# 14
# Computational Aspects in Statistical Signal Processing

*D. Kundu*

## 14.1 Introduction

Signal processing may broadly be considered to involve the recovery of information from physical observations. The received signal is usually disturbed by thermal, electrical, atmospheric or intentional interferences. The received signal can not be predicted deterministically, so statistical methods are needed to describe the signals. The problem of detecting the signal in the midst of noise arises in a wide variety of areas such as communications, radio location of objects, seismic signal processing and computer assisted medical diagnosis. Statistical signal processing is also used in many physical science applications, such as geophysics, acoustics, texture classifications, voice recognition and meteorology among many others. During the past fifteen years, tremendous advances have been achieved in the area of digital technology in general and in statistical signal processing in particular. Several sophisticated statistical tools have been used quite effectively in solving various important signal processing problems. Various problem specific algorithms have evolved in analyzing different type of signals and they are found to be quite effective in practice.

The main purpose of this chapter is to familiarize the statistical community with some of the signal processing problems where statisticians can play a very important role to provide satisfactory solutions. We provide examples of the different statistical signal processing models and describe in detail the computational issues involved in estimating the parameters of one of the most important models, namely the undamped exponential models. The undamped exponential model plays a significant role in statistical signal processing and related areas.

We also would like to mention that statistical signal processing is a huge area and it is not possible to cover all the topics in a limited space. We provide several important references for further reading.

To begin with we give several examples to motivate the readers and to have a feeling for the subject.

EXAMPLE 1: In speech signal processing, the accurate determination of the resonant frequencies of the vocal tract (i.e. the formant frequencies) in various articulatory configuration is of interest, both in synthesis and in the analysis of speech. In particular for vowel sounds, the formant frequencies play a dominant role in determining which vowel is produced by a speaker and which vowel is perceived by a listener. If it is assumed that the vocal tract is represented by a tube of varying cross sectional area and the configuration of the vocal tract varies little during a time interval of one pitch period then the pressure variation $\boldsymbol{p(t)}$ at the acoustic pick up at the time point $\boldsymbol{t}$ can be written as

$$\boldsymbol{p(t) = \sum_{i=1}^{K} c_i e^{s_i t}}$$

(see Fant 1960; Pinson 1963). Here $\boldsymbol{c_i}$ and $\boldsymbol{s_i}$ are complex parameters. Thus for a given sample of an individual pitch frame it is important to estimate the amplitudes $\boldsymbol{c_i}$'s and the poles $\boldsymbol{s_i}$'s.

EXAMPLE 2: In radioactive tracer studies the analysis of the data obtained from different biological experiments is of increasing importance. The tracer material may be simultaneously undergoing diffusion, excretion or interaction in any biological systems. In order to analyze such a complex biological system it is often assumed that the biological system is a simple compartment model. Such a model is valid to the extent that the results calculated on the basis of this model agree with those actually obtained in a real system. Further simplification can be done by assuming that the system is in a steady state, that is the inflow of the stable material is equal to the outflow. It has been observed by Sheppard and Householder (1951) that under above assumptions, specific activity at the time point $\boldsymbol{t}$ can be well approximated by

$$\boldsymbol{f(t) = \sum_{i=0}^{p} N_i e^{\lambda_i t},}$$

where $\boldsymbol{f(t)}$ is the specific density of the material at the time point $\boldsymbol{t}$, $\boldsymbol{N_i}$'s are the amplitudes, $\boldsymbol{\lambda_i}$'s are the rate constants and $\boldsymbol{p}$ represents the number of components. It is important to have a method of estimation of $\boldsymbol{N_i}$, $\boldsymbol{\lambda_i}$ as well as $\boldsymbol{p}$.

EXAMPLE 3: Next consider a signal $\boldsymbol{y(t)}$ composed of $\boldsymbol{M}$ unknown damped

or undamped sinusoidal signals , closely spaced in frequency, ***i.e.***,

$$y(t) = \sum_{k=1}^{M} A_k e^{-\delta_k t + j2\pi\omega_k t},$$

where $A_k$'s are amplitudes, $\delta_k(> 0)$'s are the damping factors, $\omega_k$'s are the frequencies and $j = \sqrt{-1}$. For a given signal $y(t)$ at finite time points, the problem is to estimate the signal parameters, namely $A_k$'s, $\delta_k$'s, $\omega_k$'s and $M$. This problem is called the spectral resolution problem and is very important in digital signal processing.

EXAMPLE 4: In texture analysis it is observed that certain metal textures can be modeled by the following two dimensional model

$$y(m,n) = \sum_{k=1}^{P} \left(A_k \cos(m\lambda_k + n\mu_k) + B_k \sin(m\lambda_k + n\mu_k)\right),$$

here $y(m,n)$ represents the gray level at the $(m,n)$-th location, $(\mu_k, \lambda_k)$ are the unknown two dimensional frequencies and $A_k$ and $B_k$ are unknown amplitudes. Given the two dimensional noise corrupted images it is often required to estimate the unknown parameters $A_k$, $B_k$, $\mu_k$, $\lambda_k$ and $P$. For different forms of textures, the readers are referred to the recent articles of Zhang and Mandrekar (2002) or Kundu and Nandi (2003).

EXAMPLE 5: Consider a linear array of $P$ sensors which receives signals, say $\mathbf{x}(t) = (x_1(t), \ldots, x_M(t))$, from $M$ sources at the time point $t$. The signals arrive at the array at angles $\theta_1, \ldots, \theta_M$ with respect to the line of array. One of the sensors is taken to be the reference element. The signals are assumed to travel through a medium that only introduces propagation delay. In this situation, the output at any of the sensors can be represented as a time advanced or time delayed version of the signals at the reference element. The output vector $\mathbf{y}(t)$ can be written in this case

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \epsilon(t); \qquad t = 1, \ldots, N,$$

where $\mathbf{A}$ is a $P \times M$ matrix of parameters which represents the time delay. The matrix $\mathbf{A}$ can be represented as

$$\mathbf{A} = [\mathbf{a}(\omega_1), \ldots, \mathbf{a}(\omega_M)],$$

where

$$\omega_k = \pi \cos(\theta_k) \quad \text{and} \quad \mathbf{a}(\omega_k) = [1, e^{-j\omega_k}, \ldots, e^{-j(P-1)\omega_k}]^T.$$

With suitable assumptions on the distributions of $\mathbf{x}(.)$ and $\epsilon(.)$, given a sample at $N$ time points, the problem is to estimate the number of sources and also the directions of arrival $\theta_1, \ldots, \theta_M$.

## 14.2 Notation and Preliminaries

Throughout this chapter, scalar quantities are denoted by regular lower or upper case letters, lower and upper case bold type faces are used for vectors and matrices. For real matrix, $\mathbf{A}^{\boldsymbol{T}}$ denotes the transpose of the matrix $\mathbf{A}$ and for complex matrix $\mathbf{A}$, $\mathbf{A}^{\boldsymbol{H}}$ denotes the complex conjugate transpose. An $\boldsymbol{n} \times \boldsymbol{n}$ diagonal matrix with diagonal entries $\boldsymbol{\lambda_1, \ldots \lambda_n}$ is denoted by $\text{diag}\{\boldsymbol{\lambda_1, \ldots, \lambda_n}\}$. Suppose $\mathbf{A}$ is a $\boldsymbol{m} \times \boldsymbol{m}$ real matrix, then the projection matrix on the column space of $\mathbf{A}$ is denoted by $\mathbf{P_A} = \mathbf{A}(\mathbf{A}^{\boldsymbol{T}}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^{\boldsymbol{T}}$.

We need the following definition and one matrix theory result. For a detailed discussions of matrix theory the readers are referred to Rao (1973), Marple (1987) and Davis (1979).

