5. Signal Modelling and Spectral Estimation

5.1 Signal Modelling

We discuss modeling of real world signals using parametric pole-zero (PZ) signal models, specifically ARMA($P,Q$) random signal models described by:

$$x(n) = -\sum_{k=1}^{P} a_k x(n-k) + w(n) + \sum_{k=1}^{Q} d_k w(n-k)$$

\[\text{Equation 5.1}\]

where $w(n) \sim N(0, \sigma_w^2)$.

**Signal Modelling Problem**

Given finite-length data $\{x(n)\}_{n=0}^{N-1}$, which can be regarded as a typical sample sequence of the signal source under consideration, we want to select [model selection] the “most reasonable” order (values of $P$ and $Q$) of the model, and then estimate [model estimation] the signal model parameters $\{a_k\}_i$, $\{d_k\}_i$, and $\sigma_w^2$ to satisfy a prescribed criterion (usually the LS error criterion). Finally we must validate the model [model validation] by investigating how well it captures the key features of the data.

5.1.1 Model Selection

To select the order of the model we estimate the autocorrelation, partial autocorrelation, or power spectrum from the available data, and compare them to the corresponding quantities obtained from the theoretical model. The lowest order which provides the closest fit is then selected.

For the special case of all-pole (AP) models where $Q = 0$ we have an AR($P$) process. The problem becomes that of linear prediction and the following result from Chapter 3 can be used:

$$k_m = 0 \quad \text{for} \ m > P$$

where $x(n)$ is an AR($P$). That is:

**Selection based on the Partial Autocorrelation**

Plot the partial autocorrelation sequence ($k_m$) for $m \geq 1$ (using Durbin’s algorithm or the Algorithm of Schur and the sampled autocorrelation sequence calculated from the available data $\{x(n)\}_{n=0}^{N-1}$) and identify the value $P$ where $k_m = 0$ for $m > P$. Then $P$ is the order of the AP model.
5.1.2 Model Estimation

The most typical model estimation problem is the principle of least-squares (LS) estimation to estimate parameters of all-pole (AP) models of order $P$. The solution for general pole-zero (PZ) models is more complicated and discussed elsewhere [1, Section 9.3].

For the full-windowing case the data matrix is given by:

$$
\mathbf{X}^T = \begin{bmatrix}
    x(0) & x(1) & \cdots & x(P) & \cdots & x(N-1) & 0 & \cdots & 0 \\
    0 & x(0) & \cdots & x(P-1) & \cdots & x(N-2) & x(N-1) & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & x(0) & \cdots & x(N-P-1) & x(N-P) & \cdots & x(N-1)
\end{bmatrix}
$$

For the no-windowing case the data matrix is given by:

$$
\mathbf{X}^T = \begin{bmatrix}
    x(P) & x(P+1) & \cdots & x(N-2) & x(N-1) \\
    x(P-1) & x(P) & \cdots & x(N-3) & x(N-2) \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    x(0) & x(1) & \cdots & x(N-P-2) & x(N-P-1)
\end{bmatrix}
$$

Following from Section 4.4 the FLP predictor co-efficients $\{\hat{a}_k\}_{k=1}^P$ and associated LS error $E_{ls}^f$ are obtained by solving the normal equations:

$$
(\mathbf{X}^T \mathbf{X}) \hat{\mathbf{a}} = \begin{bmatrix}
    E_{ls}^f \\
    0
\end{bmatrix}
$$

where $\hat{\mathbf{a}} = [1 \quad \hat{a}_1 \quad \cdots \quad \hat{a}_P]^T$. The variance of the excitation process, $\sigma_w^2$, can be obtained by:

$$
\hat{\sigma}_w^2 = \frac{1}{N + P} E_{ls}^f \quad \text{for full-windowing}
$$

$$
\sigma_w^2 = \frac{1}{N - P} E_{ls}^f \quad \text{for no-windowing}
$$

5.1.3 Model Validation

The goal of model validation is to find out whether the model

- agrees sufficiently with the observed data
- describes the “true” signal generation system
- solves the problem that initiated the design process

