Imperial College London

COURSEWORK

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

Adaptive Signal Processing and Machine Intelligence

Junyu Yan (CID: 01895990)
junyu.yan20@imperial.ac.uk

Date: February 16, 2021

Contents

1 Spectrum Estimation

1.1 Properties of Power Spectral Density (PSD)

In order to verify the equivalence with two definitions of PSD, we start from the definition 2, that is,

$$\begin{aligned}
P(\omega) &= \lim_{N\to\infty} E\left\{\frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-jn\omega}\right|^2\right\} \\
&= \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-jn\omega}\sum_{m=0}^{N-1} x^*(m)e^{jm\omega}\right\} \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} E\left\{x(n)x^*(m)\right\}e^{-j(n-m)\omega} \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r(n-m)e^{-j(n-m)\omega} \qquad (1) \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{k=-(N-1)}^{N-1} (N-|k|)r(k)e^{-jk\omega} \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{k=-(N-1)}^{N-1} Nr(k)e^{-j\omega} - \lim_{N\to\infty}\frac{1}{N}\sum_{k=-(N-1)}^{N-1}|k|\,r(k)e^{-j\omega} \\
&= \sum_{k=-(N-1)}^{N-1} r(k)e^{-j\omega} - \lim_{N\to\infty}\frac{1}{N}\sum_{k=-(N-1)}^{N-1}|k|\,r(k)e^{-j\omega}
\end{aligned}$$

It is assumed that the covariance sequence $r(k)$ decays rapidly, that is,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=-(N-1)}^{N-1}|k||r(k)| = 0 \qquad (2)$$

Therefore, the (1) can be reduced to (3), which is the definition 1 of PSD.

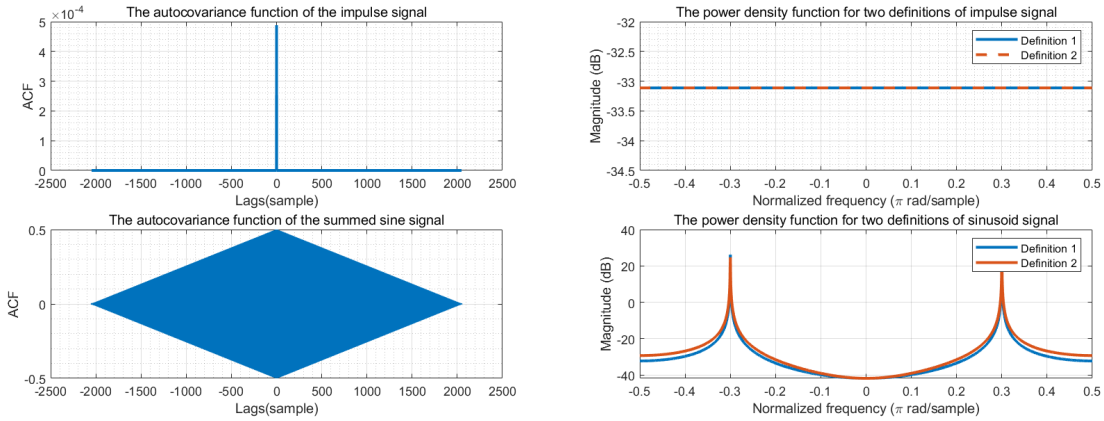
$$P(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k} \qquad (3)$$

The definition 1 represents that the PSD of the signal is the Discrete Fourier Transform of the Autocovariance Function (ACF). While the definition 2 states that the PSD is the distribution of averaged signal power over frequency. Therefore, the equivalence proves that when the covariance sequence decays rapidly (i.e., $r(k)$ is absolutely integrable), the PSD of a discrete time deterministic sequence is equal to the DTFT of its ACF.

To simulate the equivalence, we obtain the PSD of impulse signal and sinusoidal signal, respectively. As shown in Figure.1.(a), the covariance sequence of the impulse signal decays significantly rapidly, however; the ACF of the sinusoidal signal reduces slowly. Therefore, as shown in Figure.1.(b), the two definitions of PSD coincide with each other for the impulse signal, but they are different for the sinusoidal signal.

1.2 Periodogram-based Methods Applied to Real-World Data

a) The time series data usually has a non-zero mean and a long-period linear trend that will highly influence the analysis of data, and must be removed before further processing. Removing mean eliminates the effect of DC component in the original data. Linearly detrending the data reduces



(a) ACF (b) PSD

Figure 1: The autocovariance function and PSD for the impulse signal and the sinusoidal signal

the low-frequency components so that the high-frequency components can be more easily identified. Also, it can remove jitters in the data leading to less noise in the signal. Moreover, other preprocessing methods, such as the logarithm combined with the subtraction of mean, performs even better on removing jitters, and can be used to identify the change of periodicity in data.

From the Figure.2, it is clear that removing mean as well as trend centralizes and smooths the original data. Removing mean and trend decreases the PSD spectrum below the normalized frequency 0.01, and it has little influence on the high-frequency components. The logarithm preprocessing method captures more details in the PSD spectrum. For example, the rise at normalized frequency equal to 0.01 is captured, and more harmonics can be identified. Moreover, it can also capture the change of perception of periodicity in data. As shown in Figure.2, the waveform after logarithm and subtraction of mean will experience a strong oscillator if there is less periodicity.