DEFINITION 1: An $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\mathbf{J}$ is called a reflection (or exchange) matrix if

$$\mathbf{J} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{1} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

RESULT 1: (Spectral Decomposition) If an $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\mathbf{A}$ is Hermitian, then all its eigenvalues are real and it is possible to find $\boldsymbol{n}$ normalized eigenvectors $\mathbf{v_1, \ldots, v_n}$ corresponding to $\boldsymbol{n}$ eigenvalues $\boldsymbol{\lambda_1 \ldots, \lambda_n}$ such that

$$\mathbf{A} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\lambda_i} \mathbf{v}_{\boldsymbol{i}} \mathbf{v}_{\boldsymbol{i}}^{\boldsymbol{H}}.$$

If all the $\boldsymbol{\lambda_i}$'s are non-zero then from Result 1, it is immediate that

$$\mathbf{A^{-1}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \frac{\mathbf{1}}{\boldsymbol{\lambda_i}} \mathbf{v}_{\boldsymbol{i}} \mathbf{v}_{\boldsymbol{i}}^{\boldsymbol{H}}.$$

Now we provide one important result used in statistical signal processing known as Prony's method. It was originally proposed more than two hundred years back by Prony (1795), a Chemical engineer. It is described in several numerical analysis text books and papers, see for example Froberg (1969), Hildebrand (1956), Lanczos (1964) and Barrodale and Oleski (1981). Prony first observed that for arbitrary real constants $\boldsymbol{\alpha_1, \ldots, \alpha_M}$ and for distinct constants $\boldsymbol{\beta_1, \ldots, \beta_M}$, if

$$\boldsymbol{\mu_i = \alpha_1 e^{\beta_1 i} + \ldots + \alpha_M e^{\beta_M i}; \qquad i = 1, \ldots, n,} \tag{14.1}$$

then there exists $\boldsymbol{(M+1)}$ constants, $\{\boldsymbol{g_0, \ldots, g_M}\}$, such that

$$\begin{bmatrix} \boldsymbol{\mu_1} & \ldots & \boldsymbol{\mu_{M+1}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\mu_{n-M}} & \ldots & \boldsymbol{\mu_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{g_0} \\ \vdots \\ \boldsymbol{g_M} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}. \tag{14.2}$$

Note that without loss of generality we can always put a restriction on $g_i$'s such that $\sum_{i=0}^{M} g_i^2 = 1$. The above $n - M$ equations are called Prony's equations. Moreover, the roots of the following polynomial equation

$$g_0 + g_1 z + \ldots + g_M z^M = 0 \tag{14.3}$$

are $e^{\beta_1}, \ldots, e^{\beta_M}$. Therefore, there is a one to one correspondence between $\{g_0, \ldots, g_M\}$, such that

$$\sum_{i=0}^{M} g_i^2 = 1, \qquad g_0 > 0 \tag{14.4}$$

and the nonlinear parameters $\{\beta_1, \ldots, \beta_M\}$ as given above. It is also interesting to note that $g_0, \ldots, g_M$ are independent of the linear parameters $\alpha_1, \ldots, \alpha_M$. One interesting question is how to recover $\alpha_1, \ldots, \alpha_M$ and $\beta_1, \ldots, \beta_M$, if $\mu_1, \ldots, \mu_n$ are given. It is clear that for a given $\mu_1, \ldots, \mu_n$, $g_0, \ldots, g_M$ are such that, (14.2) is satisfied and they can be easily recovered by solving the linear equations (14.2). From $g_0, \ldots, g_M$, by solving (14.3), $\beta_1, \ldots, \beta_M$ can be obtained. Now to recover $\alpha_1, \ldots, \alpha_M$, we write (14.1) as

$$\mu = \mathbf{X}\alpha, \tag{14.5}$$

where $\mu = (\mu_1, \ldots, \mu_n)^T$, and $\alpha = (\alpha_1, \ldots, \alpha_M)^T$ are $n \times 1$ and $M \times 1$ vectors respectively. The $n \times M$ matrix $\mathbf{X}$ is as follows:

$$\mathbf{X} = \begin{bmatrix} e^{\beta_1} & \ldots & e^{\beta_M} \\ \vdots & \vdots & \vdots \\ e^{n\beta_1} & \ldots & e^{n\beta_M}. \end{bmatrix}.$$

Therefore, $\alpha = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\,\mu$. Since $\beta_i$'s are distinct and $M < n$, note that $\mathbf{X}^T\mathbf{X}$ is a full rank matrix.

## 14.3 Undamped Exponential Signal Parameters Estimation

The estimation of the frequencies of the sinusoidal components embedded in additive white noise is a fundamental problem in signal processing. It arises in many areas of signal detection. The problem can be written mathematically as follows:

$$y(n) = \sum_{i=1}^{M} A_i e^{j\omega_i n} + z(n), \tag{14.6}$$

where $A_1, \ldots, A_n$ are unknown complex amplitudes, $\omega_1, \ldots, \omega_n$ are unknown frequencies, $\omega_i \in (0, 2\pi)$, and $z(n)$s are independent and identically distributed random variables with mean zero and finite variance $\frac{\sigma^2}{2}$

for both the real and imaginary parts. Given a sample of size $\boldsymbol{N}$, namely $\boldsymbol{y(1), \ldots, y(N)}$, the problem is to estimate the unknown amplitudes, the unknown frequencies and sometimes the order $\boldsymbol{M}$ also. For a single sinusoid or for multiple sinusoids, where frequencies are well separated, the periodogram function can be used as a frequency estimator and it provides a optimal solution. But for unresolvable frequencies, the periodogram can not be used and also in this case the optimal solution is not very practical. The optimal solutions, namely the maximum likelihood estimators or the least squares estimators involve huge computations and usually the general purpose algorithms like Gauss-Newton algorithm or Newton-Raphson algorithm do not work. This has led to the introduction of several sub-optimal solutions based on the eigenspace approach. But interestingly most proposed methods use the idea of Prony's algorithm (as discussed in the previous section) some way or the other. In the complex model, it is observed that the Prony's equations work and in case of undamped exponential model (14.6), there exists a symmetry relations between the Prony's coefficients. These symmetry relations can be used quite effectively to develop some efficient numerical algorithms.

### 14.3.1 Symmetry Relation

If we denote $\boldsymbol{\mu_n = E(y(n)) = \sum_{i=1}^{M} A_i e^{j\omega_i n}}$, then it is clear that there exists constants $\boldsymbol{g_0, \ldots, g_M}$, with $\boldsymbol{\sum_{i=0}^{M} |g_i|^2 = 1}$, such that

$$\begin{bmatrix} \boldsymbol{\mu_1} & \boldsymbol{\cdots} & \boldsymbol{\mu_{M+1}} \\ \boldsymbol{\vdots} & \boldsymbol{\vdots} & \boldsymbol{\vdots} \\ \boldsymbol{\mu_{n-M}} & \boldsymbol{\cdots} & \boldsymbol{\mu_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{g_0} \\ \boldsymbol{\vdots} \\ \boldsymbol{g_M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\vdots} \\ \boldsymbol{0} \end{bmatrix}. \tag{14.7}$$

Moreover, $\boldsymbol{z_1 = e^{j\omega_1}, \ldots, z_n = e^{j\omega_n}}$ are the roots of the polynomial equation

$$\boldsymbol{P(z) = g_0 + g_1 z + \ldots + g_M z^M = 0.}$$

Note that

$$\boldsymbol{|z_1| = \ldots = |z_M| = 1, \quad \bar{z}_i = z_i^{-1}; \qquad i = 1, \ldots, M.} \tag{14.8}$$

Define the polynomial $\boldsymbol{Q(z)}$ by

$$\boldsymbol{Q(z) = z^{-M} \bar{P}(z) = \bar{g}_0 z^{-M} + \ldots + \bar{g}_M.}$$

Using (14.8), it is clear that $\boldsymbol{P(z)}$ and $\boldsymbol{Q(z)}$ have the same roots. Comparing coefficients of the two polynomials yields

$$\boldsymbol{\frac{g_k}{g_M} = \frac{\bar{g}_{M-k}}{\bar{g}_0}; \quad k = 0, \ldots, M.}$$

Define

$$b_k = g_k \left( \frac{\bar{g}_0}{g_M} \right)^{-\frac{1}{2}}; \quad k = 0, \ldots, M,$$

thus

$$b_k = \bar{b}_{M-k}; \quad k = 0, \ldots, M. \tag{14.9}$$

The condition (14.9) is the conjugate symmetric property and can be written compactly as

$$\mathbf{b} = \mathbf{J}\bar{\mathbf{b}},$$

here $\mathbf{b} = (b_0, \ldots, b_M)^T$ and $\mathbf{J}$ is an exchange matrix as defined before. Therefore, without loss of generality it is possible to say the vector $\mathbf{g} = (g_0, \ldots, g_M)^T$, such that $\sum_{i=0}^{M} |g_i|^2 = 1$ which satisfies (14.7) also satisfies

$$\mathbf{g} = \mathbf{J}\bar{\mathbf{g}}.$$

It will be used later on to develop efficient numerical algorithm to obtain estimates of the unknown parameters of the model (14.6).

### 14.3.2 Least Squares Estimators and Their Properties

In this subsection, we consider the least squares estimators and provide their asymptotic properties. It is observed that the model (14.6) is a nonlinear model and therefore it is difficult to obtain theoretically any small sample properties of the estimators. The asymptotic variances of the least squares estimators attain the Cramer-Rao lower bound under the assumptions that the errors are independently and identically distributed normal random variables. Therefore, it is reasonable to compare the variances of the different estimators with the asymptotic variances of the least squares estimators.

