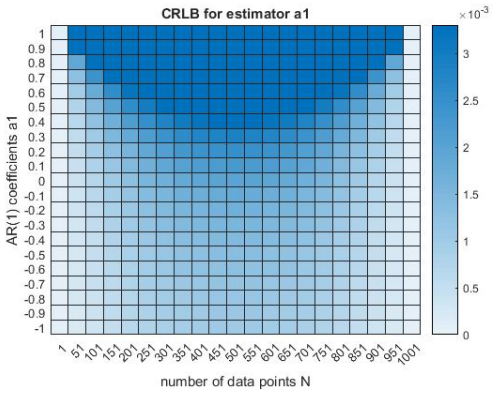

Figure 2.12: CRLB for $\hat{a}$ in [-1:0.1:1]

A physical justification is given such that the CRLB for $\hat{\sigma}^2$ is lowest when N is at maximum whereas σ is at minimum, which is the lower-right corner on the heatmap. As for the CRLB for $\hat{a}$, who is lowest when a_1^2 is 1, thus minima happen on four corners on the heatmap.

Q(d). Power spectrum density estimation of the signal is given by

$$\hat{P}_x(f;\theta) = \frac{\hat{\sigma}^2}{|1-\hat{a}_1 e^{-j2\pi f}|^2} = \frac{\hat{\sigma}^2}{A^2} \tag{2.21}$$

Partial differentiation with respect to parameter vector $\theta = [a_1 \quad \sigma^2]^T$ is

$$\frac{\partial \hat{P}_x(f;\theta)}{\partial \theta} = [\frac{\partial \hat{P}_x(f;\theta)}{\partial a_1}, \frac{\partial \hat{P}_x(f;\theta)}{\partial \sigma^2}]^T = [\frac{-2\sigma^2(A-1)}{A^3}, \frac{1}{A^2}]^T \tag{2.22}$$

given that derivative of $|A|$ is $\frac{A}{|A|}$, solved for

$$var(\hat{P}_x(f;\theta)) \geq \frac{\partial \hat{P}_x(f;\theta)}{\partial \theta}^T \mathbf{I}^{-1}(\theta) \frac{\partial \hat{P}_x(f;\theta)}{\partial \theta} = \frac{4\sigma^4(A-1)^2(1-a^2)+2\sigma^4A^2}{NA^6} \tag{2.23}$$

2.5 Real world signals: ECG

Q(a&b). To obtain a smoother estimate of the heart rate, average over every 10 samples

$$\hat{h} = \frac{1}{10}\sum_{i=1}^{10} \alpha h[i] \tag{2.24}$$

where α is a scalar

$$E[\hat{h}] = \frac{1}{10}\sum_{i=1}^{10} E[\hat{h}_i[n]] = \frac{\alpha}{10}\sum_{i=1}^{10} h = \alpha h \tag{2.25}$$

For α=1, the estimator is unbiased, for α=0.6, it will be biased. Plot the averaged estimates of heart rate as follows.

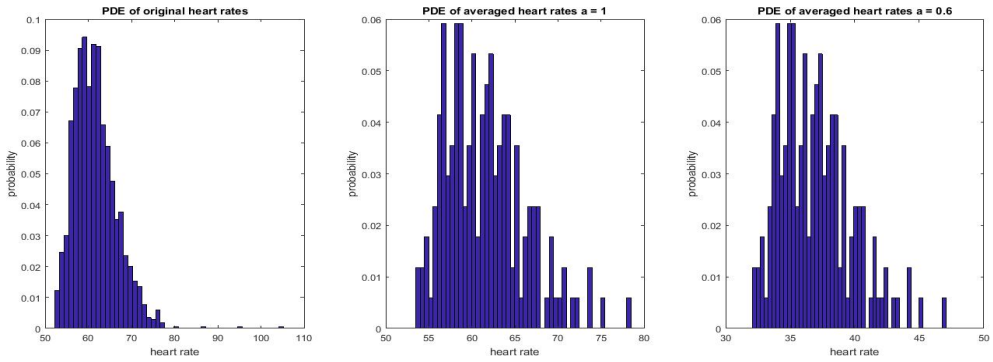

Figure 2.13: PDE of original and smoothed heart rates from trail 1

PDE of averaged heart rates have same trend but smoothed shape as the PDE of original heart rates. When α is 1, the estimator is unbiased and hence the PDE is centred around 60 beats/sec, same as the original sequence. When α is 0.6, we have a biased estimator who is centred around 38 beats/sec, deviating from the original sequence.

Q(c). Plot the autocorrelation sequence for the RRI data for three trails.

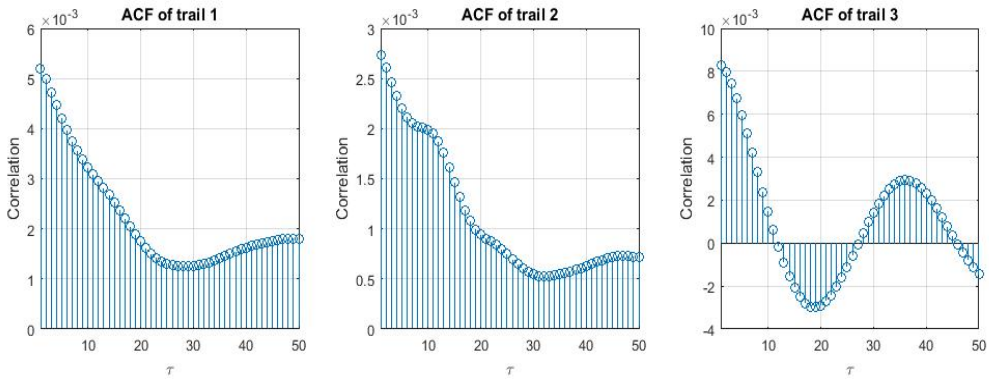

Figure 2.14: Autocorrelation of sequence trail 1,2,3

From the shape of the autocorrelation sequence, it can be deducted that the RRI data is an AR process. Because the autocorrelation sequence decreases to zero gradually while for a MA process, autocorrelation function drop to zero after a certain lag [4].

Plot partial autocorrelation function MDL and AIC, of trail 1,2,3

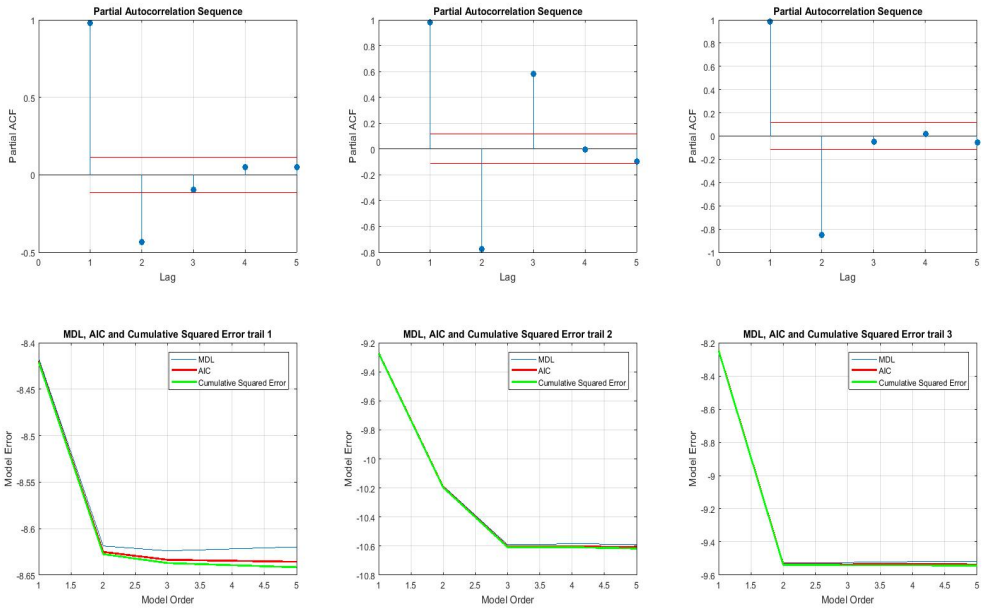

Figure 2.15: ACF,MDL,AIC for sequence trail 1,2,3

Q(d). An optimal model would be AR(3). Since partial correlation coefficients stays within confidence interval after lag 3 in trail 2. Although trail 1 and 3 can be modelled using AR(2), its penalty terms stay constant after model order 2. Thus, to achieve consistency, we choose AR(3) to model the complete RRI sequence.