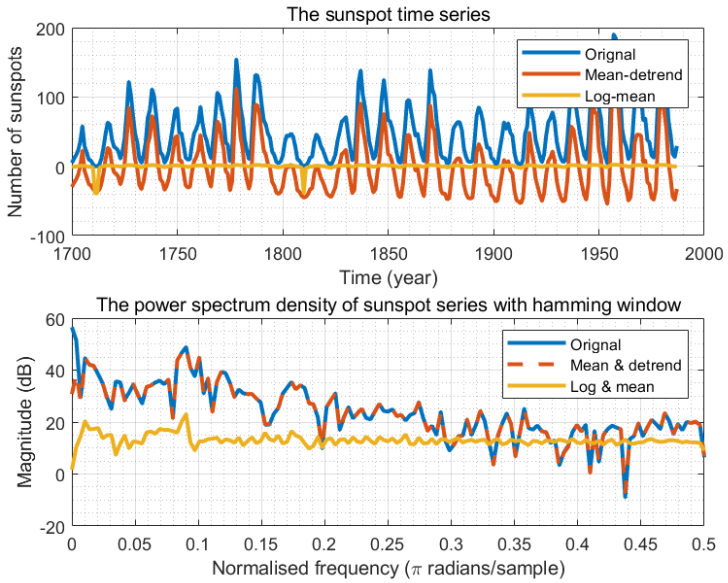


Figure 2: Top: the original and preprocessed sunspot time series. Bottom: the PSD of the sunspot time series with hamming window

b) The flashing visual stimulus flashing at a fixed integral rate will induce the response known as

steady state visual (SSVEP) in electroencephalogram (EEG) signal at the same frequency. As shown in Figure.3, both standard periodogram and averaged periodograms with different window length (i.e., 1s, 5s, 10s) are capable of identifying the peaks corresponding to SSVEP, which are the fundamental frequency at 13Hz with two harmonics at 26Hz and 39Hz, respectively. The peak between 8-10Hz, clearly captured by averaged periodogram, is the alpha-rhythm due to the tiredness of the subject during the recording. Moreover, the largest peak appearing at 50Hz is caused by the power-line interference, and it exceeds the third harmonic at 52Hz.

Compared with the standard periodogram, the averaged periodogram generates more recognizable peaks. It is clear that the averaged periodogram with window length equal to 10s has a smaller variance than the standard periodogram, but the dominant peaks can still be identified. It significantly reduces the effect of noise and highlight the main lobes, which is quite beneficial for the observation of meaningful information from PSD spectrum. For the window length equal to 1s, the spectrum is highly smoothed with clearer trend of PSD. However, it contains less details, such that the peak at 39Hz is hard to recognized. The reason is that the windowed periodogram is equal to the convolution of the PSD with the Fourier Transform of the window, which is a sinc-like function. The main lobe width of the sinc function is inversely proportional to the signal length. Therefore, truncating signal through decreasing the window length will lead to a wider main lobe, which ultimately fuses more nearby spectral peaks and reduces the frequency resolution. In conclusion, there is a trade-off between the variance reduction and the frequency resolution. Therefore, choosing a 5s window is more proper to simultaneously ensure the noise elimination and peaks distinguish-ability.

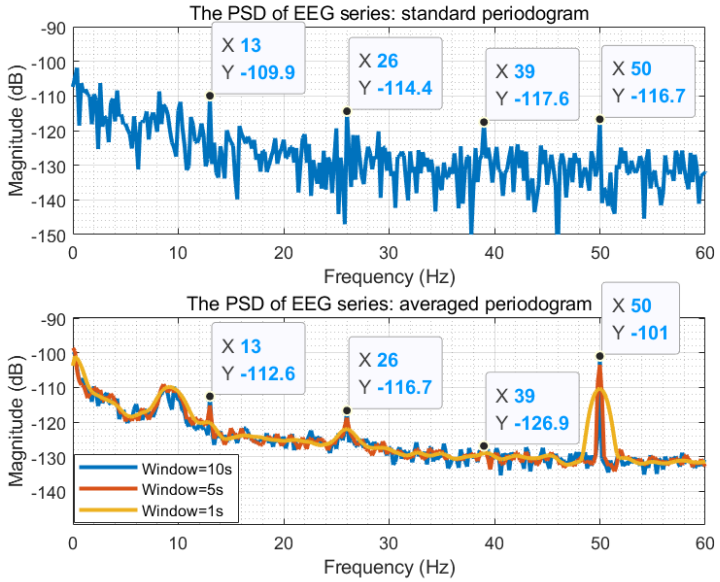


Figure 3: Top: the standard periodogram of EEG series. Bottom: the averaged periodogram of EEG series

1.3 Correlation Estimation

a) Figure.4 and Figure.5 present the correlogoram and correlation-based PSD estimate of the WGN signal, noisy sinusoidal signal and low-pass filtered WGN signal. The estimate is based on the biased and unbiased estimators given by,

$$\textbf{Biased}: \hat{\mathrm{r}}(k) = \frac{1}{N}\sum_{n=k+1}^{N} x(n)x^*(n-k) \tag{4}$$

$$\textbf{Unbiased}: \hat{\mathrm{r}}(k) = \frac{1}{N-k}\sum_{n=k+1}^{N} x(n)x^*(n-k) \qquad (0 \le k \le N-1) \tag{5}$$

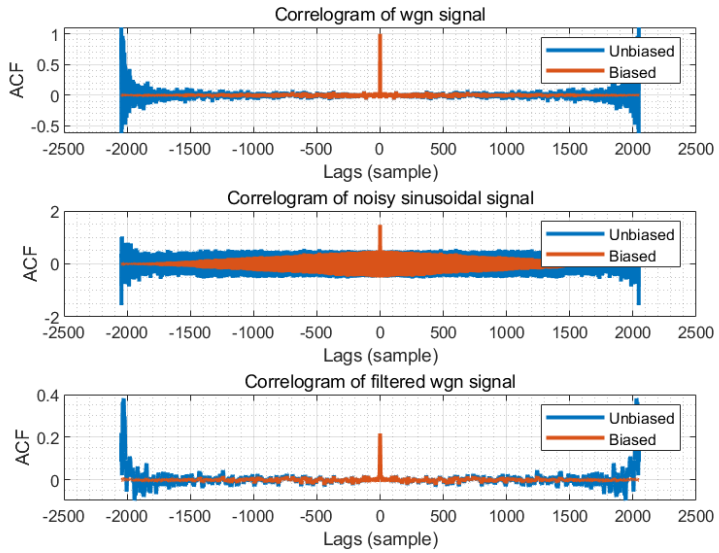


Figure 4: The correlogram of wgn signal, noisy sinusoidal signal and filtered wgn signal

