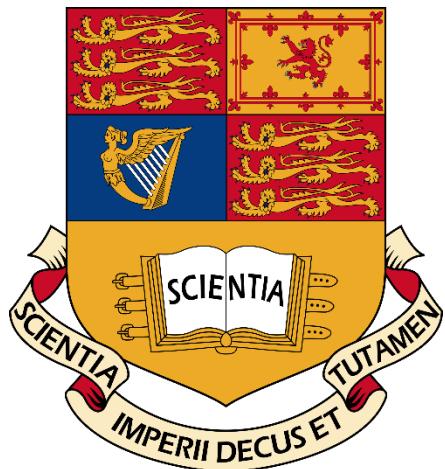


Electrical and Electronic Engineering, Imperial College London

Advanced Signal Processing

Module Leader: Prof. Danilo Mandic



Mohamed, Amin

CID: 02216105

Email: aam522@ic.ac.uk

Part 1: Random Signals and Stochastic Processes

1.1 Statistical Estimation

Uniform Distribution

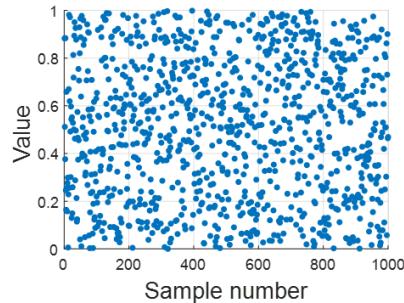


Figure: Plot of uniformly distributed random process over many realisations

Question 1

Theoretical mean:

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx$$

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x dx$$

$$\mathbb{E}\{X\} = \frac{1}{2}$$

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Sample Mean	0.5049	0.5105	0.4859	0.4985	0.4973
Percentage error	0.9850	2.0993	2.8178	0.2948	0.5317

The sample mean is an unbiased estimator of the population mean for a uniform distribution and, by the Lehmann-Scheffé theorem , is also the Minimum Variance Unbiased Estimator (MVU) in this case. According to the Cramér-Rao lower bound, as the sample size increases, the variance of the sample mean decreases ($\propto \frac{1}{N}$), improving its accuracy for the uniform distribution.

Question 2

Theoretical standard deviation given that $\mathbb{E}\{X\} = \frac{1}{2}$

$$\mathbb{E}\{X^2\} = \int_0^1 x^2 p_X(x) dx$$

$$\mathbb{E}\{X^2\} = \int_0^1 x^2 dx$$

$$\mathbb{E}\{X^2\} = \frac{1}{3}$$

$$\sigma_X^2 = \mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$\sigma_X = \frac{1}{\sqrt{12}} \approx 0.2887$$

Statistic	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Measured Std	0.2965	0.2965	0.2965	0.2965	0.2965
Percentage Error Std	2.7113	2.7113	2.7113	2.7113	2.7113

The sample standard deviation is also an unbiased estimator of the population standard deviation for a uniform distribution and, by the Lehmann-Scheffé theorem, is also the Minimum Variance Unbiased (MVU) Estimator in this case. Naturally, we also see the inversely proportional relationship between sample size and variance in sample standard deviation.

Question 3

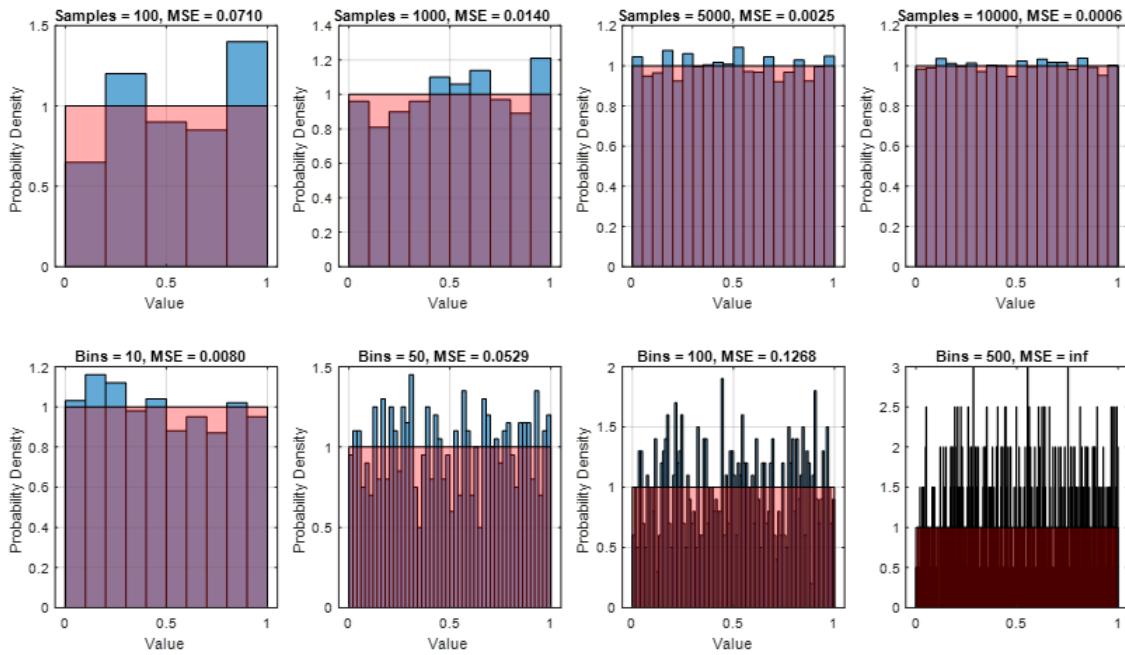
The sample standard deviations are very close to the theoretical mean of 0.5, with the values clustering around this value, indicating minimal bias in the estimator. Similarly, the sample standard deviations are close to the theoretical value of 0.2887, with small variations, suggesting that the estimator for the standard deviation is also unbiased, but with slightly higher variability in the estimates.

Bias	1	2	3	4	5	6	7	8	9	10
mean	-0.0033	-0.0053	0.0136	-0.0056	0.0054	-0.0129	0.0098	-0.0215	0.0005	-0.0070

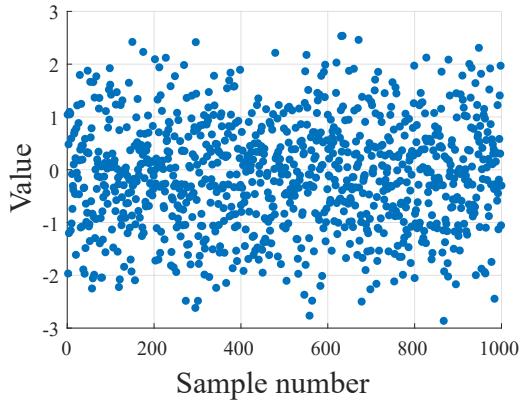
Bias	1	2	3	4	5	6	7	8	9	10
std	-0.0017	-0.0021	-0.0040	0.0040	0.0064	0.0031	0.0011	0.0006	0.0016	0.0028

Question 4

As the number of bins increases, the error and variance in the histogram estimate also increase, because the number of samples per bin decreases. This is crucial, as each bin acts as a parameter to estimate, and with fewer samples per bin, the estimation becomes less stable. According to estimation theory, increasing the sample size reduces the variance and error of the estimate, leading to a more accurate representation of the theoretical PDF.



Standard Normal Distribution



Question 1

Theoretical mean

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p_X(x) dx$$

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$\mathbb{E}\{X\} = 0$, due to odd function and symmetric interval

Question 2

Theoretical standard deviation

$$Var(X) = \sqrt{(\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2)}$$

$$Var(X) = \sqrt{\mathbb{E}\{X^2\}}$$

$$Var(X) = \sqrt{\left(\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right)}$$

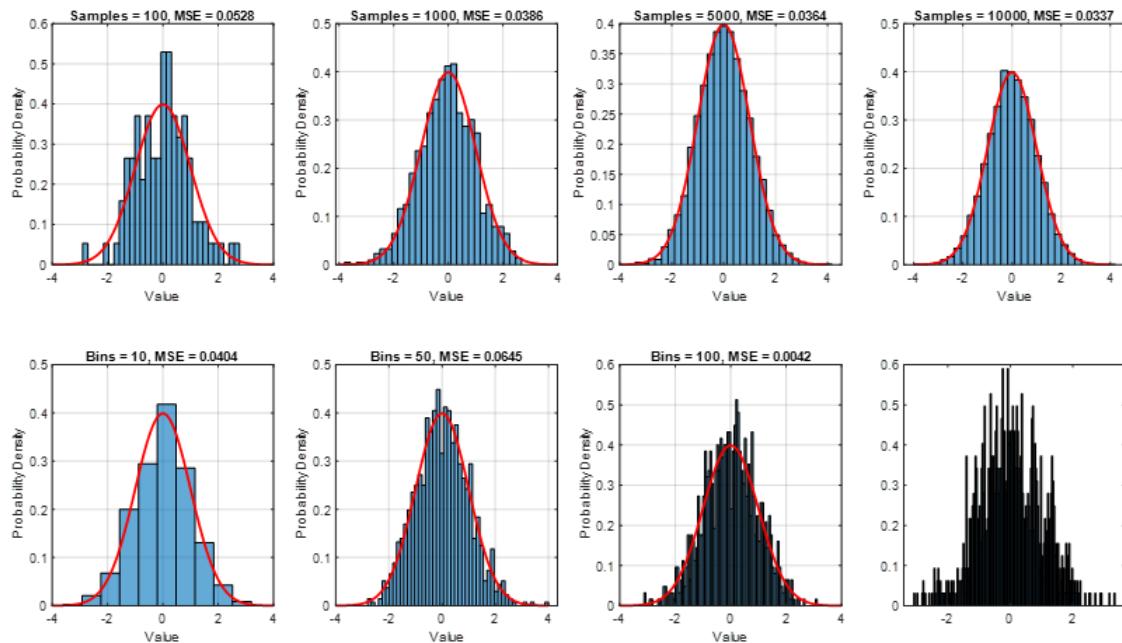
$Var(X) = 1$, using integration by parts

Both the sample mean and the sample standard deviation are unbiased estimators of the population mean and population standard deviation for a standard normal distribution and, by the Lehmann-Scheffé theorem, is also the Minimum Variance Unbiased Estimator (MVU) in this case also. According to the Cramér-Rao lower bound, as the sample size increases, the variance of decreases ($\propto \frac{1}{N}$), improving its accuracy.

Question 3

Bias	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
mean	0.0274	-0.0102	0.0424	-0.0424	-0.0264	-0.0266	-0.0170	0.0142	-0.0366	-0.0446
std	0.0151	-0.0098	-0.0051	0.0220	0.0156	0.0159	-0.0224	-0.0219	0.0122	0.0307

Question 4

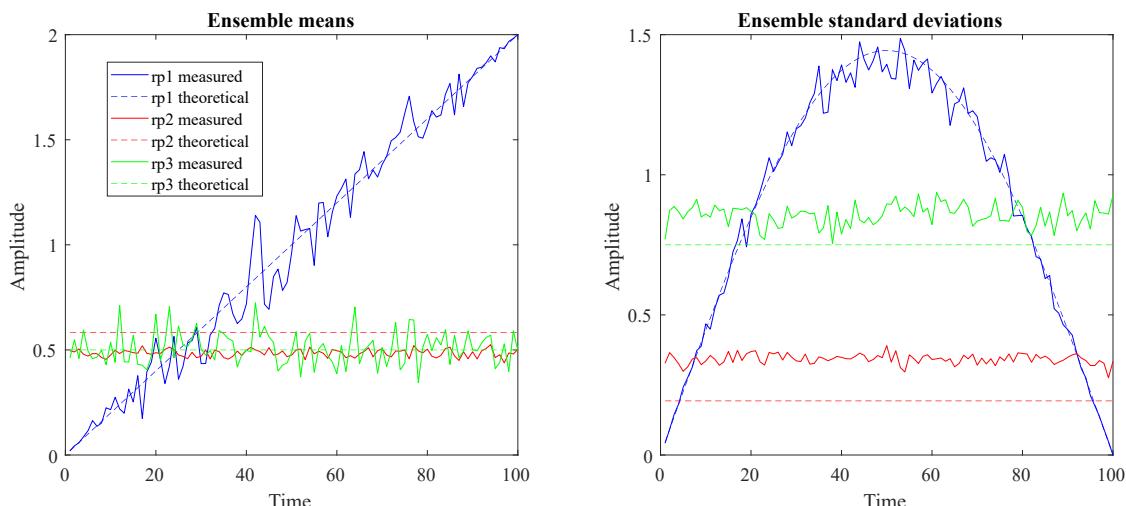


More samples is always better

The number of bins affect how complex our pdf can be. We must pick the number of bins in accordance to the number of samples we have as to minimise variance.

1.2 Stochastic Processes

Question 1



Function 1 is not a stationary process as the ensemble mean is time-dependent

Function 2 and 3 are stationary processes as the ensemble mean is a constant line across time, therefore time-independent.

Question 2

Random Process 1

Mathematical Expression

$$rp_1(t) = [u(t) - 0.5] \times 5 \sin(\pi t) + 0.02t$$

Theoretical Ensemble Mean

$$\mathbb{E}\{rp_1(t)\} = \mathbb{E}\{u(t) - 0.5\} \times 5 \sin(\pi t) + 0.02t$$

$$\mathbb{E}\{rp_1(t)\} = 0.02t$$

Therefore ergodic as mean is a function of time

Theoretical Ensemble Standard Deviation

$$\mathbb{E}\{(rp_1(t))^2\} = \mathbb{E}\{(u(t) - 0.5)^2\} \times 25 \sin^2(\pi t) + (0.02t)^2 + 2 \times 0.02t \times \mathbb{E}\{u(t) - 0.5\} \times 5 \sin(\pi t)$$

$$\mathbb{E}\{(rp_1(t))^2\} = \mathbb{E}\{(u(t) - 0.5)^2\} \times 25 \sin^2(\pi t) + (0.02t)^2$$

$$\mathbb{E}\{(u(t) - 0.5)^2\} = \frac{1}{12} = \mathbb{E}\{u^2(t) - u(t) + 0.25\} = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

$$\mathbb{E}\{rp_1^2(t)\} = \frac{1}{12}(25 \sin^2(\pi t)) + (0.02t)^2$$

$$Var(rp_1(t)) = \mathbb{E}\{rp_1^2(t)\} - \mathbb{E}\{rp_1(t)\}^2 = \frac{25}{12} \sin^2(\pi t) + (0.02t)^2 - (0.02t)^2$$

$$Var(rp_1(t)) = \frac{25}{12} \sin^2(\pi t)$$

Random Process 2

Mathematical Expression

$$rp_2(t) = [u(t) - 0.5] \times u(t) + u(t)$$

Theoretical mean

$$\mathbb{E}\{rp_2(t)\} = \mathbb{E}\{u^2(t)\} - 0.5\mathbb{E}\{u(t)\} + \mathbb{E}\{u(t)\} = \mathbb{E}\{u^2(t)\} + 0.5\mathbb{E}\{u(t)\}$$

$$\mathbb{E}\{rp_2(t)\} = \frac{1}{3} + \frac{1}{4} = \frac{7}{12} \approx 0.5833$$

Therefore ergodic as mean is not a function of time.

Theoretical standard deviation

$$\mathbb{E}\{rp_2^2(t)\} = \mathbb{E}\{u^4(t)\} + \frac{1}{4}\mathbb{E}\{u^2(t)\} + \mathbb{E}\{u^3(t)\}$$

We show that $\forall n \in \mathbb{N}^+$ we can show

$$\mathbb{E}\{u^n(t)\} = \int_0^1 x^n dx = \frac{1}{n+1}$$

Therefore:

$$\mathbb{E}\{rp_2^2(t)\} = \frac{1}{5} + \frac{1}{4}\left(\frac{1}{3}\right) + \frac{1}{4} = \frac{8}{15}$$

Finally:

$$Var(rp_2(t)) = \mathbb{E}\{rp_2^2(t)\} - \mathbb{E}\{rp_2(t)\}^2 = \frac{8}{15} - \left(\frac{7}{12}\right)^2 = \frac{417}{2160} \approx 0.1931$$

Random Process 3

Mathematical Expression

$$rp_3(t) = [u(t) - 0.5] \times 3 + 0.5$$

Theoretical Mean

$$\mathbb{E}\{rp_3(t)\} = \mathbb{E}\{u(t) - 0.5\} \times 3 + 0.5$$

$$\mathbb{E}\{rp_3(t)\} = 0.5$$

Therefore ergodic as mean is not a function of time.