Chapter 3

Spectral estimation and modelling

3.1 Average periodogram estimates

Estimate PSD of WGN with sample length 128, 256, 512 based on fast Fourier transform(FFT)

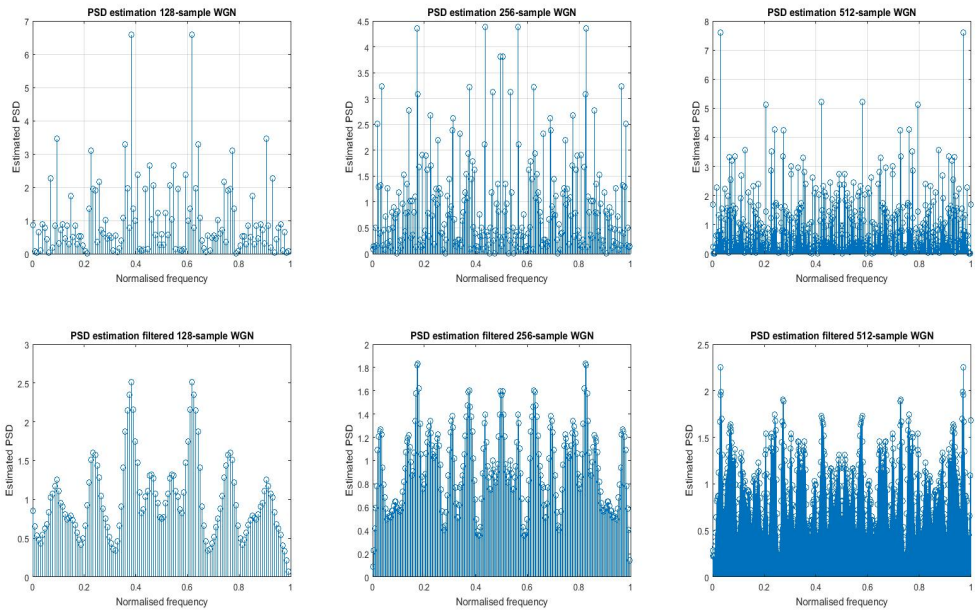

Figure 3.1: Periodogram of WGN with sample length 128, 256, 512

Q1&2. Estimated PSD is symmetrical with respect to the normalised frequency zero. Furthermore, estimated PSD deviates from the theoretical constant PSD of 1, due to finite sample numbers and finite frequency ranges. A filter can be used to smoothen out the periodogram. Zero-phase FIR filter removes the high frequency components of the periodogram and consequently leads to a more constant estimated PSD which converges more towards unity.

Q3. Because Gaussian process is stochastic ergodic, we treat estimated PSD at each frequency a random variable. Using the average over each realisation in order to create a better PSD estimator called the average periodogram, as shown in figure 3.2.

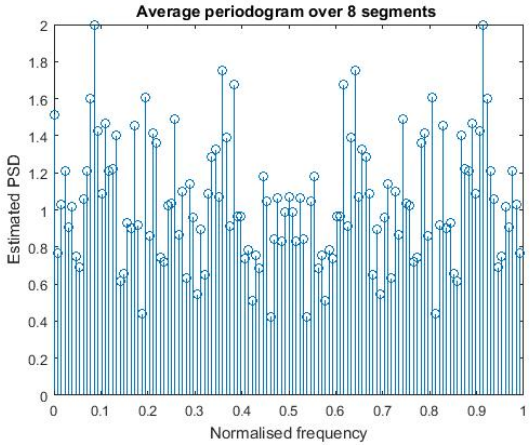

Figure 3.2: PSD estimation over eight segments

The averaged periodogram offers significant improvement from the original 128-sample segment of WGN.
Sample mean estimator?

3.2 Spectrum of autoregressive process

Taking the Fourier transform of autocorrelation estimate results in an estimate of the power spectrum, known as the Periodogram. This can also be directly expressed in terms of the data $x(n)$ using the rectangular windowed function $x_N(n)$.
As the autocorrelation function is defined as

$$\hat{r}_x(k) = \frac{1}{N}\sum_{n=-\infty}^{n=+\infty} x_N(n+k)x_N^*(k) = \frac{1}{N}x_N(k) * X_N^*(-k) \tag{3.1}$$

where

$$x_N(n) = \begin{cases} x(n) & \geq n \leq N \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

which is in fact a rectangular window.

First plot autocorrelation function of AR(1) process

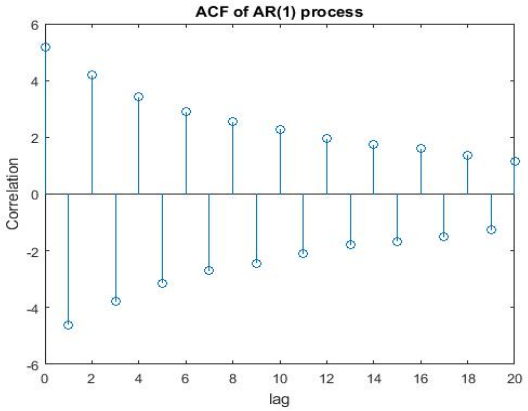

Figure 3.3: Autocorrelation function of AR(1) with a=[1,0.9]

It can be deducted that this AR(1) process acts as a high-pass filter. This is reflected in the autocorrelation function(figure 3.3), which alternates in sign as it decays to zero.

Then plot theoretical PSD and estimated periodogram

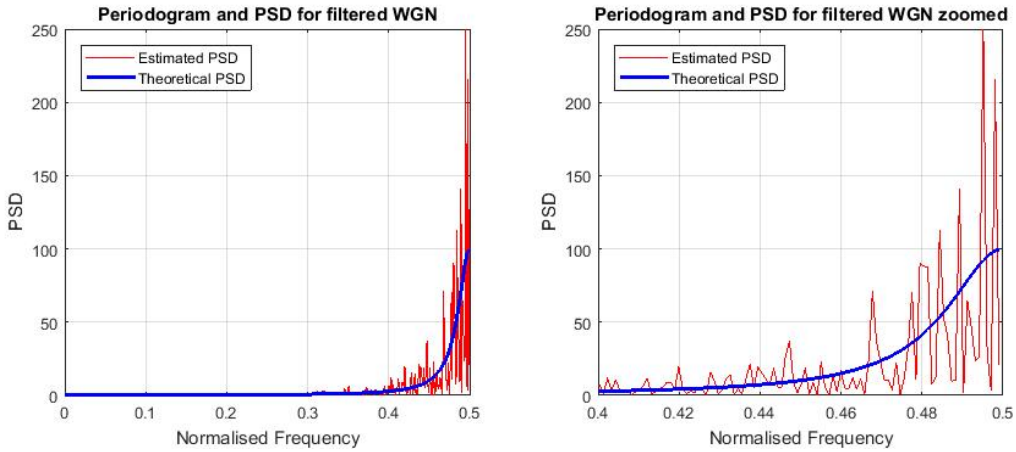

Figure 3.4: Periodogram(red) and PSD(blue) of AR(1)

Q1&2. The blue curve is the exact PSD who has a clear high pass behaviour with a cut-off frequency at approximately 0.475Hz.On the same plot, in red is the periodogram of the filtered WGN. We can observe that the periodogram and the theoretical PSD share a common trend but the periodogram tends to oscillate rapidly and have higher spikes near the cut-off frequency due to the high-pass behaviour of AR(1) process.

Q3. Figure on the left is the zoomed PSD plot on the interval $f = [0.4, 0.5]$. The oscillations boost after the cut-off frequency. These inaccuracies can be explained by the fact of using a rectangular window within the periodogram estimator, which has significant side lobes in the frequency domain. This leads to additional oscillations when convolute with WGN in the frequency domain. A solution to this issue would be to apply a Hamming window to the periodogram estimator to reduce the side lobes and subsequently the effect of oscillations.

Q4. Plot model based PSD, theoretical PSD and estimated periodogram on one figure.

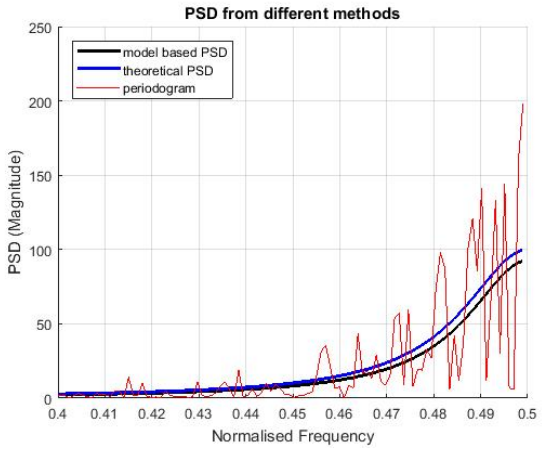

Figure 3.5: Periodogram(red), PSD(blue), and model based PSD(balck)

Compared to the periodogram estimate (red), the model based (black) is much more accurate, being nearly identical to the theoretical PSD.

Q5. Compare estimated periodogram and model based PSD(obtained from 'pyulear' function in matlab) for AR model 2, 5, 10 each with zero-mean and original sunspot series.