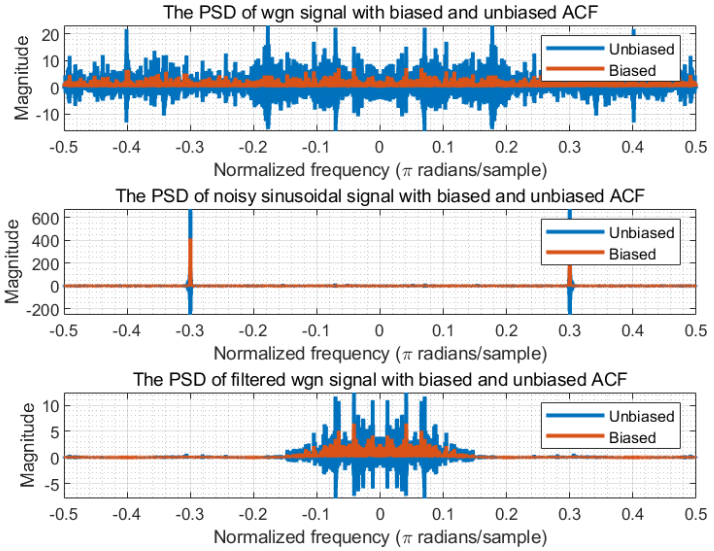


Figure 5: The PSD estimate of wgn signal, noisy sinusoidal signal and filtered wgn signal with biased estimator and unbiased estimator

By comparing the ACF estimates shown in Figure.4, we find that the biased and unbiased ACF are similar for small lags as $|k| \leq 100$. However, as k increases, the biased ACF estimate decays to 0, while the unbiased one grows. This can be explained with (4) and (5) that when k is large, the similarity between samples will decreases, leading to smaller sum and smaller unbiased ACF estimate. But for unbiased ACF, the increase in k decreases the denominator in coefficients that will still produce a large value.

Figure.5 shows that the PSD estimates with biased ACF for three kinds of signals agree with the expectation as flat for WGN signal, impulse for sinusoidal signal and rectangular for low-pass WGN. However, for unbiased PSD estimate, it is erratic and results in negative values for large k. This is due to the fact that there are fewer samples available to estimate the PSD when k is large, which will cause the estimate of ACF to be not positive definite and lead to negative power values. For this reason, the biased estimator is more widely since it has low variance and is asymptotically biased for a large N.

b) We generate the PSD estimate based on biased ACF estimator of 100 realisations for the signal composed of sinusoids and unit power WGN noise (AWGN singal) given by,

$$x(n) = 0.3\sin(2\pi 0.2n) + 0.6\sin(2\pi 0.25n) + w(n) \tag{6}$$

Figure.6 shows that the PSD estimate can successfully detect the frequencies of sinusoids contained in the signal (peaks at normalized frequency 0.2 and 0.25), although some realisations are larger than the mean at the frequency of interest. At other frequencies, the PSD estimate of 100 realisations fluctuates from 0 to 10, and the mean of them approach to 1, which is equal to the power of the WGN. In addition, the standard derivation is proportional to the PSD, therefore; the confidence interval is much greater at the peaks. Moreover, since the mean of PSD does not converge to 0 as the data length increases, and the variance approaches to the PSD that is not 0, therefore, the biased ACF estimator is not consistent.

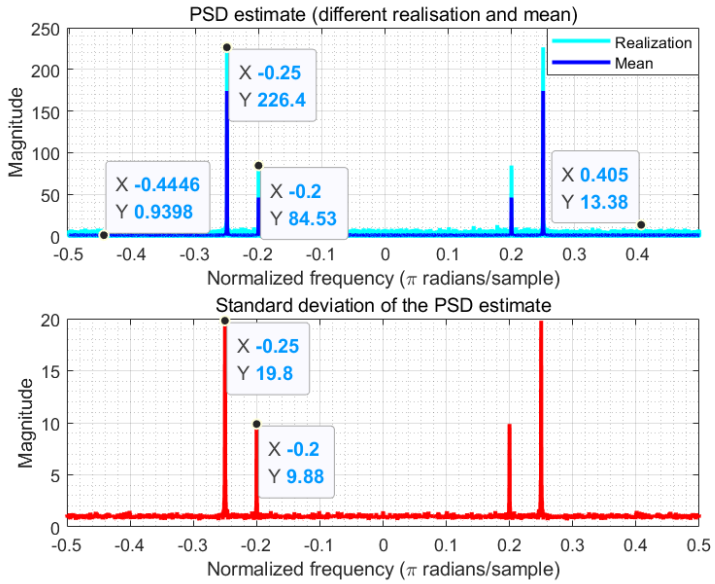


Figure 6: Top: the PSD estimate of 100 realisations and mean for the AWGN signal with biased ACF. Bottom: the standard derivation of AWGN signal with bised ACF

c) The conversion to dB can be seen as a smooth operation on parts of the spectrum with large variations. As show in Figure.7, It maintains the trend, but attenuates the transition at peaks, while amplifies the fluctuation due to noise. As a result, the dB plot enables the clearer detection of peaks with smooth and more detailed trend information about low-amplitude signals. In conclusion, this

