Chapter 8

Subspace Methods

8.1 Introduction

Principal Component Analysis (PCA) is applied to the analysis of time series data. In this context we discuss measures of complexity and subspace methods for spectral estimation.

8.2 Singular Spectrum Analysis

8.2.1 Embedding

Given a single time series x_1 to x_N we can form an *embedding* of dimension d by taking length d snapshots $\boldsymbol{x}_t = [x_t, x_{t+1}, ..., x_{t+d}]$ of the time series. We form an *embedding matrix* $\boldsymbol{X}$ with different snapshots in different rows. For $d = 4$ for example

$$\boldsymbol{X} = \frac{1}{\sqrt{N}} \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_3 & x_4 & x_5 \\ .. & .. & .. & .. \\ x_{N-3} & x_{N-2} & x_{N-1} & x_N \end{bmatrix} \tag{8.1}$$

The normalisation factor is there to ensure that $\boldsymbol{X}^T\boldsymbol{X}$ produces the covariance matrix (see PCA section).

$$\boldsymbol{C} = \boldsymbol{X}^T\boldsymbol{X} \tag{8.2}$$

We note that embedding is identical to the procedure used in autoregressive modelling to generate the 'input data matrix'. Similarly, we see that the covariance matrix of embedded data is identical to the autocovariance matrix

$$\boldsymbol{C} = \begin{bmatrix} \sigma_{xx}(0) & \sigma_{xx}(1) & \sigma_{xx}(2) & \sigma_{xx}(3) \\ \sigma_{xx}(1) & \sigma_{xx}(0) & \sigma_{xx}(1) & \sigma_{xx}(2) \\ \sigma_{xx}(2) & \sigma_{xx}(1) & \sigma_{xx}(0) & \sigma_{xx}(1) \\ \sigma_{xx}(3) & \sigma_{xx}(2) & \sigma_{xx}(1) & \sigma_{xx}(0) \end{bmatrix} \tag{8.3}$$

where $\sigma_{xx}(k)$ is the autocovariance at lag k.

The application of PCA to embedded data (using either SVD on the embedding matrix or eigendecomposition on the autocovariance matrix) is known as Singular Spectrum Analysis (SSA) [18] or PCA Embedding.

8.2.2 Noisy Time Series

If we suppose that the observed time series x_n consists of a signal s_n plus additive noise e_n of variance σ_e^2 then

$$x_n = s_n + e_n \tag{8.4}$$

If the noise is uncorrelated from sample to sample (a key assumption) then the noise autocovariance matrix is equal to $\sigma_e^2 \boldsymbol{I}$. If the signal has autocovariance matrix $\boldsymbol{C}_s$ and corresponding singular values s_k then application of SVD to the observed embedding matrix will yield the singular values (see section 8.3 for a proof)

$$\sigma_k = s_k + \sigma_e \tag{8.5}$$

Thus, the biggest singular values correspond to signal plus noise and the smallest to just noise. A plot of the singular values is known as the *singular spectrum*. The value σ_e is the *noise floor*. By reconstructing the time series from only those components above the noise floor we can remove noise from the time series.

Projections and Reconstructions

To find the projection of the data onto the kth principal component we form the projection matrix

$$\boldsymbol{P} = \boldsymbol{Q}^T \boldsymbol{X}^T \tag{8.6}$$

where $\boldsymbol{Q}$ contains the eigenvectors of $\boldsymbol{C}$ ($\boldsymbol{Q}_2$ from SVD) and the kth row of $\boldsymbol{P}$ ends up containing the projection of the data onto the kth component. We can see this more clearly as follows, for $d = 4$

$$\boldsymbol{P} = \begin{bmatrix} - & - & \boldsymbol{q}_1 & - & - \\ - & - & \boldsymbol{q}_2 & - & - \\ - & - & \boldsymbol{q}_3 & - & - \\ - & - & \boldsymbol{q}_4 & - & - \end{bmatrix} \begin{bmatrix} x_1 & x_2 & . & x_{N-3} \\ x_2 & x_3 & . & x_{N-2} \\ x_3 & x_4 & . & x_{N-1} \\ x_4 & x_5 & . & x_N \end{bmatrix} \tag{8.7}$$