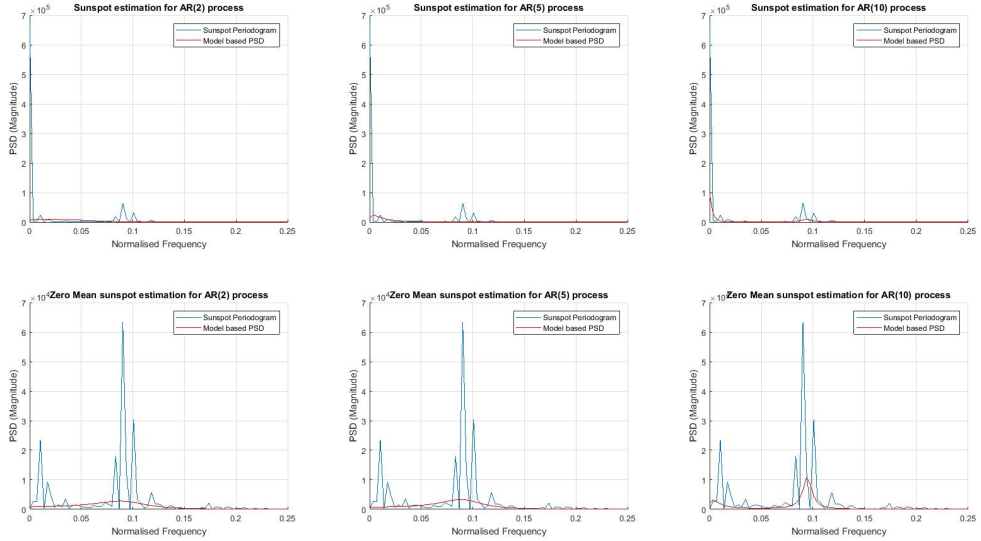

Figure 3.6: Periodogram(red) and model based PSD(blue) for sunspot

As convinced previously, the most optimal model order for sunspot series is AR(2). Low order model does not have enough freedom in the spectra since there can only be one peak, therefore the accuracy is poor. Over modeling causes too many peaks which may or may not be accurately portraying the system.
In order to compare model accuracy, calculate Mean Square Error of the PSD estimator

```
%Error function for periodogram
error = abs(normped(1:145)- pxx);
error = error.^2;
MSE(i) = sum(error)/145;
```

$MSE = [4.10 \times 10^7, 4.27 \times 10^7, 9.80 \times 10^7]$ for model order 2,5,10 respectively. Therefore, overmodelling will harm the accuracy of PSD estimation.

3.3 The Least Squares Estimation of AR Coefficients

Q1. Known that the least square cost function of finding the unknown AR coefficients $a = [a_1, a_2, ..., a_p]^T$ is given by

$$J = \sum_{k=1}^{M} \left[\hat{r}_{xx}[k] - \sum_{i=1}^{P} a_i \hat{r}_{xx}[k-i] \right]^2, \quad \text{for} \quad M \geq p \tag{3.3}$$

It is straightforward that if let $\mathbf{z} = \begin{pmatrix} \hat{r}_{xx}[1] \\ \hat{r}_{xx}[2] \\ \cdots \\ \hat{r}_{xx}[M] \end{pmatrix}$; $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_p \end{pmatrix}$ s.t. $p \leq M$;

$\mathbf{H} = \begin{pmatrix} \hat{r}_{xx}[1] & \hat{r}_{xx}[0] & \cdots & \hat{r}_{xx}[1-p] \\ \hat{r}_{xx}[2] & \hat{r}_{xx}[1] & \cdots & \hat{r}_{xx}[2-p] \\ \vdots & \vdots & & \vdots \\ \hat{r}_{xx}[M-1] & \cdots & & \hat{r}_{xx}[M-p] \end{pmatrix}$ s.t. $k \leq M$ $i \leq p$.

Then the cost function has a general form $\mathbf{J} = (\mathbf{z}-\mathbf{Ha})^T(\mathbf{z}-\mathbf{Ha})$. Differentiate the cost function with respect to $\mathbf{a}$

$$\frac{\partial \mathbf{J}}{\partial \mathbf{a}} = \mathbf{H}^T(\mathbf{z} - \mathbf{Ha}) = 0 \tag{3.4}$$

Least square estimates for AR coefficients

$$\mathbf{a} = (\mathbf{H^T H})^{-1}\mathbf{H}^T\mathbf{z} \tag{3.5}$$

Yule-Walker equations of AR(p) process

$$r_{xx}(p) = a_1 r_{xx}(p-1) + a_2 r_{xx}(p-2) + ... + a_p r_{xx}(0) \tag{3.6}$$

$$r_{xx}(p) = \sum_{i=1}^{p} a_i r_{xx}[k-i] \tag{3.7}$$

has the matrix form

$$\mathbf{r}_{xx} = \mathbf{R}_{xx}\mathbf{a} \tag{3.8}$$

where $\mathbf{a} = [a_1, ...a_p]^T$, Yule-Walker estimates for $\mathbf{a}$ is

$$\mathbf{a} = \mathbf{R}_{xx}^{-1}\mathbf{r}_{xx} = \mathbf{R}_{xx}^{-1}\mathbf{z} \tag{3.9}$$

It is intuitive that equation 3.9 and 3.5 has similar form in case of $\mathbf{R}_{xx}^{-1} = (\mathbf{H^T H})^{-1}\mathbf{H}^T$. The least squares methods can be interpreted as approximate solutions to the Yule-Walker equations by recognizing that $(\mathbf{H^T H})^{-1}\mathbf{H}^T$ is finite-sample estimates of $\mathbf{R}_{xx}^{-1}$.

Q2. The observation matrix $\mathbf{H} = \hat{r}_{xx}[k-i]$ is a deterministic matrix, by the fact that the value of $\hat{r}_{xx}$ can be determined by parameters' value and initial inputs.

Q3. Apply LSE approach calculating $\mathbf{a}$ for AR(p) model up to order 10 associated with centred mean sunspot data.

No.	1	2	3	4	5	6	7	8	9	10	MSE
$AR(1)$	-0.8219										496.8
$AR(2)$	-1.3864	0.6853									263.3
$AR(3)$	-1.3076	0.5272	0.1145								259.5
$AR(4)$	-1.3126	0.4989	0.1833	-0.0515							258.6
$AR(5)$	-1.3135	0.5020	0.1918	-0.0740	0.0171						258.5
$AR(6)$	-1.3172	0.5140	0.1613	-0.1574	0.2342	-0.1665					251.2
$AR(7)$	-1.2876	0.4702	0.1972	-0.1918	0.1211	0.1194	-0.2123				234.8
$AR(8)$	-1.2412	0.4468	0.1669	-0.1492	0.0798	0.0101	0.0841	-0.2331			222.1
$AR(9)$	-1.1936	0.4310	0.1628	-0.1660	0.1127	-0.0243	-0.0067	0.0211	-0.2085		212.6
$AR(10)$	-1.1883	0.4298	0.1654	-0.1683	0.1110	-0.0191	-0.0112	0.0105	-0.1787	-0.0249	212.4

Table 3.1: LSE of AR coefficients up to model order 10 and error

Q4. Model order 10 has lowest mean square error compared with the original data. However too many coefficients dramatically rise up system redundancy. As AR(2) has close performance as higher order models but lower computation complexity, we choose AR(2) as the optimal model, similar as section 2.3.

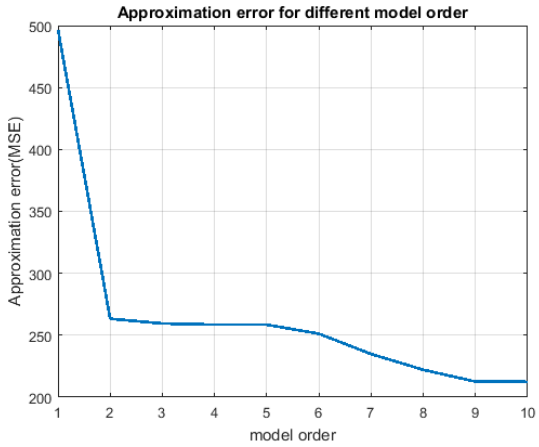

Figure 3.7: Approximation error for model order(1:10)

Q5. Plot power spectrum associated with AR(2) model with sampling frequency $1Hz$

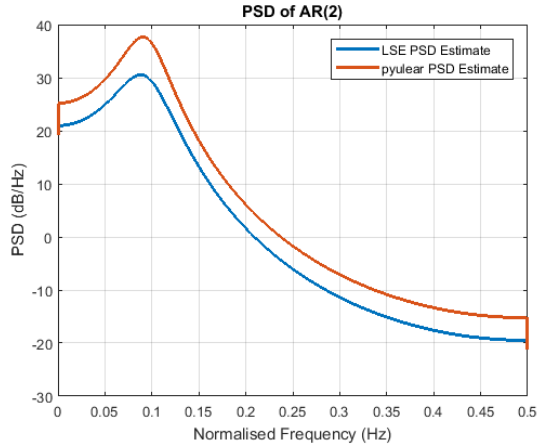

Figure 3.8: PSD from LSE and Yule-Walker

LSE approach provides an acceptable estimate of Yule-Walker method. Both two PSD estimates has pole at $0.1Hz$.

