PHASE MEASUREMENTS

AND

NARROW SPECTRAL BAND INFEERENCE

by

P.J. THOMSON

A thesis submitted to the
Australian National University
for the degree of Doctor of Philosophy.

April, 1972.
Preface

Many of the results presented in this thesis have been established in collaboration with my supervisor Professor E.J. Hannan. These joint results now appear in Hannan and Thomson (1971a), (1971b) and (1971c). Elsewhere in the thesis, unless otherwise stated, the work described is my own.

P.J. Thomson
Acknowledgements

I would like to express my gratitude to Professor E.J. Hannan for his stimulating supervision over the past three years. His friendly advice and encouragement have been greatly appreciated.

To the Australian Government I offer my thanks for the postgraduate scholarship awarded me through the Commonwealth Scholarship and Fellowship Plan. The additional financial support provided by the Australian National University is also gratefully acknowledged.
Abstract

This thesis examines the problem of obtaining spectral estimates in the situation where the analysis is to be confined to a narrow band of frequencies concentrated about some fixed frequency of interest. In particular, the standard estimation procedures which require the spectral density function to be smooth over this range of frequencies are extended to cover the case where substantial variation does exist across the band.

Chapter 2 develops a central limit theorem for the finite Fourier transforms based on N observations from a vector stationary discrete time sequence having absolutely continuous spectral distribution function. In order to arrive at a satisfactory asymptotic theory the spectral density matrix is made to depend on N in such a way that the variation modelled is reflected in the limiting distribution. This limiting distribution takes the form of a certain complex multivariate normal distribution.

Chapter 3 yields a procedure for estimating coherence in the case where the phase or cross-spectral argument is changing across the band. It is with relation to the group delay that the phase variation is modelled over the frequencies considered and this parameter is estimated together with the other spectral parameters. The approach taken is that of maximising the likelihood which is given by an appropriate form of the limiting distribution described in Chapter 2. The various statistical properties of the estimates are also established.

Chapters 4 and 5 consider the case where a signal is received together with lagged and attenuated forms of the signal. The appropriate likelihood given by the central limit theorem of Chapter 2 is then maximised with respect to the various parameters and statistical properties are again established for the resulting estimates. The estimation of a signal's characteristics using an array of recorders is also discussed in the light of these methods.
We collect here for reference some general symbols used in this thesis. In the table below the numbers refer to the pages on which the symbols are defined.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.s. = almost surely</td>
<td>11</td>
</tr>
<tr>
<td>det(·), 9</td>
<td>Q(·), 23</td>
</tr>
<tr>
<td>E, 2</td>
<td>R{·}, 5</td>
</tr>
<tr>
<td>f(λ), 3</td>
<td>σ_jk(λ), 5</td>
</tr>
<tr>
<td>f_j(λ), 4</td>
<td>[x], 6</td>
</tr>
<tr>
<td>F(λ), 2</td>
<td>′, 3</td>
</tr>
<tr>
<td>Γ(t), 2</td>
<td>θ_jk(λ), 5</td>
</tr>
<tr>
<td>h(θ), 22</td>
<td>θ_k, 19</td>
</tr>
<tr>
<td>Θ{·}, 5</td>
<td>θ_k′, 22</td>
</tr>
<tr>
<td>I_p, 12</td>
<td>tr(·), 47</td>
</tr>
<tr>
<td>L^N(θ), 20</td>
<td>u, 68</td>
</tr>
<tr>
<td>L(θ), 20</td>
<td>u(θ), 47</td>
</tr>
<tr>
<td>·</td>
<td>u(x), 7</td>
</tr>
<tr>
<td>O(x) = of order x as x → 0</td>
<td>w(λ), 6</td>
</tr>
<tr>
<td>(Ω, F, P), 1</td>
<td>w_u(λ), 6</td>
</tr>
<tr>
<td>ω_k(λ_0), 6</td>
<td>z(λ), 2</td>
</tr>
<tr>
<td>ω_k′(λ_0), 7</td>
<td></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acknowledgements</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table of Notations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter 1</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Definitions, Notations and Some Basic Spectral Theory</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Inference and a Classical Result</td>
<td>6</td>
</tr>
<tr>
<td>1.3</td>
<td>Estimation Procedures and an Associated Problem</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Structure and Scope of the Thesis</td>
<td>16</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>A Certain Central Limit Theorem</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>19</td>
</tr>
<tr>
<td>2.2</td>
<td>The Central Limit Theorem</td>
<td>24</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>The Estimation of Coherence and Group Delay</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>39</td>
</tr>
<tr>
<td>3.2</td>
<td>The Maximum Likelihood Equations and Some Fundamental Lemmas</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>A Strong Law of Large Numbers</td>
<td>54</td>
</tr>
<tr>
<td>3.4</td>
<td>A Central Limit Theorem</td>
<td>67</td>
</tr>
<tr>
<td>3.5</td>
<td>The Numerical Procedure and Some Numerical Examples</td>
<td>74</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Estimating Echo Times</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td>4.2</td>
<td>A Strong Law of Large Numbers</td>
<td>101</td>
</tr>
<tr>
<td>4.3</td>
<td>A Central Limit Theorem</td>
<td>110</td>
</tr>
<tr>
<td>4.4</td>
<td>Some Numerical Studies</td>
<td>117</td>
</tr>
</tbody>
</table>
Chapter 5: The Estimation of a Signal's Characteristics
Using an Array of Recorders

5.1 Introduction 126
5.2 The Estimation Problem 129
5.3 An Optimality Criterion for Array Structures 139

Bibliography 154
1.1 Definitions, Notations and Some Basic Spectral Theory

The Fourier analysis of time series has developed, in no small part, as a result of needs felt in various sciences other than those of the mathematical sciences and, as such, it is replete with an abundance of nomenclature and terminology. For this reason a brief description of certain classical results has been included in the first chapter. This serves the dual role of both standardising the notation that is used throughout and also that of facilitating the discussion of the problems to which the thesis is addressed. In this section and §1.2 the symbolism adopted and the results quoted conform, in the main, to those found in Hannan (1970) and, in particular, the proofs to these results can also be found in this cited reference.

We shall be concerned with making certain inferences about the underlying structure of a series of \( p \)-dimensional vector observations \( x(n), n = 1, \ldots, N \), which have been recorded sequentially through time. These observables could have arisen through sampling a continuous time function or, alternatively, they could be the result of some discrete time process. The former prescription is possibly the more pertinent since the real world abounds with such phenomena. In general we may regard the observation at time point \( t \) as being the realisation of the vector random variable \( x(t, \omega) \) where this random vector function is defined over the probability space \( (\Omega, \mathcal{F}, P) \) with \( \Omega \) indexing the family of all histories of the process and \( P \) denoting the probability measure defined on the Borel field \( \mathcal{F} \) of subsets of \( \Omega \). Suppressing reference to \( \omega \), we say that the vector process \( x(t) \) is
strictly stationary if for all $N, t_1, \ldots, t_N$ and $h$ the distributions of $x(t_1), \ldots, x(t_N)$ and of $x(t_1 + h), \ldots, x(t_N + h)$ are the same. In this definition $N$ is always integral but $t_1, \ldots, t_N$ and $h$ can only vary over the domain of the time function involved. The class of such processes is of some importance since it includes a broad range of phenomena. We henceforth require that $x(t)$ satisfy this condition and, furthermore, denoting the expectation operator by $\mathbb{E}$ we make the additional assumption that $\mathbb{E}(x(t)) = 0, \mathbb{E}(x(t)' x(t)) < \infty$, where the dash denotes vector transpose. Thus the covariance matrix function can evidently be written as

$$\mathbb{E}(x(s) x(s + t)') = \Gamma(t)$$ (1.1)

since it is dependent only on the time separation involved.

We now state the following well known representation theorems.

**Theorem 1.1**

If $x(t)$ is a continuous time process and satisfies the conditions given above then

$$\Gamma(t) = \int_{-\infty}^{\infty} e^{it\lambda} \ dF(\lambda)$$

where $F(\lambda)$ is a matrix whose increments, $F(\lambda_1) - F(\lambda_2), \lambda_1 \geq \lambda_2$, are Hermitian non-negative definite. The function $F(\lambda)$ which is called the spectral distribution matrix is uniquely defined if we require it to be continuous from the right and $\lim_{\lambda \to -\infty} F(\lambda) = 0$.