We can write the projection onto the kth component explicitly as

$$\boldsymbol{p}_k = \boldsymbol{q}_k^T \boldsymbol{X}^T \tag{8.8}$$

After plotting the singular spectrum and identifying the noise floor the signal can be reconstructed using only those components from the signal subspace. This is achieved by simply summing up the contributions from the first M chosen components

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{M} \boldsymbol{p}_k \tag{8.9}$$

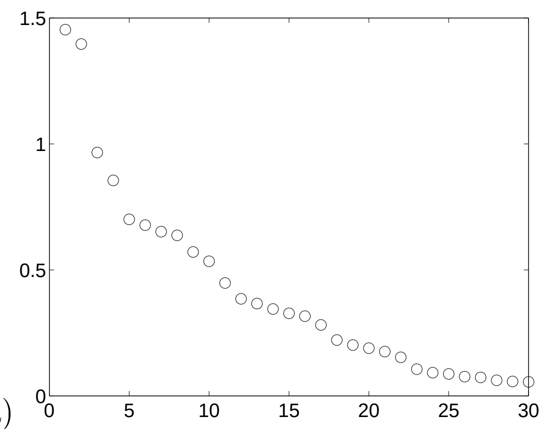
which is a row vector whose nth element, $\hat{x}_n$ contains the reconstruction of the original signal x_n.

From the section on dimensionality reduction (lecture 3) we know that the average reconstruction error will be

$$E_M = \sum_{k=M+1}^{d} \lambda_k \tag{8.10}$$

where $\lambda_k = \sigma_k^2$ and we expect that this error is solely due to the noise, which has been removed by SSA.

The overall process of projection and reconstruction amounts to a *filtering* or *denoising* of the signal. Figure 8.1 shows the singular spectrum (embedding dimension $d = 30$) of a short section of EEG. Figure 8.2 shows the original EEG data and the SSA filtered data using only the first 4 principal components.



(a)

Figure 8.1: *Singular spectrum of EEG data: A plot of λ_k versus k.*

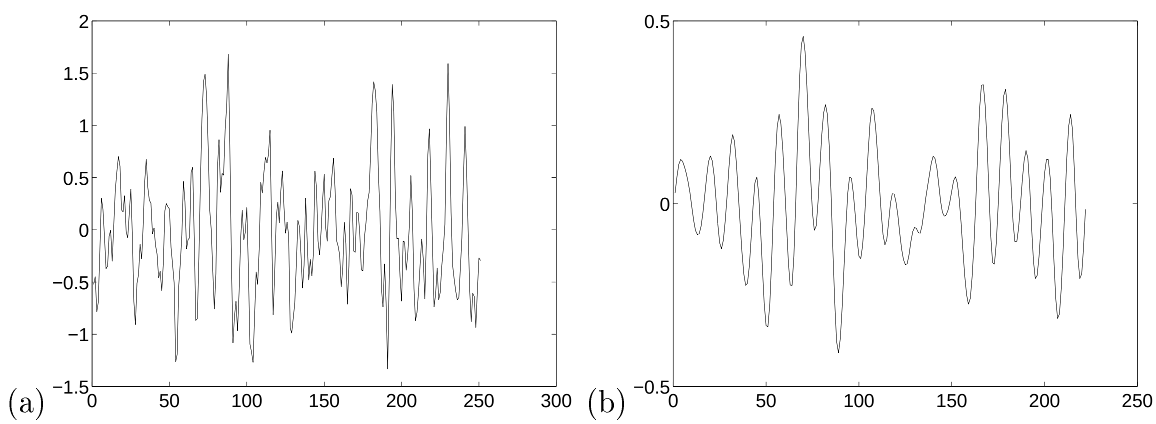


Figure 8.2: *(a) EEG data and (b) SSA-filtered EEG data.*

8.2.3 Embedding Sinewaves

A pure sinewave

If we embed a pure sinewave with embedding dimension $d = 2$ then we can view the data in the 'embedding space'. Figure 8.3 shows two such embeddings; one for a low frequency sinewave and one for a high frequency sinewave. Each plot shows that the data lie on a closed loop. There are two points to note.