Q6. Calculate approximation mean square error as $\frac{(y-y')^2}{N}$, where y' is output sequence from AR(2) model use LSE and N is data length. Plot obtained MSE over N.

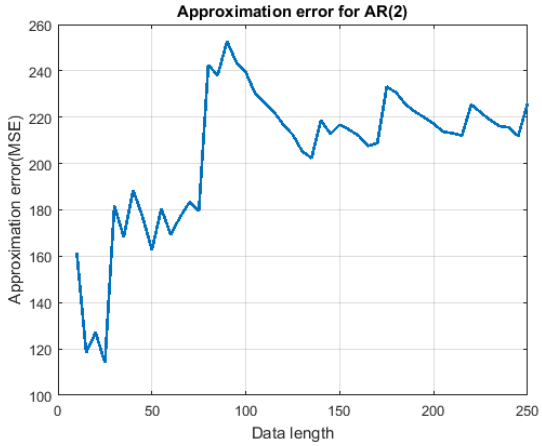

Figure 3.9: Approximation error of AR(2) against data length

In figure 3.9, approximation reaches minimum at $N = 25$, after which error boosts to maximum at $N = 100$. For longer data length, error starts fluctuating around certain values. We attain lowest mean square error at data length 25. Hitherto, we suggest 25 as an optimum data length for the AR(2) modelling of sunspot series.

3.4 Dial tone pad

Q1.A sampling rate of 32768Hz is used because $2^{15} \times \frac{1}{32768}$ is exactly 1 second, which is used to devise a precise 1Hz signal.

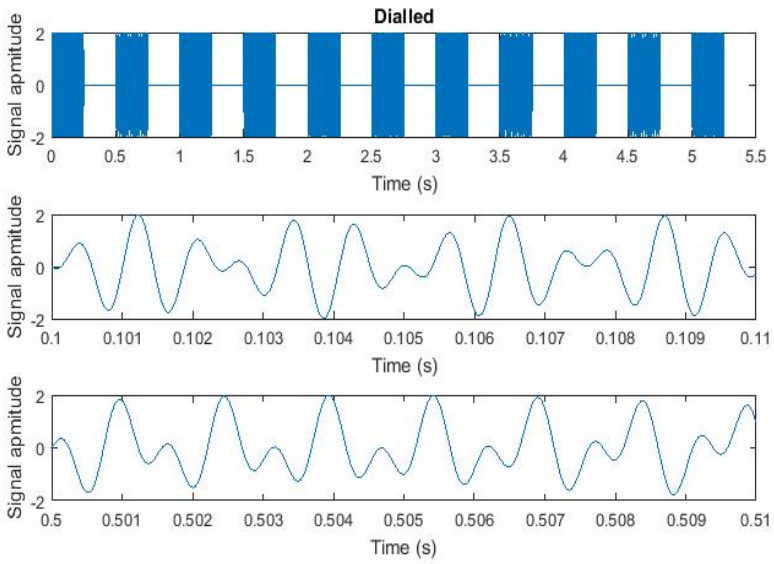

Figure 3.10: Phone number dialled 02089196025

Q2. Compute and display the spectrogram of x by the following steps:

- Divide the signal into 22 sections each of length 8192 samples.
- Windowed with a Hamming window, specify 0 samples of overlap between adjoining sections.

- Use 8192 points DFT(no zero padding), with a sampling frequency of 32768Hz.

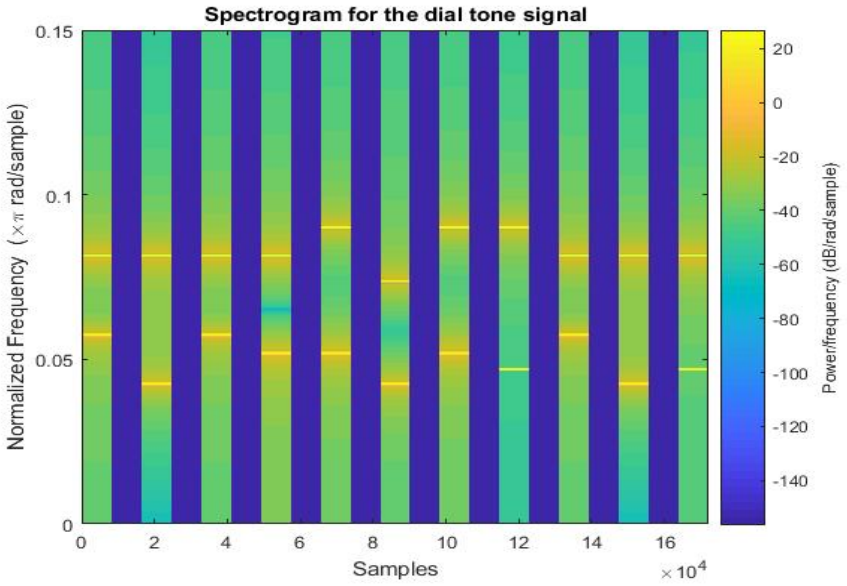

Figure 3.11: Dial tone signal spectrogram(normalised frequency)

The figure above represents the dial tone signal spectrogram. In blue is the idle time between the pressed digits. In contrast, the lime region represents the time duration when a digit is pressed. The yellow dashes are the frequency spectrum after short-time fourier transform is applied on a sinusoidal signal.

Q3. We can identify the different digits in the spectrogram above by recognise the yellow dashes. Given that normalised frequency can be calculated by $f_{norm} = f \times \frac{2\pi}{f_s}$. For example, the first pressed digit 0 has a signal with normalised frquencies 0.0574π and 0.0834π(both in units of radians per sample), these results are same as yellow dashes on the spectrogram.

Q4. Now corrupt the signal by channel noise

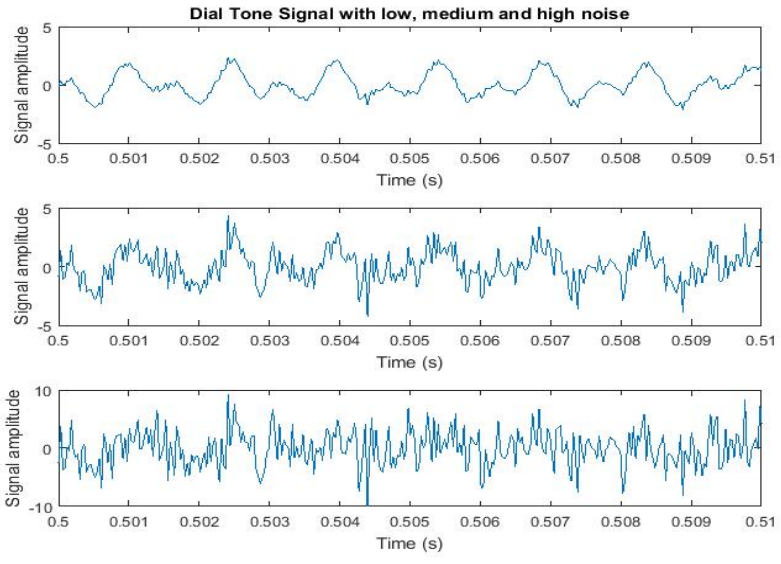

Figure 3.12: Dial tone signal with 0.2X 1X 3Xnoise

Increase noise variance until the original signal is completely corrupted as shown in figure 3.9. However the spectrogram is still distinguishable which means the tone identification is not affected. In figure 3.10, when we have 3X noise(that is 9X noise variance), we can still see the yellow dashes not submerged. But with high noise power, it has a larger opportunity making mistake during identification.