Theoretical Standard Deviation

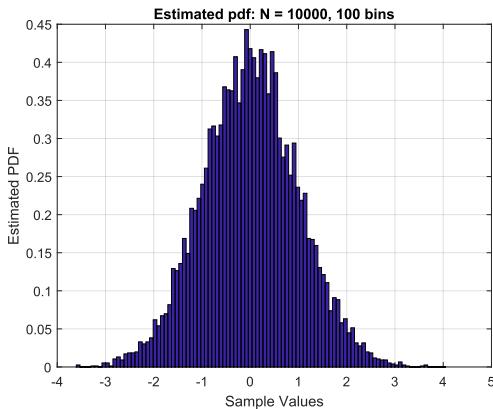
$$\mathbb{E}\{rp_3^2(t)\} = \mathbb{E}\{(3u(t) - 1)^2\} = \mathbb{E}\{9u^2(t) - 6u(t) + 1\} = 9\left(\frac{1}{3}\right) - 6\left(\frac{1}{2}\right) + 1 = 3 - 3 + 1 = 1$$

$$Var(\text{rp}_3(t)) = \mathbb{E}\{\text{rp}_3^2(t)\} - \mathbb{E}\{\text{rp}_3(t)\}^2 = 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}$$

1.3 Estimation of probability distributions

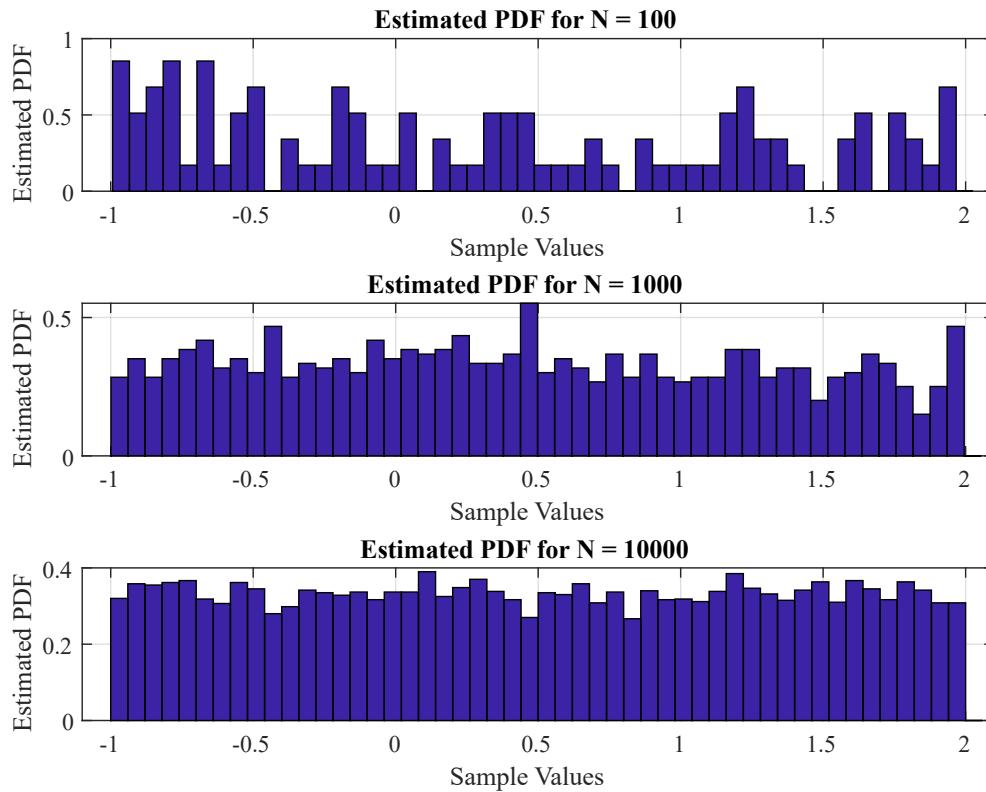
Question 1

The `pdf.m` function estimates the pdf without using `hist` or other similar MATLAB functions. Below shows the result of the function, which works as expected.



Question 2

Given that random process 3 is both stationary and ergodic, we analyse it with results in the figure below. It is clear that as the number of samples increases our estimate pdf becomes more accurate.

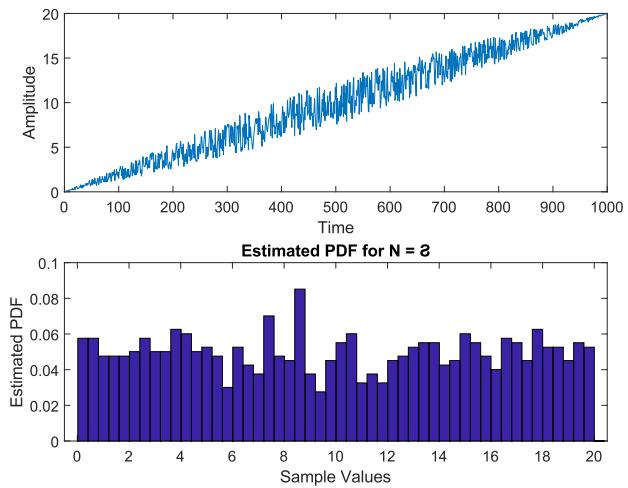


Question 3

Given that random process 1 is non-stationary we analyse as follows.

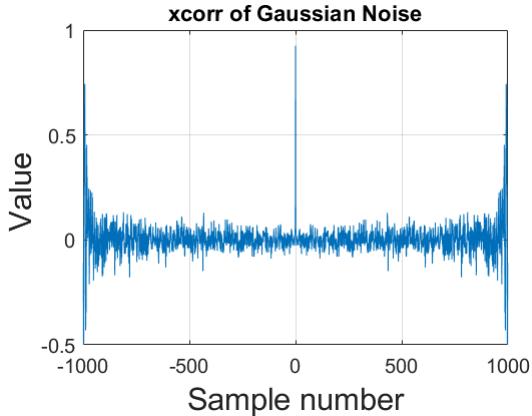
We know that random process 1 is unbounded, however, due to `pdf.m` only taking in a finite number of samples, the estimate it produces incorrectly bounds the distribution between 0 and 20.

Therefore, we lose critical information about our process that, visually, is obvious.



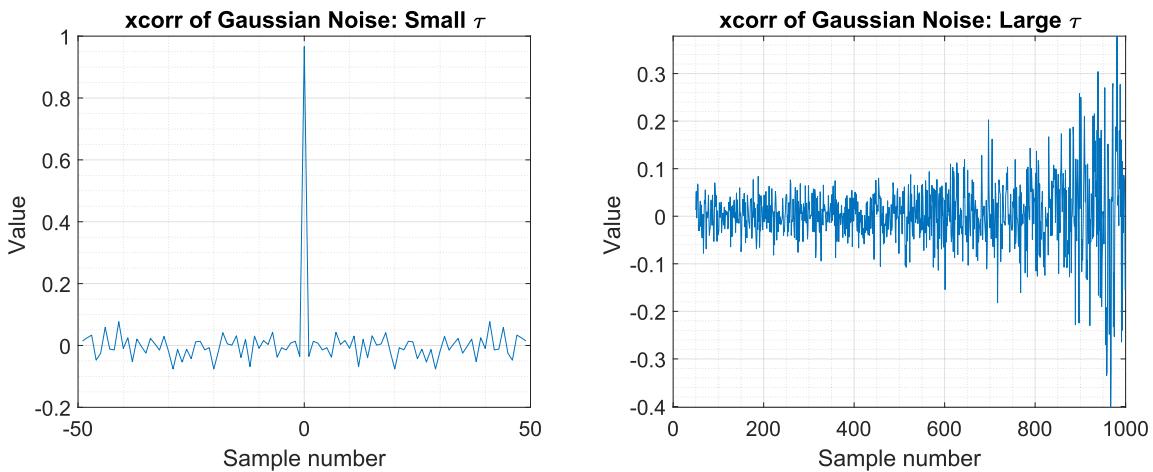
2.1 ACF of uncorrelated and correlated sequences

Question 1:



Theoretically, two Gaussian points are always uncorrelated. Theoretically, the auto-correlation should be a Dirac delta. This is reflected in our estimate.

Question 2:



The cross correlation estimate is accurate for small lag. However, the estimate diverges at large lag.

Question 3

The autocorrelation is

$$R_X(n, s) = \mathbb{E}\{X_n X_s\}$$

Assuming ergodicity, the **statistical expectation operator** can be approximated by a **time average**

$$R_X(\tau) = \lim_{n \rightarrow \infty} \sum_{n=0}^{N-1} x[n] x[n + \tau], \quad \tau \in \mathbb{Z}$$

However, due to finite data, we consider the **unbiased** estimate given by

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

From this equation, we observe that the autocorrelation estimator's output grows as the number of samples available for comparison reduces. This reduction occurs when τ increases. Intuitively, this is because the probability of finding identical samples increases as the sample size decreases.

We previously noted that the estimator starts to diverge approximately after $\tau = \pm 500$. Therefore, a conservative empirical bound would be half of this value, $\tau = 250$. Generalizing, this bound can be expressed as $\frac{N}{4}$.

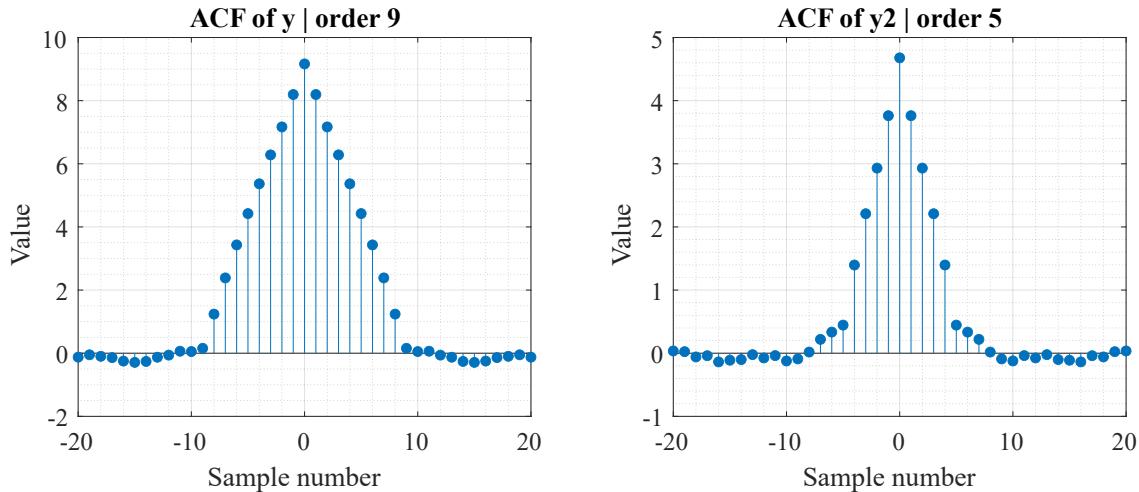
Question 4

The figure below shows the effect that a Moving Average filter would have on WGN.

Theoretical ACF is $\Lambda(\tau/N)$, where Λ is the triangle function.

This explains why order 9 has a wider domain than order 5.

The slight smoothing, which is more noticeable for order 9, can be explained given that the MA filter is a low pass filter.



Ideally the ACF should be zero outside of the ranges $-8 < \tau < 8$ for order 9 and $-4 < \tau < 4$ for order 5. However, there are slight oscillations that can be seen outside of this range, these can also be explained by low pass nature of the MA filter, specifically that infinite bandwidth is required to perfectly synthesise a discontinuous piecewise function like the triangle function.

Question 5

Given that Y_n is a filtered version of an uncorrelated process X_n , we have the relationship:

$$R_Y(\tau) = R_X(\tau) * R_h(\tau)$$

Since X_n is uncorrelated, its autocorrelation function $R_X(\tau)$ is a Dirac delta function, $\delta(\tau)$. Therefore, equation (2) simplifies to:

$$R_Y(\tau) = \delta(\tau) * R_h(\tau)$$

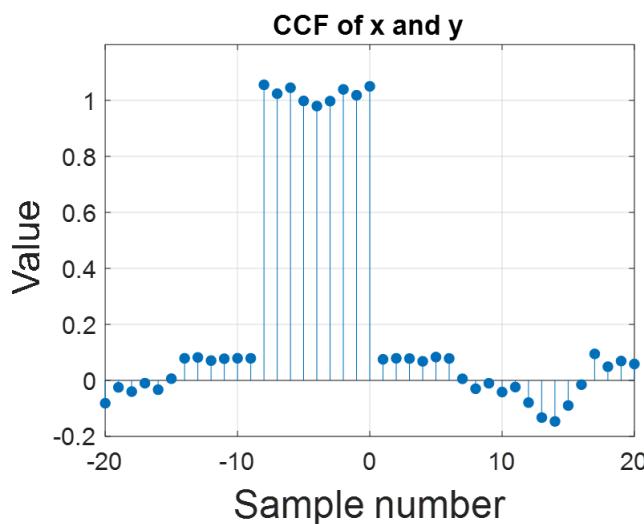
Because the convolution of any function with a Dirac delta function results in the function itself, we obtain:

$$R_Y(\tau) = R_h(\tau)$$

Thus, the autocorrelation function of Y_n is equivalent to the autocorrelation function of the impulse response, which in this case is a scaled triangle function, $N \cdot \Lambda(\frac{\tau}{N})$.

2.2 Cross-correlation function

Question 1



For a gaussian process X ,

$$R_X = E(X(t)X(t - \tau)) = \begin{cases} 1 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases} = \delta(\tau)$$

The CCF can be found

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

using the filtering property:

$$R_{XY} = h(\tau) * \delta(\tau) = \int h(u)\delta(\tau - u)du = h(\tau)$$

Question 2

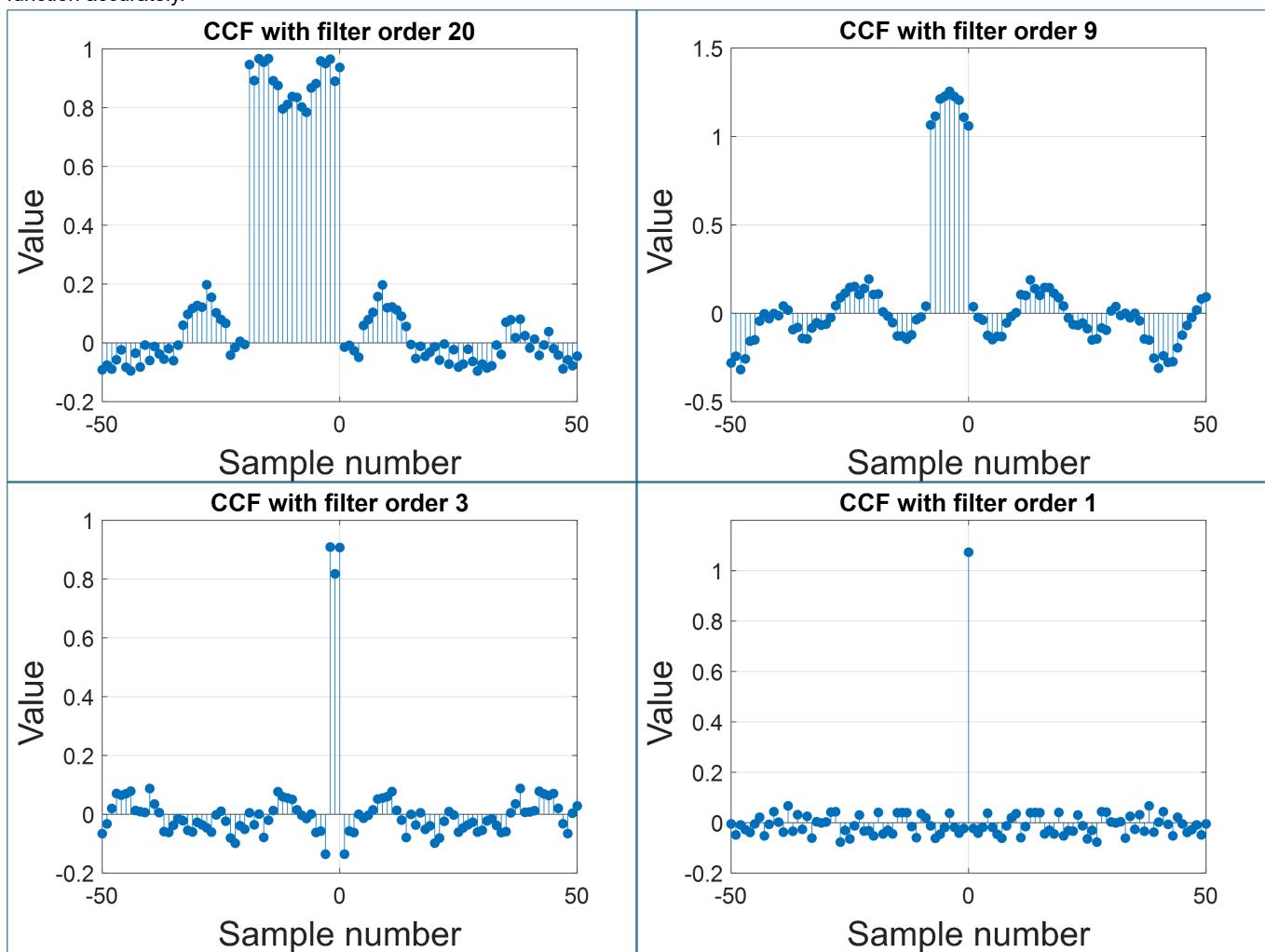
The CCF and ACF can be used in system identification to find the impulse response of a system. Primitive steps to do so are as follows:

- **Excite system** with controlled input (e.g., white noise).
- **Measure output response.**
- **Compute cross-correlation (CCF)** between input and output.
- **Compute autocorrelation (ACF)** of input.
- **Impulse response** $h[k]$ is proportional to CCF if input is white noise ($R_{XY}[k] \propto h[k] R_{XY}[k] \propto h[k]$).
- **Unbiased estimator** under ideal conditions (white noise, no measurement noise).