**Theorem 1.2**

If $x(t)$ is a continuous time process and satisfies the conditions given above then

$$x(t) = \int_{-\infty}^{\infty} e^{-it\lambda} \ dz(\lambda)$$
where $z(\lambda)$ is a complex-valued vector process of orthogonal increments with $\delta[z(\lambda) \, z(\lambda)^*] = F(\lambda)$, the star denoting transposition combined with conjugation. The function $z(\lambda)$ is uniquely determined almost everywhere if we require it to be continuous in mean square from the right.

This latter theorem implies, loosely speaking, that $x(t)$ can be decomposed into a sum of orthogonal random vectors each weighted by an appropriate sinusoidal component.

We now take $F(\lambda)$ to be absolutely continuous and thus

$$\Gamma(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) \, d\lambda \quad (1.2)$$

where $f(\lambda)$ is Hermitian and is called the spectral density matrix. Let us consider the situation where the data sequence to hand has come about through sampling a continuous time process at the time points $n\Delta$, $n = 0, \pm 1, \ldots$. Thus $\Gamma(t)$ is now replaced by

$$\Gamma(n\Delta) = \int_{-\infty}^{\infty} e^{in\Delta\lambda} f(\lambda) \, d\lambda$$

$$= \sum_{-\infty}^{\infty} \int_{-\pi/\Delta}^{\pi/\Delta + 2\pi j/\Delta} e^{in\Delta\lambda} f(\lambda) \, d\lambda$$

$$= \int_{-\pi/\Delta}^{\pi/\Delta} e^{in\Delta\lambda} f^A(\lambda) \, d\lambda$$

where

$$f^A(\lambda) = \sum_{-\infty}^{\infty} f(\lambda + 2\pi j/\Delta).$$

It is clear that all we can know from the $\Gamma(n\Delta)$ is $f^A(\lambda)$ and so two different $f(\lambda)$ yielding the same $f^A(\lambda)$ cannot be identified one from the other. This situation, which is usually described by saying that the frequencies $\lambda + 2\pi j/\Delta$, $j = 0, \pm 1, \ldots$, $\lambda \in [-\pi/\Delta, \pi/\Delta]$, are "aliased", is alleviated if $\Delta$ is chosen small enough to ensure that
f(\lambda) has negligible mass outside the range \([-\pi/\Delta, \pi/\Delta]\). A similar aliasing effect can be seen to hold in relation to Theorem 1.2, but we shall not dwell on this subject any further and shall assume that we are sampling at unit intervals (\Delta = 1) and that aliasing effects can be ignored.

With this in mind we now state the discrete time analogues of Theorems 1.1 and 1.2.

**Theorem 1.3**

If \(x(n)\) is a discrete time process and satisfies the stated assumptions then

\[ \Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} dF(\lambda) \]

where \(F(\lambda)\) is as described in Theorem 1.1 except that now \(F(-\pi)\) is null.

**Theorem 1.4**

If \(x(n)\) is a discrete time process and satisfies the stated assumptions then

\[ x(n) = \int_{-\pi}^{\pi} e^{-in\lambda} dz(\lambda) \]

where \(z(\lambda)\) is a complex-valued vector process of orthogonal increments on \((-\pi, \pi]\) with the same covariance properties as for the corresponding process in Theorem 1.2.

With reference to Theorems 1.3 and 1.4 we again impose the requirement that \(F(\lambda)\) be absolutely continuous with spectral density matrix \(f(\lambda)\).

We denote the \(j^{th}\) diagonal element of \(f(\lambda)\) by \(f_{jj}(\lambda)\) and this is called the spectrum at angular frequency \(\lambda\) of the \(j^{th}\) component whereas the off-diagonal element \(f_{jk}(\lambda)\) is called the cross-spectrum between the \(j^{th}\) and \(k^{th}\) components at this frequency. In practice
these latter quantities are often replaced by $\sigma_{jk}(\lambda)$, $\theta_{jk}(\lambda)$ which are defined as

$$\sigma_{jk}^2(\lambda) = \frac{|f_{jk}(\lambda)|^2}{f_j(\lambda) f_k(\lambda)},$$

(1.3)

$$\theta_{jk}(\lambda) = \arctan[-\Re\{f_{jk}(\lambda)\}/\Im\{f_{jk}(\lambda)\}]$$

where $\Re\{.\}$ and $\Im\{.\}$ denote the imaginary part and real part respectively of the indicated expression. The coherence, $\sigma_{jk}(\lambda)$, $0 \leq \sigma_{jk}(\lambda) \leq 1$, is an intrinsic measure of the strength of association between the $j$th and $k$th components at frequency $\lambda$. It can be interpreted as the maximum correlation that can be achieved between $dz_j(\lambda)$ and $dz_k(\lambda)$ by rephasing one of these two entities relative to the other. The phase, $\theta_{jk}(\lambda)$, describes the phase change required to accomplish this.

We conclude this section by briefly outlining the concept of a filter. Let us suppose that a vector process $y(t)$ has been constructed from $x(t)$ by means of the relation

$$y(t) = \sum A_j x(t - t_j), \quad \sum ||A_j||^2 < \infty$$

(1.4)

and $||f(\lambda)||$ is essentially bounded where by $||A||$ we mean the square root of the greatest eigenvalue of $A^*A$. This is an example of a simple linear filtering operation and the spectral density matrix of the $y(t)$ series is known to be $h(\lambda) f(\lambda) h(\lambda)^*$ where the matrix response function $h(\lambda)$ is given by

$$h(\lambda) = \sum A_j e^{it_j \lambda}.$$  

(1.5)

Thus the effect of the filter has just been that of modifying the spectral density matrix $f(\lambda)$ in this straightforward manner. Similar results hold for a large class of operations and, in general, a filter
may be regarded as a time invariant operator which yields a vector process \( y(t) \) from some vector stationary process \( x(t) \). A rigorous discussion of the theory of filters can be found in Hannan (1970).

1.2 Inference and a Classical Result

In this section we shall be primarily concerned with estimating the various spectral parameters at some fixed frequency \( \lambda_0 \). In this connection the finite Fourier transform of the observations \( x(n), n = 1, \ldots, N \), which is defined by

\[
     w(\lambda) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} x(n) \exp(in\lambda)
\]

is of central interest due to the fact that \((2\pi/N)^{\frac{1}{2}} w(\lambda)\) is an approximation to \( dz(\lambda) \), one of the orthogonal terms that form the Fourier representation of \( x(n) \). Thus we are led to consider the \( w(\lambda) \) evaluated at the \( m \) frequencies of the form \( 2\pi k/N \) nearest to \( \lambda_0 \) where \( m \) is very small compared to \( N \). We shall denote these particular \( m \) frequencies by \( \omega_k(\lambda_0) = \frac{2\pi k}{N} \), \( -\frac{m}{2} < k \leq \frac{m}{2} \), where \( \omega_0(\lambda_0) \) is the nearest frequency to \( \lambda_0 \) and by \([x]\) we mean the integral part of \( x \). Now the finite Fourier transform is sometimes replaced by an approximation (often called the Cooley-Tukey estimate) of the form

\[
     w_u(\lambda) = (2\pi N)^{-\frac{1}{2}} \sum_{n=1}^{N} u^{(N)}(n) x(n) \exp(in\lambda)
\]

where \( u^{(N)}(n) \) is a "fader" or "taper" which is near to unity for most \( n \) but fades to zero when \( n = 0 \) and \( n = N + 1 \). The \( w_u(\lambda) \) are usually evaluated at the \( m' \) frequencies of the form \( 2\pi k/N' \) where \( N \leq N' = 2^s < 2N \) and \( m' \) is chosen by requiring both bands of frequencies to have similar widths which is to say \( m'/N' = m/N \). It is not necessary to take \( N' = 2^s \) but it must be highly composite if a great reduction in computational effort is to be achieved. (See Cooley
and Tukey (1965). In an analogous fashion to the $\omega_k (\lambda_0)$ we let $\omega_k (\lambda_0)$ denote the $m'$ frequencies of the form $2\pi k / N'$.