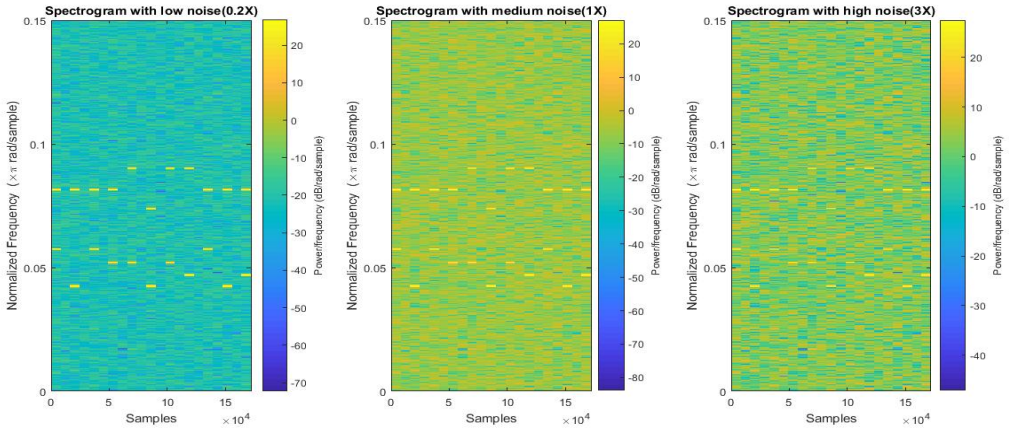

Figure 3.13: Dial tone signal spectrogram with 0.2X 1X 3Xnoise

3.5 Respiratory sinus arrhythmia from RR-Intervals

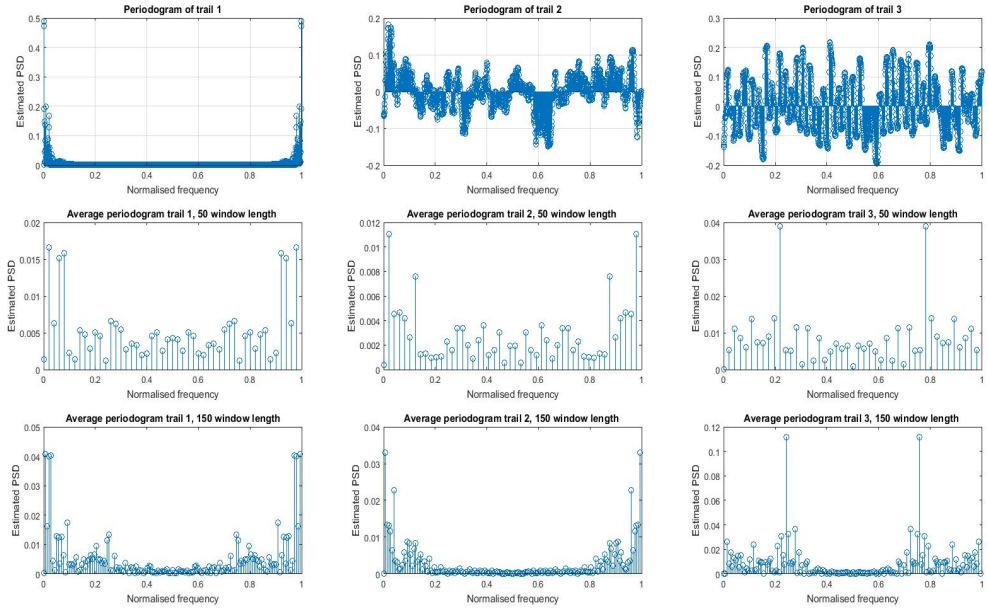

Figure 3.14: PSD estimate of RRI from origin, 50 and 150 samples window length

Hearth rate gets lower as a human constrains his breath(i.e. trail 1 to 3) and so does the RR interval. In trail 1, we observe 3 peaks at normalised frequnecy $1Hz$, the perspiration frequency is $(1-0.5)\times 4 = 2Hz$ where 0.5 is the frequency centre and 4 is the sampling frequency of RRI. Whereas in trail 2, there are two peaks corresponding to $(0.95-0.5)\times 4 = 1.8Hz$ and $2Hz$ respectively, this may due to heart rate differs during inspiration and expiration. Trail 3, however has only one peak at $(0.78-0.5)\times 4 = 1.12Hz$

Chapter 4

Optimal filtering- fixed and adaptive

4.1 Wiener filter

The optimal Wiener filter coefficients can be calculated as

$$\mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1}\mathbf{p}_{zx} \tag{4.1}$$

where $\mathbf{R}_{xx}$ is composed of autocorrelation function of input signal $x[n]$ and $\mathbf{p}_{zx}$ is composed of cross-correlation of input signal $x[n]$ and $z[n]$, which equals to system output $y[n]$+ a noise signal $\eta[n]$.

After signal $y[n]$ has been normalised, with noise power $\sigma_\eta^2 = 0.01$. Signal-to-Noise Ratio in dB for $z[n]$, that is $SNR_z = 20dB$. Solved for $\mathbf{w}_{opt} = [0.9957, 1.9782, 2.9840, 1.9826, 0.9889]^T$ which are similar to the unknown system's coefficients $[1, 2, 3, 2, 1]^T$.
Q1. Vary the scaling applied to additive noise

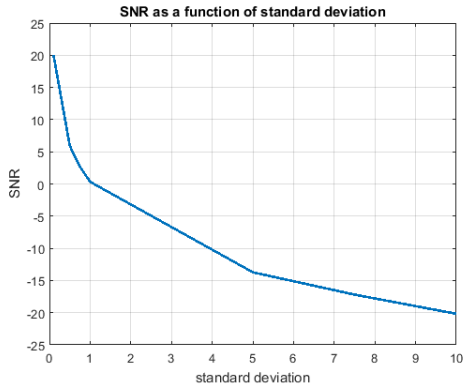

Figure 4.1: SNR varies against noise power

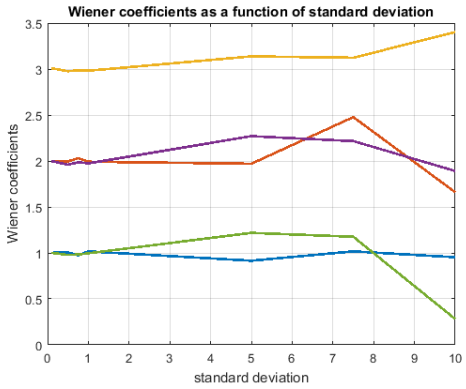

Figure 4.2: w varies against noise power

Q2. As noise power increases, SNR becomes lower indicating that the signal is corrupted. Wiener filter coefficients tend to diverge from correct values and the higher the noise power, more uncertainty are there in w.

Q3. Assume we have a q^{th} order filter and input signal $\mathbf{x}[n]$ of length N+1. Because autocorrelation is symmetric, we only need to calculate half of its samples. Number of executed iterations as follows:

Matrix	$\mathbf{R}_{xx}$	$\mathbf{r}_{zx}$	$\mathbf{R}_{xx}^{-1}$	$\mathbf{R}_{xx}^{-1}\mathbf{r}_{zx}$
no. of multiplications	N^2	$N \times (q+1)$	$O(N^3)$	$(q+1)^2$
no. of additions	N^2	$N \times (q+1)$	$O(N^3)$	$(q+1)^2$

Table 4.1: number of iterations in calculating w

4.2 The least mean square algorithm

Q1. Apply LMS algorithm on calculating Wiener coefficients. Use same input and filter order as in part 1, but adaptive gain μ to be 0.01. Solved for $\mathbf{w}_{opt} = [1.0009, 1.9901, 3.0043, 1.9812, 1.0037]^T$, which are again, close to the correct values.

Q2. Choose μ to be 0.002 and 0.01. Plot so obtained w_{opt} and estimation error(squared) over number of iterations.

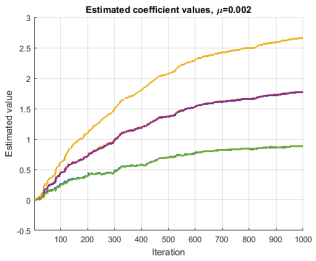

Figure 4.3: w_{opt} varies over time, $\mu = 0.002$

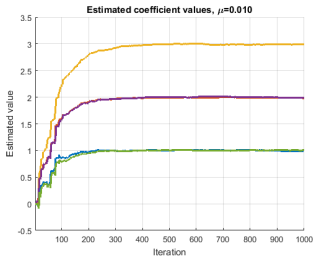

Figure 4.4: w_{opt} varies over time, $\mu = 0.01$

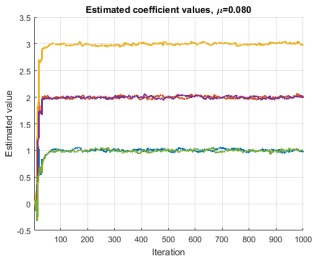

Figure 4.5: w_{opt} varies over time, $\mu = 0.08$

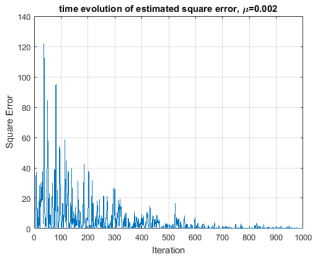

Figure 4.6: Estimation error varies over time, $\mu = 0.002$

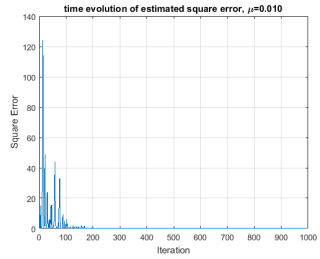

Figure 4.7: Estimation error varies over time, $\mu = 0.01$

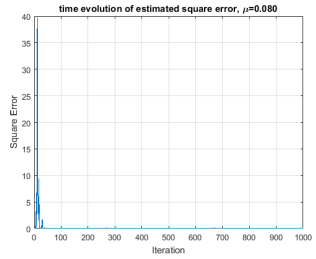

Figure 4.8: Estimation error varies over time, $\mu = 0.08$

The optimum learning rate μ would be 0.01 under which w_{opt} /converges to a stable constant, close to the correct filter coefficients $[1, 3, 3, 2, 10]^T$ meanwhile the estimation error reduces to zero.
Decreasing μ lead to very slow convergence as w_{opt} does not settle in a constant value after 1000 iterations. The trajectory traced by w_{opt} for number of iterations tends to be monotonic increasing. Error is till larger than zero after 1000 iterations.
Increasing μ has the opposite effect, leading to oscillating w_{opt}, this is because the step size for updating w is too large that it does not converge to local or global minima.