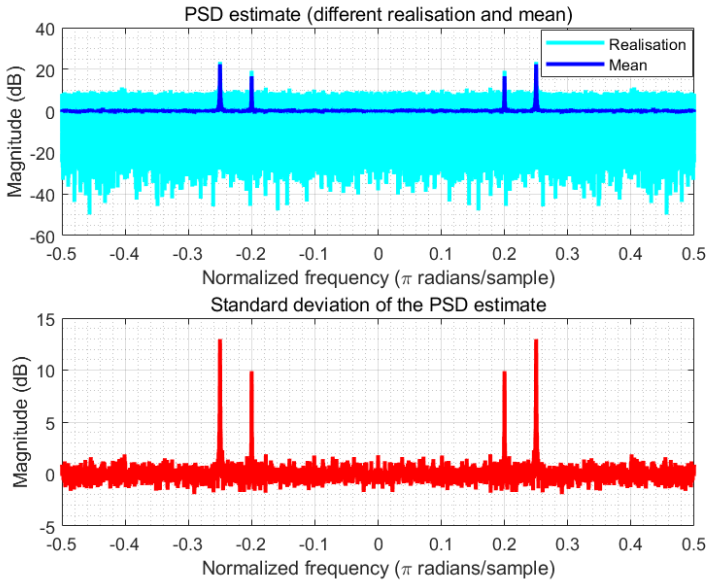


Figure 7: Top: the PSD estimate in dB of 100 realisations and mean for the AWGN signal with biased ACF. Bottom: the standard derivation in dB of AWGN signal with bised ACF

presentation is effective since it provides more trend information while cares less about the meaningless transitions.

d) We generate three complex-valued signal of different frequencies and length as expressed in (7), (8) and (9),

$$x(n) = e^{2\pi(0.3)n} + w(n) \tag{7}$$

$$x(n) = e^{2\pi(0.3)n} + e^{2\pi(0.32)n} + w(n) \tag{8}$$

$$x(n) = e^{2\pi(0.3)n} + e^{2\pi(0.32)n} + e^{2\pi(0.32)n} + w(n) \tag{9}$$

where $w(n) \sim N(0, 0.02)$. The periodogram estimates with rectangular window for these three signals of different length N are shown in Figure.8, where the spectrum becomes sharper and closer to the ideal line spectral as N increases. The reason is that when the periodogram resolution, which is proportional to 1/N, is greater than the separation between signal frequencies, then the line spectra will not be able to be correctly detected. More specifically, the periodogram estimate is the convolution of the true PSD with the Fourier Transform of rectangular window, which is a sinc-like function, therefore, the periodogram resolution is determined by the main lobe width of the window. Take signal in (8) as an example, the frequency separation is $0.02Hz$, and the main lobe width of the rectangular window is $0.89/N$. Therefore, the minimum signal length to separate frequencies clearly is $N = 0.89/0.02 = 45.5$. As a result, for $N > 45$, the peaks can be divided by window's main lobe, otherwise, other methods like subspace method should be taken into consideration.

e) The code related is shown below. For the first line of code, the `corrmtx` is to obtain the autocorrelation matrix estimate for the signal 'x'. The input signal 'x' is a linear combination of complex exponential signals with the WGN noise. The '14' defines the dimension of the autocorrelation matrix that is $(14+1)\times(14+1)$. 'mod' is set to take the modified covariance method so that the result can be more accurate. For the output of the `corrmtx`, 'R' is the autocorrelation matrix, and 'X' is a rectangular Toeplitz matrix, where X'X is a biased estimate of autocorrelation matrix.

For the second line of code, it estimates the periodogram based on the Multiple Signal Classification (MUSIC) algorithm. Input 'R' is the autocorrelation matrix, '2' is the dimensionality of the signal subspace, '[]' represents to use the default FFT points, '1' the normalized sampling frequency, and

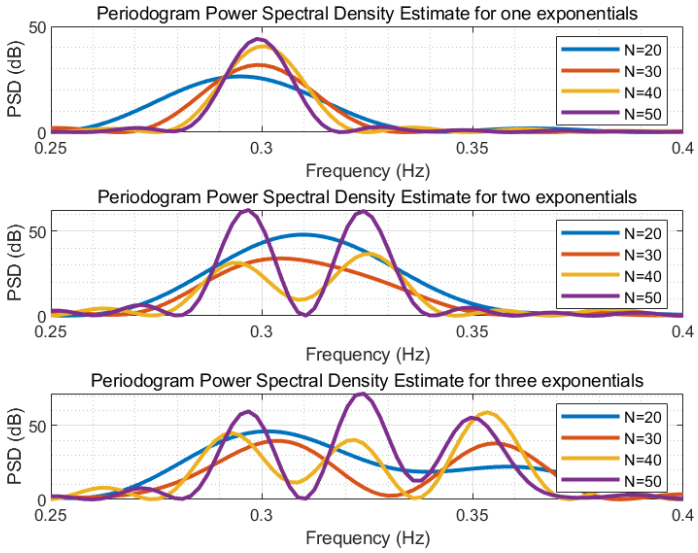
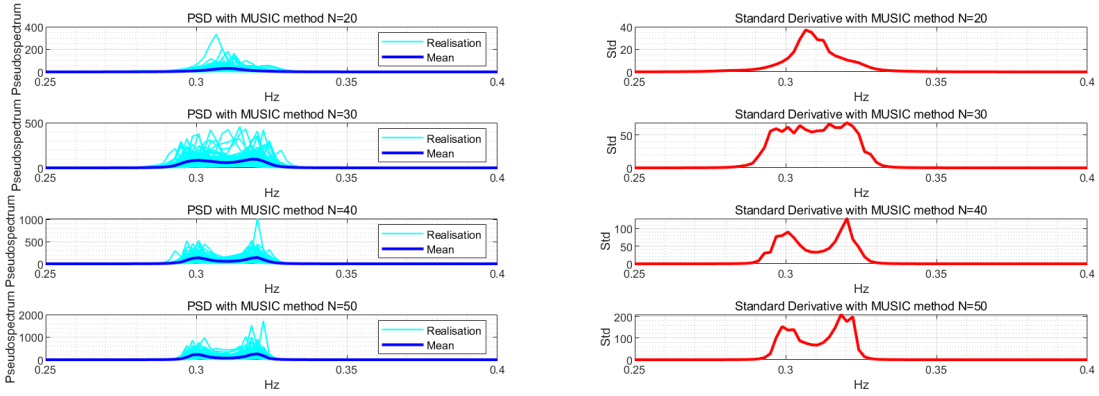