The least squares estimators of the unknown parameters of the model (14.6) can be obtained by minimizing

$$Q(\mathbf{A}, \boldsymbol{\omega}) = \sum_{n=1}^{N} \left| y(n) - \sum_{i=1}^{M} A_i e^{j\omega_i n} \right|^2 \tag{14.10}$$

with respect to $\mathbf{A} = (A_1, \ldots, A_M) = (A_{1R} + jA_{1C}, \ldots, A_{MR} + jA_{MC})$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)$, $A_{kR}$ and $A_{kC}$ denote the real and imaginary parts of $A_k$ respectively. The least squares estimator of $\boldsymbol{\theta} = (A_{1R}, A_{1C}, \omega_1, \ldots, A_{MR}, A_{MC}, \omega_M)$ will be denoted by $\hat{\boldsymbol{\theta}} = (\hat{A}_{1R}, \hat{A}_{1C}, \hat{\omega}_1, \ldots, \hat{A}_{MR}, \hat{A}_{MC}, \hat{\omega}_M)$. We use $\hat{\sigma}^2 = \frac{1}{N} Q(\hat{\mathbf{A}}, \hat{\omega})$ an estimator of $\sigma^2$ Unfortunately, the least squares estimators can not be obtained very easily in this case. Most of the general purpose algorithms, for example the Gauss-Newton algorithm or the Newton-Raphson algorithm do not work well in this case. The main reason is that the function $Q(\mathbf{A}, \boldsymbol{\omega})$ has many local minima and therefore most of the general purpose algorithms have the tendency to terminate at an early stage to a local minimum rather than

the global minimum even if the initial estimates are very good. We provide some special purpose algorithms to obtain the least squares estimators, but before that we state some properties of the least squares estimators. We introduce the following notations. Consider the following $\mathbf{3M \times 3M}$ block diagonal matrices

$$\mathbf{D} = \begin{bmatrix} \mathbf{D_1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{D_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{D}_M \end{bmatrix},$$

where each $\mathbf{D_k}$ is a $\mathbf{3 \times 3}$ diagonal matrix and

$$\mathbf{D_k} = \text{diag}\left\{ \boldsymbol{N}^{\frac{1}{2}}, \boldsymbol{N}^{\frac{1}{2}}, \boldsymbol{N}^{\frac{3}{2}} \right\}.$$

Also

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma}_M \end{bmatrix}$$

where each $\mathbf{\Sigma_k}$ is a $\mathbf{3 \times 3}$ matrix as

$$\mathbf{\Sigma_k} = \begin{bmatrix} \frac{1}{2} + \frac{3}{2}\frac{A_{kC}^2}{|A_k|^2} & -\frac{3}{2}\frac{A_{kR}A_{kC}}{|A_k|^2} & 3\frac{A_{kC}^2}{|A_k|^2} \\ -\frac{3}{2}\frac{A_{kR}A_{kC}}{|A_k|^2} & \frac{1}{2} + \frac{3}{2}\frac{A_{kR}^2}{|A_k|^2} & -3\frac{A_{kR}^2}{|A_k|^2} \\ 3\frac{A_{kC}^2}{|A_k|^2} & -3\frac{A_{kR}^2}{|A_k|^2} & \frac{6}{|A_k|^2} \end{bmatrix}.$$

Now we can state the main result from Rao and Zhao (1993) or Kundu and Mitra (1997).

RESULT 2: Under very minor assumptions, it can be shown that $\hat{\mathbf{A}}, \hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$ are strongly consistent estimators of $\mathbf{A}$, $\boldsymbol{\omega}$ and $\boldsymbol{\sigma}^{\mathbf{2}}$ respectively. Moreover, $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\mathbf{D}$ converges in distribution to a $\mathbf{3M \times 3M}$-variate normal distribution with mean vector zero and covariance matrix $\boldsymbol{\sigma}^{\mathbf{2}}\ \mathbf{\Sigma}$, where $\mathbf{D}$ and $\mathbf{\Sigma}$ are as defined above.

## 14.4 Different Iterative Methods

In this subsection we discuss different iterative methods to compute the least squares estimators. The least squares estimators of the unknown parameters can be obtained by minimizing (14.10) with respect to $\mathbf{A}$ and $\boldsymbol{\omega}$. The expression $\boldsymbol{Q(\mathbf{A}, \omega)}$ as defined in (14.10) can be written as

$$\boldsymbol{Q(\mathbf{A}, \omega)} = (\mathbf{Y} - \mathbf{C}(\boldsymbol{\omega})\mathbf{A})^{\boldsymbol{H}}(\mathbf{Y} - \mathbf{C}(\boldsymbol{\omega})\mathbf{A}). \tag{14.11}$$

Here the vector $\mathbf{A}$ is same as defined in the subsection (14.3.2). The vector $\mathbf{Y} = (y(1), \ldots, y(N))^T$ and $\mathbf{C}(\omega)$ is an $N \times M$ matrix as follows

$$\mathbf{C}(\omega) = \begin{bmatrix} e^{j\omega_1} & \ldots & e^{j\omega_M} \\ \vdots & \ddots & \vdots \\ e^{jN\omega_1} & \ldots & e^{jN\omega_M} \end{bmatrix}.$$

Observe that here the linear parameters $A_i$s are separable from the nonlinear parameters $\omega_i$s. For a fixed $\omega$, the minimization of $Q(\mathbf{A}, \omega)$ with respect to $\mathbf{A}$ is a simple linear regression problem. For a given $\omega$, the value of $\mathbf{A}$, say $\hat{\mathbf{A}}(\omega)$, which minimizes (14.11) is

$$\hat{\mathbf{A}}(\omega) = (\mathbf{C}(\omega)^H \mathbf{C}(\omega))^{-1} \mathbf{C}(\omega)^H \mathbf{Y}.$$

Substituting back $\hat{\mathbf{A}}(\omega)$ in (14.11), we obtain

$$R(\omega) = Q(\hat{\mathbf{A}}(\omega), \omega) = \mathbf{Y}^H (\mathbf{I} - \mathbf{P_C}) \mathbf{Y}, \tag{14.12}$$

where $\mathbf{P_C} = \mathbf{C}(\omega)(\mathbf{C}(\omega)^H \mathbf{C}(\omega))^{-1} \mathbf{C}(\omega)^H$ is the projection matrix on the column space spanned by the columns of the matrix $\mathbf{C}(\omega)$. Therefore, the least squares estimators of $\omega$ can be obtained by minimizing $R(\omega)$ with respect to $\omega$. Since minimizing $R(\omega)$ is a lower dimensional optimization problem than minimizing $Q(\mathbf{A}, \omega)$, therefore minimizing $R(\omega)$ is more economical from the computational point of view. If $\hat{\omega}$ minimizes $R(\omega)$, then $\hat{\mathbf{A}}(\hat{\omega})$, say $\hat{\mathbf{A}}$, is the least squares estimator of $\omega$. Once $(\hat{\mathbf{A}}, \hat{\omega})$ is obtained, the estimator of $\sigma^2$, say $\hat{\sigma^2}$, becomes

$$\hat{\sigma^2} = \frac{1}{N} Q(\hat{\mathbf{A}}, \hat{\omega}).$$

Now we discuss how to minimize (14.12) with respect to $\omega$. It is clearly a nonlinear problem and one expects to use some standard nonlinear minimization algorithm like Gauss-Newton algorithm, Newton-Raphson algorithm or any of their variants to obtain the required solution. Unfortunately, it is well known that most of the standard algorithms do not work well in this situation. Usually, they take long time to converge even from good starting values and sometimes they converge to a local minimum. For this reason, several special purpose algorithms have been proposed to solve this optimization problem and we discuss them in the subsequent subsections.

### 14.4.1 Iterative Quadratic Maximum Likelihood Estimators

The Iterative Quadratic Maximum Likelihood (IQML) method was proposed by Bresler and Macovski (1986). This algorithm mainly suggests how to minimize $R(\omega)$ with respect to $\omega$. Let us write $R(\omega)$ as follows;

$$R(\omega) = \mathbf{Y}^H (\mathbf{I} - \mathbf{P_C}) \mathbf{Y} = \mathbf{Y}^H \mathbf{P_G} \mathbf{Y}, \tag{14.13}$$

where $\mathbf{P_G} = \mathbf{G}(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H$ and

$$\mathbf{G} = \begin{bmatrix} \bar{g}_0 & \dots & 0 \\ \vdots & \dots & 0 \\ \bar{g}_M & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \bar{g}_0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \bar{g}_M \end{bmatrix}. \tag{14.14}$$

Here $\mathbf{G}$ is an $N \times (N-M)$ matrix and $\mathbf{G}^H\mathbf{C}(\omega) = \mathbf{0}$, therefore $\mathbf{I} - \mathbf{P_C} = \mathbf{P_G}$ and that implies the second equality of (14.13). Moreover, because of the structure of the matrix $\mathbf{G}$, $R(\omega)$ can be written as

$$R(\omega) = \mathbf{Y}^H\mathbf{P_G}\mathbf{Y} = \mathbf{g}^H\mathbf{Y_D}^H(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{Y}_D\mathbf{g}.$$

Here the vector $\mathbf{g}$ is same as defined before and $\mathbf{Y}_D$ is an $(N-M) \times (M+1)$ data matrix as follows

$$\mathbf{Y}_D = \begin{bmatrix} y(1) & \dots & y(M+1) \\ \vdots & \vdots & \vdots \\ y(N-M) & \dots & y(N) \end{bmatrix}.$$

The minimization of (14.13) can be performed by using the following algorithm.

**Algorithm IQML**

[1] Suppose at the $k$-th step the value of the vector $\mathbf{g}$ is $\mathbf{g}_{(k)}$.