A relevant model for such narrow band inference problems is one in which $N$ is allowed to become infinitely large while $m$ remains fixed. Unless drastic assumptions are made concerning the structure of the data it is virtually impossible to achieve a useful analysis for fixed $N$ and for $m$ very small compared to $N$ the consideration of $m$ fixed is not unreasonable. In the light of this model it is evident that we need to discuss the asymptotic behaviour of $u^N (n)$ and $N'$. Now, examples of faders might be

$$u^N (n) = \begin{cases} \frac{n}{R} & \text{or} \quad \frac{1}{2} [1 + \cos (\pi n/R)] \\ 1 & R \leq n \leq N - R \\ \frac{(N-n)}{R} & \text{or} \quad \frac{1}{2} [1 + \cos (\pi (N-n)/R)] \\ & N - R \leq n \leq N \end{cases}$$

and these suggest introducing a sequence $u^N (x)$ of continuous functions and a continuous function $u(x)$ such that

$$u^N (n) = u^N (n/N), \quad |u^N (x)| \leq b < \infty,$$  

$$\lim_{N \to \infty} |u^N (x) - u(x)| = 0, \quad 0 < x < 1.$$  

In the majority of cases, $b$ can be taken as unity as the above example indicates. One could dispense with the $u^N (x)$ and simply define

$$u^N (n) = u(n/N),$$

but in fact the choice of fader would probably depend on $N$ in the way we have indicated. Indeed in many cases $u(x)$ might reasonably be taken as identically unity to provide a good approximation to the actual fader used. However, since the fader used in practice might differ greatly from unity we have maintained the additional generality. (For example, in Munk and Cartwright (1966) the fader used is $u^N (n) = u(n/N), u(x) = 1 + \cos [\pi (2x - 1)]$.) We shall always
assume that $u^{(N)}(n)$ satisfies the above conditions and, in addition, we require $N'$ to be an integer valued function of $N$ such that

$$\lim_{N \to \infty} \frac{N}{N'} = a, \quad \frac{1}{2} < a \leq 1. \quad (2.4)$$

Of course, as we have said, $N'$ will be highly composite, but this fact is important only computationally and does not affect any of the mathematical requirements given above.

With these considerations we see that

$$\delta[w_u(\lambda) w_u(\lambda)^*] = \int_{-\pi}^{\pi} \left\{ (2\pi N)^{-1} \sum_{s=1}^{N} \sum_{t=1}^{N} u^{(N)}(s) u^{(N)}(t) e^{i(t-s)(\theta-\lambda)} \right\}$$

$$\times f(\theta) \, d\theta$$

$$= \int_{-\pi}^{\pi} |\phi_N(\theta - \lambda)|^2 f(\theta) \, d\theta$$

where

$$\phi_N(\theta) = (2\pi N)^{-1/2} \sum_{n=1}^{N} u^{(N)}(n) \exp(in\theta).$$

Thus the purpose of a fader can be seen by noting that it will be advantageous to have, for any given $N$, $|\phi_N(\theta)|^2$ as small as possible for $\theta$ away from zero. If this condition is not fulfilled then there is always the possibility that $f(\theta)$ will be large at a value away from $\lambda$ that happens to coincide with a subsidiary peak, say, of $|\phi_N(\theta - \lambda)|^2$ and the combined effect will introduce a bias into $\delta[w_u(\lambda) w_u(\lambda)^*]$ away from $f(\lambda)$. It is known that the higher the order of differentiability of the function $u(x)$ the faster $\phi_N(\theta)$ decreases as $|\theta|$ increases.

However the use of the fader introduces correlation between the $w_u(\omega_u'(\lambda_0))$ in their asymptotic distribution as will shortly be indicated. Moreover, the usual procedure for obtaining spectral estimates is to average $w_u(\lambda) w_u(\lambda)^*$ over the $m'$ frequencies $\omega_u'(\lambda_0)$ where $m'$ has been chosen so that $f(\theta)$ does not vary greatly over this
range. In this context, therefore, discrepant values of \( f(\theta) \) would be unlikely to occur just outside this band of frequencies. Hence it would appear that a major use of the fader would be to prevent any bias arising from large \( f(\theta) \) values a fair way from \( \lambda_0 \).

In order that the statement of a certain central limit theorem for the vectors \( w_u(\omega_k(\lambda_0)) \) is adequately described we now consider the complex multivariate normal distribution. This may be introduced as follows. We consider \( 2p \) random variables, \( x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_p \) having zero means and a real multivariate normal distribution with the covariance matrix

\[
\begin{bmatrix}
C & Q \\
-Q & C
\end{bmatrix}
\]

wherein all matrices are \( p \times p \). Here the partition makes \( C \) the covariance matrix of \( x_1, \ldots, x_p \) and also of \( y_1, \ldots, y_p \). Since this is a covariance matrix \( C = C' \) and \( Q = -Q' \). Now if we introduce the complex random variables \( z_j = \frac{1}{2}(x_j + iy_j), j = 1, \ldots, p \) and the matrix \( A = \frac{1}{2}(C - iQ) \) then the density of the \( x_j \) and \( y_j \) may be rearranged as

\[
\pi^{-p} \det(A^{-1}) \exp(-z^*A^{-1}z)
\]

where \( \det(A) \) denotes the determinant of \( A \).

We introduce the following conditions some of which have already been given in §1.1, but these are restated here for the sake of completeness.

(i) The process \( x(n) \) is to be strictly stationary with zero means and finite variances.

(ii) The spectral distribution matrix is to be absolutely continuous with the spectral density matrix \( f(\lambda) \) continuous at the fixed, but arbitrary, frequency \( \lambda_0 \). Also, \( \|f(\lambda)\| \) is assumed to be essentially bounded.
(iii) The $x(n)$ process is to satisfy the "uniform mixing condition", shortly to be described, introduced into this type of problem by Rosenblatt (1956). (Rozanov (1967), p.180, calls such a process completely regular whereas Billingsley (1968), p.166, calls it $\phi$-mixing and gives a detailed treatment of sequences of this type.)

With reference to §1.1 and for any integers $n_0$, $n_1$, $n_0 < n_1$ we introduce the Borel subfield $\mathcal{B}$ of $\mathcal{A}$ determined by all observables up to the time $n_0$ and the Borel subfield $\mathcal{F}$ of $\mathcal{A}$ determined by all observables after the time $n_1$. Then the uniform mixing condition requires that

$$\sup_{B \in \mathcal{B}, F \in \mathcal{F}} |P(B \cap F) - P(B) P(F)| < \psi(n_1 - n_0),$$

(2.6) where $\psi(x)$ is a positive even function of $x$ which decreases to zero as $x$ increases. We refer to these conditions collectively as Condition A.

It is clear that (2.6) implies that events far apart in time approach independence at a rate given by $\psi(x)$ and the time separation involved and is independent of the particular $B$ and $F$ considered. This condition is intuitively appealing since, in many areas of data analysis, one is more than ready to concede that this does represent a good approximation to the true state of affairs. We give below a certain central limit theorem for the $\omega_u(\omega_k' \lambda_0)$ which utilises this concept. It is noted that this result and the attendant proof are presented in Hannan (1970), Chapters 4 and 5. Many authors have established various central limit theorems for stationary processes and we briefly mention Moran (1947), Diananda (1954), Grenander and Rosenblatt (1957), Hannan (1961), Eicker (1967), Billingsley (1968), Brillinger (1969) and Ibragimov and Linnik (1971). Brillinger (1969) has developed related results to those about to be given but the conditions and type of situation considered differ from those postulated here.
Let us call \( w \) the vector with \( w_{u,j}(\omega_k(\lambda_0)) \) as the \([(k + \frac{1}{2}(m' - 1)) p + j]^\text{th} \) component with \( p \) representing the dimensionality of \( x(n) \) and, furthermore, we define the matrix \( Q \) by requiring its typical element to be

\[
Q_{jk} = \int_0^1 u^2(x) \, e^{i2\pi ax(j-k)} \, dx , \quad 1 \leq j,k \leq m' . \tag{2.7}
\]

By \( A \otimes B \) we mean the tensor product of these two matrices yielding the partitioned matrix with typical sub-matrix \( a_{jk}B \). We now state the following theorem.