Figure 8: Periodogram estimate with rectangular for one exponential, two exponential and three exponential signals with different length

'corr' indicates that 'R' is the correlation matrix estimate. For the output, 'S' is the pseudospectrum, while 'F' is the normalized signal frequency.

For the third line, its aim is to plot the pseudospectrum versus the normalized signal frequency with the line-width equal to 2, while also set the frequency range in (0.25, 0.4).

```
[X,R] = corrmtx(x,14,'mod');
[S,F] = pmusic(R,2,[],1,'corr');
plot(F,S,'c','linewidth',2); set(gca,'xlim',[0.25,0.40]);
grid on; xlabel('Hz'); ylabel('Pseudospectrum');
```



(a) PSD estimation

(b) Standard derivation

Figure 9: The PSD estimate and standard derivations of two complex exponentials with different lengths

As shown in Figure.9, the two line spectra can be identified for $N = 30$, while the correlation-based estimator can not distinguish them well with the same signal length. Therefore, the MUSIC method is able to capture more details about the spectrum and outperforms the correlation-based method for

fewer signal samples (form of superresolution). The MUSIC method generally has a higher resolution (lower bias) than the periodogram, which will decrease as the Signal-Noise-Ratio (SNR) increases. Moreover, MUSIC method also has higher variances than the periodogram method, and the variances are still proportional to the PSD. As a result, when SNR increases, the bias on peak decreases while the variance increase, which will benefit the detection of peaks locations.

The main disadvantage of the MUSIC method is that, as a subspace method, it requires the dimensions p to be known or estimated in advance. The estimation of p need be determined by Minimum Description Length (MDL) or Akaike Information Criterion (AIC), which significantly increases the computation complexity.

1.4 Spectrum of Autoregressive Processes

a) The AR parameters $\mathbf{a}$ is obtained from the Yule-Walker equation given by:

$$\mathbf{R_{xx}a} = \mathbf{r_{xx}} \Rightarrow \mathbf{a} = \mathbf{R_{xx}^{-1}r_{xx}} \tag{10}$$

where, $\mathbf{R_{xx}}$ is the autocorrelation matrix. From (10), it is clear that determination of $\mathbf{a}$ requires the $\mathbf{R_{xx}}$ to be invertible. For the biased estimator, $\mathbf{R_{xx}}$ is a positive definite Toeplitz matrix, which is definitely invertible. On the other hand, for the unbiased estimator, as shown in Figure.4, the ACF does not decay to 0 for large lags so that the positive definiteness may not be satisfied. Therefore, there may be non-positive eigenvalues of ACF matrix, and the inversion of autocorrelation matrix will not exist. As a result, the parameters $\mathbf{a}$ cannot be solve.

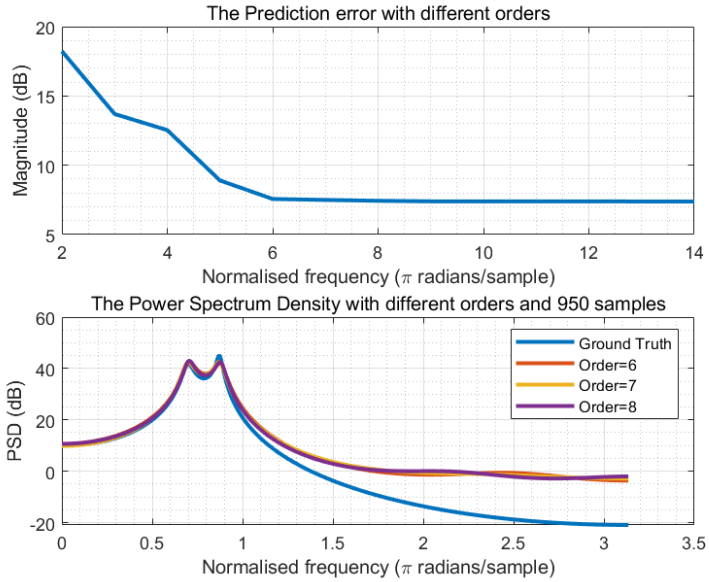


Figure 10: Top: the prediction error with different orders of the AR model with 1000 samples. Bottom: the estimated PSD with AR models of different order and the true PSD with 1000 samples.

b) According to Figure.10, as the model order increases, the prediction error decreases, and the AR model will have more degrees of freedom to accurately capture the fully underlying structures of data to make more adequate estimation. However, increasing the model order will increase the computation complexity. Figure.10 suggests that the prediction error reaches a plateau at $p = 6$, therefore, the order $p = 7$ will be optimal for the trade-off between the prediction error and the complexity. Moreover, the prediction error at the actual order $p = 4$ is high, which illustrates that it performs not well on the estimation of PSD. Also, different samples will produce different estimation result. These are due to the bias of the ACF and the small sample size. As a result, the model accuracy can be

improved by increasing the model order or using more samples.

c) As shown in Figure.11, compared with the aforementioned 1000-sample condition, the Mean Squared Errors (MSEs) of prediction are reduced for all orders after the sample size increased to 10000. The error is significantly low at the actual order $p = 4$, and the PSD estimation with it correctly detects the two peaks. In addition, the estimated PSD approaches to the true PSD better. For the chosen order is lower the correct order $p = 4$, then the AR model will not have enough freedoms to capture all variations of interest, which results in a high prediction error. On the other hand, for a high order, the model will be over-fitting, leading to great computation complexity.