[2] Compute the matrix $\mathbf{C}^{(k)} = \mathbf{Y}_D^H\,(\mathbf{G}_{(k)}^H\mathbf{G}_{(k)})^{-1}\mathbf{Y}_D$; here $\mathbf{G}_{(k)}$ is the matrix $\mathbf{G}$ in (14.14) obtained by replacing $\mathbf{g}$ with $\mathbf{g}_{(k)}$.

[3] Solve the quadratic minimization problem

$$\min_{\mathbf{x}:||x||=1} \mathbf{x}^H\mathbf{C}^{(k)}\mathbf{x}$$

and suppose the solution is $\mathbf{g}_{(k+1)}$.

[4] Check the convergence condition $|\mathbf{g}_{(k+1)} - \mathbf{g}_{(k)}| < \epsilon$ (some pre assigned value). If the convergence condition is met, go to step [5] otherwise $k = k+1$ and go to step [1].

[5] Put $\hat{\mathbf{g}} = (\hat{g}_0, \dots, \hat{g}_M) = \mathbf{g}_{(k+1)}$, solve the polynomial equation

$$\hat{g}_0 + \hat{g}_1 z + \dots + \hat{g}_M z^M = 0. \tag{14.15}$$

Obtain the $M$ roots of (14.15) as $\hat{\rho}_1 e^{j\hat{\omega}_1} \dots, \hat{\rho}_M e^{j\hat{\omega}_M}$. Consider $(\hat{\omega}_1, \dots, \hat{\omega}_M)$ as the least squares estimators of $(\omega_1, \dots, \omega_M)$.

### 14.4.2 Constrained Maximum Likelihood Method

In the IQML method the symmetric structure of the vector $\mathbf{g}$ is not used. The Constrained Maximum Likelihood (CML) method was proposed by Kannan and Kundu (1994). The basic idea of the CML method is to minimize $\mathbf{Y}^{H}\mathbf{P}_{G}\mathbf{Y}$ with respect to the vector $\mathbf{g}$ when it satisfies the symmetric structure (14.9). It is clear that $\boldsymbol{R}(\boldsymbol{\omega})$ can also be written as

$$\boldsymbol{R}(\boldsymbol{\omega}) = \mathbf{Y}^{H}(\mathbf{I} - \mathbf{P}_{\mathbf{C}})\mathbf{Y} = \mathbf{Y}^{H}\mathbf{P}_{B}\mathbf{Y}, \tag{14.16}$$

where $\mathbf{P}_{B} = \mathbf{B}(\mathbf{B}^{H}\mathbf{B})^{-1}\mathbf{B}^{H}$ and

$$\mathbf{B} = \begin{bmatrix} \bar{b}_0 & \dots & 0 \\ \vdots & \dots & 0 \\ \bar{b}_M & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \bar{b}_0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \bar{b}_M \end{bmatrix}.$$

Moreover $\boldsymbol{b_i}$'s satisfy the symmetric relation (14.9). Therefore the problem reduces to minimize $\mathbf{Y}^{H}\mathbf{P}_{B}\mathbf{Y}$ with respect to $\mathbf{b} = (b_0, \dots, b_M)$ such that $\mathbf{b}^{H}\mathbf{b} = \mathbf{1}$ and $b_k = \bar{b}_{M-k}$ for $k = 0, \dots, M$. The constrained minimization can be done in the following way. Let us write $\mathbf{b} = \mathbf{c} + \mathrm{j}\ \mathbf{d}$, where

$$\begin{aligned} \mathbf{c}^{T} &= (c_0, c_1, \dots, c_{\frac{M}{2}}, \dots, c_1, c_0), \\ \mathbf{d}^{T} &= (d_0, d_1, \dots, d_{\frac{M}{2}-1}, 0, -d_{\frac{M}{2}-1}, \dots, -d_1, -d_0), \end{aligned}$$

if $M$ is even and

$$\begin{aligned} \mathbf{c}^{T} &= (c_0, c_1, \dots, c_{\frac{M-1}{2}}, c_{\frac{M-1}{2}}, \dots, c_1, c_0), \\ \mathbf{d}^{T} &= (d_0, d_1, \dots, d_{\frac{M-1}{2}}, -d_{\frac{M-1}{2}}, -d_{\frac{M-1}{2}-1}, \dots, -d_1, -d_0), \end{aligned}$$

if $M$ is odd. The matrix $\mathbf{B}$ can be written as $\mathbf{B} = \left[\bar{\mathbf{b}}^{1}, \dots, \bar{\mathbf{b}}^{N-M}\right]$, where $(\bar{\mathbf{b}}^{i})^{T}$ is of the form $[\mathbf{0}, \bar{\mathbf{b}}, \mathbf{0}]$. Let $\mathbf{U}$ and $\mathbf{V}$ denote the real and imaginary parts of the matrix $\mathbf{B}$. Assuming that $M$ is odd, $\mathbf{B}$ can be written as

$$\mathbf{B} = \sum_{\alpha=0}^{\frac{M-1}{2}} \left(c_{\alpha}\mathbf{U}_{\alpha} + j d_{\alpha}\mathbf{V}_{\alpha}\right),$$

where $\mathbf{U}_{\alpha}$, $\mathbf{V}_{\alpha}$, $\alpha = 0, 1, \dots, \frac{M-1}{2}$ are $N \times (N - M)$ matrices with entries 0 and 1 only. The minimization of $\mathbf{Y}^{H}\mathbf{P}_{B}\mathbf{Y}$ (as given in (14.16)) can be obtained by differentiating $\mathbf{Y}^{H}\mathbf{P}_{B}\mathbf{Y}$ with respect to $c_{\alpha}$ and $d_{\alpha}$;

for $\boldsymbol{\alpha = 0, 1, \ldots, \frac{M-1}{2}}$, which is equivalent to solving a matrix equation of the form

$$\mathbf{D}(\tilde{\mathbf{c}}, \tilde{\mathbf{d}}) \begin{bmatrix} \tilde{\mathbf{c}} \\ \tilde{\mathbf{d}} \end{bmatrix} = \mathbf{0},$$

where $\mathbf{D}$ is an $\boldsymbol{(M+1) \times (M+1)}$ matrix. Here $\tilde{\mathbf{c}}^{\boldsymbol{T}} = (\boldsymbol{c_0, \ldots, c_{\frac{M-1}{2}}})$ and $\tilde{\mathbf{d}}^{\boldsymbol{T}} = (\boldsymbol{d_0, \ldots, d_{\frac{M-1}{2}}})$. The matrix $\mathbf{D}$ can be written in a partitioned form as

$$\mathbf{D} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{B}}^{\boldsymbol{H}} & \tilde{\mathbf{C}} \end{bmatrix}, \tag{14.17}$$

where $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$ are all $\frac{M+1}{2} \times \frac{M+1}{2}$ matrices. The $(\boldsymbol{i, k})$-th elements of the matrices $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$ are as follows;

$$\begin{aligned}
\tilde{\boldsymbol{A}}_{\boldsymbol{ik}} &= \mathbf{Y}^{\boldsymbol{H}}\mathbf{U}_{\boldsymbol{i}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{U}_{\boldsymbol{k}}^{\boldsymbol{H}}\mathbf{Y} - \mathbf{Y}^{\boldsymbol{H}}\mathbf{B}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}(\mathbf{U}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{U}_{\boldsymbol{k}} + \mathbf{U}_{\boldsymbol{k}}^{\boldsymbol{T}}\mathbf{U}_{\boldsymbol{i}}) \\
&\quad (\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{B}^{\boldsymbol{H}}\mathbf{Y} + \mathbf{Y}^{\boldsymbol{H}}\mathbf{U}_{\boldsymbol{k}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{U}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{Y} \\
\tilde{\boldsymbol{B}}_{\boldsymbol{ik}} &= -\boldsymbol{j}\mathbf{Y}^{\boldsymbol{H}}\mathbf{U}_{\boldsymbol{i}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{V}_{\boldsymbol{k}}^{\boldsymbol{H}}\mathbf{Y} - \boldsymbol{j}\mathbf{Y}^{\boldsymbol{H}}\mathbf{B}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}(\mathbf{U}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{V}_{\boldsymbol{k}} - \mathbf{V}_{\boldsymbol{k}}^{\boldsymbol{T}}\mathbf{U}_{\boldsymbol{i}}) \\
&\quad (\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{B}^{\boldsymbol{H}}\mathbf{Y} + \boldsymbol{j}\mathbf{Y}^{\boldsymbol{H}}\mathbf{V}_{\boldsymbol{k}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{U}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{Y} \\
\tilde{\boldsymbol{C}}_{\boldsymbol{ik}} &= \mathbf{Y}^{\boldsymbol{H}}\mathbf{V}_{\boldsymbol{i}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{V}_{\boldsymbol{k}}^{\boldsymbol{H}}\mathbf{Y} - \mathbf{Y}^{\boldsymbol{H}}\mathbf{B}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}(\mathbf{V}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{V}_{\boldsymbol{k}} + \mathbf{V}_{\boldsymbol{k}}^{\boldsymbol{T}}\mathbf{V}_{\boldsymbol{i}}) \\
&\quad (\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{B}^{\boldsymbol{H}}\mathbf{Y} + \mathbf{Y}^{\boldsymbol{H}}\mathbf{V}_{\boldsymbol{k}}(\mathbf{B}^{\boldsymbol{H}}\mathbf{B})^{-1}\mathbf{V}_{\boldsymbol{i}}^{\boldsymbol{T}}\mathbf{Y}
\end{aligned}$$