The effect of filter length on the shape of the CCF of y

The higher the filter order, the longer, in time, that the CCF "pulse" is present for.

The shape becomes more sinusoidal, as mentioned previously, this can be explained using the low pass nature of the MA filter. Order 20 is a stronger low pass filter than Order 1, hence is more band limited and thus struggles to approximate the discontinuous rectangle function accurately.



2.3 Autoregressive Modelling

Question 1

Stability analysis would require that one investigates the roots of the characteristic polynomial.

For the random process:

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + w[n]$$

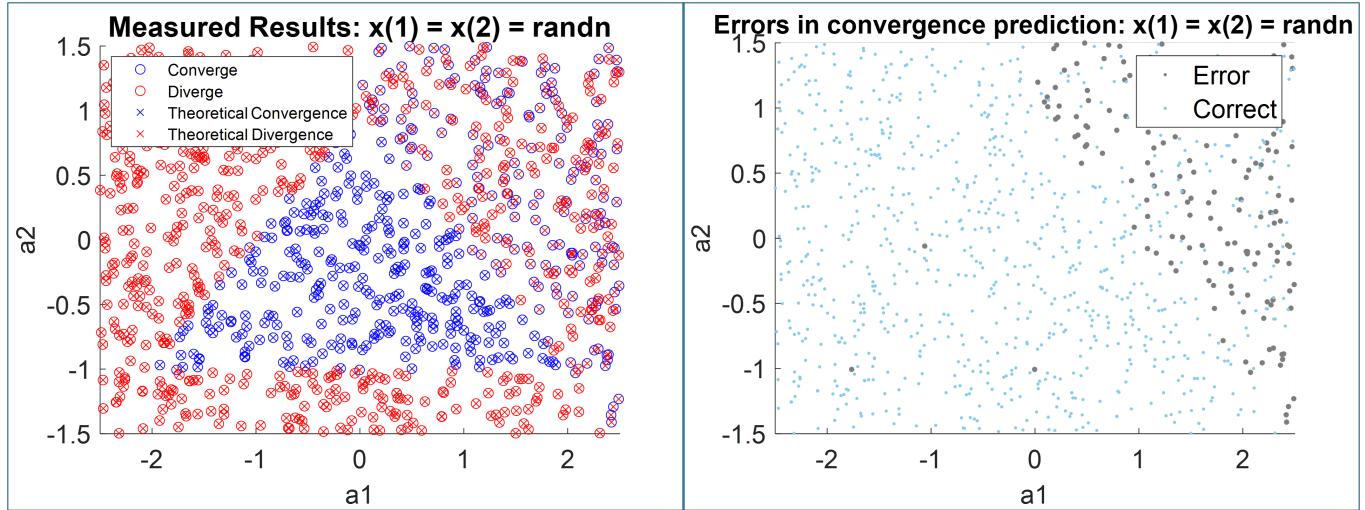
The characteristic equation becomes:

$$1 - a_1 z^{-1} - a_2 z^{-2} = 0$$

For stability, the **all the roots must exist within the unit circle of the z-plane**

This is done in the code with the following code:

```
a = [1, -a1, -a2];
z = roots(a);
theoretical_convergence = all(abs(z) < 1);
```



The system's poles must lie within the unit circle for stability. The roots of the characteristic equation $C(z) = z^2 - a_1z - a_2$ are

$z = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2}$. For the positive root, we have:

$$\begin{aligned} \frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} &< 1 \\ \sqrt{a_1^2 + 4a_2} &< (2 - a_1) \\ a_1^2 + 4a_2 &< 4 + a_1^2 - 4a_1 \\ a_1 + a_2 &< 1 \end{aligned}$$

Similar calculations for the negative root give $a_2 - a_1 < 1$.

The characteristic equation can now be written as:

$$\begin{aligned} C(z) &= (z - z_1)(z - z_2) = 0 \\ &= z^2 - (z_1 + z_2)z + z_1z_2 = 0 \\ \implies a_1 &= -(z_1 + z_2) \\ a_2 &= z_1z_2 \end{aligned}$$

Since both roots are less than 1, $|a_2| < 1$.

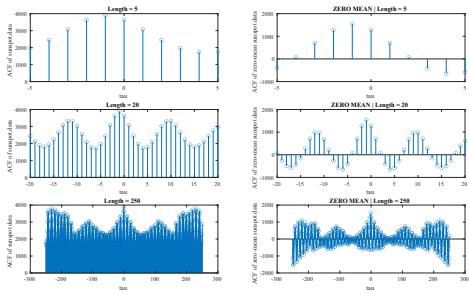
Thus the triangular plane of stability is defined by the following inequalities:

$$\begin{aligned} a_1 + a_2 &< 1 \\ a_1 - a_2 &< 1 \\ |a_2| &< 1 \end{aligned}$$

Question 2

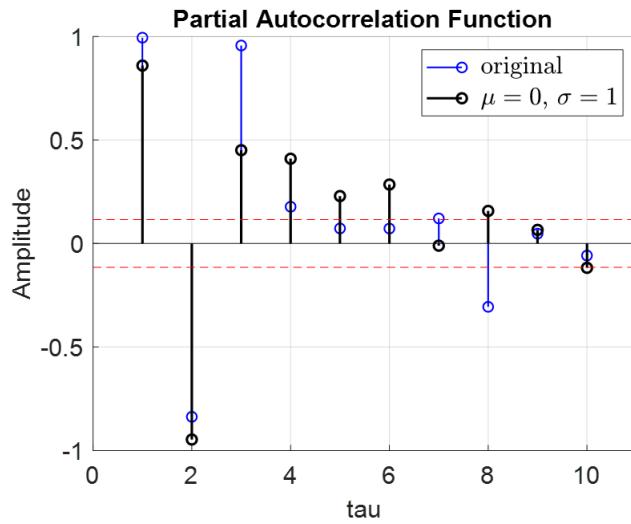
For $N = 5$, no recurrence is seen. For $N = 20$, a 13-year pattern emerges. For $N = 250$, irregularities appear due to exceeding the $\frac{N}{4}$ bound. Within this bound, a sinusoid is clear.

Zero-mean data prevents inflated values from shifted element products, revealing clearer trends. It corrects misleading similarities and clarifies periodicity.



Question 3

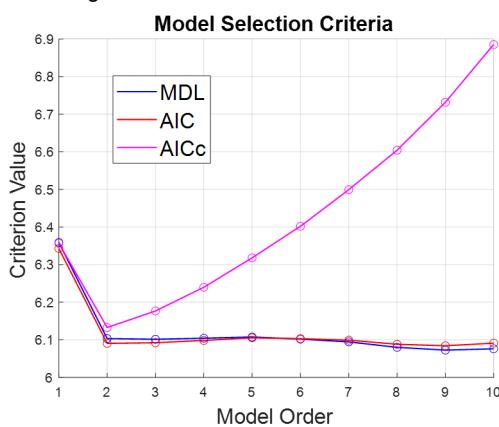
The figure below shows that model order 2 is optimal, as other coefficients are significantly closer to 0. The standardized PCF is more accurate, removing offsets that obscure trends. The PCF's sharp drop after order 2 indicates weak correlation beyond 2 time units, defining an AR(2) process.



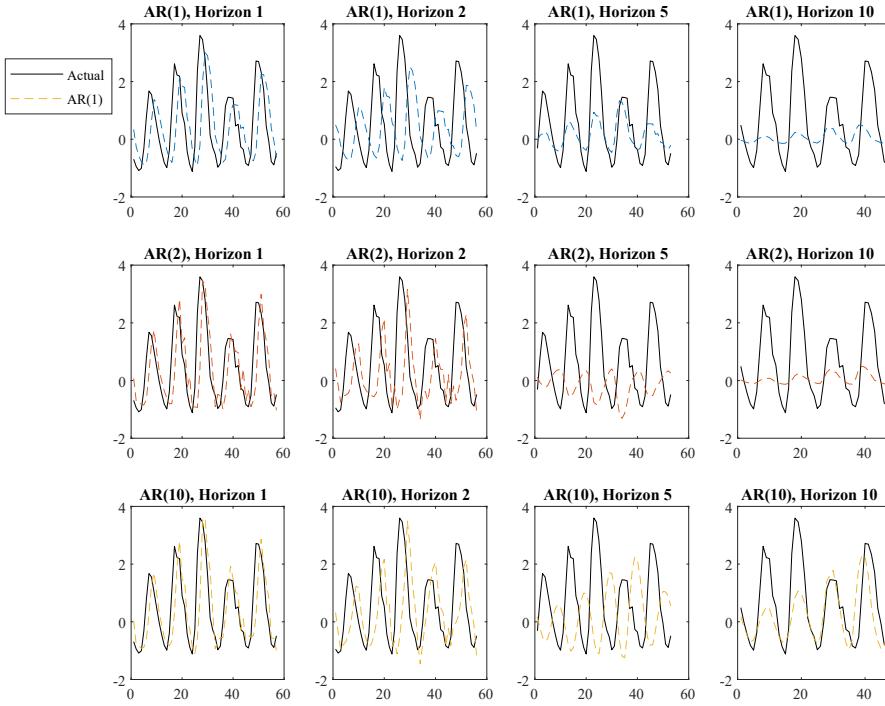
Question 4

All three metrics aim to balance the reduction in error with the increase in complexity, they do so by different penalty terms.

For Minimum description length (MDL) and Akaike Information Criterion (AIC), the global minimum would indicate that model order of 9 is ideal. However, the first minimum is at the order of 2 and as expected, with the corrected AIC which is better for smaller sample sizes, the global minimum becomes more clear.



Question 5



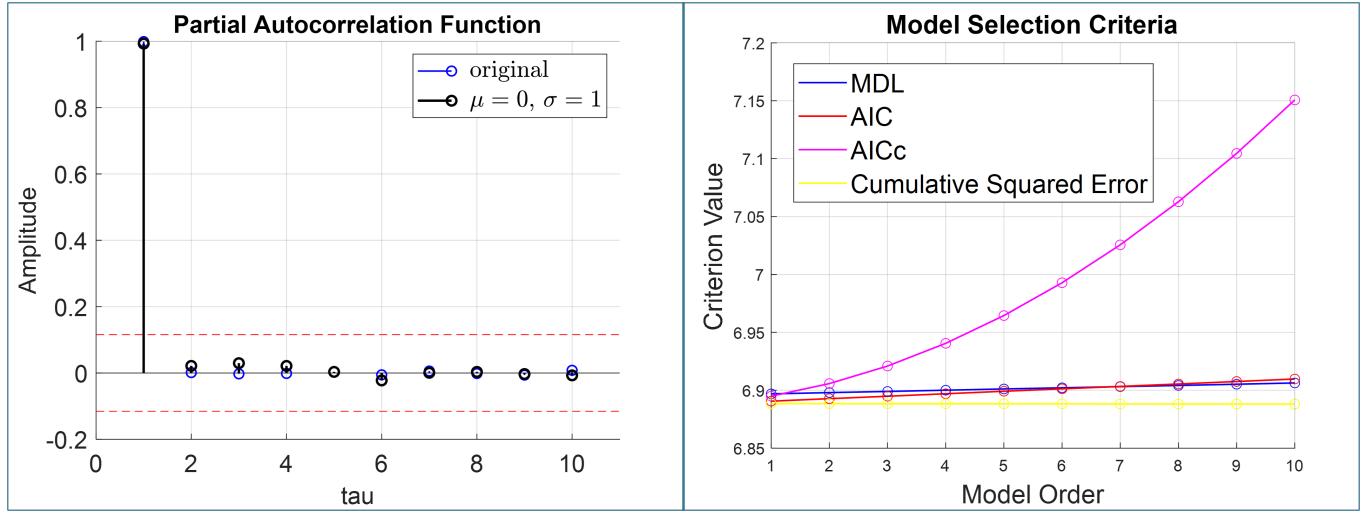
The results show that increasing model order allows for more accurate predictions over a longer horizon. AR(1) and AR(2) both fail at larger horizons, as we can see a model fail to capture the pattern.

This can be justified as in multi-step forecasting, error accumulation increases with prediction horizon, paradoxically favoring higher complexity models that overfit less in longer predictions.

The tradeoff between model order and prediction error becomes clear: extrapolation favours higher model orders, however, it increases the chances of overmodelling, where increased variance makes predictions more sensitive to noise.

2.4 Cramer-Rao Lower Bound

Question 1a.



The PCF drops of far below the threshold of $\pm 1.96/\sqrt{N}$ for model order, $p > 1$.

The information theoretic criteria also show the same. The MDL and AIC do not show any obvious global minimum and the AIC corrected for smaller model order shows a clear optimal model order of $p = 1$

Question 1b.

The following derivations have been sourced from Stephen Kay's 'Fundamentals of Signal Processing: Estimation Theory'. The PSD implied by the AR model is:

$$P_{xx}(f; \theta) = \frac{\sigma_u^2}{|A(f)|^2} \quad (1)$$

where $\theta = [a[1], a[2], \dots, a[p], \sigma_u^2]^T$ and $A(f) = \sum_{m=1}^p a[m] \exp(j2\pi f m)$. The partial derivatives are:

$$\frac{\partial \ln P_{xx}(f; \theta)}{\partial a[k]} = \frac{\partial \ln |A(f)|^2}{\partial a[k]} \quad (2)$$

$$= \frac{1}{|A(f)|^2} [A(f) \exp(j2\pi fk) + A^*(f) \exp(-j2\pi fk)] \quad (3)$$

$$\frac{\partial \ln P_{xx}(f; \theta)}{\partial \sigma_u^2} = \frac{1}{\sigma_u^2} \quad (4)$$

For $k = 1, 2, \dots, p$ and $l = 1, 2, \dots, p$, we have:

$$I[\theta]_{kl} = \frac{N}{2} \int_{-0.5}^{0.5} \frac{1}{|A(f)|^4} [A(f) \exp(j2\pi fk) + A^*(f) \exp(-j2\pi fk)] \times [A(f) \exp(j2\pi fl) + A^*(f) \exp(-j2\pi fl)] df \quad (5)$$

$$= \frac{N}{2} \int_{-0.5}^{0.5} \frac{1}{A^*(f)^2} \exp(j2\pi f(k+l)) + \frac{1}{|A(f)|^2} \exp(j2\pi f(k-l)) + \frac{1}{|A(f)|^2} \exp(j2\pi f(l-k)) + \frac{1}{A^2(f)} \exp(-j2\pi f(l+k)) df \quad (6)$$

Using the Hermitian property of the integrand ($A(f) = A^*(f)$), we have

$$\int_{-0.5}^{0.5} \frac{1}{A^*(f)^2} \exp(j2\pi f(k+l)) df = \int_{-0.5}^{0.5} \frac{1}{A^2(f)} \exp(-j2\pi f(l+k)) df \quad (7)$$

$$\int_{-0.5}^{0.5} \frac{1}{|A(f)|^2} \exp(j2\pi f(k-l)) df = \int_{-0.5}^{0.5} \frac{1}{|A(f)|^2} \exp(j2\pi f(l-k)) df \quad (8)$$

Therefore,

$$I[\theta]_{kl} = N \int_{-0.5}^{0.5} \frac{1}{A^*(f)^2} \exp(j2\pi f(k+l)) df + N \int_{-0.5}^{0.5} \frac{1}{|A(f)|^2} \exp(j2\pi f(k-l)) df \quad (9)$$

The second integral is the inverse Fourier transform of $\frac{1}{A^*(f)^2}$ evaluated at $n = k+l > 0$. This term is 0 since the sequence is the convolution of 2 anti-causal sequences, i.e.,

$$\mathcal{F}^{-1} \left[\frac{1}{A(f)} \right] = \begin{cases} h[n], & n \geq 0 \\ 0, & n < 0 \end{cases} \quad (10)$$

$$\mathcal{F}^{-1} \left[\frac{1}{A^*(f)^2} \right] = \begin{cases} h[n] * h[n], & n < 0 \\ 0, & n \geq 0 \end{cases} \quad (11)$$

Therefore,

$$I[\theta]_{kl} = \frac{N}{\sigma_u^2} r_{xx}(k-l) \quad (12)$$

We are given I_{11} , I_{12} and I_{21} . For $k = p+1$ and $l = p+1$:

$$I[\theta]_{kl} = \frac{N}{2} \int_{-0.5}^{0.5} \frac{1}{\sigma_u^4} df = \frac{N}{2\sigma_u^4} \quad (13)$$

Therefore,

$$I(\theta) = \begin{bmatrix} \frac{N}{\sigma_u^2} R_{xx} & \mathbf{0} \\ \mathbf{0}^T & \frac{N}{2\sigma_u^4} \end{bmatrix} \quad (14)$$

where $[R_{xx}]_{ij} = r_{xx}(i-j)$ is a $p \times p$ Toeplitz autocorrelation matrix and $\mathbf{0}$ is a $p \times 1$ vector of zeros.