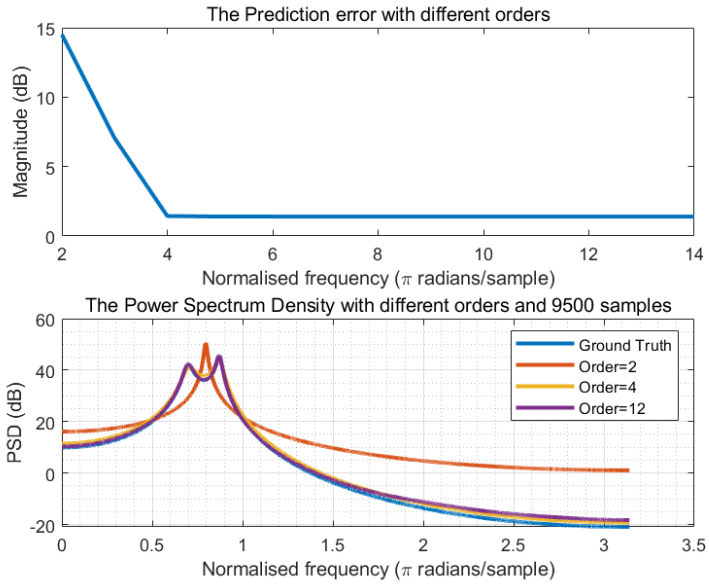


Figure 11: Top: the prediction error with different orders of the AR model with 10000 samples. Bottom: the estimated PSD with AR models of different order and the true PSD with 10000 samples.

1.5 Real World Signals: Respiratory Sinus Arrhythmia from RR-Intervals

a) Respiratory sinus arrhythmia (RSA) is the modulation of cardiac function with respiratory function, which can be observed through speeding up or slowing down the heart rate. The R-R-Interval(RRI) in a cardiac (ECG) data recording will be shortened during inspiration (breathing out) or extended during expiration (breathing in). Therefore, in order to study the features of RSA and ECG signal, we obtain the RRI data for three trials, as normal breathing, fast breathing at 25bpm and the slow breathing at 7.5bpm. The periodograms of these three-trial RRI data are obtained with the standard method as well as Bartlett's Method with different window lengths (i.e., 50s, 150s). The hamming window are applied for the wider main-lobe and more bounded side-lobes. The result is shown in Figure.12.

b) Figure.12 presents that there is a trade-off between the variance reduction and the frequency resolution. Comparing the standard periodogram (window length more than 200s) and averaged periodogram with the $150s$ window, it is clear that the averaged periodogram is more smooth than the standard one with smaller variance. Moreover, the $50s$-window periodogram only provides the trend for the PSD, while omit many meaningless variations. Although using a small window will reduce the amplitude of fluctuation, the peaks of interest can still be correctly located. In addition, the decrease in samples leads to the lower frequency resolution, allowing peaks become more wider and harmonics to be identified more easily.

The peaks of interest in Figure.12 occurs at $f_1 = 0.1465Hz$, $f_2 = 0.418Hz$ and $f_3 = 0.123Hz$. The respiratory rate corresponding to the RRI frequency is equal to $BreathsPerMinute(BPM) = 60 \times f$.

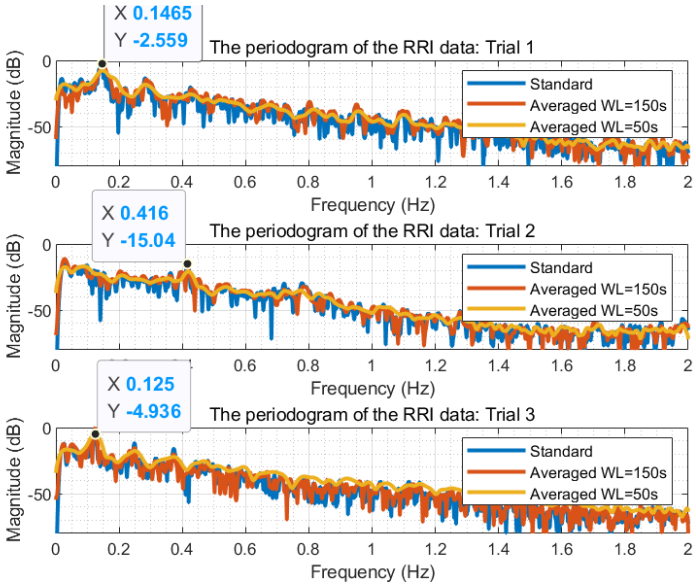


Figure 12: The standard and averaged periodogram with 50s and 150s windows for the RRI data of three trials

Therefore, the breathing rates we obtained are $9bpm$, $25bpm$ and $7.5bpm$, respectively. Since the standard and average periodogram are approximately coincident at peaks, the rates acquired are valid and satisfy our initial assumptions for normal, fast and slow breaths.