for $\boldsymbol{i, k = 0, 1, \ldots, \frac{M-1}{2}}$. Similarly, when $\boldsymbol{M}$ is even, let us define $\tilde{\mathbf{c}}^{\boldsymbol{T}} = (\boldsymbol{c_0, \ldots, c_{\frac{M}{2}}})$ and $\tilde{\mathbf{d}}^{\boldsymbol{T}} = (\boldsymbol{d_0, \ldots, d_{\frac{M}{2}}})$. In this case also the matrix $\mathbf{D}$ can be partitioned as (14.17), but here $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$ are $\left(\frac{M}{2}+1\right) \times \left(\frac{M}{2}+1\right)$, $\left(\frac{M}{2}+1\right) \times \left(\frac{M}{2}\right)$ and $\left(\frac{M}{2}\right) \times \left(\frac{M}{2}\right)$ matrices respectively. The $(\boldsymbol{i, k})$-th elements of $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$ are the same as before with appropriate ranges of $\boldsymbol{i}$ and $\boldsymbol{k}$. The matrix $\mathbf{D}$ is a real symmetric matrix in both cases and we need to solve a matrix equation of the form

$$\mathbf{D(x)x = 0}, \qquad \text{such that} \quad \mathbf{||x|| = 1}. \tag{14.18}$$

The constrained minimization problem is transformed to a real valued nonlinear eigenvalue problem. If $\hat{\mathbf{x}}$ satisfies (14.18), then $\hat{\mathbf{x}}$ should be an eigenvector of the matrix $\mathbf{D}(\hat{\mathbf{x}})$. The CML method suggests the following iterative technique to solve (14.18)

$$\left(\mathbf{D}(\mathbf{x}^{(k)}) - \boldsymbol{\lambda}^{(k+1)}\mathbf{I}\right)\mathbf{x}^{(k+1)} = \mathbf{0}, \qquad ||\mathbf{x}^{(k+1)}|| = \mathbf{1},$$

where $\boldsymbol{\lambda}^{(k+1)}$ is the eigenvalue which is closest to zero of the matrix $\mathbf{D}(\hat{\mathbf{x}})$ and $\mathbf{x}^{(k+1)}$ is the corresponding eigenvector. The iterative process should be stopped when $\boldsymbol{\lambda}^{(k+1)}$ is small compared to $||\mathbf{D}||$, the largest eigenvalue of the matrix $\mathbf{D}$. The CML estimators can be obtained using the following algorithm:

**CML Algorithm:**

[1] Suppose at the $\boldsymbol{i}$-th step the value of the vector $\mathbf{x}$ is $\mathbf{x}_{(\boldsymbol{i})}$. Normalize $\mathbf{x}$, ***i.e.*** $\mathbf{x}^{(\boldsymbol{i})} = \frac{\mathbf{x}^{(i)}}{||\mathbf{x}^{(i)}||}$.

[2] Calculate the matrix $\mathbf{D}(\mathbf{x}^{(\boldsymbol{i})})$.

[3] Find the eigenvalue $\boldsymbol{\lambda}^{(\boldsymbol{i+1})}$ of $\mathbf{D}(\mathbf{x}^{(\boldsymbol{i})})$ closest to zero and normalize the corresponding eigenvector $\mathbf{x}^{(\boldsymbol{i+1})}$.

[4] Test the convergence by checking whether $|\boldsymbol{\lambda}^{(\boldsymbol{i+1})}| < \boldsymbol{\epsilon}||\mathbf{D}||$.

[5] If the convergence condition in step 4 is not met, then $\boldsymbol{i := i+1}$ and go to step 2.

### 14.4.3 Expectation Maximization Algorithm

The expectation maximization (EM) algorithm, developed by Dempster, Laird and Rubin (1977) is a general method for solving the maximum likelihood estimation problem when the data are incomplete. The details on the EM algorithm can be found in Chapter 4. Although the EM algorithm has been originally used for incomplete data, it can be used also when the data is complete. The EM algorithm can be used quite effectively in estimating the unknown amplitudes and the frequencies of the model (14.6) and the method was proposed by Feder and Weinstein (1988). In developing the EM algorithm it is assumed that the errors are ***i.i.d*** Gaussian random variables, although it can be relaxed and it is not pursued here. But in formulating the EM algorithm one needs to know the distribution of the error random variables.

For better understanding we explain the EM algorithm briefly over here. Let $\mathbf{Y}$ denote the observed (may be incomplete) data possessing the probability density function $\boldsymbol{f}_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})$ indexed by the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbf{R}^{\boldsymbol{k}}$ and let $\mathbf{X}$ denote the 'complete' data vector related to $\mathbf{Y}$ by

$$\boldsymbol{H}(\mathbf{X}) = \mathbf{Y},$$

where $\boldsymbol{H}(.)$ is a many to one non-invertible function. Therefore, the density function of $\mathbf{X}$, say $\boldsymbol{f}_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ can be written as

$$\boldsymbol{f}_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) = \boldsymbol{f}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta})\boldsymbol{f}_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) \qquad \forall \boldsymbol{H}(\mathbf{x}) = \mathbf{y}, \tag{14.19}$$

where $\boldsymbol{f}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta})$ is the conditional probability density function of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$. Taking the logarithm on both sides of (14.19)

$$\ln \boldsymbol{f}_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = \ln \boldsymbol{f}_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) - \ln \boldsymbol{f}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta}).$$

Taking the conditional expectation given $\mathbf{Y} = \mathbf{y}$ at the parameter value $\boldsymbol{\theta'}$ gives,

$$\ln f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = E\{\ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y},\boldsymbol{\theta'}\} - E\{\ln f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y},\boldsymbol{\theta'}\}. \tag{14.20}$$

If we define

$$\begin{aligned} L(\boldsymbol{\theta}) &= \ln f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}), \\ U(\boldsymbol{\theta},\boldsymbol{\theta'}) &= E\{\ln f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y},\boldsymbol{\theta'}\}, \quad \text{and} \\ V(\boldsymbol{\theta},\boldsymbol{\theta'}) &= E\{\ln f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y},\boldsymbol{\theta'}\}, \end{aligned}$$

then (14.20) becomes

$$L(\boldsymbol{\theta}) = U(\boldsymbol{\theta},\boldsymbol{\theta'}) - V(\boldsymbol{\theta},\boldsymbol{\theta'}).$$

Here $L(\boldsymbol{\theta})$ is the log-likelihood function of the observed data and that needs to be maximized to obtain the maximum likelihood estimators of $\boldsymbol{\theta}$. Since $V(\boldsymbol{\theta}, \boldsymbol{\theta'}) \leq V(\boldsymbol{\theta'}, \boldsymbol{\theta'})$ (Jensen's inequality), therefore, if

$$U(\boldsymbol{\theta},\boldsymbol{\theta'}) > U(\boldsymbol{\theta'},\boldsymbol{\theta'})$$

then

$$L(\boldsymbol{\theta}) > L(\boldsymbol{\theta'}). \tag{14.21}$$

The relation (14.21) forms the basis of the EM algorithm. The algorithm starts with an initial guess and let us denote by $\hat{\boldsymbol{\theta}}^{(m)}$ the current estimate of $\boldsymbol{\theta}$ after $m$ iterations. Then $\hat{\boldsymbol{\theta}}^{(m+1)}$ can be obtained as follows

$$\begin{aligned} \text{E Step} &: \text{Compute} \quad U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(m)}) \\ \text{M Step} &: \hat{\boldsymbol{\theta}}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \; U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(m)}). \end{aligned}$$

It is possible to use the EM algorithm to obtain the estimates of the unknown parameters of the model (14.6) under the assumption that the errors are Gaussian random variables. Under these assumptions, the log-likelihood function takes the form;

$$L(\boldsymbol{\theta}) = c - \frac{1}{2\sigma^2} \sum_{n=1}^{N} \left| y(n) - \sum_{i=1}^{M} A_i e^{j\omega_i n} \right|^2.$$