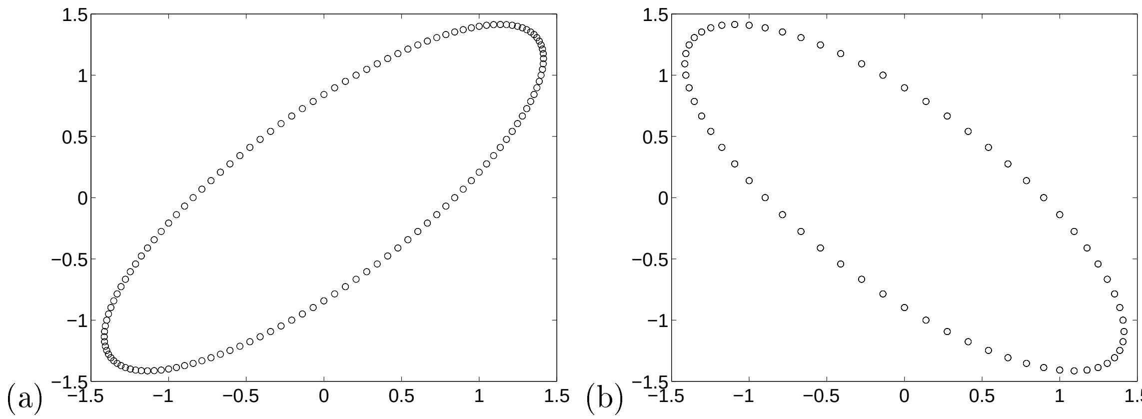


Figure 8.3: *Embedding Sinewaves: Plots of* x_{n+1} *versus* x_n *for sinewaves at frequencies of (a) 13Hz and (b) 50Hz.*

Firstly, whilst a loop is intrinsically a 1-dimensional object (any point on the loop can be described by a single number; how far round the loop from an agreed reference point) in terms on linear bases (straight lines and planes) we need *two* basis vectors. If the embedding took place in a higher dimension ($d > 2$) we would still need two basis vectors. Therefore, if we embed a pure sinewave in d dimensions the number of corresponding singular values will be 2. The remaining singular values will be zero.

Secondly, for the higher frequency signal we have fewer data points. This will become relevant when we talk about spectral estimation methods based on SVD.

Multiple sinewaves in noise

We now look at using SSA on data consisting of multiple sinusoids with additive noise. As an example we generated data from four sinusoids of different ampltidues and additive Gaussian noise. The amplitudes and frequencies were $a_1 = 2, a_2 = 4, a_3 = 3, a_4 = 1$ and $f_1 = 13, f_2 = 29, f_3 = 45, f_4 = 6$ and the standard deviation of the noise was $\sigma_e = 2$. We generated 3 seconds of data and sampled at 128Hz. We then embedded the data in dimension $d = 30$. Application of SVD yielded the singular spectrum shown in Figure 8.4; we also show the singular spectrum obtained for a data set containing just the first two sinewaves. The pairs of singular values constitutng the signal are clearly visible. Figure 8.5 shows the Power Spectral Densities (computed using Welch's modified periodogram method; see earlier) of the projections onto the

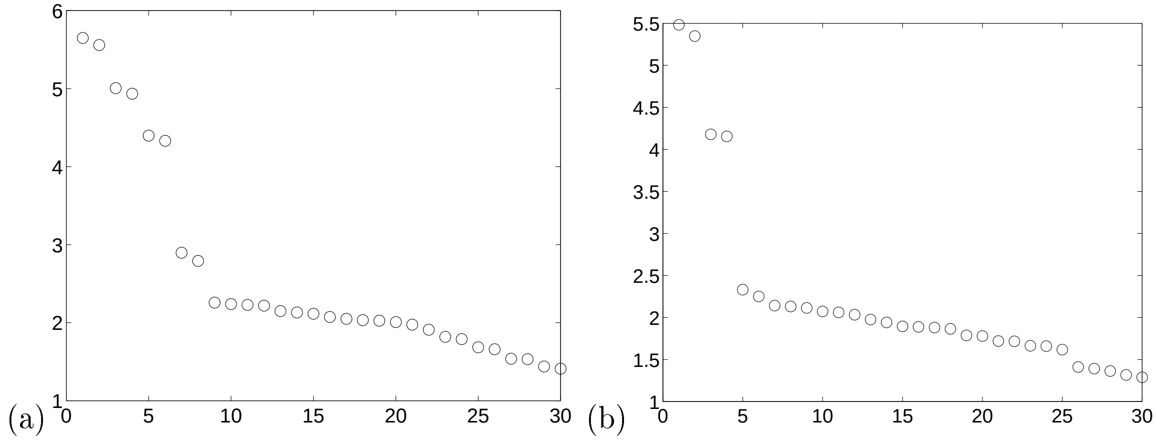


Figure 8.4: *The singular spectrums for (a) $p = 4$ and (b) $p = 2$ sinewaves in additive noise.*

first four pairs of principal components. They clearly pick out the corresponding sinewaves.