**Theorem 1.5**

If \( x(n) \) satisfies Condition A, a necessary and sufficient condition that, for \( \lambda_0 \neq 0,\pi \), the distribution of \( w \) should approach, as \( N \to \infty \), a \( pm' \)-variate complex multivariate normal distribution with zero mean vector and covariance matrix \( Q \otimes f(\lambda_0) \) is that the random variable

\[
X(\lambda_0,N) = \|(2\pi N)^{-\frac{1}{2}} \sum_{i=1}^{N} x(n) \exp(in\lambda_0)\|^2
\]

should be uniformly integrable, i.e.,

\[
\lim_{N \to \infty} \lim_{x \to \infty} \int_{X(\lambda_0,N) > x} X(\lambda_0,N) \, dP = 0 .
\]

If, for \( \lambda_0 = 0 \), the vector \( w_{u}(0) \) is excluded from \( w \) and for \( N \) even and \( \lambda_0 = \pi \), the vector \( w_{u}(\pi) \) is not included in \( w \) then the theorem remains true.

**Corollary 1.1**

In the important central case where there is no fader and the entries in \( w \) are of the form given by (2.1) we observe that \( Q \) becomes the unit matrix. Thus the vectors \( w_{k}(\lambda_0) \) are asymptotically
independently distributed each with p-variate complex multivariate normal distribution having zero mean and covariance matrix \( f(\lambda_0) \).

In the case of the Cooley-Tukey procedure with \( a < 1 \) we note that even if \( u(x) = 1 \) the matrix \( Q \) does not become the unit matrix.

1.3 Estimation Procedures and an Associated Problem

When one maximises the likelihood given by Corollary 1.1 with respect to the spectral parameters \( f_{jk}(\lambda_0) \) the maximum likelihood estimates that result are described by

\[
\hat{f}_{jk}(\lambda_0) = \frac{1}{m} \sum_{s} w_j(\omega_s(\lambda_0)) \frac{w_k(\omega_s(\lambda_0))}{w_s(\lambda_0)}
\]

(3.1)

where the summation is over \( s \) such that \(-\frac{1}{2}m < s \leq \frac{1}{2}m\). If the parameterisation chosen is of the form given by (1.3) then, for \( j \neq k \), the statistics given by (3.1) are replaced by

\[
\hat{\theta}_{jk}(\lambda_0) = \frac{\hat{f}_{jk}(\lambda_0)}{|\hat{f}_{jk}(\lambda_0)|} \frac{\hat{f}_j(\lambda_0) \hat{f}_k(\lambda_0)}{\hat{f}_j(\lambda_0) \hat{f}_k(\lambda_0)}
\]

(3.2)

which are, respectively, the sample coherence and phase between the \( j^{th} \) and \( k^{th} \) components of the \( x(n) \) process at frequency \( \lambda_0 \). These estimates form the basis for most spectral analyses and it is with reference to these quantities that one tends to make inferential statements concerning the nature of the time series being studied.

If the Cooley-Tukey technique is considered the estimates obtained by maximising the likelihood given by Theorem 1.5 are identical to (3.1) except that now \( w \) is replaced by \((Q^{-\frac{1}{2}} \otimes I_p)w\) where \( I_p \) is the \( p \)-rowed unit matrix and \( m \) is replaced by \( m' \). Because of the work entailed in effecting this transformation and because a may be
near to unity and \( u^{(N)}(n) \) not far from identically unity this procedure is not likely to be used and the estimate

\[
\hat{f}(\lambda_0) = \left[ N^{-1} \sum_{n=1}^{N} \{u^{(N)}(n)\}^2 \right]^{-1} (m')^{-1} \sum_{k} w_u(\omega_k(\lambda_0)) w_u(\omega_k'(-\lambda_0))^* (3.3)
\]

is often chosen instead since this yields a more tractable approximate alternative. However, in any event, because of the intervention of the \( Q \) matrix into the analysis when the Cooley-Tukey procedure is used the variances of these spectral estimates are greater than those achieved for the corresponding estimates given by (3.1) where faders have not been implemented.

One principal drawback to all of these formulae is that they presuppose, by virtue of Theorem 1.5, that the spectral density matrix \( f(\lambda) \) does not vary to any great extent over the narrow band of frequencies under study. However there are many situations where substantial variation does exist across the band. It is known, for instance, that whereas spectra, for the most part, are fairly well behaved, cross-spectra, which incorporate the various phases, are not so easily dealt with. Rephasing can occur in a multitude of ways and this can result in rapid cross-spectral change which could lead to spurious results. It has been found, for example, that the conventional method for estimating the coherence, given by (3.2), will yield too low a value if, near \( \lambda_0 \), there is rapid phase variation present. The bias arises because the methods under discussion estimate \( |f_{jk}(\lambda_0)| \) by averaging estimates of \( f_{jk}(\lambda) \) across the band so that if \( |f_{jk}(\lambda)| \) is near to constant across that band, but \( \theta_{jk}(\lambda) \) is not, then the modulus being estimated is of the approximate form

\[
(2\delta)^{-1} \int_{|\lambda-\lambda_0|<\delta} \exp(-i\theta_{jk}(\lambda)) |f_{jk}(\lambda_0)| d\lambda
\]

which may be small if \( \theta_{jk}(\lambda) \) is varying rapidly near \( \lambda_0 \).
particular problem has been discussed by a number of authors including Pierson and Dalzell (1960), Akaike and Yamanouchi (1963), Tick (1967) and Jones (1969), but, although many estimation procedures were proposed, no satisfactory general theory eventuated. Thus the problem that presents itself is that where there is variation in the spectral density function across the narrow band in question. Clearly one could increase N until the bandwidth was sufficiently small for an ordinary analysis to be applied, but this is not always possible since there is usually a practical limit, dictated by other than mathematical considerations, which determines this value. Similarly, for m, we could decrease this quantity, but in so doing we might be discarding information that could well be utilised given some knowledge of the variation involved. It would seem, then, that this is the "Achilles heel" of Theorem 1.5 and some extension is required in order that the spectral density function is allowed to exhibit some form of fluctuation across the band.

There are many avenues of approach to this problem. One could, for example, make an (unrealistic) Gaussian assumption concerning the data and then, taking a fixed band of frequencies, establish some likelihood criterion which, when maximised, would yield estimates of the desired spectral parameters. It might then be possible to establish the distributions of these estimates under much weaker conditions than those initially postulated and thus proceed to develop some near optimal procedure. Even this, however, would be difficult to obtain. The theory that will be constructed here resembles the classical theory of Theorem 1.5 except that now an accounting is taken of the variational properties of the spectral density. It is clear that some mathematical device will be required since Theorem 1.5 in its present form is of little use because, as N becomes large, the band of
width $2\pi m/N$ will become so narrow that the spectral components may eventually be regarded as constant over the band.

We thus consider our spectral density matrix to be of the form $h^{(N)}(\lambda) f(\lambda) h^{(N)}(\lambda)^* \text{ where } f(\lambda) \text{ is a spectral density and } h^{(N)}(\lambda)$ is the matrix response function of a filter. Moreover, it will now be assumed that the variation across the band that is to be modelled can be ascribed to $h^{(N)}(\lambda)$ whereas $f(\lambda)$ is to be continuous at $\lambda_0$ and smooth over the frequencies considered. These mild conditions would appear to represent the very least that one might require in order to distinguish between the rapidly varying and slowly varying components of a spectral density function. However, considering the asymptotic theory, we must now retain this variation as $N$ becomes large and so the matrix $h^{(N)}(\lambda)$ is made to depend on $N$ to the extent that as $N$ increases $h^{(N)}(\lambda) f(\lambda) h^{(N)}(\lambda)^*$ varies increasingly rapidly across the band so that some limiting form of variation enters the resulting asymptotic distribution. Hence our observations $x(n), n = 1, \ldots, N$ are considered to be imbedded in a sequence of such sets of observations where a typical set is indexed by $N$ and has spectral density $h^{(N)}(\lambda) f(\lambda) h^{(N)}(\lambda)^*$. We experience, of course, only one such data set and one value of $N$, but to develop some relevant approximation to the actual distribution of the vector $w$ requirements such as those given above must be imposed.