Question 1c.

$$\text{var}(a[\hat{k}]) \geq \frac{\sigma_u^2}{N} [R_{xx}^-]_{kk} \quad \text{for } k = 1, 2, \dots, p \quad (15)$$

$$\text{var}(\sigma_u^2) \geq \frac{2\sigma_u^4}{N} \quad (16)$$

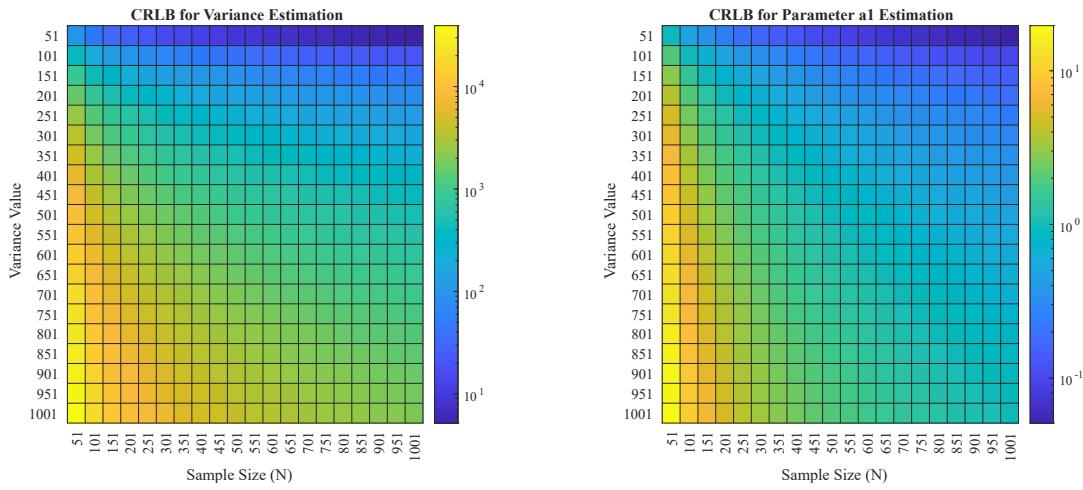
$$\text{Using } p = 1, (a[\hat{1}]) \geq \frac{\sigma_u^2}{Nr_{xx}[0]} \quad (17)$$

$$\text{But } r_{xx}[0] = \frac{\sigma_u^2}{1 - a^2[1]} \quad (18)$$

$$\text{So that } \text{var}(a[\hat{1}]) \geq \frac{1}{N} (1 - a^2[1]) \quad (19)$$

Estimation of the filter parameter improves when $a[1]$ nears 1, not 0; because the filter's pole resides at $-a[1]$, processes with PSDs displaying sharp peaks offer enhanced parameter estimation.

Cramer-Rao Lower Bound Analysis of NASDAQ Data



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.

Question 1d.

For $p = 1$,

$$\hat{P}_x(f; \theta) \Big|_{a_1} = \frac{\sigma^2}{2} [e^{j2\pi f} A(f) + e^{-j2\pi f} A(f)] \frac{1}{|A(f)|^4} \quad (20)$$

$$= \frac{\sigma^2}{2} [e^{j2\pi f} a_1 + e^{-j2\pi f} a_1] \frac{1}{|A(f)|^4} \quad (21)$$

$$= \frac{\sigma^2}{2} [2 \cos(2\pi f) a_1] \frac{1}{|A(f)|^4} \quad (22)$$

$$= \frac{\sigma^2 2 \cos(2\pi f) a_1}{|A(f)|^4} \quad (23)$$

and

$$\hat{P}_x(f; \theta) \Big|_{\sigma^2} = \frac{1}{|A(f)|^2} \quad (24)$$

Thus,

$$\text{var}(\hat{P}_x(f; \theta)) \leq \left[\frac{\sigma^2 2 \cos(2\pi f) a_1}{|A(f)|^4} \right] \left[\frac{1}{|A(f)|^2} \right] [1 - a_1^2] \frac{1}{N} \quad (25)$$

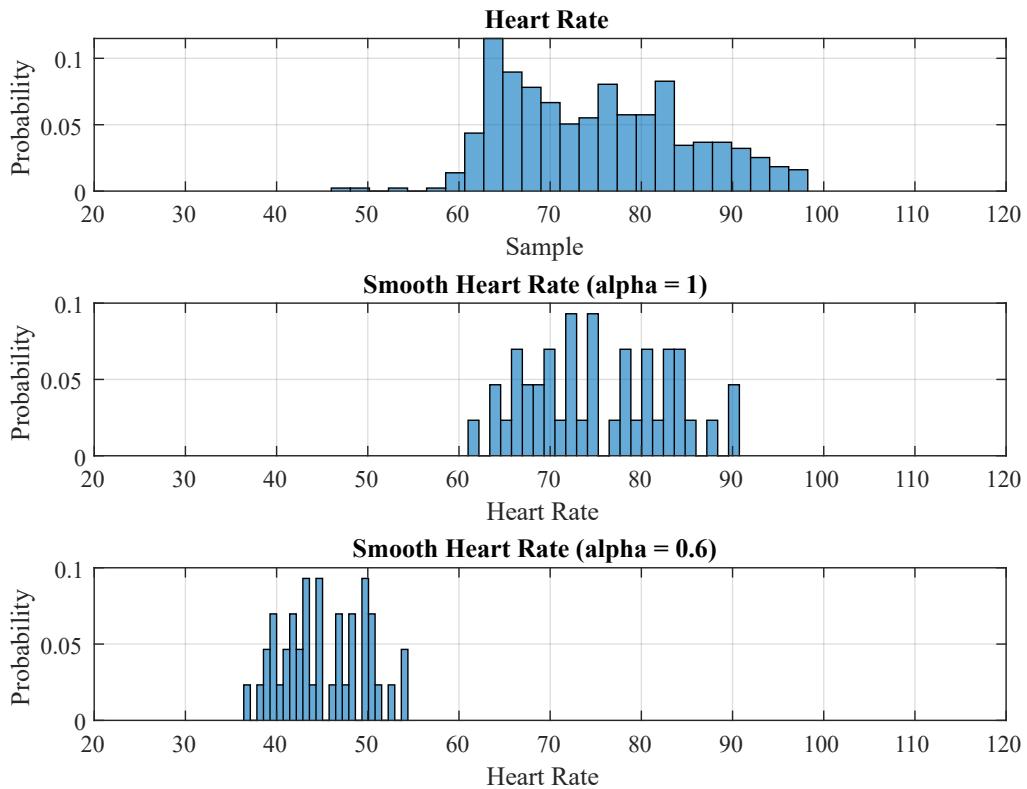
$$= \left[\frac{\sigma^4 2 \cos(2\pi f) a_1}{|A(f)|^4} \right] \left[\frac{1}{|A(f)|^2} \right] [1 - a_1^2] \frac{1}{N} \quad (26)$$

$$= \frac{1}{N} \left[\frac{\sigma^4 4 (\cos(2\pi f) a_1)^2 (1 - a_1^2)}{|A(f)|^8} + \frac{\sigma^4}{|A(f)|^4} \right] \quad (27)$$

2.5 Real World Signals: ECG from iAMP experiment

Part a.

The recorded data is impacted by my illness during the session, resulting in an unstable heart rate and light-headedness. Consequently, the heart rate distribution deviates from a normal distribution.



Part b.

Raw heart rate exhibits higher variance than smoothed data, suggesting lower Fisher information.

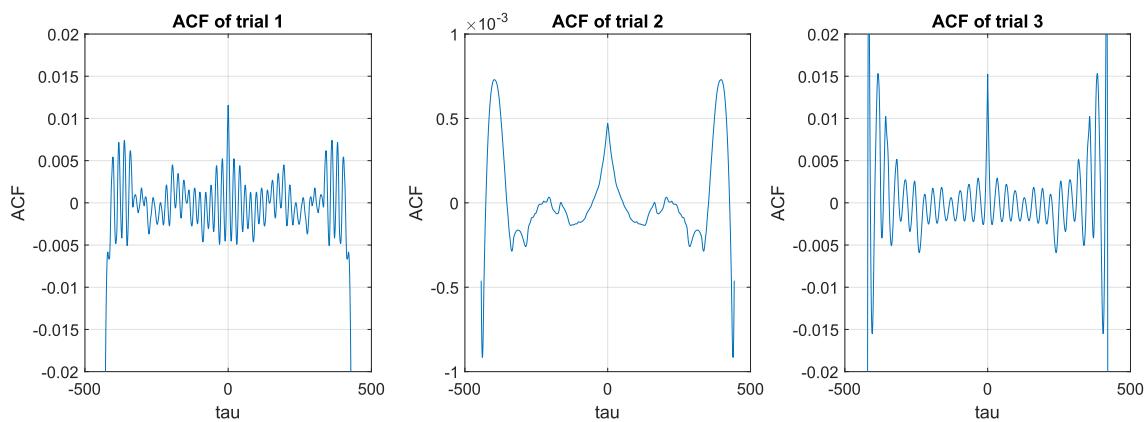
Averaging by a factor of 10 reduces variance to $\sigma/10$, visually confirmed by the compact distribution at $\alpha = 0.6$.

The α factor also reduces noise bias by shifting the averaged heart rate mean, \hat{h} . Notably, when $\alpha = 1$, the means of h and \hat{h} coincide, implying lower mean and variance for $\alpha = 0.6$.

Part c. - AR modelling of heart rate

The process is autoregressive due to it being sinusoidal and infinite in length.

A moving average process have a finite length autocorrelation sequence.



Part d. PCF and Information theoretic criterion

Trial 1

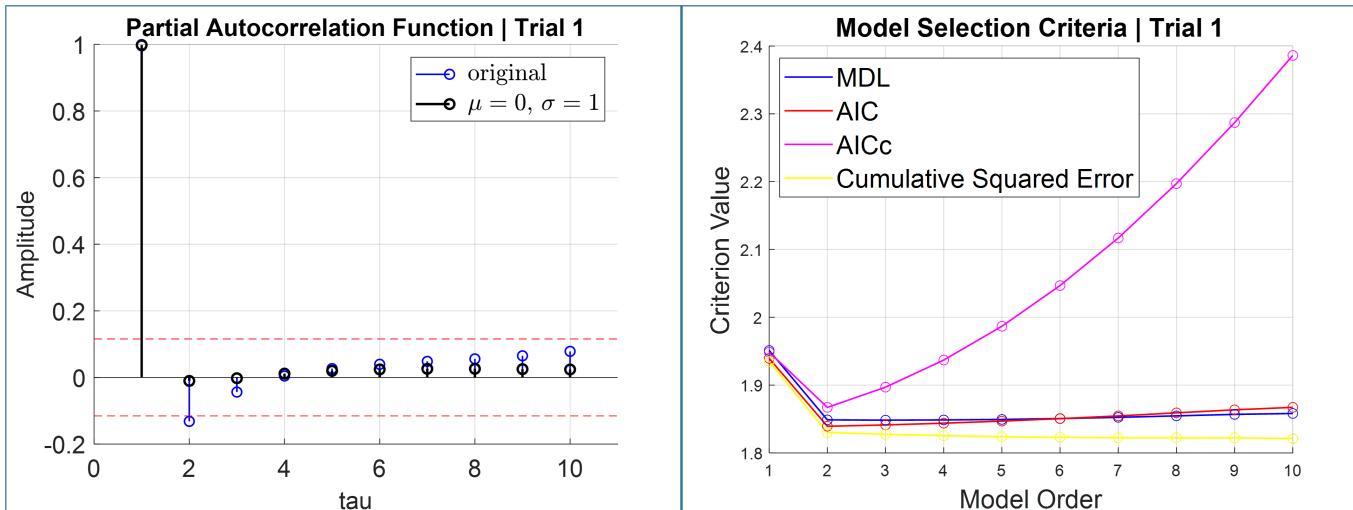
The partial autocorrelation function implies that the ideal model order is 1.

The global minimum of the different metrics are as follows:

- MDL is order 3
- AIC is order 2
- corrected AIC is order 2

- Cumulative MSE is order 10

Given that the only clear elbow is from the corrected AIC at an order of 2, and that the PACF strongly favours a lower model order, this process can be modelled as an AR(2) system.



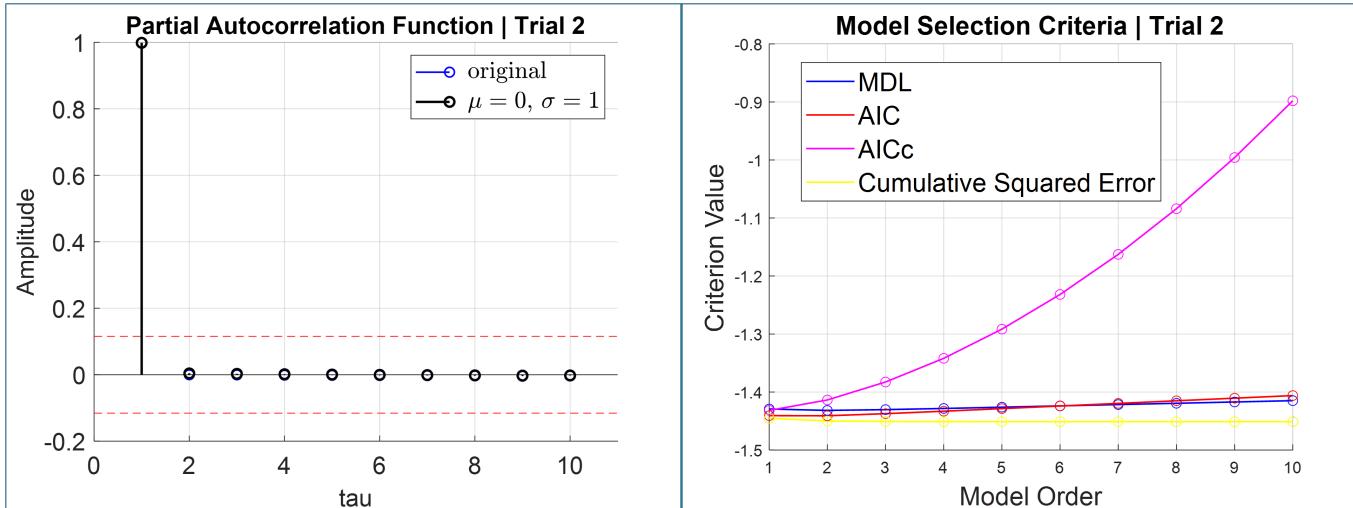
Trial 2

The partial autocorrelation function implies that the ideal model order is 1.

The global minimum of the different metrics are as follows:

- MDL is order 2
- AIC is order 2
- corrected AIC is order 1
- Cumulative MSE is order 10

Given the results, we can conclude that this process can be modelled as an AR(1) process.