One approach based on the LS error estimation of a PZ model is to note that in the ideal modelling case the residual is:

$$
e(n) = x(n) - \hat{x}(n) = w(n)
$$

that is, the residual is a realization of a white noise process. If the modelling is incorrect or unrepresentative then the residual will contain “unmodelled structure” which will detract from the white noise assumption. Thus statistical techniques based on the autocorrelation, power spectrum and partial autocorrelation of the residual signal can be used to ascertain whether the residual signal is “statistically significantly” close to a white noise signal.
5.1.4 Sunspot Data Example

**Problem**
Find the AP signal model that best describes the sunspot data in [Figure 5-1].

![Figure 5-1 Sunspot numbers from 1760 to 1869 (Figure 9.3[1])](image)

**Model Selection**
The Partial AutoCorrelation Sequence (PACS) of the sunspot data is plotted in [Figure 5-2]. The horizontal dashed lines represent the significance threshold for determining if \( k_m = 0 \). We see that \( k_m = 0 \) for \( m > 2 \) hence we select an AP model with \( P = 2 \).

![Figure 5-2 The PACS values of the sunspot numbers (Figure 9.6[1])](image)

**Model Estimation**
Using LS error analysis with full-windowing as described in Section 5.1.2 yields the following AR(2) model for the sunspot data:

\[
\hat{x}(n) = 1.318x(n-1) - 0.634x(n-2) + w(n) \quad \hat{\sigma}_w^2 = 289.2
\]

**Model Validation**
The residual \( e(n) = x(n) - \hat{x}(n) \) is computed and the following tests are performed to check whether the residual is statistically white:

- **autocorrelation test**: A plot of the normalised samples autocorrelation sequence, \( \hat{\rho}(l) = \hat{r}(l)/\hat{r}(0) \), of the residual should yield \( \hat{\rho}(0) = 1 \) and \( \hat{\rho}(l) \approx 0 \quad l > 0 \). This test is confirmed in [Figure 5-3].
- **PSD test**: A plot of the standardised cumulative periodogram of the residual should yield \( \tilde{I}(f) \propto f \). This is confirmed in [Figure 5-3].
• **PACS test**: A plot of the partial autocorrelation sequence of the residual should yield \( k_m \simeq 0 \quad m \geq 0 \). This is confirmed in Figure 5.3

![Residual samples](image1.png) ![Autocorrelation test](image2.png) ![PSD test](image3.png) ![Partial autocorrelation test](image4.png)

**Figure 5.3** Validation tests on the \( P=2 \), AP model fit to the sunspot numbers (Figure 9.7[1])

### 5.2 Spectral Estimation

The power spectral density (PSD) of a zero-mean stationary stochastic process was defined as:

\[
R_x(e^{j\omega}) = \sum_{l=-\infty}^{\infty} r_x(l)e^{-j\omega l}
\]

where the autocorrelation sequence \( r_x(l) \) is absolutely summable (i.e. \( \sum_{l=-\infty}^{\infty}|r_x(l)| < \infty \)). Our problem is to form an estimate of the power spectrum, \( \hat{R}_x(e^{j\omega}) \), of a stationary process \( x(n) \) from a finite record of observations \( \{x(n)\}_{0}^{N-1} \) of a single realisation.

#### 5.2.1 Non-parametric Methods

**Periodogram and Modified Periodogram**

The *periodogram* of the data segment \( \{x(n)\}_{0}^{N-1} \) is defined by:

\[
\hat{R}_x(e^{j\omega}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} v(n)e^{-j\omega n} \right|^2 = \frac{1}{N} |V(e^{j\omega})|^2
\]

*Equation 5.2*
is the Discrete-Time Fourier Transform (DTFT) of the windowed sequence:
\[ v(n) = x(n)w(n) \quad 0 \leq n \leq N - 1 \]
The window \( w(n) \), of length \( N \), is the data window.

If the data window is the rectangular window:
\[
w(n) = \begin{cases} 1 & 0 \leq n \leq N - 1 \\ 0 & \text{elsewhere} \end{cases}
\]

then Equation 5.2 is referred to as the periodogram.