In essence we are making the premise, given these circumstances, that in order to measure fluctuation in the spectral density function it must be possible to break off the components that are causing the fluctuation in the form of a filter and then to assume that over the band $f(\lambda)$ may be regarded as nearly constant whereas the variation caused by $h^{(N)}(\lambda)$ is discernible. Without such an assumption it may be impossible to disentangle the required variational effects.
from the observed spectral density. A rigorous development of this concept and the related central limit theorem are given in Chapter 2.

1.4 Structure and Scope of the Thesis

As has already been indicated, Chapter 1 contains a brief introduction to some classical results and a certain central limit theorem. In §1.3 the basic problem is stated and the discussion that follows describes the line of approach that is to be adopted in deriving a solution.

The second chapter contains the statement and proof of a central limit theorem (Theorem 2.1) which represents a generalisation of Theorem 1.5 in that now the variation exhibited by the spectral density function across the band is reflected in the limiting distribution. The asymptotic theory is established with $N$, the sample size, becoming large whereas $m$, the number of fundamental frequencies in the band, remains fixed. The contents of this chapter are based on Hannan and Thomson (1971a).

In Chapter 3 we consider the problem of estimating coherence in a situation where there is rapid phase change over the narrow band of frequencies of interest. Given this state of affairs, conventional methods for measuring the coherence give too low an estimate and this would indeed be so if the group delay or derivative of phase with respect to frequency is large relative to the band width considered. It is the group delay which is assumed to contribute the major part of the phase variation across the band and it is this quantity which is estimated along with the coherence and other spectral parameters. However the group delay is also an important parameter in its own right since it directly determines the lead or lag of one component relative to the other over the band of frequencies considered. The technique
used is that of maximising the likelihood given by an appropriate form of Theorem 2.1. However the various statistical properties of the derived estimates do not follow automatically since the situation considered does not fit into the standard maximum likelihood framework and so an asymptotic approach is developed with m becoming large under the assumption that Theorem 2.1 represents the true likelihood. The theorems presented give no indication as to the relative sizes of m and N for a good approximation to result and so §3.5 contains a numerical example and some Monte-Carlo studies which serve to throw some light on this matter. The major part of Chapter 3 appears in Hannan and Thomson (1971b).

Chapter 4 establishes a procedure for estimating the various spectral parameters in an echo type situation where both the signal is received as well as lagged and attenuated forms of the signal. The lag parameters are assumed to be frequency dependent and large enough to cause discernible variation in the spectral density over the band. Thus Theorem 2.1 is again applied and we develop the asymptotic properties of the associated maximum likelihood estimates in an analogous manner to that given in Chapter 2. In §4.4 the procedure is illustrated with reference to a simple numerical example. The basis for this chapter is to be found in Hannan and Thomson (1971c).

The last chapter considers certain approximations to the maximum likelihood estimates with special reference to the situation where the observed signal is received by an array of recorders. This array, which is usually of a symmetrical nature, generates a pattern of lags which are reflected in the appropriate likelihood function given, once again, by Theorem 2.1. Each recorder also receives noise so that we are basically concerned with an interferometer in which noise effects are not negligible. The characteristics of an "optimum" array
are discussed and certain array structures are examined in the light of a proposed optimality criterion.

It is emphasised that the results developed in this thesis are primarily for the case where the variation is narrow band in nature. This is not to say that if a common fluctuation existed over adjacent bands and the various estimates were calculated and combined in some appropriate way that this would not be near optimal, but this aspect of the problem is not formally covered herein.

We note here that all numerical work was carried out on the I.B.M. 360/50 computer at the Australian National University and hence all the figures pertaining to times and other such information relate to this particular machine. Also, equations that need to be referenced will be labelled in the form (i.j.k) which is to be interpreted as the $k^{th}$ equation in the $j^{th}$ section of the $i^{th}$ chapter. If no confusion will result the $i$ is often suppressed if the equation that is referenced is in the current chapter. All theorems, lemmas and corollaries will be numbered according to the chapters in which they occur and so Theorem 1.3 refers to the third theorem in Chapter 1.
2.1 Introduction

We proceed to examine this situation where there is variation in the spectral components across a narrow band of frequencies centred about $\lambda_0$. In order that we might obtain a satisfactory asymptotic theory we need to make the spectral density matrix of our observed process depend on $N$ so that it varies increasingly rapidly across the band. We therefore consider our vector observations to be of the form $x^{(N)}(n)$ where $x^{(N)}(n)$ is obtained from $x(n)$ by means of a filter having matrix response function $h^{(N)}(\lambda)$. Thus the spectral density of $x^{(N)}(n)$ is given by $f^{(N)}(\lambda) = h^{(N)}(\lambda) \ast f(\lambda) h^{(N)}(\lambda)^*$ and at this stage $h^{(N)}(\lambda)$ is only subjected to the requirement that it be composed of measurable functions and the elements of $f^{(N)}(\lambda)$ be integrable with

$$\int_{-\pi}^{\pi} \|f^{(N)}(\lambda)\| \, d\lambda$$

uniformly bounded. The process $x(n)$ will be required to satisfy Condition A of §1.2. Writing $w^{(N)}(\lambda)$ for $w(\lambda)$ computed from $x^{(N)}(n)$ we shall now relocate the grid of values at which these are computed taking this to be $\theta_k = \lambda_0 + \omega_k$ where $\omega_k = 2\pi k/N$, $-\frac{1}{2}m < k \leq \frac{1}{2}m$. Of course $w^{(N)}(\lambda)$ is usually evaluated at $\omega_k(\lambda_0)$, but in the circumstances considered where $f^{(N)}(\lambda)$ is varying rapidly near $\lambda_0$ this will not lead to a useful result. However the difference between the two grids is not large since $|\theta_k - \omega_k(\lambda_0)| \leq \pi/N$ and so it may be true that $f^{(N)}(\lambda)$ is varying sufficiently slowly, in practice, for that variation to be neglected over a band of width not greater than $\pi/N$. In this case one might choose to use $\omega_k(\lambda_0)$ in place of $\theta_k$ or, equivalently, take the central frequency of interest to be $\omega_0(\lambda_0)$. Nevertheless, for a precise statement, the relocation of the grid is necessary.
Now we have

\[ \phi \left[ w^{(N)}(\theta_k) w^{(N)}(\theta_k)^* \right] \]

\[ = (2\pi)^{-1} \int_{-\pi}^{\pi} f^{(N)}(\theta) \left\{ -1 \sum_{s=1}^{N} \sum_{t=1}^{N} e^{-i(s-t)(\theta-\theta_k)} \right\} d\theta \]

\[ = (2\pi)^{-1} \int_{-\pi}^{\pi} h^{(N)}(\theta+\lambda_0) f(\theta+\lambda_0) h^{(N)}(\theta+\lambda_0)^* L_N(\theta-\omega_k) d\theta \]

where

\[ L_N(\theta) = \sin^2 \frac{\pi \theta}{N} \sin^2 \frac{\pi \theta}{N} . \]

In (1.1) we have defined \( h^{(N)} \), \( f \), \( L_N \) outside \((-\pi, \pi]\) by periodicity. The function \( L_N(\theta) \), Fejér's kernel, is substantially concentrated at the origin. Rewriting (1.1) as

\[ (2\pi)^{-1} \int_{-N\pi}^{N\pi} h^{(N)}(\theta/N+\lambda_0) f(\theta/N+\lambda_0) h^{(N)}(\theta/N+\lambda_0)^* N^{-1} L_N((\theta-2\pi k)/N) d\theta \]

we are led to require that \( h^{(N)}(\theta/N+\lambda_0) \) should approach a limiting matrix function \( h(\theta) \) in some suitable manner. Thus we now proceed, in a somewhat heuristic fashion, to establish a convergence criterion for \( h^{(N)}(\theta/N+\lambda_0) \) whose sufficiency will later be proved. It would appear from the above that terms of the form

\[ (2\pi)^{-1} \int_{-N\pi}^{N\pi} \| h^{(N)}(\theta/N+\lambda_0) - h(\theta) \|^2 \| f(\theta/N+\lambda_0) \| N^{-1} L_N((\theta-2\pi k)/N) d\theta \]