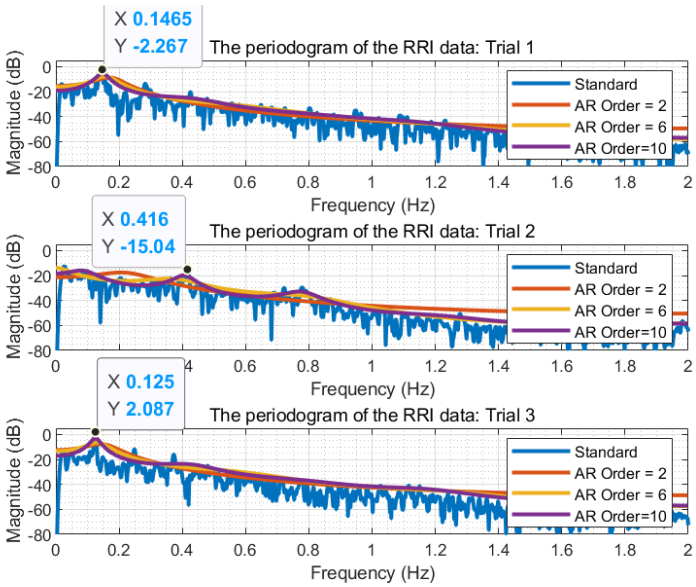


Figure 13: The standard and AR model periodogram with different orders for the RRI data of three trials

c) Figure.13 shows that the spectrum estimated by the AR model is much more smooth than the standard one with only the trend of PSD, and little information about noise and harmonics. The main characteristics of PSD, such as peaks, are much more easily to be identified. Additionally, the orders $p = 2$, $p = 6$ and $p = 10$ are the optimal orders for the normal, fast and slow breathing AR models, respectively. Although the AR model method does not suffer from the trade-off of bias and variance, it

still need a prudent choice for the order of model. A small model order will introduce high estimation error, and a large order will lead to over-modelling and great complexity.

1.6 Robust Regression

a) As shown in Figure.14, non-zero singular values input signal $\mathbf{X}$ is 3, indicating the rank of $\mathbf{X}$ is equal to 3. The rank of the signal corrupted by noise $\mathbf{X}_{noise}$ is 10, with 10 non-zero singular values. This is because that the input signal will only have 3 subspace; however, the noise signal will occupy the whole subspace. Due to the noise, the singular values for signal subspace will be slightly increased, but for other remaining subspace, the singular values are increased significantly by noise. More specifically, there are quite small squared errors at $\mathbf{X}$ subspace, but much greater deviations at the others. Therefore, when the Signal-Noise-Ratio (SNR) is small, the errors in irrelevant subspace will make the singular values in all dimensions similar, then it will be much difficult to identify the rank of $\mathbf{X}_{\text{noise}}$.

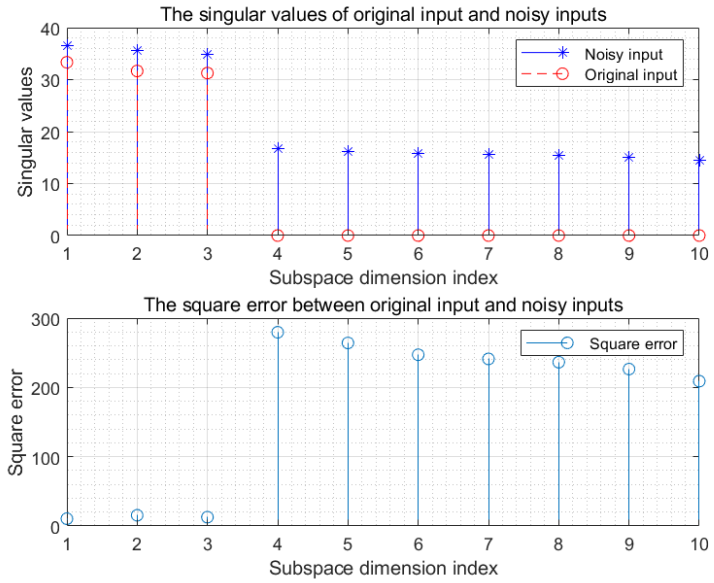


Figure 14: Top: the singular values of input signal and noisy input signal. Bottom: the square error between each singular value of original input and noisy input

b) Figure.15 suggests that the low-rank $\tilde{\mathbf{X}}_{\text{noise}}$, which uses only the 3 most significant principle components, is a good approximation of the original since the square error between $\mathbf{X}$ and $\tilde{\mathbf{X}}_{\text{noise}}$ are all low enough in all dimensions. This proves that the meaningful information of the original signal is only stored in a few principle components, while the others are the additive noise. The reason why squared errors are not reduced to 0 is that some information is lost due to the truncated Singular Value Decomposition (SVD). Also, it is quite interesting that not all errors reduced at a same scale. This is due to different SNRs, which means that a higher SNR will make it more difficult to remove noise as a result of overfitting of noise.

c) The output data are obtained as

$$\mathbf{Y} = \mathbf{XB} + \mathbf{N}_{\mathbf{Y}} \tag{11}$$

where the regression matrix $\mathbf{B}$ can be calculated based on two methods as Ordinary Least Squares (OLS) and Principle Component Regression (PCR). The $\hat{\mathbf{B}}_{\text{OLS}}$ and $\hat{\mathbf{B}}_{\text{PCR}}$ are given by (12) and (13). Figure.16 shows the square error between the original output and the reproduced output using OLS or PCR method with $r = 3$ largest principle components on both training and testing data. Table.1 shows the summation error. Comparing the results shown in Figure.16 and Table.1, we conclude that