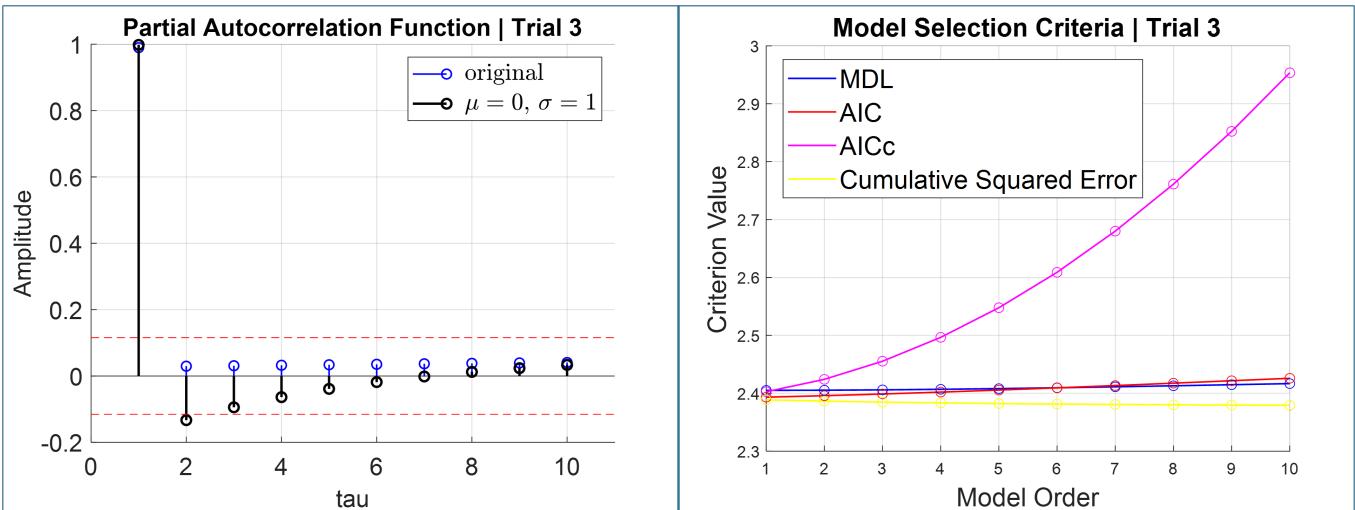
Trial 3

The partial autocorrelation function implies that the ideal model order is 1.

The global minimum of the different metrics are as follows:

- MDL is order 1
- AIC is order 1
- corrected AIC is order 1
- Cumulative MSE is order 10

Given the results, we can conclude that this process can be modelled as an AR(1) process.

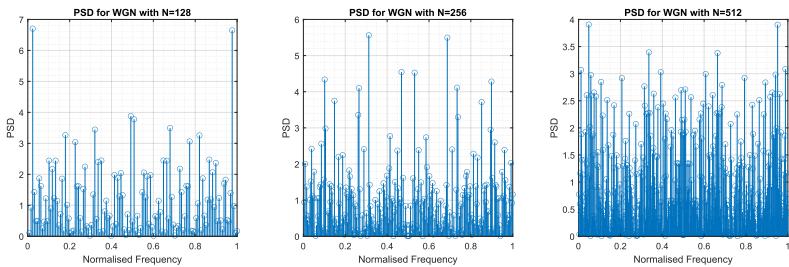


Question 1 - Periodogram function

The periodogram accurately reflects the symmetric PSD of real WGN input.

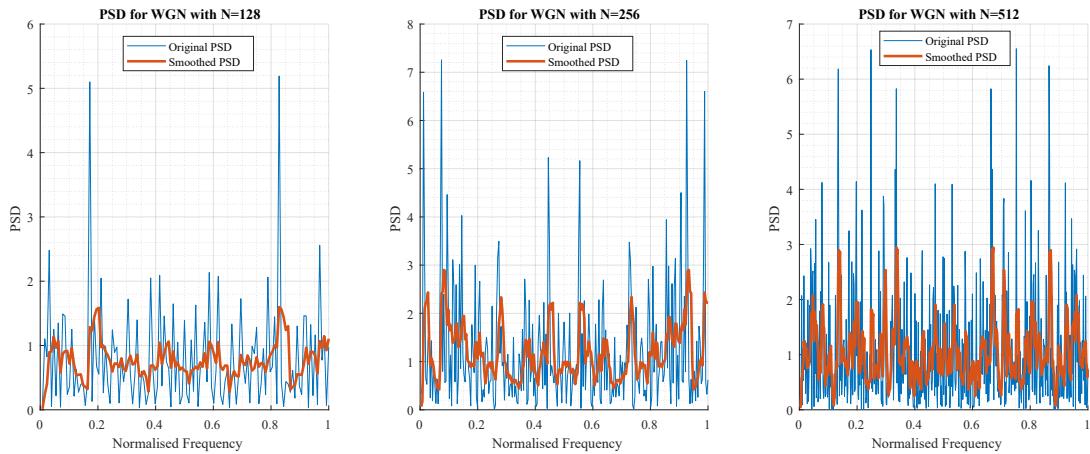
However, it deviates from the ideal flat PSD. The PSD is derived from the ACF's FFT and the estimator's variance increases with lag. This impacts the periodogram's variance.

Increased sample size improves mean convergence to the ideal PSD. However, the variance remains. Therefore, the periodogram estimator is inconsistent."



3.1 Averaged periodogram estimates

Question 1



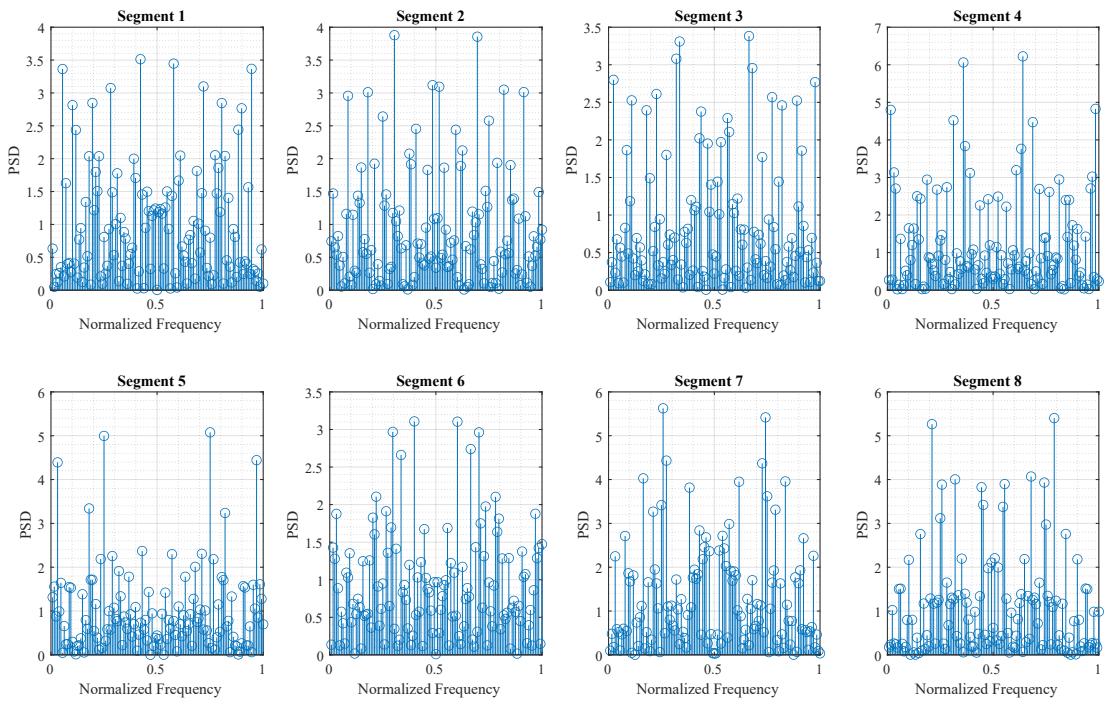
Smoothing approximates the function, retaining important trends while removing noise and other sharp details.

Visually, the smoothed periodogram shows lower variance and thus appears to be a better estimate.

Question 2

Non-overlapping segments of WGN are random, and completely uncorrelated. There is no pattern between each plot and this can be attributed to the large variance of the data.

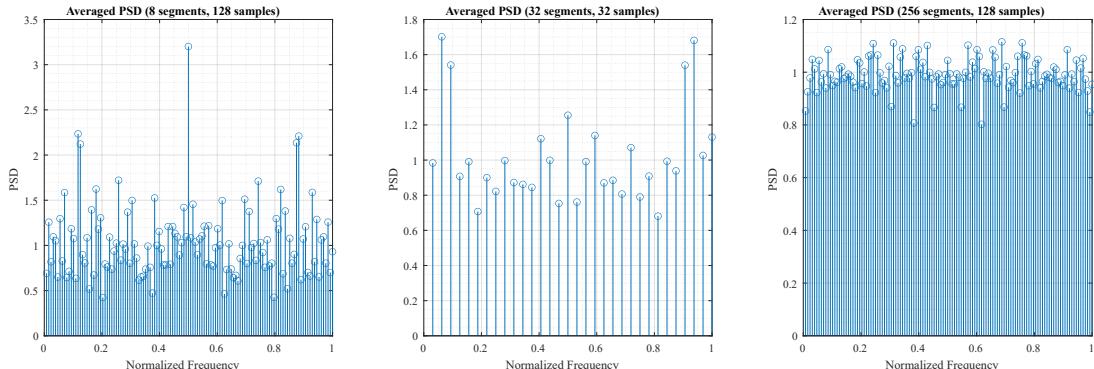
Each segment is an unreliable estimate for the ideal PSD.



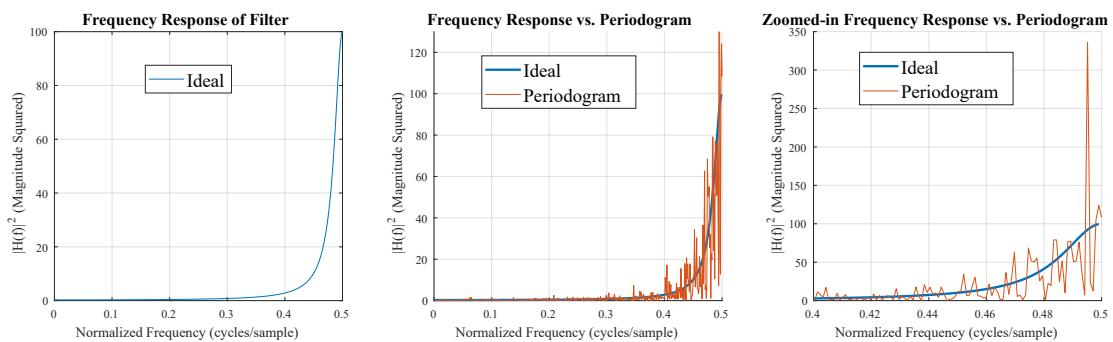
Question 3

As the number of segments increase, variance decreases linearly at the cost of lower frequency resolution for fixed number of samples (as shown by comparing 8 segment \times 128 samples and 32 segments \times 32 samples).

However, if we keep samples per segment constant and increase number of segments (and thus the total number of samples), we get a very accurate estimate of the PSD (as shown by in the, near ideal, 256 segments \times 128 samples case).



3.2 Spectrum of autoregressive process



Question 1

As signals x and y are real, the PSD is symmetric about 0.5 Hz.

Inspecting the ideal PSD of the output signal y . The normalised cut-off is roughly 0.4 Hz and the maximum PSD value is at 0.5 Hz.

Question 2

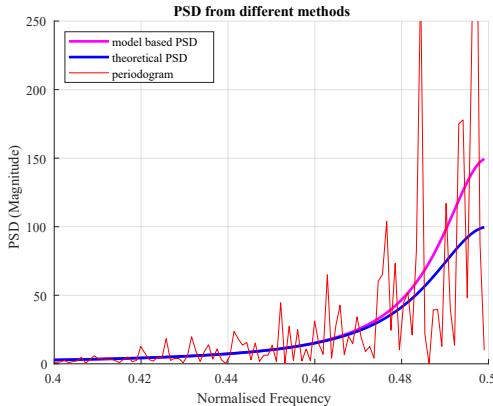
The periodogram has some error, this is expected as the estimate is based on a finite number of samples. Despite this, the periodogram manages to capture the general trend, suppressing lower frequencies and allowing high frequencies to pass.

Question 3

Zooming into the periodogram's PSD at $f = [0.4, 0.5]$ reveals significant oscillations due to the inherent rectangular windowing. This window's side lobes, equivalent to sinc function convolution $F_{rect}(\frac{t}{\alpha})(f) = a \cdot \text{sinc}(af)$, minimally impact low frequencies but distort higher frequencies, like those in $[0.4, 0.5]$, by spectral leakage. Applying a Hamming window, which reduces side lobes, would mitigate these oscillations, improving PSD accuracy.

Question 4

Compared to the periodogram, the model based estimate for the PSD (in magenta) is a lot more accurate. In particular, the variance of the estimate is a lot lower.

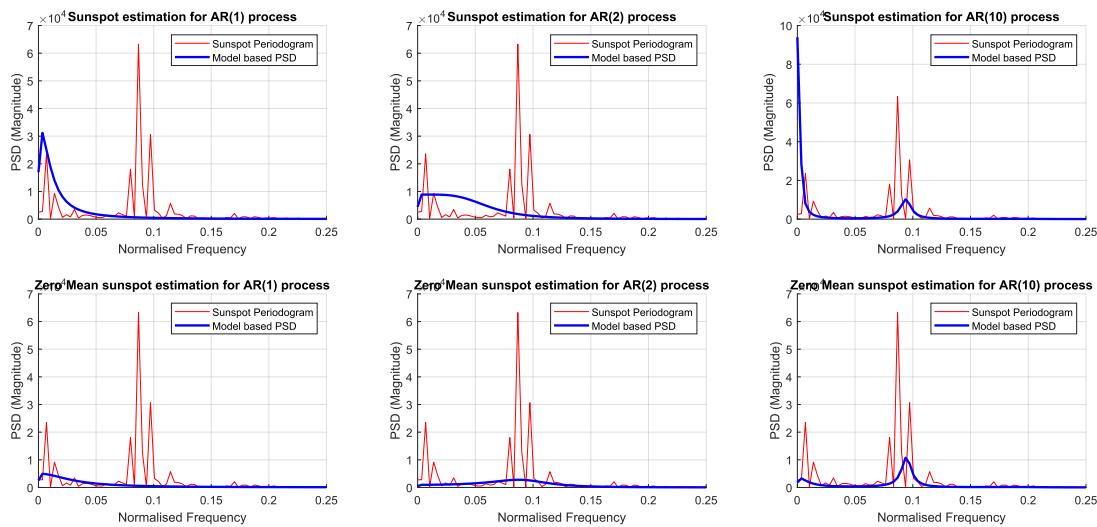


Question 5

The results support our earlier conclusion that the optimal model order is AR(2).

Low model order can only have one peak, hence it is inflexible in the spectral domain and the accuracy is poor.

High model order causes too many peaks.



Comparing the PSD estimates using mean squared error gives the following results

- order 1, $MSE = 4.06 \times 10^7$
- order 2, $MSE = 4.27 \times 10^7$
- order 10, $MSE = 9.51 \times 10^7$

Hence, clearly shown that over-modelling hurts performance.

3.3 The Least Squares Estimation of AR Coefficients

Question 1

An autoregressive moving average (ARMA) model assumes a power spectral density (PSD) of:

$$P_{xx}(f) = \frac{\sigma_u^2 |B(f)|^2}{|A(f)|^2}$$

where

$$B(f) = 1 + \sum_{k=1}^q b[k]e^{j2\pi fk}$$

$$A(f) = 1 + \sum_{k=1}^p a[k]e^{j2\pi fk}$$

Here, the $b[k]$ terms are the MA filter parameters, and the $a[k]$ terms are the AR filter parameters. The LSE approach only focuses on the AR parameters. We now take the inverse z -transform of the PSD to find the autocorrelation function (ACF):

$$P_{xx}(z) = \frac{\sigma^2 B(z)B(z^{-1})}{A(z)A(z^{-1})}$$

where $B(f) = B(e^{j2\pi f})$ and $A(f) = A(e^{j2\pi f})$.

$$\mathcal{Z}^{-1}[A(z)P_{xx}(z)] = \mathcal{Z}^{-1}\left[\sigma^2 \frac{B(z^{-1})}{A(z)}\right]$$

Since the filter impulse response is causal,

$$h[n] = \mathcal{Z}^{-1}\left[\frac{B(z)}{A(z)}\right] = 0, \quad \text{for } n < 0$$

and

$$h[n] = \mathcal{Z}^{-1}\left[\frac{B(z^{-1})}{A(z^{-1})}\right] = 0, \quad \text{for } n > 0$$

Thus, the system is anti-causal, and we have:

$$\begin{aligned} \mathcal{Z}^{-1}[\sigma^2 B(z)B(z^{-1})A(z^{-1})] &= \sigma^2 b[n] * h[n] \\ \mathcal{Z}^{-1}[\sigma^2 B(z)B(z^{-1})A(z^{-1})] &= \begin{cases} \sigma^2 b[n] * h[n], & n \leq q \\ 0, & n > q \end{cases} \\ \mathcal{Z}^{-1}[A(z)P_{xx}(z)] &= \mathcal{Z}^{-1}[\sigma^2 B(z)B(z^{-1})A(z^{-1})] = 0, \quad \text{for } n > q \end{aligned}$$

The difference equation of the ACF for $n > q$ can now be written as:

$$\sum_{k=0}^p a[k]r_{xx}[n-k] = 0, \quad \text{for } n > q$$

where $a[0] = 1$. These equations are known as the **modified Yule-Walker equations** since they are identical to the original Yule-Walker equations except for the fact that they hold for $n > 0$. Assuming $x[n]$ is available from $0, 1, \dots, (N-1)$, and the ACF is estimated for lags $n = 0, 1, \dots, M$, where $M \leq N-1$, then the LSE of $a[k]$ will minimize:

$$J = \sum_{n=q+1}^M \left[\hat{r}_{xx}[n] - \sum_{k=1}^p a[k]\hat{r}_{xx}[n-k] \right]^2$$

which can be written in matrix form as:

$$J = (x - H\theta)^T(x - H\theta)$$

where

$$\begin{aligned} x &= \begin{bmatrix} \hat{r}_{xx}[q+1] \\ \hat{r}_{xx}[q+2] \\ \vdots \\ \hat{r}_{xx}[M] \end{bmatrix} \\ \theta &= \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} \\ H &= \begin{bmatrix} \hat{r}_{xx}[q] & \hat{r}_{xx}[q-1] & \cdots & \hat{r}_{xx}[q-p+1] \\ \hat{r}_{xx}[q+1] & \hat{r}_{xx}[q] & \cdots & \hat{r}_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}[M] & \hat{r}_{xx}[M-1] & \cdots & \hat{r}_{xx}[M-p] \end{bmatrix} \end{aligned}$$

Question 2

The least squares estimate is given by, $\hat{a} = (H^T H)^{-1} H^T x$.