To use the EM algorithm, let us write

$$\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T,$$

here

$$x_k(n) = A_k e^{j\omega_k n} + z_k(n).$$

Here $\boldsymbol{z_k(n)}$'s are obtained by arbitrarily decomposing the total noise $\boldsymbol{z(n)}$ into $\boldsymbol{M}$ components, so that

$$\sum_{k=1}^{M} \boldsymbol{z_k(n) = z(n).}$$

Therefore, the model (3.1) can be written as

$$\boldsymbol{y(n) = \sum_{k=1}^{M} x_k(n) = \mathbf{H}x(n)}$$

and $\mathbf{H} = \boldsymbol{[1, \ldots, 1]}$. Let us choose $\boldsymbol{z_k(n)}$ to be independent zero mean Gaussian random variables with variance $\boldsymbol{\beta_k}$ such that

$$\sum_{k=1}^{M} \boldsymbol{\beta_k = 1} \qquad \boldsymbol{\beta_k \geq 0.}$$

With the above notation, the EM algorithm takes the following form. If $\boldsymbol{\hat{A}_k^{(m)}}$ and $\boldsymbol{\hat{\omega}_k^{(m)}}$ denote the estimates of $\boldsymbol{A_k}$ and $\boldsymbol{\omega_k}$ respectively after $\boldsymbol{m}$ iterations, then

$$\text{E Step} \quad : \quad \boldsymbol{\hat{x}_k^{(m)}(n) = \hat{A}_k^{(m)} e^{j\hat{\omega}_k^{(m)} n} + \beta_k \left[ y(n) - \sum_{i=1}^{M} \hat{A}_i^{(m)} e^{j\hat{\omega}_i^{(m)} n} \right]}$$

$$\text{M Step} \quad : \quad \boldsymbol{(\hat{A}_k^{(m+1)}, \hat{\omega}_k^{(m+1)}) = \arg \max_{A_k, \omega_k} \left| \hat{x}_k^{(m)}(n) - A_k e^{j\omega_k n} \right|^2 .}$$

It is interesting to note that $\boldsymbol{\hat{A}_k^{(m+1)}}$ and $\boldsymbol{\hat{\omega}_k^{(m+1)}}$ are the maximum likelihood estimators based on $\boldsymbol{\hat{x}_k^{(m)}(n)}$. The most important feature of this algorithm is that it decomposes the complicated multiparameter optimization problem into $\boldsymbol{M}$ separate simple one dimensional optimization problems. This particular method can be used for on line implementation purposes. $\boldsymbol{M}$ different processors can be used to obtain the estimates of the unknown parameters very efficiently.

There are several other iterative algorithms like the Iterative Inverse Filtering method of Matausek, Stankovic and Radovic (1983), Dynamic Programming method of Yau and Bresler (1993) or 7-step efficient algorithm by Bai, Rao, Chow and Kundu (2003) available in the literature but they are not pursued here. Now we present some important non-iterative algorithms which are used in practice.

## 14.5 Different Non-Iterative Methods

### 14.5.1 Modified Forward Backward Linear Prediction Method

The modified Forward Backward Linear Prediction (MFBLP) method was proposed by Tufts and Kumaresan (1982). Assuming that the data are

noiseless in the model (14.6), it can be easily seen that for any $M < L < N - M$, there exists $b_1, \ldots, b_L$, such that

$$\begin{bmatrix} y(L) & y(L-1) & \ldots & y(2) & y(1) \\ \vdots & \vdots & \vdots & \vdots & \\ y(N-1) & y(N-2) & \ldots & y(N-L+1) & y(N-L) \\ \bar{y}(2) & \bar{y}(3) & \ldots & \bar{y}(L) & \bar{y}(L+1) \\ \vdots & \vdots & \vdots & \vdots & \\ \bar{y}(N-L+1) & \bar{y}(N-L+2) & \ldots & \bar{y}(N-1) & \bar{y}(N) \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_L \end{bmatrix}$$

$$= - \begin{bmatrix} y(L+1) \\ \vdots \\ y(N) \\ \bar{y}(1) \\ \vdots \\ \bar{y}(N-L) \end{bmatrix} \quad \text{or} \quad \mathbf{Y_{TK}b} = -\mathbf{h} \quad \text{(say)}.$$

Unlike Prony's equations, here the constants $\mathbf{b} = (b_1, \ldots, b_L)$ may not be unique. The vector $\mathbf{b}$ should be chosen such that $||\mathbf{b}||^2 = \sum_{i=1}^{L} |b_i|^2$ is minimum (minimum norm solution). In this case it has been shown by Kumaresan (1982) that out of the $L$ roots of the polynomial equation

$$x^L + b_1 x^{L-1} + \ldots b_L = 0, \tag{14.22}$$

$M$ of them are of the form $e^{j\omega_k}$, for $k = 1, \ldots, M$ and the polynomial equation (14.22) has $L - M$ other roots and they have magnitudes strictly less than one. The above finding helps to distinguish the $M$ signal zeros from the $L - M$ noise zeros. Since in practice the data are noisy, the following procedure can be obtained to estimate the vector $\mathbf{b}$ as

$$\hat{\mathbf{b}} = - \sum_{k=1}^{M} \sigma_k^{-1} [\mathbf{u}_k^H] \mathbf{v}_k.$$

Here $\sigma_k$'s are the singular values of the matrix $\mathbf{Y_{TK}}$, $\mathbf{v}_k$ for $k = 1, \ldots L$ and $\mathbf{u}_k$ for $k = 1, \ldots, N - L$ are the eigenvectors of $\mathbf{Y}_{TK}^H \mathbf{Y_{TK}}$ and $\mathbf{Y_{TK}} \mathbf{Y}_{TK}^H$ respectively. The main reason to use the truncated singular value decomposition is to increase the signal to noise ratio in the data prior to obtaining the solution vector $\mathbf{b}$. Once $\hat{\mathbf{b}}$ is obtained using (14.22), the estimates of the frequencies can be obtained very easily.

### 14.5.2 Estimation of the Signal Parameters via Rotation Invariant Technique

The Estimation of the Signal Parameters via Rotation Invariant Technique (ESPRIT) was proposed by Roy (1987) and it was used to estimate the

parameters of the Direction of Arrival (DOA) model. Although ESPRIT method was originally proposed to estimate the parameters of the DOA model but it can be used quite effectively to estimate the frequencies of the sinusoidal model (14.6) also. The method can be briefly described as follows.

For a given $\boldsymbol{L}$, $\boldsymbol{M < L < N - M}$ construct two data matrices

$$\mathbf{Y_{R1}} = \begin{bmatrix} \boldsymbol{y(1)} & \boldsymbol{\ldots} & \boldsymbol{y(L)} \\ \vdots & \vdots & \vdots \\ \boldsymbol{y(N-L)} & \boldsymbol{\ldots} & \boldsymbol{y(N-1)} \end{bmatrix} \quad \text{and}$$

$$\mathbf{Y_{R2}} = \begin{bmatrix} \boldsymbol{y(2)} & \boldsymbol{\ldots} & \boldsymbol{y(L+1)} \\ \vdots & \vdots & \vdots \\ \boldsymbol{y(N-L+1)} & \boldsymbol{\ldots} & \boldsymbol{y(N)} \end{bmatrix}.$$

Suppose $\mathbf{C_{11}} = \left[\mathbf{Y_{R1}^H Y_{R1} - \sigma^2 I}\right]$ and $\mathbf{C_{12}} = \left[\mathbf{Y_{R2}^H Y_{R1} - \sigma^2 J}\right]$, where $\mathbf{I}$ and $\mathbf{J}$ are the identity matrix and exchange matrix respectively, each of order $\boldsymbol{L \times L}$. Consider the singular values of the matrix pencil (see Pillai 1989)

$$\mathbf{C_{11} - \gamma C_{12}}. \tag{14.23}$$

It can be shown (Pillai 1989) that the $\boldsymbol{M}$ non-zero eigenvalues of the matrix pencil (14.23) will be of the form $\boldsymbol{e^{j\omega_k}}$ for $\boldsymbol{k = 1, \ldots, M}$. But since $\boldsymbol{\sigma^2}$ is unknown, in practice, therefore $\boldsymbol{\sigma^2}$ needs to be estimated to construct $\mathbf{C_{11}}$ and $\mathbf{C_{12}}$ by replacing $\boldsymbol{\sigma^2}$ with $\boldsymbol{\hat{\sigma}^2}$. To avoid this situation Roy and Kailath (1989) proposed the Total Least Squares Estimation of Signal Parameters via Rotation Invariant Technique (TLS-ESPRIT), which does not require the estimation of $\boldsymbol{\sigma^2}$.

### 14.5.3 TLS-ESPRIT

Under the same set up as in the previous case, consider two $\boldsymbol{2L \times 2L}$ matrices $\mathbf{R}$ and $\mathbf{\Sigma}$ as follows;

$$\mathbf{R} = \begin{bmatrix} \mathbf{Y_{R1}^H} \\ \\ \mathbf{Y_{R2}^H} \end{bmatrix} \left[\mathbf{Y_{R1} : Y_{R2}}\right], \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{I} & \mathbf{J} \\ \mathbf{J}^T & \mathbf{I} \end{bmatrix}.$$

Let $\mathbf{e_1, \ldots, e_M}$ be the generalized eigenvectors (Rao 1973), corresponding to the largest $\boldsymbol{M}$ generalized eigenvalues of $\mathbf{R}$ with respect to the known matrix $\mathbf{\Sigma}$. Construct the two $\boldsymbol{L \times M}$ matrices $\mathbf{E_1}$ and $\mathbf{E_2}$ from $\mathbf{e_1, \ldots, e_M}$ as

$$\left[\mathbf{e_1 : \ldots : e_M}\right] = \begin{bmatrix} \mathbf{E_1} \\ \\ \mathbf{E_2} \end{bmatrix},$$

and then obtain the unique $\mathbf{2}\boldsymbol{M} \times \boldsymbol{M}$ matrix $\mathbf{W}$ and the two $\boldsymbol{M} \times \boldsymbol{M}$ matrices $\mathbf{W_1}$ and $\mathbf{W_2}$ as

$$[\mathbf{E_1} : \mathbf{E_2}]\,\mathbf{W} = \mathbf{0}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W_1} \\ \\ \mathbf{W_2} \end{bmatrix}.$$

Finally obtain the $\boldsymbol{M}$ eigenvalues of -$\mathbf{W_1 W_2^{-1}}$. In the noise less situation $\boldsymbol{M}$ eigenvalues will be of the form $\boldsymbol{e^{j\omega_k}}$, for $\boldsymbol{k = 1, \ldots, M}$ (Pillai 1989). Therefore, the frequencies can be estimated form the eigenvalues of the matrix -$\mathbf{W_1 W_2^{-1}}$.