8.3 Spectral estimation

If we assume that our signal consists of p complex sinusoids

$$s_k = \exp(i2\pi f_k n) \tag{8.11}$$

where $k = 1..p$ then the signal autocovariance function, being the inverse Fourier transform of the Power Spectral Density, is

$$\sigma_{xx}(m) = \sum_{k=1}^{p} P_k \exp(i2\pi f_k m) \tag{8.12}$$

where m is the lag, P_k and f_k are the power and frequency of the kth complex sinusoid and $i = \sqrt{-1}$. If the signal embedding dimension is d, where $d > p$, then we can compute $\sigma_{xx}(m)$ for $m = 0..d-1$. The corresponding autocovariance matrix, for $d = 4$, for example is given by

$$\boldsymbol{C}_{xx} = \begin{bmatrix} \sigma_{xx}(0) & \sigma_{xx}(1) & \sigma_{xx}(2) & \sigma_{xx}(3) \\ \sigma_{xx}(1) & \sigma_{xx}(0) & \sigma_{xx}(1) & \sigma_{xx}(2) \\ \sigma_{xx}(2) & \sigma_{xx}(1) & \sigma_{xx}(0) & \sigma_{xx}(1) \\ \sigma_{xx}(3) & \sigma_{xx}(2) & \sigma_{xx}(1) & \sigma_{xx}(0) \end{bmatrix} \tag{8.13}$$

The kth sinusoidal component of the signal at these d points is given by the d-dimensional vector

$$\boldsymbol{s}_k = [1, \exp(i2\pi f_k), \exp(i4\pi f_k), ..., \exp(i2\pi(M-1)f_k)]^T \tag{8.14}$$

The autocovariance matrix can now be written as follows

$$\boldsymbol{C}_{xx} = \sum_{k=1}^{p} P_k \boldsymbol{s}_k \boldsymbol{s}_k^H \tag{8.15}$$