Q3. The LMS is based on the use of instantaneous estimates of the autocorrelation function $\hat{r}_x(j,x)$ and the cross-correlation function $\hat{r}_{yx}(k)$, substitute these into steepest gradient descent and we have.

$$w(n+1) = w(n) + \mu e(n)x(n) \tag{4.2}$$

given that

$$\hat{y}(n) = w^T(n)x(n) = \sum_{i=0}^{M} w_i(n)x(n-i) \tag{4.3}$$

$$e(n) = z(n) - \hat{y}(n) \tag{4.4}$$

whereas $w(n)$ and $x(n)$ are column vectors of length (M+1). Computation complexity for each iteration as follows

Matrix	$\hat{y}(n)$	$e(n)$	$w(n)$
no. of multiplications	$M+1$		$2M+3$
no. of additions	$M+1$	$M+1$	$2(M+1)$

Table 4.2: number of iterations in calculating w

4.3 Gear shifting

The main drawback of the ordinary LMS algorithm is that it is sensitive to the scaling of its input $x(n)$. This makes it very hard (if not impossible) to choose a learning rate μ that guarantees stability of the algorithm. [3] As discussed in part 2, high adaptive gain implies fast rise time and low gain implies good steady state accuracy. Gear shifting benefits from both by making adaptive gain high for large error and lower for small error.
We use Normalised Least Mean Square Algorithm as adaptive gain policy. That is,

$$w(n+1) = w(n) + \frac{\mu}{\sigma + ||x(n)||^2 + \epsilon} \tag{4.5}$$

Plot time evolution of w_{opt} under normalised LMS.

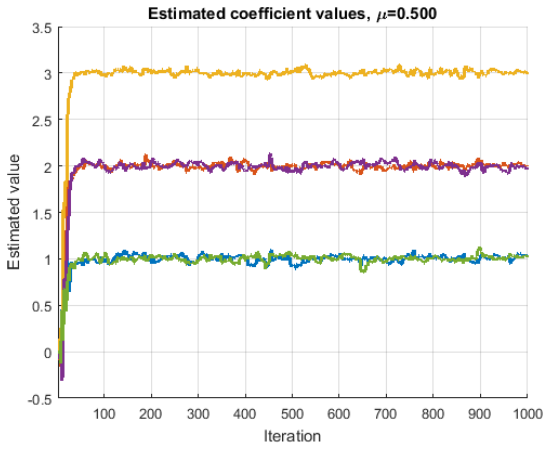

Figure 4.9: Normalised LMS, $\mu = 0.5$

The rise time is less than 50 samples in this case, confirmed that normalised LMS significantly reduces the rise time while retaining low overshooting in $w(n)$.

4.4 Identification of AR process

Apply 2^{nd} adaptive filtering in Autoregressive system identification.

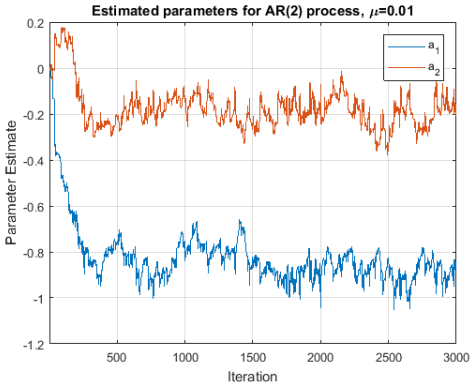

Figure 4.10: Estimated a varies over time, $\mu = 0.01$

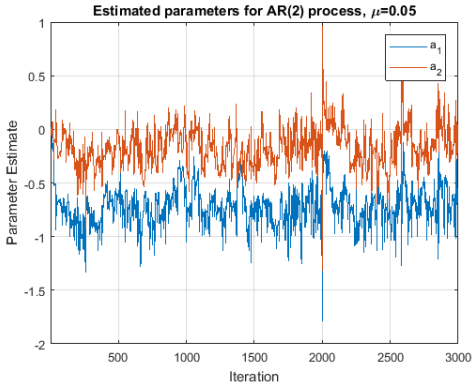

Figure 4.11: Estimated a varies over time, $\mu = 0.05$

Figure 4.10 and 4.11 illustrate that the AR coefficients converges towards -0.2 and -0.9 (in fact they are positive but due to MATLAB format). As adaptive gain μ increases, estimated a tend to oscillate heavily making the estimation irrational.

4.5 Speech recognition

Q1. In order to investigate effect of adaptive gain and filter order on the predictor. Use sounds "a" as an example. Plot results of gain=1,order=3, gain=1,order=5, gain=5,order=5, respectively. (Other sounds are excluded form the report due to page limits)

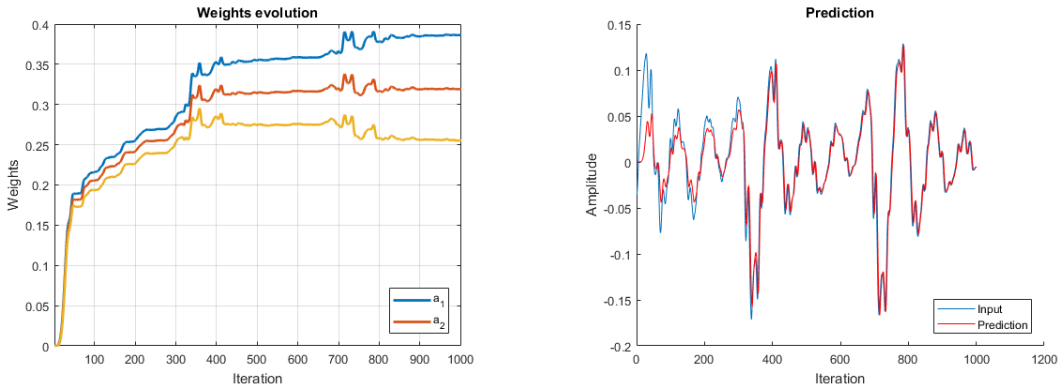

Figure 4.12: weights evolution and prediction versus actual gain =1 order=3

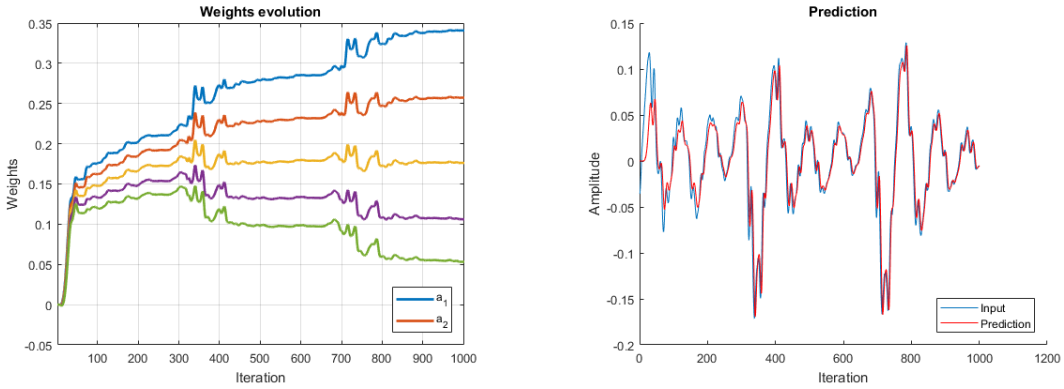

Figure 4.13: weights evolution and prediction versus actual gain=1 order=5

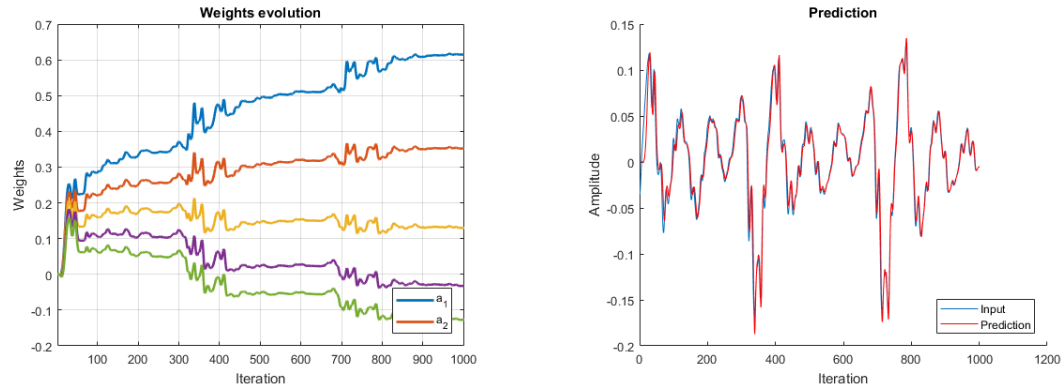

Figure 4.14: weights evolution and prediction versus actual gain=5 order=5

By comparing figure 4.5 and figure 4.5 , the graphs do not show much improvement for increasing predictor order. Figure 4.5 and figure 4.5 illustrate that increasing adaptive gain will enhance the performance of the predictor. As adaptive gain increases, the predictor output converges more quickly to the original signal. Nevertheless, increasing the adaptive gain by too much will lead to oscillations and inaccuracies.