must tend to zero as \( N \) becomes large. Writing \( \chi_N(\theta) \) for the function which is unity over \((-N\pi, N\pi]\) and zero otherwise we observe that

\[ \chi_N(\theta) N^{-1} L_N((\theta-2\pi k)/N) \]

\[ = L(\theta-2\pi k) \left\{ \chi_N(\theta) \left( \frac{1}{2}(\theta-2\pi k)/N \right)^2 / \sin^2 \left( \frac{1}{2}(\theta-2\pi k)/N \right) \right\} \]

converges boundedly to \( L(\theta-2\pi k) \) where

\[ L(\theta) = \sin^2 \frac{\theta}{2}/(\frac{\pi}{2})^2 . \]
Since \( \|f(\theta)\| \) is essentially bounded and
\[
\chi_N(\theta) N^{-1} L_N((\theta-2\pi k)/N) \leq \left\{ \left( \frac{2(1 + \frac{m}{N})}{\sin^2 \left( \frac{\pi}{2}(1 + \frac{m}{N}) \right)} \right) \right\} L(\theta-2\pi k)
\]
\[
\leq K L(\theta-2\pi k)
\]
for some constant \( K \) it is evident that (1.2) is dominated by
\[
[K \sup_{\theta} \|f(\theta)\|] (2\pi)^{-1} \int_{-\infty}^{\infty} \|h^{(N)}(\theta/N+\lambda_0) - h(\theta)\|^2 L(\theta-2\pi k) \, d\theta.
\]
Hence we now require \( h^{(N)}(\theta/N+\lambda_0) \) to converge to \( h(\theta) \) in the sense
\[
\lim_{N \to \infty} (2\pi)^{-1} \int_{-\infty}^{\infty} \|h^{(N)}(\theta/N+\lambda_0) - h(\theta)\|^2 L(\theta-2\pi k) \, d\theta = 0 \quad (1.3)
\]
for all \( k \) such that \(-\frac{1}{2}m < k \leq \lfloor \frac{1}{2}m \rfloor\). We note that \( h(\theta) \) is also a function of \( \lambda_0 \), but since this latter quantity is kept fixed we have not explicitly denoted this dependence. What we have required is that the variation of \( h^{(N)} \) with \( N \), in the neighbourhood of \( \lambda_0 \), should be such that when the scale is changed by dilating the \( \theta \) axis near \( \lambda_0 \) by \( N \), \( h^{(N)} \) will converge in the sense of (1.3) to a function independent of \( N \). We say "in the neighbourhood of \( \lambda_0 \)" since \( \lambda_N(0) \) is such that only these points are important.

However something needs to be said concerning the function \( h(\theta) \). Let us consider, for example, a simple filtering situation such as that given by (1.1.4) and (1.1.5). In these circumstances \( h^{(N)}(\theta+\lambda_0) \) may be written as \( \sum A_j e^{it_j \theta} \) where the \( A_j \) are constant matrices and we shall, without loss of generality, take the \( t_j \) positive with \( t_0 \) identically zero. Furthermore, the \( t_j \), \( j \neq 0 \), are assumed to be large enough to cause discernible variation across the band. Thus
\[
h^{(N)}(\theta/N+\lambda_0) = \sum A_j e^{i(t_j/N)\theta}
\]
and we may take account of the size requirement for the \( t_j \) by regarding
these quantities to be functions of \( N \) such that

\[
\lim_{N \to \infty} \left[ t_j/N \right] = v_j
\]

for some positive constants \( v_j \). However, if we are examining a record that has been sampled at \( N \) time points, the maximum lag that can be identified from the record is clearly less than \( N \). Hence we shall also take the \( t_j \) to be less than \( N \) and this implies that \( 0 \leq v_j < 1 \). Now, in this case, \( h(\theta) \) will be given by

\[
h(\theta) = \sum A_j e^{iv_j\theta}, \quad 0 \leq v_j < 1.
\] (1.4)

Reverting to a more general specification we shall require that

\[
h(\theta) = \int_0^1 e^{iv\theta} dA(v)
\] (1.5)

where \( A(v) \) is composed of complex functions whose real and imaginary parts define signed measures. In relation to the example given by (1.4) and taking \( dA(v) \) to be \( a(v) \, dv \), it is apparent in this case that

\[
a(v) = \sum A_j \delta(v-v_j)
\]

where \( \delta(v) \) is the Dirac delta function.

Let us now elucidate the more general situation where faders are involved and note that it includes the \( w^{(N)}(\lambda) \) as a special case.

We introduce \( w_u^{(N)}(\lambda) \) as the Cooley-Tukey estimate based on the observations \( x^{(N)}(n) \) and evaluate this function over the grid of values

\[
\theta_k = \lambda_0 + \omega_k
\]

where \( \omega_k = 2\pi k/N' \), \(-\pi < k < \pi\). The comments made with respect to the choice of the grid \( \theta_k \) apply here in relation to the \( \theta_k \) grid. However, additional requirements need to be imposed on the fader functions \( u_N(x) \) and the sequence \( N/N' \) over and above those already given by (1.2.3) and (1.2.4). The definitions of \( u(x) \) and \( u_N(x) \) are modified so that
\[ \sup |u_N(x_1) - u(x_2)| = O(5) \]  \hspace{1cm} (1.6)

whenever \(|x_1 - x_2| < 6\). Moreover, the \(N'\) must satisfy

\[ |N/N' - a| = O(N^{-1}) \]  \hspace{1cm} (1.7)

Now, in almost precisely the same way as we introduced (1.3), we are led to require that

\[ \lim (2\pi)^{-1} \int_{-\infty}^{\infty} \|h(N(\theta/N+\lambda_0)) - h(\theta)\|^2 |q(\theta-2\pi ak)|^2 d\theta = 0 \]  \hspace{1cm} (1.8)

for \(-\frac{3}{2}m' < k \leq \lceil \frac{3}{2}m' \rceil\) with

\[ q(\theta) = \int_0^1 u(x) e^{-ix\theta} dx. \]

This condition implies (1.3) since \(|q(\theta)|^2 = L(\theta)\) when \(u(x)\) is identically unity. Defining the matrix function \(Q(\theta)\) by requiring its typical element to be \(q(\theta-2\pi aj)\overline{q(\theta-2\pi ak)}\), \(-\frac{3}{2}m' < j, k \leq \lceil \frac{3}{2}m' \rceil\), we note that

\[ \|Q(\theta)\| = \sum_j |q(\theta-2\pi aj)|^2 . \]

Hence we may replace (1.8) by the neater condition

\[ \lim (2\pi)^{-1} \int_{-\infty}^{\infty} \|h(N(\theta/N+\lambda_0)) - h(\theta)\|^2 \|Q(\theta)\| d\theta = 0 . \]  \hspace{1cm} (1.9)

If \(h(N(\theta))\) satisfies (1.9) with \(h(\theta)\) described by (1.5) and \(u_N(x)\), \(N'\) satisfy (1.6) and (1.7) respectively, then we shall say that

Condition B holds. Thus, to summarise, these relatively innocuous conditions require \(h(N(\theta))\) to behave as \(h(N(\theta-\lambda_0))\) in the immediate neighbourhood of \(\lambda_0\). In this manner we may preserve the variation that exists across the band since it can now be taken through to the limiting distribution.
2.2 The Central Limit Theorem

We introduce the vector \( w^{(N)} \) which has \( w_{u,j}^{(N)}(\theta_k) \) as the \( ((k + [\frac{1}{2}(m'-1)])p + j)^{th} \) component.

Theorem 2.1

Let \( x(n) \) satisfy Condition A and \( h^{(N)} \) satisfy Condition B. A necessary and sufficient condition that, for \( \lambda_0 \neq 0, \pi \) and all such \( h^{(N)} \), the distribution of the vector \( w^{(N)} \) should approach, as \( N \to \infty \), a \( pm' \)-variate complex multivariate normal distribution with zero mean vector and covariance matrix

\[
(2\pi)^{-1} \int_{-\infty}^{\infty} Q(\theta) \otimes \{h(\theta)^* f(\lambda_0) h(\theta)\} \, d\theta
\]

is that the random variable

\[
X(\lambda_0, N) = \| (2\pi N)^{-\frac{1}{2}} \sum_{1}^{N} x(n) e^{i \lambda_0 n} \|^2
\]

should be uniformly integrable, i.e.

\[
\lim_{\lambda_0 \to \infty} \lim_{N \to \infty} \int_{X(\lambda_0, N) > X} X(\lambda_0, N) \, dP = 0 .
\]

If, for \( \lambda_0 = 0 \), the vector \( w_u^{(N)}(0) \) is excluded from \( w^{(N)} \) and for \( N \) even and \( \lambda_0 = \pi \), the vector \( w_u^{(N)}(\pi) \) is not included in \( w \) then the theorem remains true.