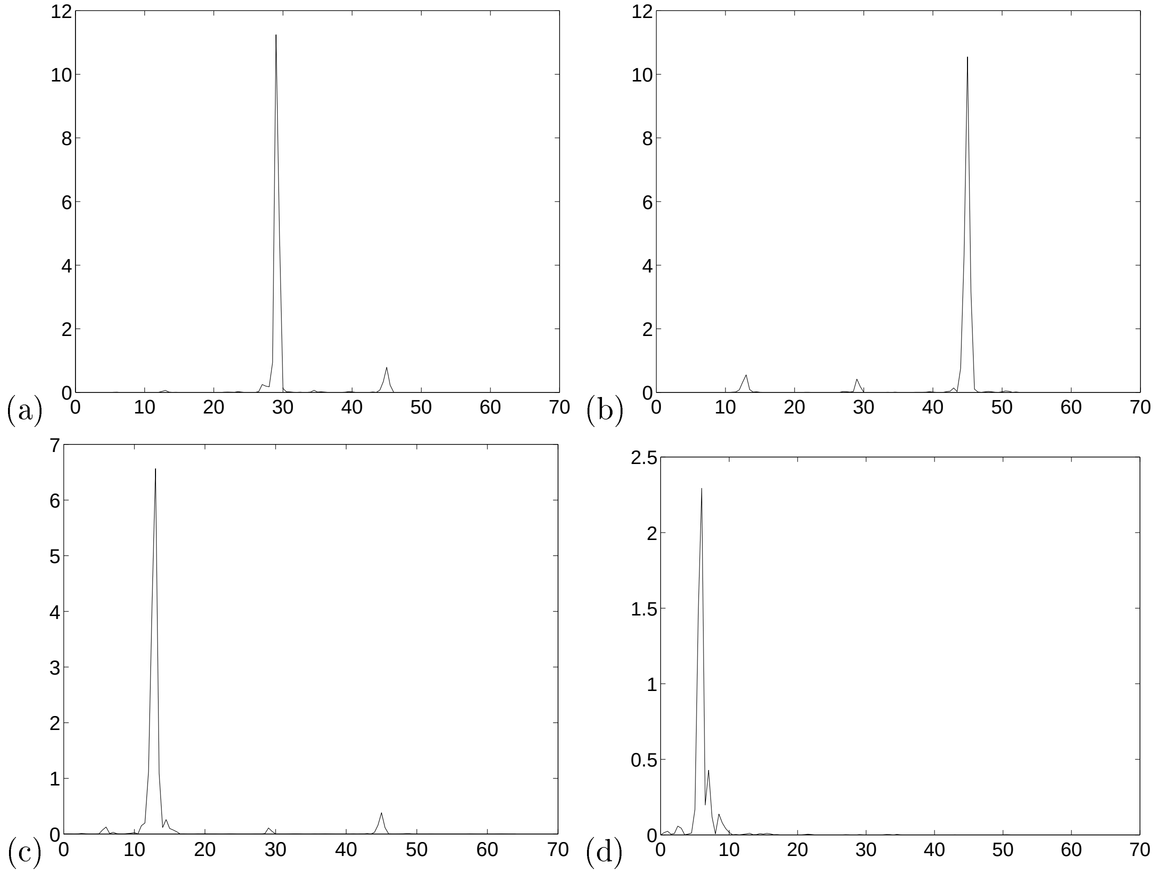


Figure 8.5: *The Power Spectral Densities of the (a) first (b) second (c) third and (d) fourth pairs of projections. They clearly correspond to the original pure sinewaves which were, in order of amplitude, of frequencies 29, 45, 13 and 6Hz. The Fourier transform of the data is the sum of the Fourier transforms of the projections.*

where H is the Hermitian transpose (take the conjugate and then the transpose).

We now model our time series as signal plus noise. That is

$$y[n] = x[n] + e[n] \tag{8.16}$$

where the noise has variance σ_e^2. The autocovariance matrix of the observed time series is then given by

$$\boldsymbol{C}_{yy} = \boldsymbol{C}_{xx} + \sigma_e^2 \boldsymbol{I} \tag{8.17}$$

We now look at an eigenanalysis of $\boldsymbol{C}_{yy}$ where the eigenvalues are ordered $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_M$ where M is the embedding dimension. The corresponding eigenvectors are $\boldsymbol{q}_k$ (as usual, they are normalised). In the absence of noise, the eigenvalues $\lambda_1, \lambda_2, .., \lambda_p$ will be non-zero while $\lambda_{p+1}, \lambda_{p+2}, .., \lambda_M$ will be zero (this is because there are only p degrees of freedom in the data - from the p sinusoids).

The signal autocovariance matrix can therefore be written as

$$\boldsymbol{C}_{xx} = \sum_{k=1}^{p} \lambda_k \boldsymbol{q}_k \boldsymbol{q}_k^H \tag{8.18}$$

(this is the usual $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H$ eigendecomposition written as a summation) where the sum runs only over the first p components.

In the presence of noise, $\lambda_1, \lambda_2, .., \lambda_p$ and $\lambda_{p+1}, \lambda_{p+2}, .., \lambda_M$ will be non-zero. Using the orthogonality property $\boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{I}$ we can write the noise autocovariance as

$$\sigma_e^2 \boldsymbol{I} = \sigma_e^2 \sum_{k=1}^{M} \boldsymbol{q}_k \boldsymbol{q}_k^H \tag{8.19}$$

where the sum runs over *all* M components.

Combining the last two results allows us to write the observed autocovariance matrix as

$$\boldsymbol{C}_{yy} = \sum_{k=1}^{p} (\lambda_k + \sigma_e^2) \boldsymbol{q}_k \boldsymbol{q}_k^H + \sum_{k=p+1}^{M} \sigma_e^2 \boldsymbol{q}_k \boldsymbol{q}_k^H \tag{8.20}$$

We have two sets of eigenvectors. The first p eigenvectors form a basis for the *signal subspace* while the remaining eigenvectors form a basis for the *noise subspace*. This last name is slightly confusing as the noise also appears in the signal subspace; the *signal*, however, does not appear in the noise subspace. In fact, the signal is orthogonal to the eigenvectors constituting the noise subspace. This last fact can be used to estimate the frequencies in the signal.