This is known as the least squares modified Yule-Walker equation, where LSE is applied to estimated ACF values instead of actual data. While H is typically deterministic, here it becomes random due to finite-sample ACF estimation. This randomness increases for larger M , as ACF estimates at higher lags suffer from greater variance due to averaging over fewer terms.

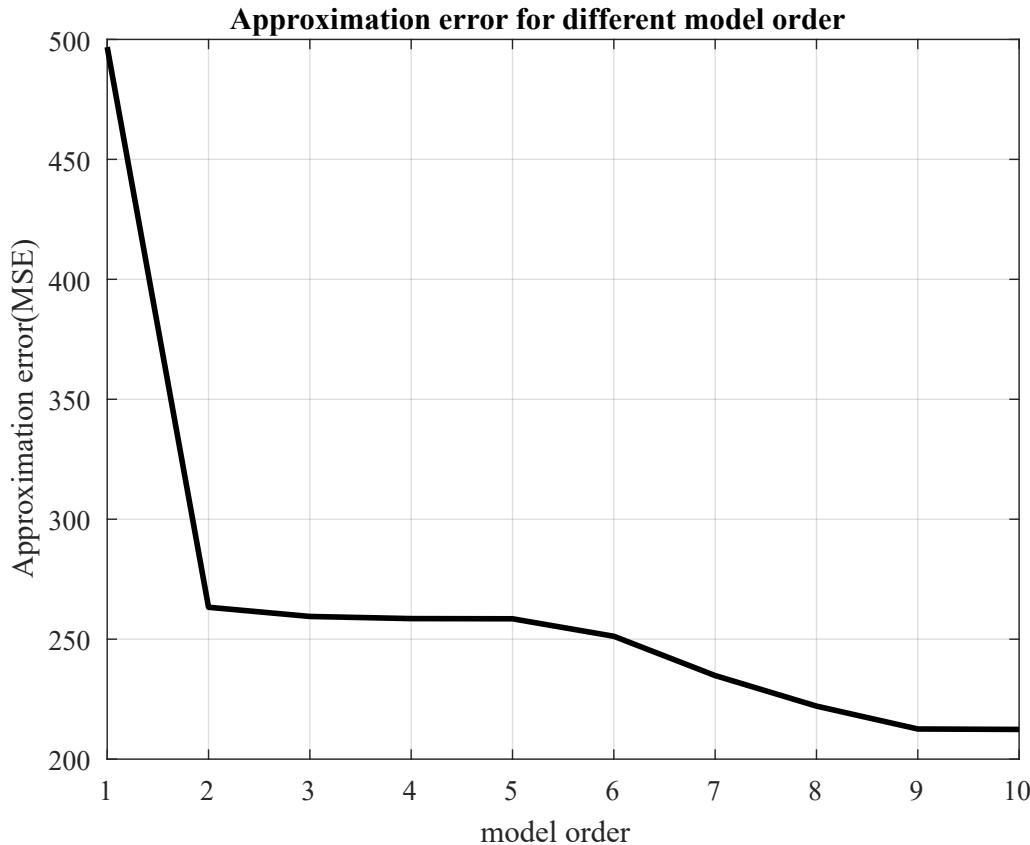
Question 3

The following table shows the coefficients for model orders up to 10.

No.	1	2	3	4	5	6	7	8	9	10	11	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	

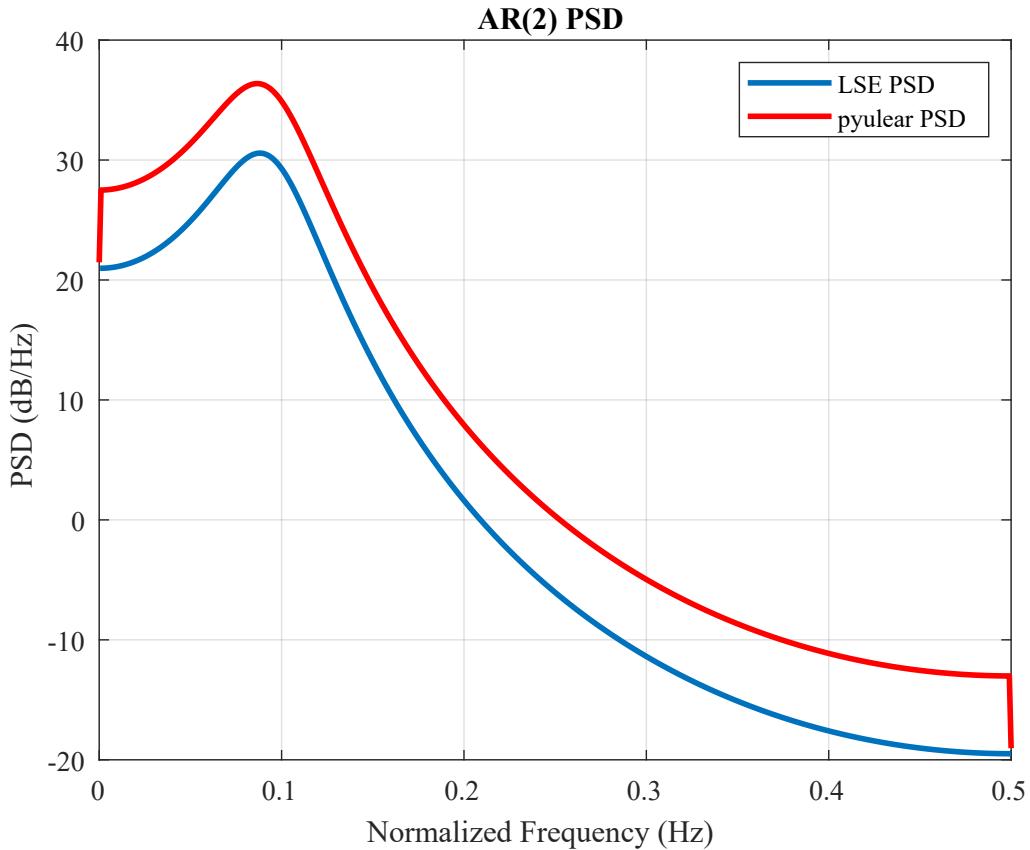
Question 4

Model order 10 has the lowest mean squared error. However, we know that lower model orders are more favourable for generalisability. Given that model order 2 has a close enough error with significantly less computational complexity, we once again choose model AR(2) at the optimal.



Question 5

The LSE approach provides a good estimate of the Yule-Walker method, sharing the same pole location at 0.1 Hz

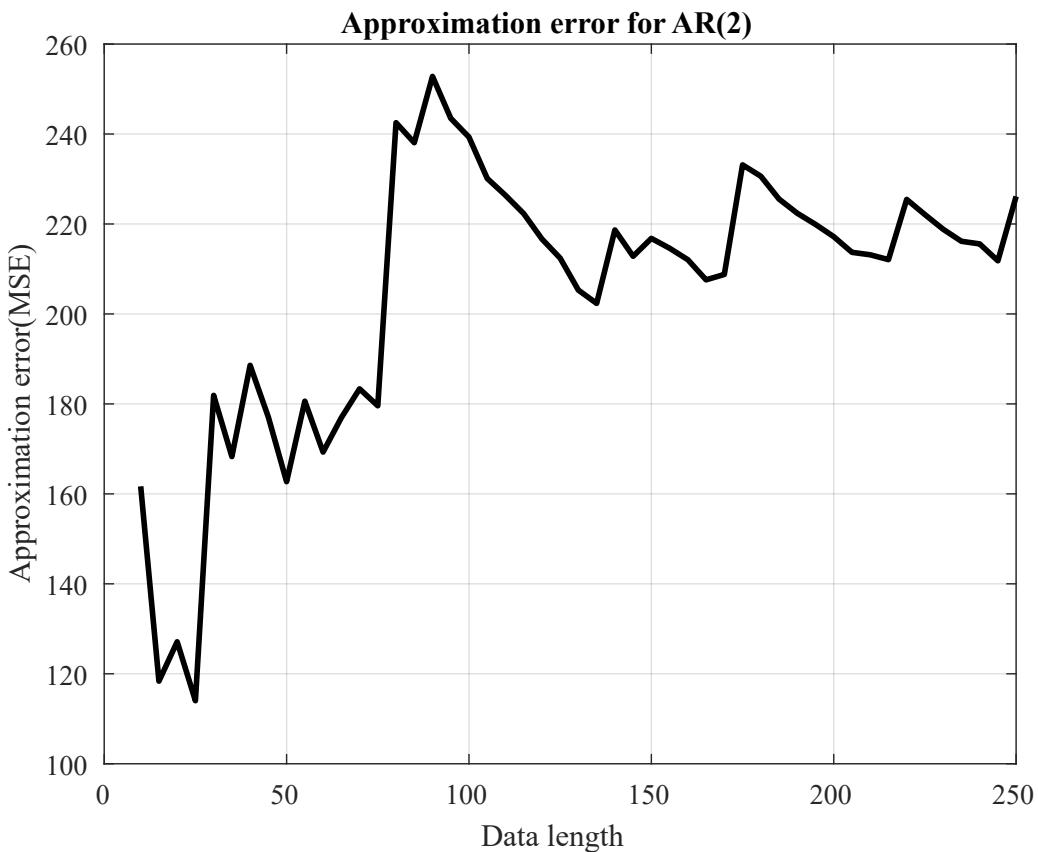


Question 6

The minimum is found at $N = 25$ and the maximum is found at $N = 100$.

The MSE plateaus for $N > 150$, implying that increase in data length will not improve performance. This implies that the AR(2) model has 'reached the limit' of what it can capture.

There we suggest that 25 as the optimal data length for the AR(2) modelling of sunspot time series data.



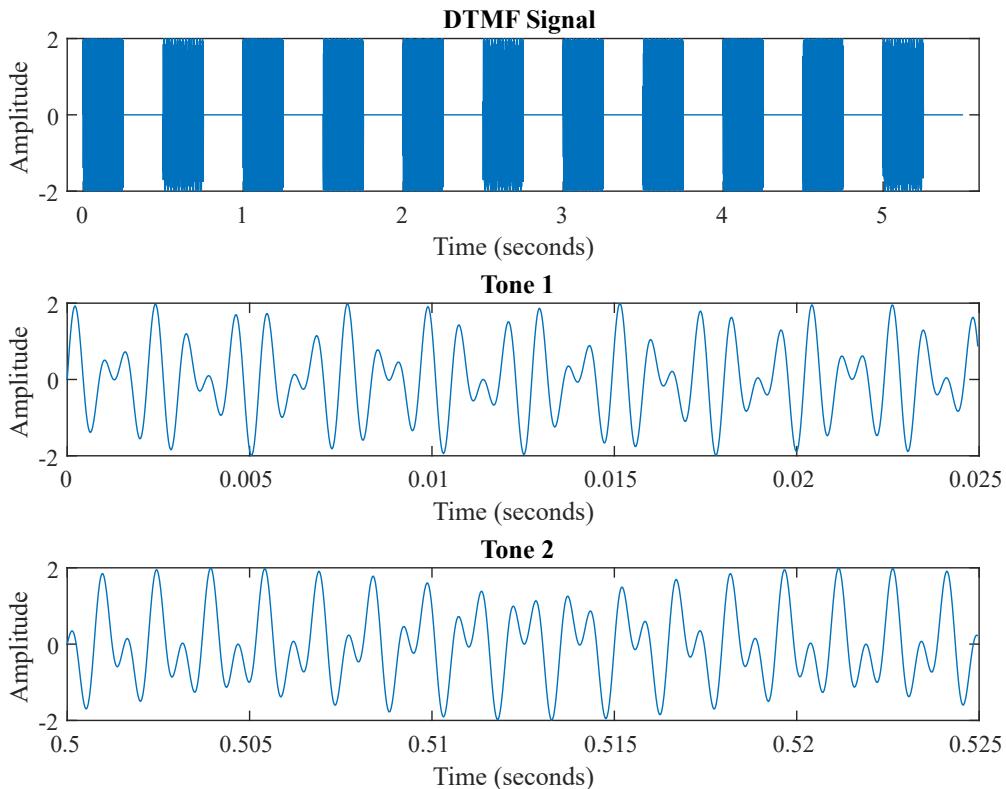
3.4 Spectrogram for time-frequency analysis: dial tone pad

Phone number used for analysis is 02018211092

Question 1

The Nyquist frequency must be $2 \times 1477 = 2954$ Hz.

Oversampling with a frequency of 32768 Hz will boost SNR. The sampling frequency is also a direct power of 2, $2^{15} = 32768$, which will make the FFT computationally efficient, for example during the bit-reversal permutation.



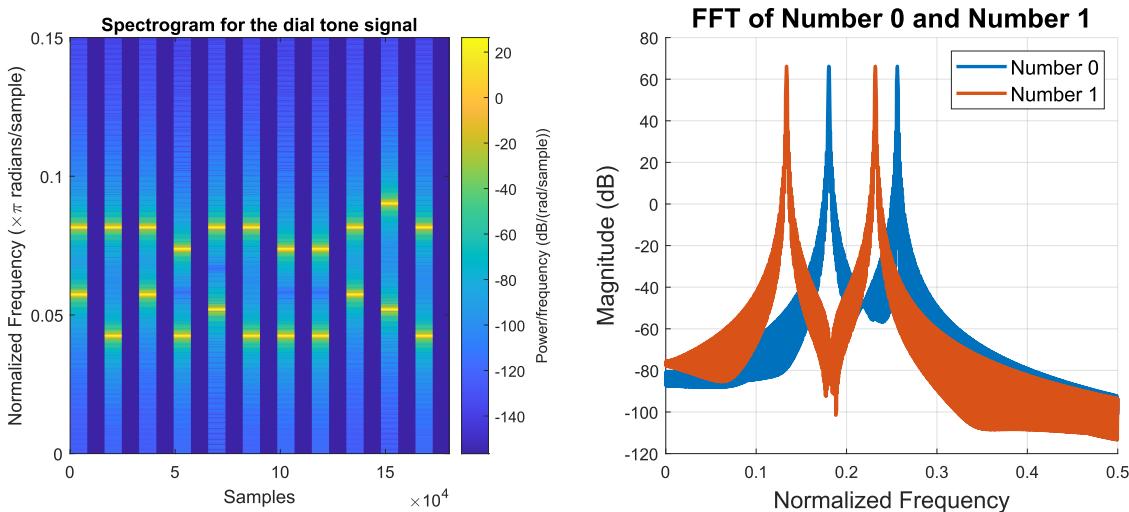
Question 2

The spectrogram clearly shows peaks in power at discrete frequencies for each digit being pressed.

The two horizontal bars per tone accurately captures the DTMF system.

Without noise, the idle time has zero power, which is reflected in the uniformity of the dark blue.

The Hann window's side lobes cause spectral spreading in the spectrogram. Using a finite 8192-sample window, derived from a 0.25-second key duration at 32768 samples per second, shifts peak frequencies from their ideal values. This displacement, influenced by the window length, is a limitation of the Discrete Fourier Transform and the time-frequency uncertainty principle.