Before giving the proof of this result we state the following corollaries.

Corollary 2.1

Under Conditions A and B the condition given by (2.2) is necessary and sufficient in order that \( w^{(N)} \), having component \( w_j^{(N)}(\theta_k) \), should have an asymptotic complex multivariate normal distribution with covariance matrix (2.1) wherein the typical entry in \( Q(\theta) \) is

\[
sin^2 \frac{1}{2} \theta / ((\frac{1}{2} \theta - \pi j)(\frac{1}{2} \theta - \pi k)).
\]
Corollary 2.2

If \( h^{(N)}(\lambda) \equiv I_p \) then Theorem 2.1 establishes the result of Theorem 1.5.

This second corollary holds since

\[
(2\pi)^{-1} \int_{-\infty}^{\infty} Q_{jk}(\theta) \, d\theta = (2\pi)^{-1} \int_{-\infty}^{\infty} q(\theta - 2\pi aj) \overline{q(\theta - 2\pi ak)} \, d\theta
\]

\[
= \int_{0}^{1} u^2(x) e^{i2\pi ax(j-k)} \, dx
\]

where the last equality follows from Plancherel's Theorem and so, with reference to (1.2.7),

\[
Q = (2\pi)^{-1} \int_{-\infty}^{\infty} Q(\theta) \, d\theta.
\]

In this situation where there is no discernible variation across the band we observe that if the \( \theta_k' \) are replaced by the \( \omega_k'(\lambda_0) \) then the proof is not affected by this change since, essentially, \( f^{(N)}(\omega_k'(\lambda_0)) = f(\omega_k'(\lambda_0)) \) is asymptotically equivalent to \( f(\lambda_0) \). Thus Theorem 1.5 and its corollaries are special cases of Theorem 2.1.

Corollary 2.3

If

\[
x^{(N)}(n) = \sum_{-\infty}^{\infty} A^{(N)}(j) \varepsilon(n-j), \quad \sum_{-\infty}^{\infty} \|A^{(N)}(j)\|^2 < \infty, \quad (2.3)
\]

where the \( \varepsilon(n) \) are independently and identically distributed with covariance matrix \( I_p \) and

\[
h^{(N)}(\lambda) = \sum_{-\infty}^{\infty} A^{(N)}(j) e^{ij\lambda}
\]

satisfies Condition B then the result of Theorem 2.1 holds since \( \| (2\pi N)^{-\frac{1}{2}} \sum_{N} \varepsilon(n) \exp(\in N\lambda_0) \|^2 \) is uniformly integrable. If, in this case, the \( A^{(N)}(j) \) are independent of \( N \) then, by Corollary 2.2, Theorem 1.5 holds.
This corollary merely shows that the central limit theorem is valid for a process of the form (2.3) which is usually called a generalised linear process. This is in a more general form than is usually proposed since now the \( A^{(N)}(j) \) can depend on \( N \).

**Proof of Theorem 2.1**

We first point out that the covariance matrix of \( w^{(N)} \) converges to that of the limiting distribution. Taking \( \lambda_0 \neq 0, \pi \) it is evident that

\[
\delta \left\{ w^{(N)}(\theta_j) w^{(N)}(\theta_k) \right\} = (2\pi)^{-1} \int_{-\pi}^{\pi} f^{(N)}(\theta) \times \left\{ N^{-1} \sum_{s=1}^{N} \sum_{t=1}^{N} u^{(N)}(s) u^{(N)}(t) e^{-is(\theta-\theta_j)} e^{it(\theta-\theta_k)} \right\} d\theta
\]

\[
= (2\pi)^{-1} \int_{-\pi}^{\pi} f^{(N)}(\theta+\lambda_0) q^{(N)}(\theta-\theta_j') q^{(N)}(\theta-\theta_k') d\theta
\]

where

\[
q^{(N)}(\theta) = N^{-\frac{1}{2}} \sum_{n=1}^{N} u^{(N)}(n) e^{-in\theta}
\]

and so

\[
\delta \left\{ w^{(N)} w^{(N)*} \right\} = (2\pi)^{-1} \int_{-\pi}^{\pi} Q^{(N)}(\theta) \otimes f^{(N)}(\theta+\lambda_0) d\theta
\]

with the matrix \( Q^{(N)}(\theta) \) having typical entry \( q^{(N)}(\theta-\omega_j') q^{(N)}(\theta-\omega_k') \), \(-\frac{1}{2}m' < j,k \leq \lfloor \frac{1}{2}m' \rfloor \). Now

\[
N^{\frac{1}{2}} q(N\theta' - 2\pi aj) = N^{\frac{1}{2}} \int_{0}^{1} u(x) e^{-ix(N\theta' - 2\pi aj)} dx
\]

\[
= N^{\frac{1}{2}} \sum_{n=1}^{N} \left\{ \int_{(n-1)/N}^{n/N} (u(x) e^{i2\pi ajx} - u_{N}(n/N) e^{i\omega_j'}) e^{-in\theta} dx \right\} + \left\{ \int_{-\frac{1}{2}}^{0} u_{N}(n/N) e^{-in\theta} dx \right\}
\]

\[
= q^{(N)}(\theta-\omega_j') e^{i\theta/2} \sin\frac{\theta}{2}(\frac{1}{2}) - N^{\frac{1}{2}} \int_{0}^{1} u_{N,j}(x) e^{-iN\theta} dx
\]
where

\[ \varepsilon_{N,j}(x) = \sum_{n=1}^{N} u_{N}(n/N) e^{in\omega_{j}} \chi_{N}(n,x) - u(x) e^{i2\pi ajx} \]

and \( \chi_{N}(n,x) \) assumes the value unity if \( (n-1)/N < x \leq n/N \) and is zero otherwise. Hence we may replace \( q^{(N)}(\theta - \omega_{j}) q^{(N)}(\theta - \omega_{k}) \) in (2.4) by \( N q^{(N_0 - 2\pi aj)} q^{(N_0 - 2\pi ak)} (\frac{1}{2}\theta / \sin \frac{1}{2}\theta)^2 \) since

\[
(2\pi)^{-1} \int_{-\pi}^{\pi} \| f^{(N)}(\theta + \lambda_0) \| \| N^{\frac{1}{2}} \int_{0}^{1} \varepsilon_{N,j}(x) e^{-iN\theta x} dx \|^{2} (\frac{1}{2}\theta / \sin \frac{1}{2}\theta)^2 d\theta \\
\leq \left( \frac{\pi}{2} \right) N^{\frac{1}{2}} \sup_{x} |\varepsilon_{N,j}(x)|^{2} (2\pi)^{-1} \int_{-\pi}^{\pi} \| f^{(N)}(\theta + \lambda_0) \| d\theta \\
= O(N^{-1})
\]

This latter follows from the integrability of \( f^{(N)} \) and the fact that over the interval \( (n-1)/N < x \leq n/N \)

\[ |u(x) e^{i2\pi ajx} - u_{N}(n/N) e^{iN\omega_{j}}| \]

\[ \leq |u(x) - u_{N}(n/N)| + 2|u(x)| |\sin \pi j(ax-n/N')| \\
\leq |u(x) - u_{N}(n/N)| + 2\pi j u(x) |a|x-n/N'| + (n/N) |a-N/N'| \\
= O(N^{-1})
\]

by conditions (1.6) and (1.7). Thus we need now only consider

\[
(2\pi)^{-1} \int_{-\pi}^{\pi} f^{(N)}(\theta + \lambda_0) N q^{(N_0 - 2\pi aj)} q^{(N_0 - 2\pi ak)} (\frac{1}{2}\theta / \sin \frac{1}{2}\theta)^2 d\theta
\]

for \(-\frac{1}{2}m' < j,k \leq \frac{1}{2}m'\). The above may be rewritten in matrix form as

\[
(2\pi)^{-1} \int_{-\frac{\pi}{2N}}^{\frac{\pi}{2N}} Q(\theta) \otimes h^{(N)}(\theta/N+\lambda_0) f(\theta/N+\lambda_0) h^{(N)}(\theta/N+\lambda_0)^* \\
\times (\frac{1}{2}\theta/N)^2 / \sin^2(\frac{1}{2}\theta/N) d\theta \quad (2.5)
\]