Q2. An heuristic way of find optimal filter length would be using the prediction gain.

$$R_p = 10log_{10}(\frac{\sigma_x^2}{\sigma_e^2}) \tag{4.6}$$

The larger the prediction gain, the better the performance of the predictor. Plot prediction gain against model order below.

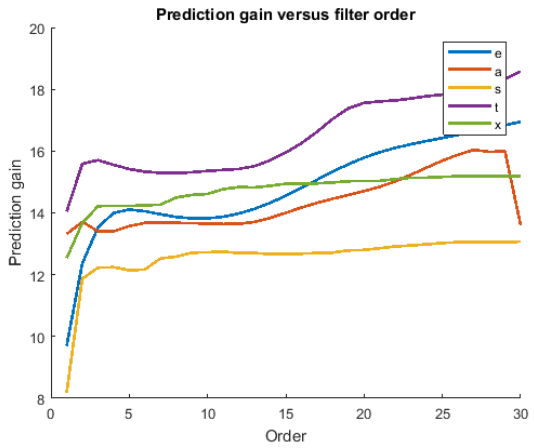

Figure 4.15: prediction gain, fs=44100Hz

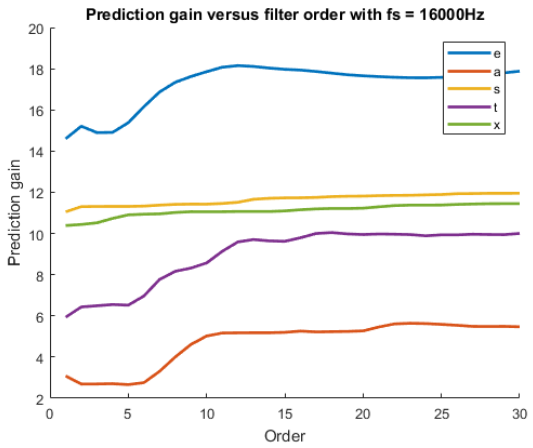

Figure 4.16: prediction gain, fs=16000Hz

As discussed in question 2, larger model order within in an acceptable range will enhance the predictor performance. However a very larger order will increase model complexity dramatically. Therefore an appropriate trade off between model order and complexity is desired. From figure 4.15, an appropriate model order would be 10 for all sound signals. Furthermore an analytical approach would use tools like the Minimum Description Length (MDL) or Akaike Information Criterion (AIC) in order to get a more intuitive view of the trade off.

Q3. With a sampling frequency of 16000Hz, the prediction gain drops as shown in figure 4.16. This is reasonable because a lower sampling rate leads to lower precision and larger noise power indeed. To conclude, obtaining quasi-stationary vowel sounds requires less data points at lower sampling frequency, but we would still require more data points for weights evolution to converge.

4.6 Dealing with computational complexity: sign algorithms

In this section, we apply several sign algorithms to estimate AR(2) parameters in section 4.4.

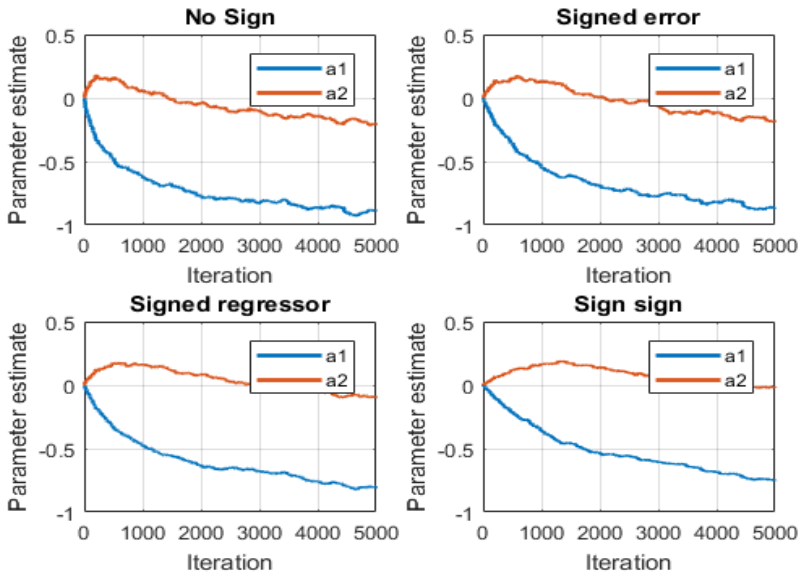

Figure 4.17: AR parameter estimation using sign algorithms

All sign algorithms lead to the same desired output with a1 = -0.9 and a2 = -0.2 but they all take different converging trends. Compared to conventional LMS algorithm, signed algorithms have slower learning curves among which sign-sign has slowest convergence rate.

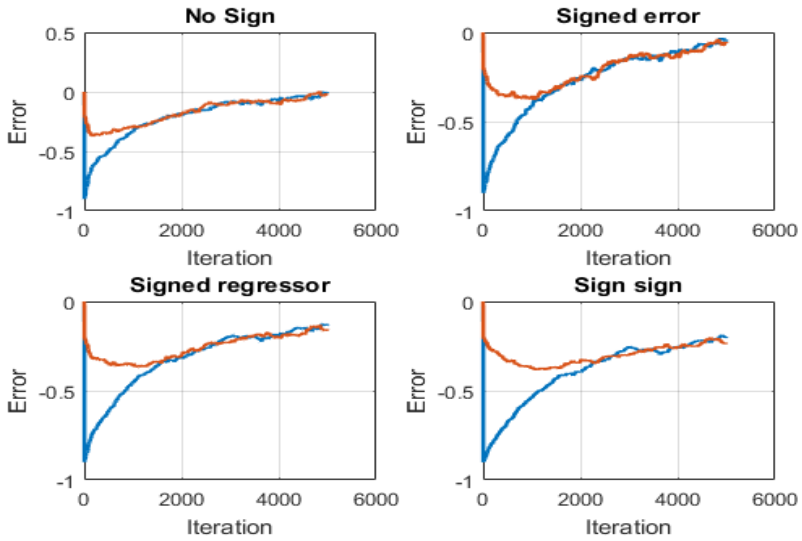

Figure 4.18: Estimation error of sign algorithms

Figure 4.18 is plot of the estimation errors. All signed algorithms give lower accuracy compared to conventional LMS. Among all signed algorithms, signed error has best accuracy. Again, test sign algorithms with adaptive gain 0.01, filter order 5 on sound file “a”. First it is worthwhile to be noticed that due to the very small adaptive gain, the conventional algorithm does not converge to desired values within 1000 samples. As shown in figure 4.19, signed error and signed regressor have close convergence rate. Sign-sign algorithm does not converge. We can conclude that generally, signed error has best accuracy and convergence performance.