If the data window is a nonrectangular window, e.g. the Hamming window:
\[
w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N - 1 \\ 0 & \text{elsewhere} \end{cases}
\]

then Equation 5.2 is referred to as the modified periodogram.

The values of the periodogram at the discrete set of frequencies \( \{\omega_k = 2\pi k/N\}_{k=0}^{N-1} \) can be calculated by:
\[
\hat{R}_x(k) = \hat{R}_x(e^{j2\pi k/N}) = \frac{1}{N} |\hat{V}(k)|^2 \quad k = 0, 1, \ldots, N - 1
\]

where:
\[
\hat{V}(k) = \sum_{n=0}^{N-1} v(n)e^{-j(2\pi k/n)}
\]
is the \( N \)-point DFT of the windowed sequence \( v(n) \).

In MATLAB, the modified periodogram is implemented by the function:
\[
Rx = \text{psd}(x, \text{Nfft}, \text{Fs}, \text{window}(N), \text{"none"});
\]
where:
- \( x \) is the \( N \)-length data segment
- \( \text{Nfft} \) is the size of the DFT, which is chosen such that \( \text{Nfft} > N \) to obtain a high-density spectrum by zero-padding.
- \( \text{Fs} \) is the sampling frequency
- \( \text{window}(N) \) is the name of any MATLAB-provided window function (e.g. \text{hamming}(N));

**Mean of \( \hat{R}_x(e^{j\omega}) \)**

It can be shown that:
\[
E\{\hat{R}_x(e^{j\omega})\} \neq R_x(e^{j\omega})
\]
that is, the periodogram is a biased estimate of the true power spectrum. This is because the expected value of the periodogram is obtained by convolving the true spectrum \( R_x(e^{j\omega}) \) with the spectrum \( R_w(e^{j\omega}) = |W(e^{j\omega})|^2 \) of the data window. However in cases where \( \lim_{N \to \infty} R_w(e^{j\omega}) \)
approaches an “impulse-like” function (as is the case with the rectangular window) and the
window is properly normalised such that \( \sum_{n=1}^{N} |w(n)|^2 = N \), we have the following result:

\[
\lim_{N \to \infty} E\{ \hat{R}_x(e^{j\omega}) \} = R_x(e^{j\omega})
\]

that is, the periodogram is an asymptotically unbiased estimator if the true power spectrum.

**Note**

The reason for choosing a nonrectangular window is that the spectrum of a rectangular
window exhibits a “sidelobe” structure that introduces false peaks in the periodogram when
convolved with the true power spectrum and other windows of the same length (e.g. Hamming) exhibit less prominent sidelobes and produce smoother and more accurate
modified periodograms.

---

**Variance of \( \hat{R}_x(e^{j\omega}) \)**

It can be shown that:

\[
\text{cov}\{ \hat{R}_x(e^{j(2\pi/N)k_1}), \hat{R}_x(e^{j(2\pi/N)k_2}) \} \equiv 0 \quad k_1 \neq k_2
\]

\[
\text{var}\{ \hat{R}_x(e^{j\omega}) \} \equiv R_x^2(e^{j\omega}) \left[ 1 + \left( \frac{\sin \omega N}{N \sin \omega} \right)^2 \right]
\]

that is, different realisations and values of the periodogram spaced in frequency by integer
multiples of \( 2\pi/N \) are uncorrelated, and, even for large \( N \), the variance of the estimate
remains at the level of \( R_x^2(e^{j\omega}) \).

**Thus the periodogram is a very poor estimator of the power spectrum function since it
exhibits high variability for different realisations of the signal.**

---

**Smoothing the Periodogram**

To reduce the variability of the periodogram one can make use of the fact that the variance of
a sum of \( K \) uncorrelated realisations of the same random variable is \( 1/K \) times the variance of
the random variable itself. Thus we can reduce the variability of the periodogram by
performing a smoothing operation. Two approaches are:

1. **Averaging contiguous values of the same periodogram.** The well-known method that
implements this approach is the *Blackman-Tukey* method [1, pages 222-227]. In MATLAB
this method is implemented by the SASP function:

\[
[R, \text{var}x] = \text{bt}_\text{psd}(x, L, \text{Nfft});
\]

where \( L \) is the window length or lag such that \( L < N \), \( N \) is the data segment length and \( \text{Nfft} > N \)
to ensure a high-density spectrum.