Suppose, for example, that $d = p + 1$. This means there will be a single vector in the noise subspace and it will be the one with the smallest eigenvalue. Now, because the signal is orthogonal to the noise we can write

$$\boldsymbol{s}_k^H \boldsymbol{q}_{p+1} = 0 \tag{8.21}$$

If we write the elements of $\boldsymbol{q}_{p+1}$ as q_{p+1}^k then we have which can be written as

$$\sum_{k=1}^{d} q_{p+1}^k \exp(-i2\pi(k-1)f_k) = 0 \tag{8.22}$$

Writing $z_k = \exp(-i2\pi k f_k)$ allows the above expression to be written in terms of a polynomial in z. The roots allow us to identify the frequencies. The amplitudes can then be found by solving the usual AR-type equation. This method of spectral estimation is known as *Pisarenko's harmonic decomposition* method.

More generally, if we have $d > p + 1$ (ie. p is unknown) then we can use the Multiple Signal Classification (MUSIC) algorithm. This is essentially the same as Pisarenko's method except that the noise variance is estimated as the average of the $d - p$ smallest eigenvalues. See Proakis [51] for more details. Figure 8.6 compares spectral estimates for the MUSIC algorithm versus Welch's method on synthetic data containing 5 pure sinusoids and additive Gaussian noise.

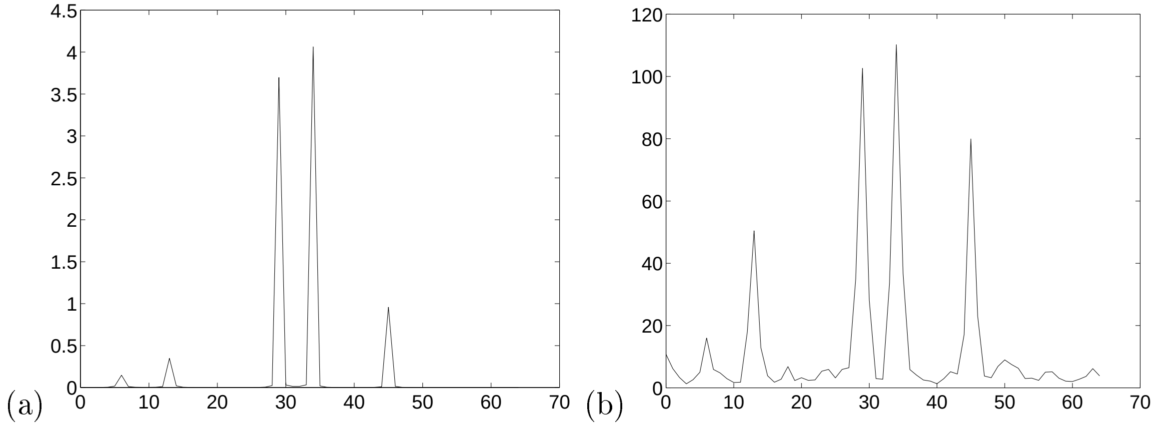


Figure 8.6: *Power Spectral Density estimates from (a) MUSIC and (b) Welch's modified periodogram.*

8.3.1 Model Order Selection

Wax and Kailath [62] suggest the Minimum Description Length (MDL) criterion for selecting p

$$MDL(p) = -N \log \left(\frac{G(p)}{A(p)} \right) + E(p) \tag{8.23}$$

where

$$\begin{aligned} G(p) &= \prod_{k=p+1}^{d} \lambda_k \\ A(p) &= \left[\frac{1}{d-p} \sum_{k=p+1}^{d} \lambda_k \right]^{d-p} \\ E(p) &= \frac{1}{2} p(2d-p) \log N \end{aligned} \tag{8.24}$$

where d is the embedding dimension, N is the number of samples and λ_k are the eigenvalues. The optimal value of p can be used as a measure of signal *complexity*.

8.3.2 Comparison of methods

Kay and Marple [31] provide a comprehensive tutorial on the various spectral estimation methods. Pardey *et. al* [45] show that the AR spectral estimates are typically better than those obtained from periodogram or autocovariance-based methods. Proakis and Manolakis (Chapter 12) [51] tend to agree, although for data containing a small number of sinusoids in additive noise, they advocate the MUSIC algorithm and its relatives.