It is known that TLS-ESPRIT works better than ESPRIT. The main computation in the TLS-ESPRIT is the computation of the eigenvectors of an $\boldsymbol{L} \times \boldsymbol{L}$ matrix. It does not require any root finding computation like MFBLP. The consistency property of the ESPRIT or the TLS-ESPRIT is not yet known.

### 14.5.4 Quinn's Method

Quinn (1994) proposed a method of estimating the frequencies of the model (14.6) by interpolation using the Fourier coefficients. Although, it has been proposed originally for single frequency model (***i.e*** for $\boldsymbol{M = 1}$), but it can be extended easily for the multiple frequencies also (Kundu and Mitra 1998). First we describe the Quinn's method for single frequency only. Suppose,

$$\boldsymbol{X(i) = \sum_{n=1}^{N} y(n) e^{-\frac{j2\pi in}{N}}}.$$

Then Quinn's algorithm becomes:

[1] Let $\boldsymbol{|X(\tau_T)|^2 \geq |X(i)|^2}$, for $\mathbf{1} \leq \boldsymbol{i} \leq \left[\frac{\boldsymbol{N-1}}{\mathbf{2}}\right]$.

[2] Let $\hat{\boldsymbol{\alpha}}_\mathbf{1}$ = Real$\frac{\boldsymbol{X(\tau_T - 1)}}{\boldsymbol{X(\tau_T)}}$, $\hat{\boldsymbol{\alpha}}_\mathbf{2}$ = Real$\frac{\boldsymbol{X(\tau_T + 1)}}{\boldsymbol{X(\tau_T)}}$ and $\hat{\boldsymbol{\delta}}_\mathbf{1} = \frac{\hat{\boldsymbol{\alpha}}_1}{(\mathbf{1} - \hat{\boldsymbol{\alpha}}_1)}$ and $\hat{\boldsymbol{\delta}}_\mathbf{2}$ = - $\frac{\hat{\boldsymbol{\alpha}}_2}{(\mathbf{1} - \hat{\boldsymbol{\alpha}}_2)}$. If $\hat{\boldsymbol{\delta}}_\mathbf{1}$ and $\hat{\boldsymbol{\delta}}_\mathbf{2}$ are both $> \mathbf{0}$, put $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_\mathbf{2}$, otherwise $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_\mathbf{1}$.

[3] Estimate $\boldsymbol{\omega}$ by $\hat{\boldsymbol{\omega}} = \frac{\boldsymbol{2\pi(\tau_T + \hat{\delta})}}{\boldsymbol{N}}$.

If more than one frequency is present continue with the second largest $\boldsymbol{|X(i)|^2}$ and so on. Computationally Quinn's method is very easy to implement and it is observed that the estimates are consistent and the asymptotic mean squared errors of the estimates are of the order $\boldsymbol{N^{-3}}$, the best possible convergence rate in this case.

### 14.5.5 Noise Space Decomposition Method

The Noise Space Decomposition (NSD) method was proposed by Kundu and Mitra (1995). The method can be briefly described as follows. Consider the $\boldsymbol{N-L\times L+1}$ data matrix $\mathbf{A}_{\boldsymbol{NSD}}$ as

$$\mathbf{A}_{\boldsymbol{NSD}}=\begin{bmatrix} \boldsymbol{y(1)} & \dots & \boldsymbol{y(L+1)} \\ \dots & \dots & \dots \\ \boldsymbol{y(N-L)} & \dots & \boldsymbol{y(N)} \end{bmatrix},$$

for any integer $\boldsymbol{L}$, such that $\boldsymbol{M\le L\le N-M}$. Construct the $\boldsymbol{L+1\times L+1}$ matrix $\mathbf{T}=\frac{\mathbf{1}}{\boldsymbol{N}}\left[\mathbf{A}_{\boldsymbol{NSD}}^{\boldsymbol{H}}\mathbf{A}_{\boldsymbol{NSD}}\right]$. Let the spectral decomposition of the matrix $\mathbf{T}$ be as

$$\mathbf{T}=\sum_{\boldsymbol{i=1}}^{\boldsymbol{L+1}}\hat{\boldsymbol{\sigma}}_{\boldsymbol{i}}^{\boldsymbol{2}}\mathbf{u}_{\boldsymbol{i}}\mathbf{u}_{\boldsymbol{i}}^{\boldsymbol{H}},$$

where $\hat{\boldsymbol{\sigma}}_{\boldsymbol{1}}^{\boldsymbol{2}}>\dots>\hat{\boldsymbol{\sigma}}_{\boldsymbol{L+1}}^{\boldsymbol{2}}$ are the ordered eigenvalues of the matrix $\mathbf{T}$ and $\mathbf{u}_{\boldsymbol{i}}$'s are orthogonal eigenvectors corresponding to $\hat{\boldsymbol{\sigma}}_{\boldsymbol{i}}^{\boldsymbol{2}}$. Construct the $\boldsymbol{L+1\times L+1-M}$ matrix $\mathbf{B}_{\boldsymbol{NSD}}$ as

$$\mathbf{B}_{\boldsymbol{NSD}}=\left[\mathbf{u}_{\boldsymbol{M+1}}:\dots:\mathbf{u}_{\boldsymbol{L+1}}\right].$$

Partition the matrix $\mathbf{B}_{\boldsymbol{NSD}}$ as

$$\mathbf{B}_{\boldsymbol{NSD}}^{\boldsymbol{H}}=\left[\mathbf{B}_{\boldsymbol{1k}}^{\boldsymbol{H}}:\mathbf{B}_{\boldsymbol{2k}}^{\boldsymbol{H}}:\mathbf{B}_{\boldsymbol{3k}}^{\boldsymbol{H}}\right]$$

for $\boldsymbol{k=0,\dots,L-M}$ where $\mathbf{B}_{\boldsymbol{1k}}^{\boldsymbol{H}}$, $\mathbf{B}_{\boldsymbol{2k}}^{\boldsymbol{H}}$, and $\mathbf{B}_{\boldsymbol{3k}}^{\boldsymbol{H}}$ are of the orders $\boldsymbol{L+1-M\times k}$, $\boldsymbol{L+1-M\times M+1}$ and $\boldsymbol{L+1-M\times L-k-M}$ respectively. Find a vector $\boldsymbol{X_k}$ such that

$$\begin{bmatrix} \mathbf{B}_{\boldsymbol{1k}} \\ \\ \mathbf{B}_{\boldsymbol{3k}} \end{bmatrix}\mathbf{X}_{\boldsymbol{k}}=\mathbf{0}.$$

Let us denote the vector $\mathbf{c}^{\boldsymbol{k}}=\mathbf{B}_{\boldsymbol{2k}}\mathbf{X}_{\boldsymbol{k}}$ for $\boldsymbol{k=0,\dots,L-M}$. Now consider the average of the vectors $\mathbf{c}^{\boldsymbol{k}}$'s as the vector $\mathbf{c}$, ***i.e.***

$$\mathbf{c}=\frac{\mathbf{1}}{\boldsymbol{L-M+1}}\sum_{\boldsymbol{k=0}}^{\boldsymbol{L-M}}\mathbf{c}^{\boldsymbol{k}}=(\boldsymbol{c_1},\dots,\boldsymbol{c_{M+1}}).$$

Construct the polynomial equation

$$\boldsymbol{c_1+c_2x+\dots+c_{M+1}x^M=0}. \tag{14.24}$$

From (14.24) obtain the $\boldsymbol{M}$ roots and estimate the frequencies from there. It is observed that the estimated frequencies are strongly consistent. The main computation involved in the NSD method is the spectral decomposition of the $\boldsymbol{L+1\times L+1}$ matrix and the root finding of an $\boldsymbol{M}$ degree polynomial.

It is well known that (Harvey 1981, Chapter 4.5) when a regular likelihood is maximized through any iterative procedure, the estimators obtained after one single round of iteration already have the same asymptotic properties as the least squares estimators. This holds, if the starting values are chosen $\sqrt{N}$ consistently. Now since NSD estimators are strongly consistent, the NSD estimators can be combined with one single round of scoring algorithm. The modified method is called Modified Noise Space Decomposition Method (MNSD). As expected, MNSD works much better than NSD.