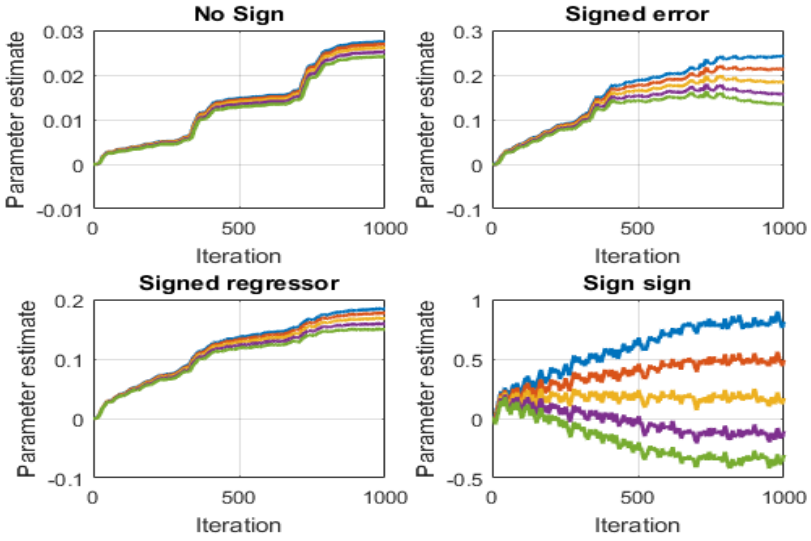

Figure 4.19: Weights parameter

Chapter 5

5.1 MLE for the frequency of signal (Extra Credit)

Q1. The maximum likelihood estimate $\hat{f_0}$ of the frequency is found by minimising

$$\mathbf{J}(\theta) = \sum_{n=0}^{N-1} (x[n] - A\cos(2\pi f_0 n + \phi))^2 \tag{5.1}$$

Expand the cosine to yield

$$\mathbf{J}(\theta) = \sum_{n=0}^{N-1} (x[n] - A\cos(2\pi f_0 n)\cos(\phi) + A\sin(2\pi f_0 n)\sin(\phi))^2 \tag{5.2}$$

Use non-linear transformation such that

$$\alpha_1 = A\cos\alpha$$

$$\alpha_2 = -A\sin\alpha$$

And the inverse is given by

$$A = \sqrt{\alpha_1^2 + \alpha_2^2}$$

$$\phi = \arctan(\frac{-\alpha_2}{\alpha_1})$$

Also, define

$$\mathbf{c} = [1, \cos(2\pi f_0) ... \cos(2\pi f_0 (N-1))]^T$$

$$\mathbf{s} = [1, \sin(2\pi f_0) ... \sin(2\pi f_0 (N-1))]^T$$

Therefore

$$\mathbf{J}'(\alpha, f_0) = (\mathbf{x} - \alpha_1\mathbf{c} - \alpha_2\mathbf{s})^T(\mathbf{x} - \alpha_1\mathbf{c} - \alpha_2\mathbf{s}) = (\mathbf{x} - \mathbf{H}\alpha)^\mathbf{T}(\mathbf{x} - \mathbf{H}\alpha) \tag{5.3}$$

Q2. The least square solution of equation 5.1 is

$$\alpha = (\mathbf{H^T H})^{-1}\mathbf{H^T x} \tag{5.4}$$

where $(\mathbf{H^T H})^{-1}\mathbf{H^T}$ is the pseudo inverse, which is idempotent. Substitute α in to $\mathbf{J}'(\alpha, f_0)$ and by the idempotency of $\mathbf{I} - (\mathbf{H^T H})^{-1}\mathbf{H^T}$

$$\begin{aligned}\mathbf{J}'(\alpha, \mathbf{f_0}) &= (\mathbf{x} - \mathbf{H}(\mathbf{H^T H})^{-1}\mathbf{H^T x})^{\mathbf{T}}(\mathbf{x} - \mathbf{H}(\mathbf{H^T H})^{-1}\mathbf{H^T x}) \\ &= \mathbf{x^T}(\mathbf{I} - \mathbf{H}(\mathbf{H^T H})^{-1}\mathbf{H^T})\mathbf{x}\end{aligned}$$

We intend to minimise $\mathbf{J}'$, which is equivalent to maximise

$$\mathbf{x^T H}(\mathbf{H^T H})^{-1}\mathbf{H^T x} \tag{5.5}$$

Q3. Known that $\mathbf{H}$ is composed of two column vectors $\mathbf{c}$ and $\mathbf{s}$ who have same dimension as $\mathbf{x}$.

$$\mathbf{x}^T\mathbf{H} = [x_1, x_2, ...x_i] \cdot \begin{pmatrix} c_1 & s_1 \\ c_2 & s_2 \\ ... & ... \\ c_i & s_i \end{pmatrix} = [\mathbf{c^T x} \quad \mathbf{s^T x}]$$

$$\mathbf{H^T H} = \begin{pmatrix} c_1 & c_2 & ... & c_i \\ s_1 & s_2 & ... & s_i \end{pmatrix} \cdot \begin{pmatrix} c_1 & s_1 \\ c_2 & s_2 \\ ... & ... \\ c_i & s_i \end{pmatrix} = \begin{pmatrix} \mathbf{c^T c} & \mathbf{c^T s} \\ \mathbf{s^T c} & \mathbf{s^T s} \end{pmatrix}$$

Therefore maximising equation 5.5 equivalents to maximising

$$\begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix}^T \begin{pmatrix} \mathbf{c^T c} & \mathbf{c^T s} \\ \mathbf{s^T c} & \mathbf{s^T s} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix} \tag{5.6}$$

Under the condition that f_0 is not close to 0 or $\frac{1}{2}$ and because of periodicity of sinusoid.

$$\begin{aligned}\mathbf{c^T s} &= \sum_{n=0}^{N-1} \cos 2\pi f_0 n \sin 2\pi f_0 n \\ &= \frac{1}{2}\sum_{n=0}^{N-1} \sin 4\pi f_0 n \\ &= Imj[\frac{1}{2}\sum_{n=0}^{N-1} exp(4\pi f_0 n)] \\ &\approx 0\end{aligned}$$

$$\begin{aligned}\mathbf{c^T c} &= \sum_{n=0}^{N-1} \cos 2\pi f_0 n \cos 2\pi f_0 n \\ &= \frac{1}{2}\sum_{n=0}^{N-1} \cos 4\pi f_0 2n + 1 \\ &= Re[\frac{1}{2}\sum_{n=0}^{N-1} exp(4\pi f_0 2n)] + \frac{N}{2} \\ &\approx \frac{N}{2}\end{aligned}$$

Similarly, $\mathbf{s^T s} \approx \frac{N}{2}$ Hence, expression 5.6 becomes approximately

$$\begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix}^T \begin{pmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix} \tag{5.7}$$

Q4. When f_0 is close to 0 or $\frac{1}{2}$, $A\cos(2\pi f_0 n + \phi)$ in equation 5.1 will be a constant, ans there is no meaning to differentiate $\mathbf{J}(\theta)$ with respect to f_0. Expression 5.7 becomes approximately

$$\begin{aligned} \begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix}^T \begin{pmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{c^T x} \\ \mathbf{s^T x} \end{pmatrix} &= \frac{2}{N}[(\mathbf{c^T x})^2 + (\mathbf{s^T x})^2] \\ &= \frac{2}{N}[(\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n)^2 + (\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n)^2] \\ &= \frac{2}{N}|\sum_{n=0}^{N-1} x[n]exp(-j2\pi f_0 n)|^2 \end{aligned} \tag{5.8}$$

Therefore MLE of frequency is obtained by maximizing the periodogram.

$$\hat{P}_x(f) = \frac{1}{N}|\sum_{n=0}^{N-1} x[n]exp(-j2\pi f n)|^2 \tag{5.9}$$

Q5. The exact MLE is obtained by maximising expression 5.5 while approximation MLE is obtained by maximising the periodogram. Plot the periodogram of given signal with rectangular window and MLE estimate for $f_0 = 0.25$ and 0.05.

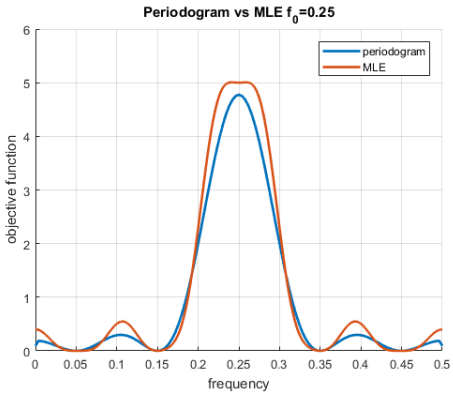

Figure 5.1: periodogram and MLE estimate 1

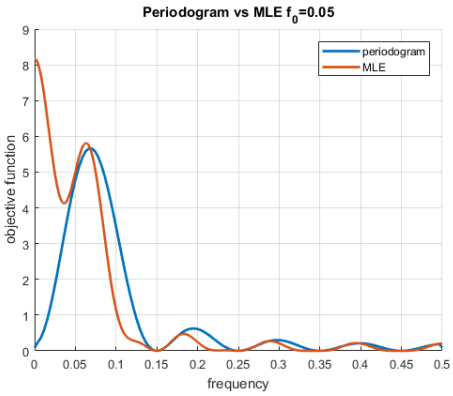

Figure 5.2: periodogram and MLE estimate 2

In figure 5.1, periodogram and MLE estimate both have peak at normalised frequency $0.25Hz$. But in figure 5.2, for $f_0 = 0.05$ the peak of periodogram(Blue plot) shift away from $0.05Hz$. This is because periodogram is an asymptotically unbiased estimator of the power spectral density, it may be biased when number of samples is small.

Bibliography

[1] George E. P Box. *Time series analysis*. Wiley, 2016. pages 4

[2] Steven M kay. *Fundamentals of Statistical Signal Processing*. Prentice Hall, 2013. pages 14

[3] Unknown. Normalised least mean squares filter, May 2016. pages 34

[4] Unknown. Identifying the numbers of ar or ma terms in an arima model, January 2017. pages 20

[5] Wikipedia. White noise, February 2018. pages 11