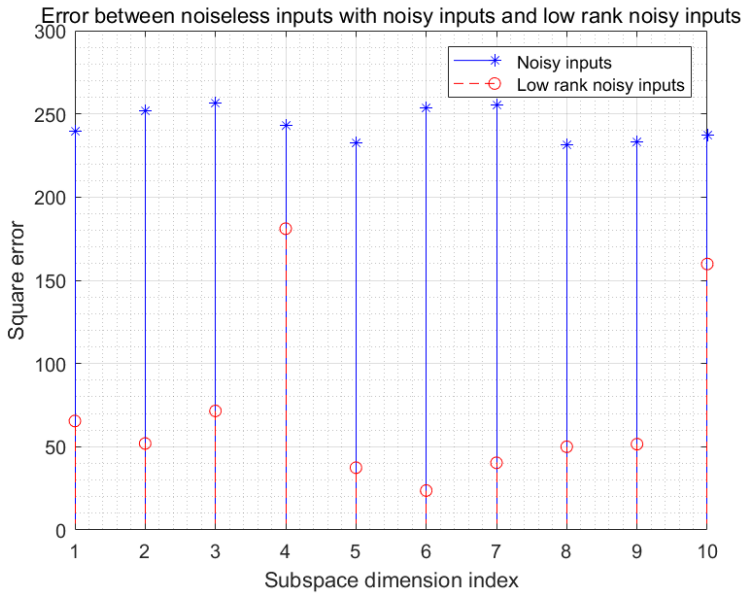


Figure 15: Square errors between input signal and noisy input signal or low-rank noisy input signal

Table 1: Summation of square error on training and testing data

	Training data	Testing data
OLS	3551.6569	2377.3501
PCR	3577.1496	2351.5383

the OLS method fits training data better, while the PCR method performs better on testing data. This can be explained that the PLS considered the noise components, therefore, it over-fits the training set. Also, since it fits much on the random noise in training data, therefore, paying less attention on the relationships between signal samples, which results in a poor performance on testing data. As for PCR method, it ignores the noise components so that it does not over-fit the training set, but is able to predict test data better with more knowledge of signal samples.

$$\hat{\mathbf{B}}_{OLS} = (\mathbf{X}_{noise}^{\mathbf{T}}\mathbf{X}_{noise})^{-1}\mathbf{X}_{noise}^{\mathbf{T}}\mathbf{Y} \tag{12}$$

$$\hat{\mathbf{B}}_{PCR} = \mathbf{V}_{1:r}(\Sigma_{1:r})^{-1}\mathbf{U}_{1:r}^{T}\mathbf{Y} \tag{13}$$

d) The best way to evaluate the OLS and PCR method is to test the estimated regression coefficients $\hat{\mathbf{B}}$, over an ensemble of test data. Then, we generate 100 realisations of output based on regression coefficients, and compute the MSE error between each realisation $\mathbf{Y}$ and its estimate $\hat{\mathbf{Y}}$ given by

$$MSE = \mathrm{E}\{\| \mathbf{Y} - \hat{\mathbf{Y}} \|_2^2\} \tag{14}$$

to compare the effectiveness of these two methods. The MSE errors are listed in Figure.17, and the summation errors for OLS and PCR are 2351.6783 and 2327.469, respectively. It is clear that PCR outperforms the OLS by a 0.5% smaller MSE, but both methods provides confident results.

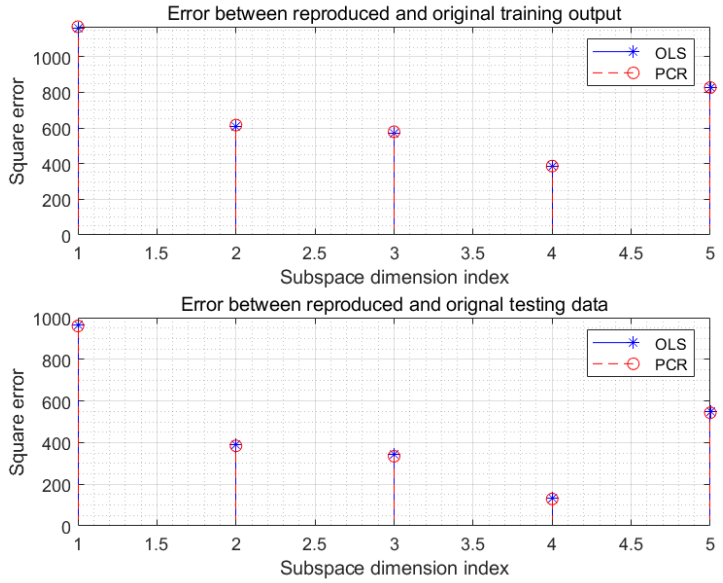


Figure 16: The estimation error by OLS and PCR

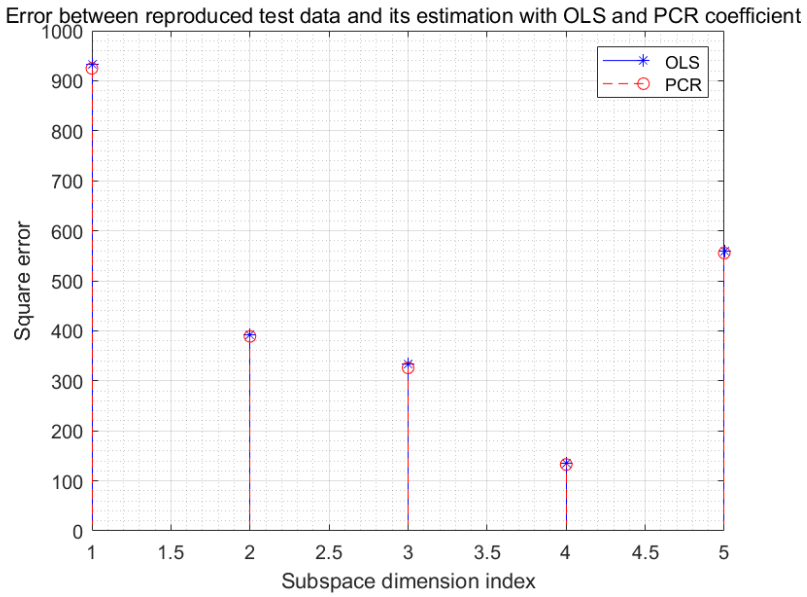


Figure 17: The MSE error by OLS and PCR with 100 realisations