Question 3

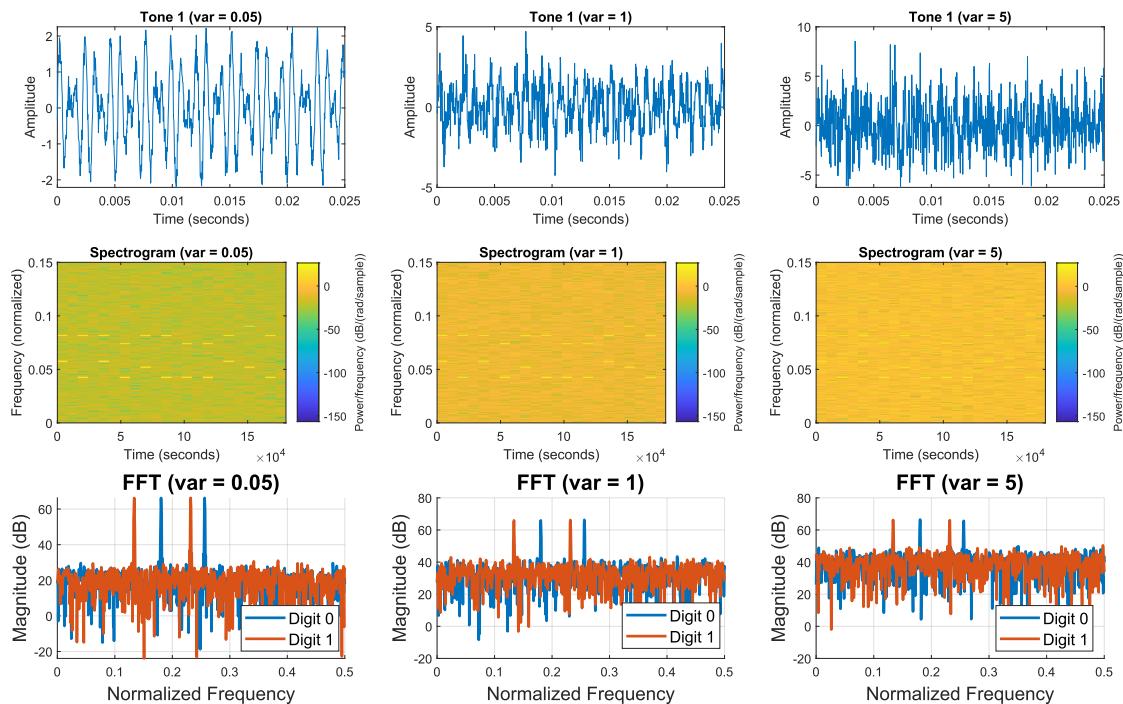
DTMF digit identification leverages the distinct frequency pairs within each 8192-sample block. By establishing tolerance bands (e.g., ± 30 Hz) around the standard DTMF frequencies (697 Hz to 1477 Hz), we accurately determine the pressed key, as no frequency components fall outside the 675 Hz to 1500 Hz range. Spectrogram analysis further confirms this, where yellow dashes correspond to the expected normalized frequencies (e.g., 0.0574π and 0.0834π radians per sample for digit '0'), validating the frequency-based detection.

Question 4

Low, medium and high noise is introduced with variances, σ_N^2 , of 0.05, 1 and 5.

Analytically, the PSD should be offset by a DC level for all frequencies. This is because the Fourier Transform is a linear transform and therefore the resulting noisy PSD = ideal PSD + PSD of noise.

The PSD of WGN is ideally a flat line with amplitude σ_N^2 , therefore the noisy PSD should be an offset from the original PSD. Since the FFT plot is in decibels, a constant increase in power across all frequencies from the added white noise will appear as a proportionally larger increase in the dB value of originally low-amplitude frequencies compared to those with high initial amplitudes. Consequently, as the noise power increases, the noise floor and the signal peaks will tend towards a similar dB level, effectively reducing the signal-to-noise ratio.



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

Respiratory sinus arrhythmia (RSA) refers to the modulation of cardiac function influenced by respiratory effort, where heart rate increases during inspiration and decreases during expiration. The strength of RSA serves as a useful indicator of cardiovascular health and can be observed in electrocardiographic (ECG) data when breathing follows a regular pattern.

The power spectral density (PSD) of the R-R interval (RRI) data is analyzed by plotting both the standard periodogram and the averaged periodogram using different window lengths (50s and 100s).

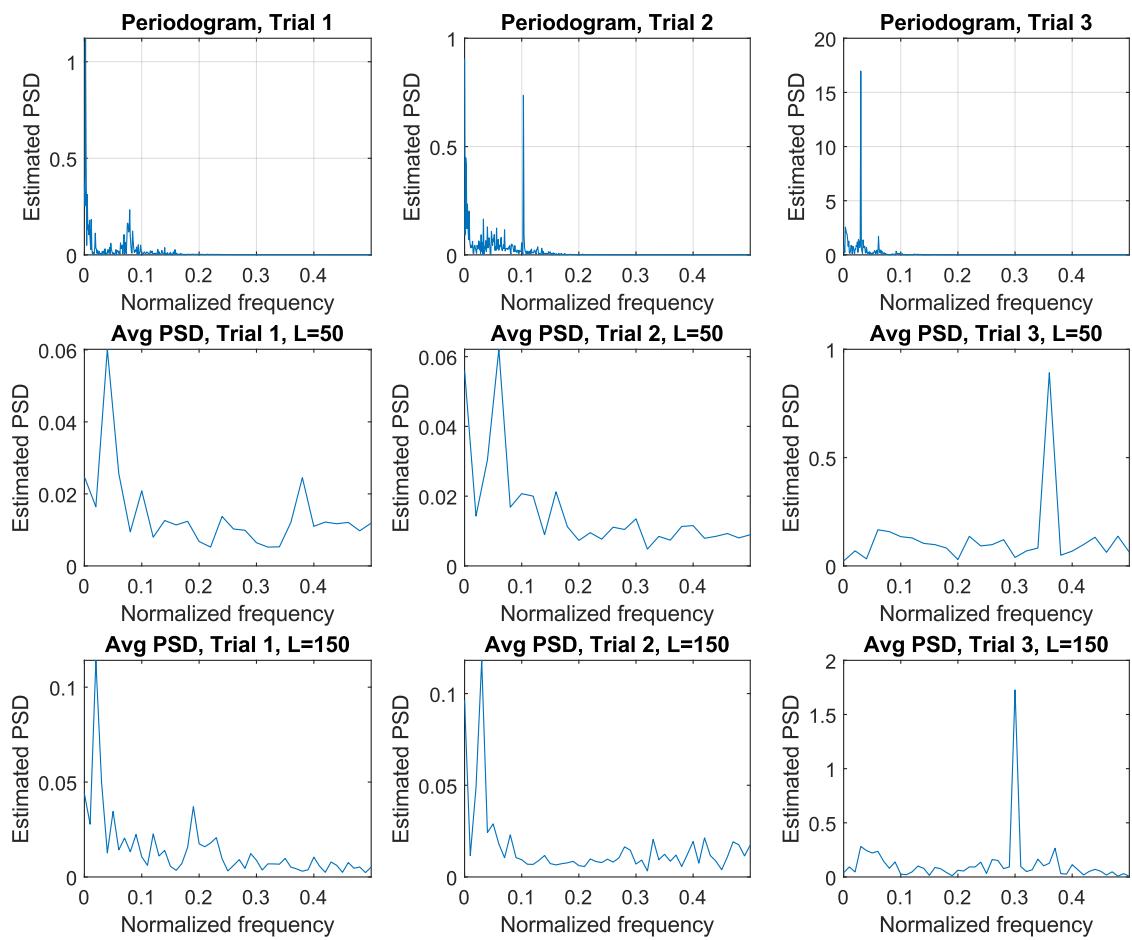
The figure below shows the standard periodogram, an average periodogram with 50 second window and an average periodogram with a 100 second window for all three trials.

For Trial 1, the unconstrained breathing allows for variable respiratory rate. The RSA-related frequency components appears as a broad peak up to $0.04 \times 4 = 0.16$ Hz.

For Trial 2, the constrained breathing at 50 bpm, the dominant peak is roughly at $0.6 \times 4 = 0.42$ Hz which slightly less than expected 0.83Hz ($50/60 = 0.833$ Hz).

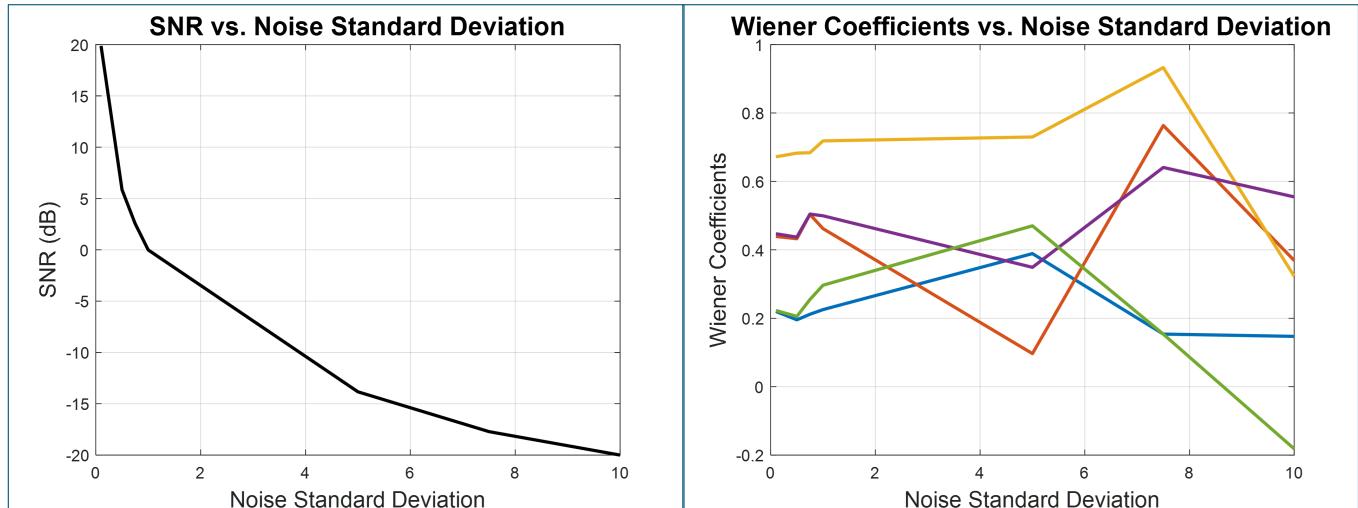
For Trial 3, the constrained breathing at around 15 bpm, the dominant peak is around $0.03 \times 4 = 0.12$ Hz which, once again, is slightly

lower than expected ($15/60 = 0.25$ Hz).



4.1 Wiener Filter

Question 1



The given filter coefficients for the unknown system are $b = [1 \ 2 \ 3 \ 2 \ 1]$. The ideal SNR using normalized data is:

$$SNR = \frac{\sigma_{y[n]}^2}{\sigma_{\eta[n]}^2} = \frac{1}{0.1^2} = 100 = 20 \text{ dB}$$

We use R_{xx} and p_{zx} to find the optimal weight coefficients $w_{opt} = R_{xx}^{-1} \cdot p_{zx}$, where R_{xx} is the ACF of the input signal $x[n]$ and p_{zx} is composed of the cross-correlation of input signal $x[n]$ and $z[n] = y[n] + \eta[n]$ ($y[n]$ is system output and $\eta[n]$ is the noise signal).

The normalized optimal coefficients for noise power of 0.1 are $w_{opt} = [0.2351 \ 0.4629 \ 0.6964 \ 0.4628 \ 0.2335]$. While normalizing the filter output ensures power matching between the input and output, it scales the data. Hence the coefficients do not resemble those of the given unknown system.

Without normalization, the coefficients are $w = [0.9479 \ 1.9333 \ 2.9437 \ 1.9330 \ 0.9411]$, which resemble the given coefficients. The SNR is 20.2 dB. Increasing the sample size and/or decreasing the noise variance give a better estimate of the coefficients.

Question 2

The table below shows how the SNR and optimal coefficients vary with noise power. The clearest observation is that as noise power increases, SNR decreases, which is expected by definition.

While Weiner filter coefficients can diverge at higher noise power, in this experiment, this isn't the case.

Noise standard deviation	w1	w2	w3	w4	w5	SNR (dB)
0.01	0.9995	2.0020	3.0045	2.0077	1.0029	53.3791
0.1	1.0025	2.0030	3.0065	2.0098	1.0001	32.6505
0.5	0.9809	1.9969	3.0013	2.0290	1.0137	19.3272
1.4	1.0091	2.0935	3.0062	2.0244	0.9721	9.1208
2.5	1.0130	2.0269	2.9294	2.0924	1.1046	5.3107
3.1	0.8381	1.8543	2.8922	2.0316	1.1164	2.5721

Question 3

Operation	Multiplications/Additions	Complexity
Autocorrelation (R_{xx})	N^2	$O(N^2)$ or $O(NN_w)$ if $N \gg N_w$
Cross-correlation (p_{zx})	$N(q+1)$ or $N(N_w+1)$	$O(N(q+1))$ or $O(NN_w)$ if $N \gg N_w$
Matrix Inversion (R_{xx}^{-1})	$O(N^3)$	$O(N^3)$ or $O(N_w^3)$ if $N_w \gg 1$
Matrix Multiplication ($R_{xx}^{-1} \cdot p_{zx}$)	$(q+1)^2$ or $(N_w+1)^2$	$(q+1)^2$ or $(N_w+1)^2$

Autocorrelation and Cross-correlation:

To compute the autocorrelation function (ACF) of x and the cross-correlation of x and z , denoted as R_{xx} and p_{zx} respectively, we require

$N_w + 1$ correlations. If calculating a single correlation has a complexity of $O(N)$, then for N samples, the complexity for each statistic is $O(N(N_w + 1))$.

Multiplications and Additions:

Both the ACF and cross-correlation calculations involve $N(N_w + 1)$ multiplications and $N(N_w + 1)$ additions. According to Big O notation, the combined complexity for both is $O(N(N_w + 1)) + O(N(N_w + 1)) \approx O(N(N_w + 1))$. When N is much larger than N_w ($N \gg N_w$), this simplifies to approximately $O(NN_w)$.

Matrix Inversion:

Inverting an $N \times N$ matrix has a complexity of $O(N^3)$. Therefore, the inverse in $w_{opt} = R_{xx}^{-1} \cdot p_{zx}$ requires $O((N_w + 1)^3)$ operations. If N_w is much greater than 1 ($N_w \gg 1$), this complexity approximates to $O(N_w^3)$.

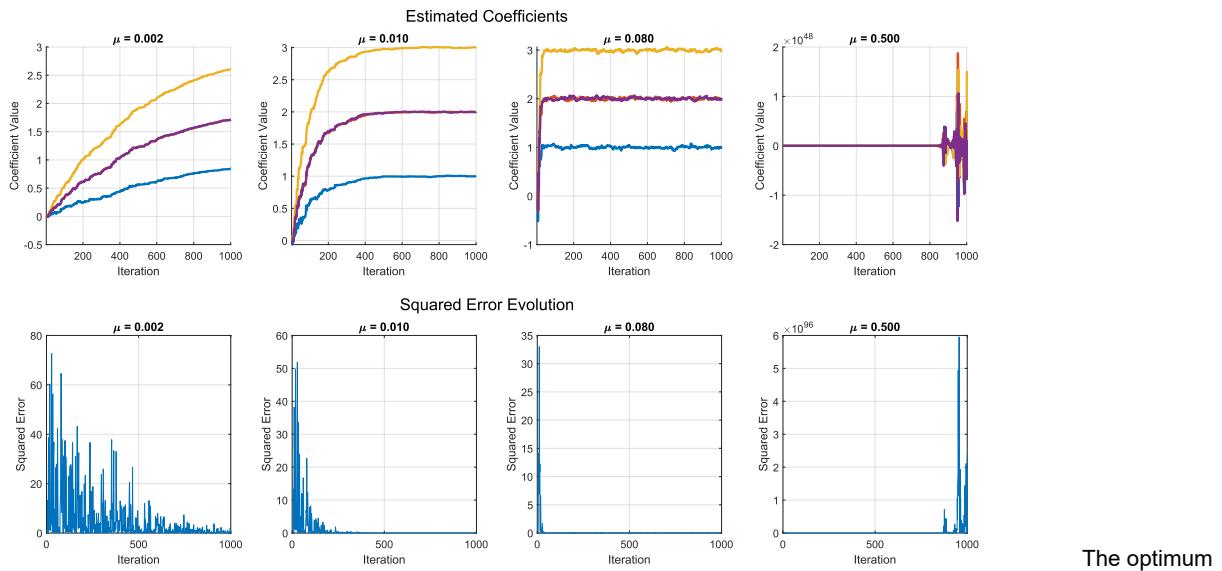
4.2 The least mean square (LMS) algorithm

Question 1

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 $w_{opt} = [1.0025, 2.0037, 3.0153, 1.9985, 1.0014]^T$, which are again, close to the correct values.