2. **Averaging periodograms obtained from multiple data segments.** The well-known method
that implements this approach is the *Welch-Bartlett* method [1, pages 227-232]. In MATLAB
this method is implemented by the function:

\[
R_x = \text{psd}(x, \text{Nfft}, \text{Fs}, \text{window}(L), \text{Noverlap}, \text{'none'});
\]

where \( L \) is the window or sub-segment length, \( N \) is the data segment length, \( \text{Noverlap} \)
specifies the number of overlapping samples and \( \text{Nfft} > N \) to ensure a high-density spectrum.
5.2.2 Parametric Methods

The non-parametric methods for spectrum estimation rely on the DFT and consequently exhibit a frequency resolution \( \Delta f \approx 1/N \) where \( N \) is the length of the data segment. Furthermore with smoothed periodograms the frequency resolution is reduced to \( \Delta f \approx 1/L \) where \( L < N \) is the window length. Increasing the data segment length is usually not possible if data has been collected over a short finite interval or the signal may only be considered stationary over limited intervals (e.g. for speech the typical analysis frame length of 20 - 30 ms corresponds to the interval of the largest unit of time that the speech articulators remain static, longer intervals will include movement of the speech articulators (i.e. non-stationary dynamics) which result in changes to the spectral characteristics of the signal).

Parametric methods rely on an underlying model that adequately describes the generation of the sampled data. The spectrum calculated from the model produces a high-resolution, smoothed spectrum due to the structure imposed by the model. However the success of such methods very much depends on how accurate the model is, how accurate the model parameters can be estimated, and how sensitive the model estimation is to perturbations (e.g. noise) present in the data or deficiencies in the model (e.g. incorrect order of model).

**LS all-pole method**

An AP model using LS estimation (see Section 5.1.2) is selected and the parameters estimated. The PSD is then given by:

\[
\hat{R}_s(e^{j\omega}) = \hat{\sigma}_n^2 \frac{1}{1 + \sum_{k=1}^{P} \hat{\alpha}_k e^{-j\omega k}}^2
\]

The following considerations should be noted:

- Unlike the non-parametric methods the LS all-pole method produces a continuous spectrum and provides values of the PSD at any frequency of interest.
- The estimator is consistent only if the analysed process is AR\((P_0)\) with \( P_0 \leq P \).
- In practice the value of \( P \) is limited by the amount of available data (usually \( P < N/3 \) for meaningful estimates).
- Order selection is important when there is model mismatch (i.e. the analysed process is not AR\((P_n)\)): if \( P \) is too large, the obtained spectrum exhibits spurious peaks, if \( P \) is too small, the PSD is smoothed over and resolution is lost.

The LS all-pole method can be generalised to the LS pole-zero method where a PZ model is selected but the model parameters estimated by this approach are limited by computational and other practical difficulties.

**Example of parametric spectral estimation using the LS AP/PZ methods**

The PSD of an ARMA(4,2) process is estimated by using the LS AP/PZ methods of order AP(10) and PZ(4,2) on a 300 sample segment. The actual PSD and estimated PSD’s are plotted in Figure 5-4. The effect of the model mismatch is evident when using the AP(10) method (especially the spectral valley between the two spectral peaks and the high-frequency response) compared to accuracy of the estimated PSD when using the correct model in the
PZ(4,2) method. However even with a model mismatch the AP(10) accurately identifies the spectral peaks in the spectrum. Indeed it can be shown that the AP models tend to accurately model spectral peaks but not spectral valleys.

![Figure 5-4 Actual PSD and estimated PSD from the AP(10) and PZ(4,2) methods for an ARMA(4,2) process (Figure 9.13[1])](image)

5.3 References