## 14.6 Conclusions

In this chapter we have provided a brief introduction to one of the most important signal processing models, namely the sum of exponential or sinusoidal model. We provide different computational aspects of this particular model. In most algorithms described in this chapter it is assumed that the number of signals $M$ is known. In practical applications however, estimation of $M$ is an important problem. The estimation of the number of components is essentially a model selection problem. This has been studied quite extensively in the context of variable selection in linear and nonlinear regressions and multivariate analysis. Standard model selection techniques may be used to estimate the number of signals of the multiple sinusoids model. It is usually assumed that the maximum number of signals can be at most a fixed number, say $K$. Since, model (14.6) is a nonlinear regression model, the $K$ competing models are nested. The estimation of $M$ is obtained by selecting the model that best fits the data. Several techniques have been used to estimate $M$ and all of them are quite involved computationally. Two most popular ones are by information theoretic criteria and by cross validation method. Information theoretic criteria like Akaike Information Criteria, Minimum Description Length Criterion have been used quite effectively to estimate $M$ by Wax and Kailath (1985), Reddy and Biradar (1993) and Kundu and Mitra (2001). Cross validation technique has been proposed by Rao (1988) and it has been implemented by Kundu and Mitra (2000). Cross validation technique works very well in this case, but it is quite time consuming and not suitable for on line implementation purposes.

We have discussed quite extensively about the one-dimensional frequency estimation problem, but recently some significant results have been obtained for two dimensional sinusoidal model as described in Example 4, see for example Rao, Zhao and Zhou (1994) or Bai, Kundu and Mitra (1996). Several computational issues have not yet been resolved. Interested readers are referred to the following books for further reading; Dudgeon and Merseresau (1984), Kay (1988), Haykin (1985), Pillai (1989), Bose and Rao (1993) and Srinath, Rajasekaran and Viswanathan (1996).

**Acknowledgments** The author would like to thank Dr Swagata Nandi and Dr. Amit Mitra for carefully reading this manuscript and providing some valuable comments.

# References

Bai, Z. D., Kundu, D. and Mitra, A. (1996). A theorem in probability and its applications in multidimensional signal processing, *IEEE Transactions on Signal Processing*, **44**, 3167–3169.

Bai, Z. D., Rao, C. R., Chow, M. and Kundu, D. (2003). An efficient algorithm for estimating the parameters of superimposed exponential signals, *Journal of Statistical Planning and Inference*, **110**, 23-34.

Barrodale, I. and Oleski, D. D. (1981). Exponential approximations using Prony's method, *The Numerical Solution of Non-Linear Problems*, Baker, C. T. H. and Phillips, C. eds., Academic Press, New York, 258–269.

Bose, N. K. and Rao, C. R. (1993). *Signal Processing and its Applications*, Handbook of Statistics, **10**, North-Holland, Amsterdam.

Bresler, Y. and Macovski, A. (1986). Exact maximum likelihood parameters estimation of superimposed exponential signals in noise, *IEEE Transactions on Acoustics Speech and Signal Processing*, **34**, 1081–1089.

Davis, P. J. (1979). *Circulant Matrices*, Wiley, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **B 39**, 1–38.

Dudgeon, D. E. and Merseresau, R. M. (1984). *Multidimensional Digital Signal Processing*, Englewood Cliffs, NJ: Prentice Hall Inc.

Fant, C. G. M. (1960). *Acoustic Theory of Speech Foundation*, Mouton and Co. The Hague.

Feder, M. and Weinstein, E. (1988). Parameter estimation of superimposed signals using the EM algorithm, *IEEE Transactions on Acoustics Speech and Signal Processing*, **36**, 477–489.

Froberg, C. E. (1969). *Introduction to Numerical Analysis*, 2nd ed., Addison-Wesley, Reading, Massachusetts.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*, Philip Allan, New York.

Haykin, S. (1985). *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, New York.

Hildebrand, F. B. (1956). *Introduction to Numerical Analysis*, McGraw-Hill, New York.

Kannan, N. and Kundu, D. (1994). On modified EVLP and ML methods for estimating superimposed exponential signals, *Signal Processing*, **39**, 223–233.

Kay, S. M. (1988). *Modern Spectral Estimation: Theory and Applications*, Prentice Hall, New Jersey.

Kumaresan, R. (1982). *Estimating the Parameters of Exponentially Damped or Undamped Sinusoidal Signals in Noise*, Ph.D. Thesis, The University of Rhode Island, RI.

Kundu, D. and Mitra, A. (1995). Consistent method of estimating the superimposed exponential signals, *Scandinavian Journal of Statistics*, **22**, 73–82.

Kundu, D. and Mitra, A. (1997). Asymptotic theory of the least squares estimators of sinusoidal signals, *Statistics*, **30**, 221–238.

Kundu, D. and Mitra, A. (1998). Different non-iterative methods to estimate sinusoidal frequencies: a numerical comparison, *Journal of Statistical Computation and Simulation*, **62**, 9-28.

Kundu, D. and Mitra, A. (2000). Detecting the number of signals using cross validation approach, *Signal Processing*, **80**, 525–534.

Kundu, D. and Mitra, A. (2001). Estimating the number of signals of the damped exponential models, *Computational Statistics and Data Analysis*, **36**, 245–256.

Kundu, D. and Nandi, S. (2003). Determination of discrete spectrum in a random field, *Statistica Neerlandica*, **57**, 258–283.

Lanczos, C. (1964). *Applied Analysis*, Prentice Hall, London.

Marple Jr., S. L. (1987). *Digital Spectral Analysis*, Prentice-Hall, Englewood Cliffs, N.J.

Matausek, M., Stankovic, S. S. and Radovic, D. V. (1983). Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids, *IEEE Transactions on Acoustics Speech and Signal Processing*, **31**, 1456–1463.

Pillai, S. U. (1989). *Array Signal Processing*, Springer-Verlag, New York.

Pinson, E. N. (1963). Pitch synchronous time-domain estimation of formant frequencies and bandwidths, *Journal of the Acoustics Society of America*, **35**, 1264–1273.

Prony, R. (1795). Essai experimentale et analytique, *Journal of Ecole Polytechnique* (Paris), **1**, 24–76.

Quinn, B. G. (1994). Estimating frequency by interpolation using Fourier coefficients, *IEEE Transactions on Signal Processing*, **42**, 1264–1268.

Rao, C. R. (1973). *Linear Statistical Inferences and its Applications*, John Wiley & Sons, 2nd ed., New York.

Rao, C. R. (1988). Some recent results in signal detection, *Statistical Decision Theory and Related Topics IV*, Gupta, S.S. and Berger, J.O. eds., Springer Verlag, **2**, 319–332.

Rao, C. R. and Zhao, L. C. (1993). Asymptotic behavior of the maximum likelihood estimators of superimposed exponential signals, *IEEE Transactions on Acoustics Speech and Signal Processing*, **41**, 1461–1463.

Rao, C. R., Zhao, L. C. and Zhou, B. (1994). Maximum likelihood estimators of 2-D superimposed exponential signals, *IEEE Transactions on Acoustics Speech and Signal Processing*, **42**, 1795–1802.

Reddy, V. U. and Biradar, L. S. (1993). SVD-based information theoretic criteria for detection of the number of damped/ undamped sinusoids and their performance analysis, *IEEE Transactions on Acoustics Speech and Signal Processing*, **41**, 2872–2881.

Roy, R. H. (1987). *ESPRIT: Estimation of Signal Parameters via Rotational Invariance Technique*, Ph.D. Thesis, Stanford University, Stanford, California.

Roy, R. H. and Kailath, T. (1989). ESPRIT: estimation of signal parameters via rotational invariance technique, *IEEE Transactions on Acoustics Speech and Signal Processing*, **43**, 984–995.

Sheppard, C. W. and Householder, A. S. (1951), The mathematical basis of the interpretation of tracer experiments in closed steady state systems, *Journal of Applied Physics*, **22**, 510–520.

Srinath, M. D., Rajasekaran, P. K. and Viswanathan, R. (1996). *Introduction to Statistical Signal Processing with Applications*, Prentice-Hall, Englewood Cliffs, New York.

Tufts, D. W. and Kumaresan, R. (1982). Estimation of frequencies of multiple sinusoids: making linear prediction and perform like maximum likelihood, *Proceedings of IEEE*, **70**, 975–989.

Wax, M. and Kailath, T. (1985). Detection of signals by information theoretic criteria, *IEEE Transactions on Acoustics Speech and Signal Processing*, **30**, 387–392.

Yau, S. F. and Bresler, Y. (1993). Maximum likelihood parameter estimation of superimposed signals by dynamic programming, *IEEE Transactions on Signal Processing*, **41**, 804–807.

Zhang, H. and Mandrekar, V. (2002). Estimation of hidden frequencies for 2D stationary processes, *Journal of Time Series Analysis*, **22**, 613–629.