Question 2



learning rate μ would be 0.01 under which w_{opt} converges to a stable constant, close to the correct filter coefficients $[1, 2, 3, 2, 1]^T$ 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 still 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.

Question 3

The computational complexity of the LMS algorithm can be determined by analyzing its three core equations. The weight update equation, $w(n+1) = w(n) + \mu e[n]x(n)$, requires $N_w + 1$ additions, which is denoted as $O(N_w + 1)$.

Output Estimation:

The transpose operation in the output estimation, $\hat{y}[n] = w^T(n)x(n)$, has a complexity of approximately $O(N_w)$. The multiplication of an $N_w \times (N_w + 1)$ matrix with an $(N_w + 1) \times 1$ column vector requires $O(N_w + 1) + O(N_w)$ operations.

Error Calculation:

Calculating the error, $e[n] = d[n] - \hat{y}[n]$, involves a simple scalar subtraction and does not significantly impact the overall complexity.

Total Complexity:

Assuming $N_w \gg 1$, the total complexity of the LMS algorithm is of the order $4O(N_w)$, which simplifies to approximately $O(N_w)$.

4.3 Gear shifting

The standard LMS algorithm's sensitivity to input scaling presents a challenge in selecting a stable learning rate μ . As previously discussed, a high adaptive gain leads to fast convergence, while a low gain provides better steady-state accuracy. Gear shifting

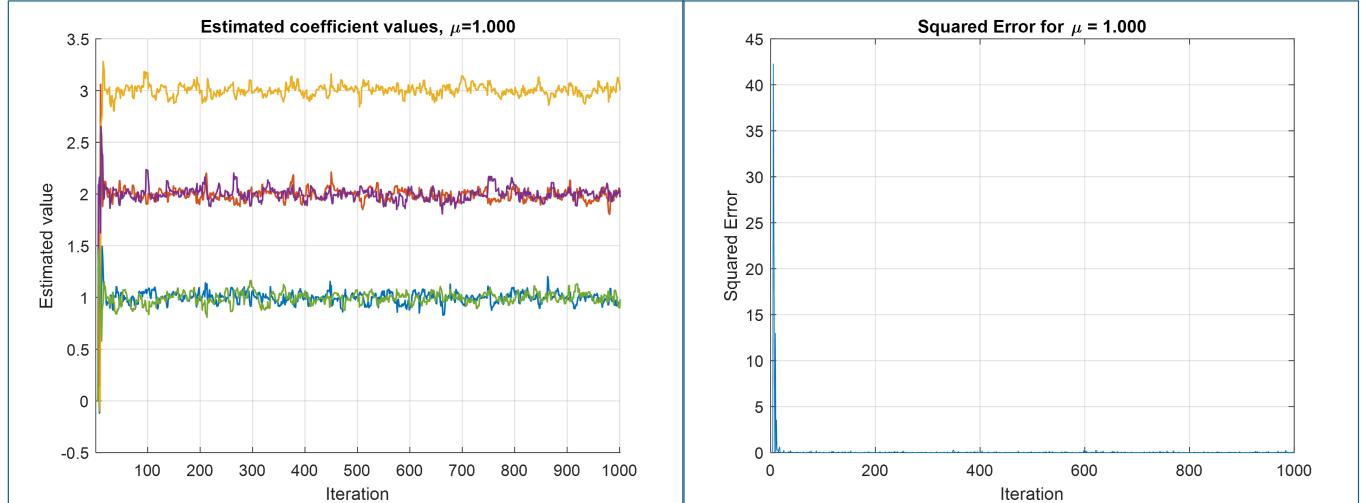
addresses this by dynamically adjusting the gain: increasing it for large errors and decreasing it for small errors.

A common implementation of this adaptive gain policy is the Normalized Least Mean Squares (NLMS) algorithm, which updates the weights according to:

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

where σ and ϵ are small constants to prevent division by zero. This normalization ensures stability and improves convergence compared to the fixed-gain LMS, especially when the input signal's power varies significantly.

This approach leverages the benefits of both high and low gains, leading to significantly faster convergence, as demonstrated below with $\mu = 1.0$.



4.4 Identification of AR processes

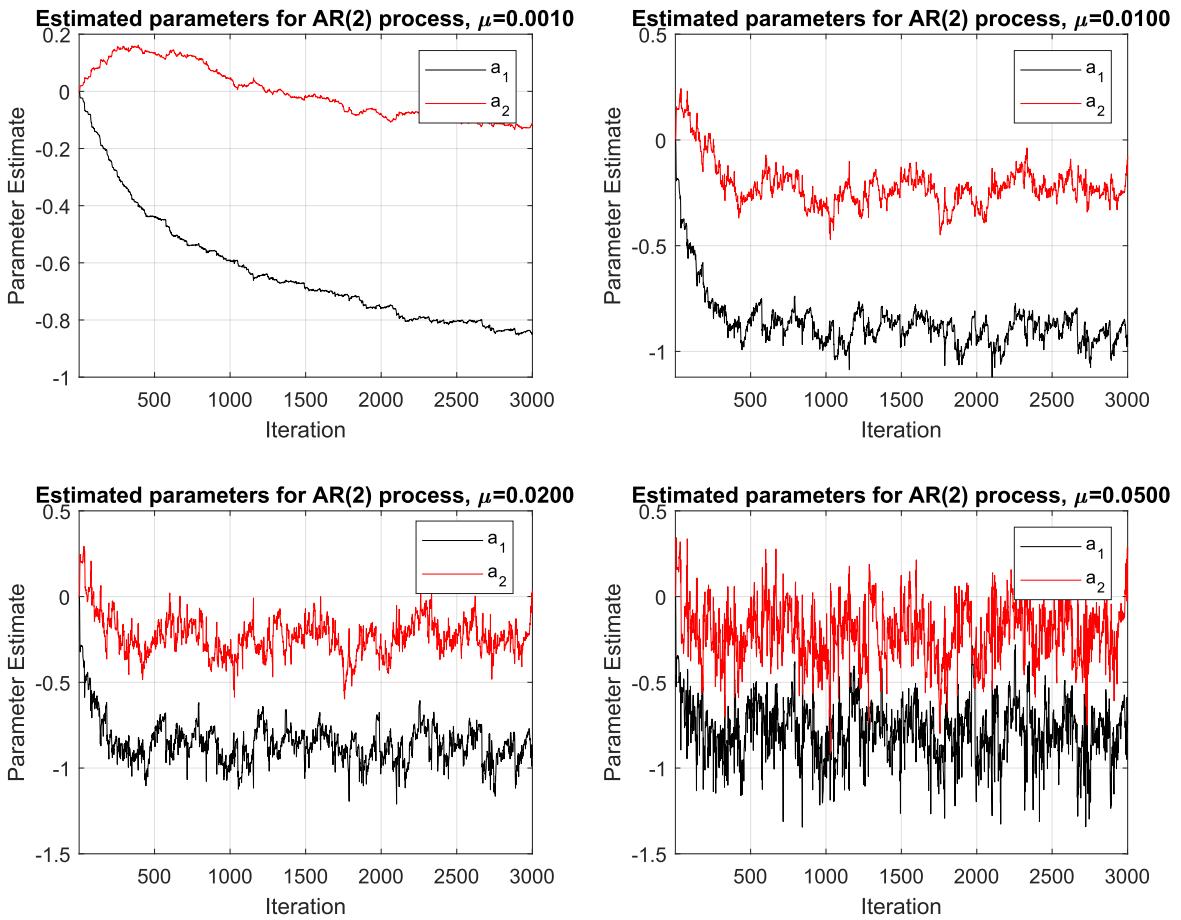
Question 1

Since the given system is an AR(2) process, the LMS algorithm needs to estimate a_1 and a_2 . Figure 4b shows that the coefficients converge to -0.2 and -0.9. This is because the filter output is defined by:

$$y(n) = \sum_{i=0}^M b_i x(n-i) - \sum_{j=0}^N a_j x(n-j) = 0.9x(n-1) - 0.2x(n-2)$$

Question 2

The plots in Figure 4 have been extended to 3000 samples so that the convergence is clear. The convergence is too slow for $\mu = 0.001$ and the estimate is too distorted for $\mu = 0.05$. There is a good balance between accuracy and speed of convergence at $\mu = 0.01$ and $\mu = 0.02$.



4.5 Speech recognition

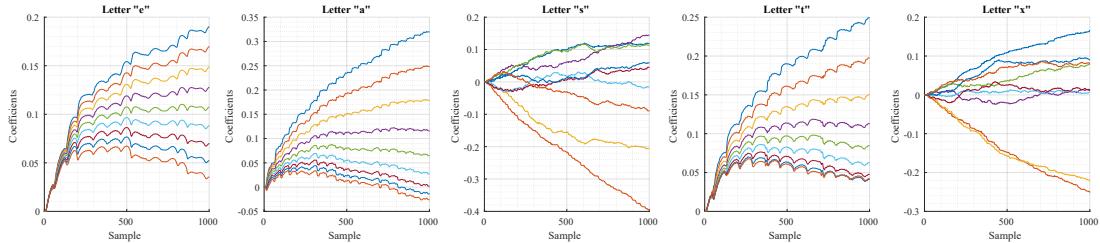
Question 1

Information theoretic criterion, namely MDL and AIC, are used to determine optimal order of the predictor. The results are shown in the table below.

Order	e	a	s	t	x
MDL	13	34	36	14	21
AIC	13	34	49	15	21

The calculations in the figure below use 100 samples, adaptation gain of 0.01, and a filter order of 10. Low gain and high order is an intentional choice to ensure accuracy. Given that the speech signal is non stationary, gear shifting is not recommended and introduce errors in the estimates.

Increasing the adaptation gain past 0.5 causes divergence of estimates.

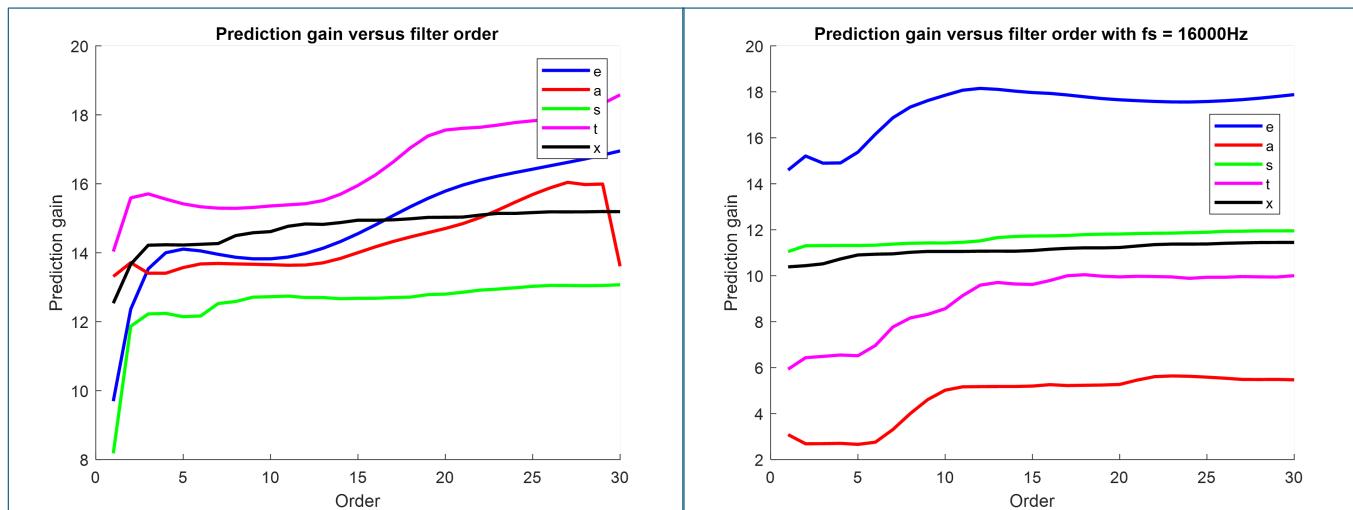


Question 2

To find optimal filter length, we look for the combination of μ and p that gives the maximum prediction gain, R_p , balancing against computational complexity.

As discussed in part 2, larger model order tends to enhance predictor performance. However, too large a model order will increase

model complexity dramatically. Finding the optimal balance is therefore desired.
From the figure below, the appropriate model order would be between 8 and 12.



Question 3

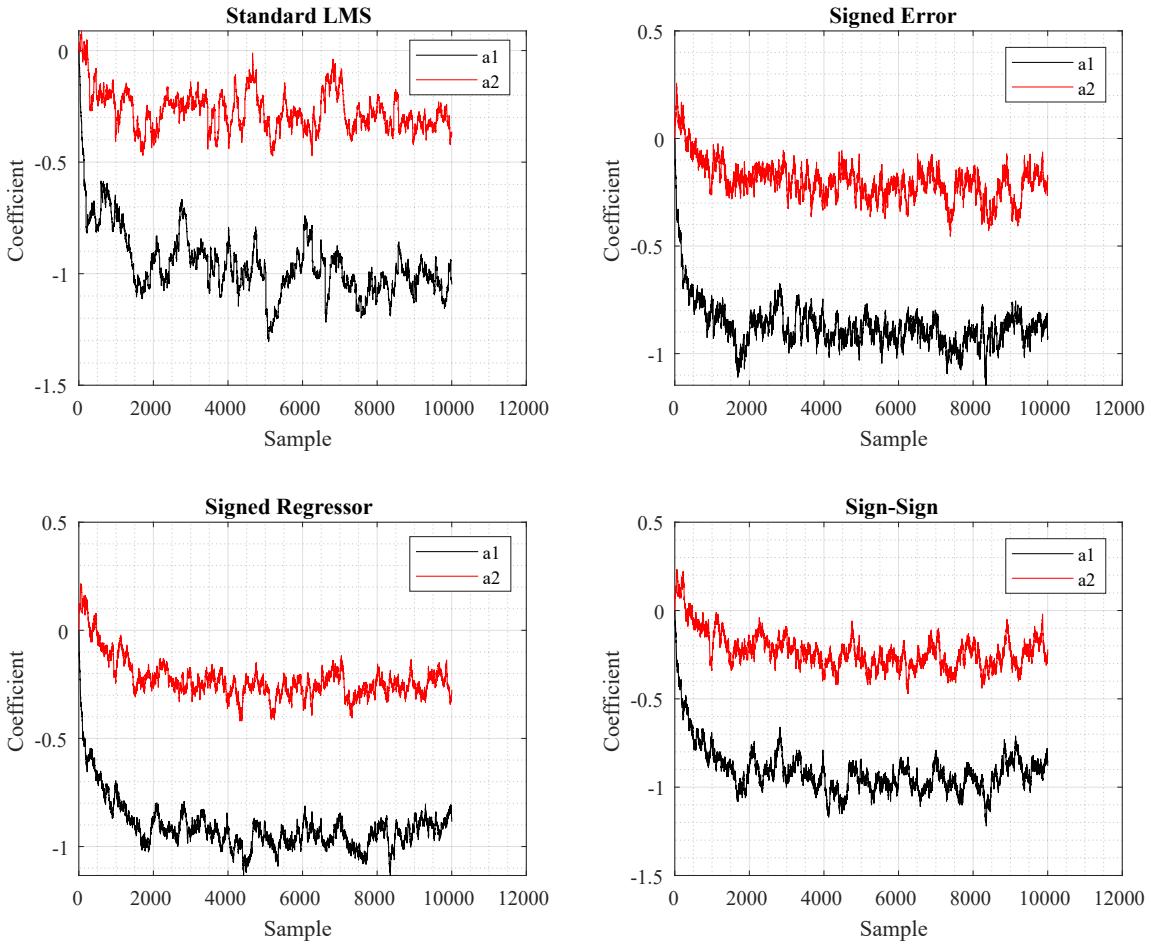
With 16 kHz sampling frequency, the prediction gain drops as shown in the figure above.

This is expected because lower sampling rate tends to imply lower precision, higher quantisation error, and larger noise power
To accurately represent quasi-stationary vowel sounds and achieve convergence in learning curves, the sampling frequency must be high enough to capture the necessary signal detail, and the number of data samples must be sufficient to represent stable signal segments and allow for adequate algorithm adaptation.

4.6 Dealing with computational complexity: sign algorithms

The figure below shows that while the signed LMS converges to the known values ($a_1 = -0.2$ and $a_2 = -0.9$) faster than standard LMS at the expense of more oscillations at convergence.

The convergence of the signed-error algorithm is the sharpest. Both the sign-sign and signed regressor is the slowest to converge but both show lower variance than signed error.



As shown in the figure below, signed error and signed regressor have close convergence rates. Sign-sign algorithm does not converge. We can conclude that, generally, signed error has the best accuracy and convergence performance.

The figure shows the performance on the audio for letter "a".