Noting that
\[(2\pi)^{-1} \int_{-NT}^{NT} \|Q(\theta)\| \|h^{(N)}(\theta/N + \lambda_0) - h(\theta)\|^2 \|f(\theta/N + \lambda_0)\|^2 (\frac{1}{2}e/N + 1/N^2) \sin^2(\frac{1}{2}e/N) d\theta\]

is dominated by

\[\left(\frac{\pi^2}{4}\right) \sup_{\theta} \|f(\theta)\| (2\pi)^{-1} \int_{-\infty}^{\infty} \|h^{(N)}(\theta/N + \lambda_0) - h(\theta)\|^2 \|Q(\theta)\| d\theta\]

we see, with reference to (1.9), that (2.5) differs from

\[(2\pi)^{-1} \int_{-NT}^{NT} Q(\theta) \otimes h(\theta) f(\theta/N + \lambda_0) h(\theta)^* \left(\frac{1}{2}e/N\right)^2 \sin^2(\frac{1}{2}e/N) d\theta\]

by terms which converge to zero. Since \(f(\theta)\) is continuous at \(\lambda_0\) the above expression converges to the stated limit by a direct application of Lebesgue's dominated convergence theorem.

Now we may establish the necessity since, using \(F_N\) for the distribution function of \(X(X^*, N)\) and \(F\) for its limiting distribution and taking the case \(h^{(N)}(\lambda) = I\) \(\rho\), we have

\[\lim_{N \to \infty} \int_0^\infty x dF_N(x) = \int_0^\infty x dF(x)\]

However, for all \(b \geq 0\), if the theorem holds,

\[\lim_{N \to \infty} \int_0^b x dF_N(x) = \int_0^b x dF(x)\]

and these two relations together imply (2.2).

To establish sufficiency we first observe that, for each \(\epsilon > 0\), we may find matrices \(A_j\) and constants \(v_j\), \(-M \leq j \leq M\), so that

\[\int_{-\infty}^\infty \|h(\theta)\| - \sum_{-M}^M A_j e^{j\theta} \|Q(\theta)\| d\theta < \epsilon\]

This follows from (1.5). (We shall use \(\epsilon\) for a small positive constant, not always the same one.) Next we put \(N_j = \lfloor N v_j \rfloor\) so that

\[\lim_{N \to \infty} N_j/N = v_j\]
and indeed $|N_j/N - v_j| < N^{-1}$. Thus we shall have, for suitable $M, A_j, v_j$ and $N$ sufficiently large,

$$\int_{-\infty}^{\infty} \| h(\theta) - \sum_{-M}^{M} A_j e^{jN_j\theta} \|^2 \| Q(\theta) \| d\theta < \epsilon.$$  

We put

$$\tilde{h}(\theta) = \sum_{-M}^{M} A_j e^{jN_j\theta}, \quad \tilde{h}^{(N)}(\theta) = \sum_{-M}^{M} A_j e^{jN_j\theta},$$

and now observe that the above results imply that

$$(2\pi)^{-1} \left| \int_{-\pi}^{\pi} Q^{(N)}(\theta) \otimes \left\{ \left[ h^{(N)}(\theta + \lambda_0) - \tilde{h}^{(N)}(\theta) \right] \otimes f(\theta + \lambda_0) \left[ h^{(N)}(\theta + \lambda_0) - \tilde{h}^{(N)}(\theta) \right]^* \right\} d\theta \right| < \epsilon.$$  

Hence, putting

$$\tilde{w}^{(N)}(\omega_k^*) = (2\pi)^{-1} \sum_{N} u^{(N)}(n) \tilde{x}^{(N)}(n) e^{in\omega_k^*},$$

$$\tilde{x}^{(N)}(n) = \sum_{-M}^{M} A_j x(n-N_j) e^{j(n-N_j)\lambda_0}$$

and defining $\tilde{w}^{(N)}$ in terms of these as was $w^{(N)}$ in terms of the $w_u^{(N)}(\omega_k^*)$, we have

$$\| \phi(\tilde{w}^{(N)} - \tilde{w}^{(N)}) \tilde{w}^{(N)} - \tilde{w}^{(N)} \| < \epsilon.$$  

Thus we are reduced to establishing the sufficiency of (2.2) for $\tilde{w}^{(N)}$ with covariance matrix (2.1) having $h(\theta)$ replaced by $\tilde{h}(\theta)$.

For this purpose it is sufficient to consider the scalar quantity

$$\zeta_N = \sum_{k} \alpha_k^* \tilde{w}^{(N)}(\omega_k'),$$

where the $\alpha_k$ are arbitrary vectors of complex numbers. Assuming that

$$v_{-M} < v_{-M+1} < \cdots < v_M$$

we consider the set
$I(N) = \{n| -N_M < n \leq N-N_{-M}\}$

where the integers in $I(N)$ lie in an interval whose length does not exceed $N(1 + v_M - v_{-M} + \epsilon)$ and $\epsilon$ may be taken arbitrarily small for $N$ sufficiently large. We decompose $I(N)$ into $4M+1$ subsets by the points of subdivision, $-N_j$, $N-N_j$, $-M \leq j \leq M$. Thus the first of these might be

$$\{n| -N_M < n \leq -N_{M-1}\}, \{n| -N_{M-1} < n \leq N-N_M\}, \ldots,$$

this being the case where $N-N_M$ is to the left of all other points of subdivision other than $-N_M$ and $-N_{M-1}$. We shall call these sets $I_t(N)$ so that $I_1(N), I_2(N)$ are shown just above for the case there described. We now elaborate a technique due to Rosenblatt (1956) and developed by him for the case when all the $N_j$ are identically zero. We choose $a_N$, $b_N$, $c_N$ so that

$$\lim_{N \to \infty} a_N = \lim_{N \to \infty} b_N = \lim_{N \to \infty} c_N = \infty,$$  

$$\lim_{N \to \infty} (b_N / a_N) = 0,$$

and $c_N$ is the integral part of $(N + N_M - N_{-M})/(a_N + b_N)$. We now form the sets

$$S_j = \{n| -N_M + j(a_N + b_N) < n \leq -N_M + (j + 1)a_N + jb_N\},$$

$$0 \leq j < c_N,$$

$$T_j = \{n| -N_M + (j + 1)a_N + jb_N < n \leq -N_M + (j + 1)(a_N + b_N)\},$$

$$0 \leq j < c_N,$$

$$T_{c_N} = \{n| -N_M + c_N(a_N + b_N) < n \leq N - N_M\}.$$

Then

$$I(N) = (\bigcup S_j) \cup (\bigcup T_j).$$
Let us call $S$ the union of all those $S_j$ lying wholly within some subset $I_t$ and let us call $T$ the union of all $T_k$ and those $S_j$ not in $S$. There are $c_N$ sets $S_j$ lying in $S$ where $c_N - 4M \leq \tilde{c}_N \leq c_N$. Furthermore, if a member of $T$ does not lie within some $I_t$ then that interval is subdivided into intervals which do lie within some $I_t$ and, with this added refinement, we see that $T$ can now be considered to be composed of intervals $T'_k$ such that $T'_k$ lies within some $I_t$. Thus $I(N) = S \cup T$ and we may decompose $\zeta_N$ as $\zeta_N(S) + \zeta_N(T)$ where $\zeta_N(S)$ is the sum of $\tilde{c}_N$ components of which a typical one is

$$
\sum_k \alpha_k^* \sum_j^{(r)} A_j \exp\{iN\omega'_k + if_r(\lambda_0 + \omega'_k)\} \tilde{\xi}_{j,r,k}^{(N)} \tag{2.6}
$$

where

$$
\tilde{\xi}_{j,r,k}^{(N)} = (2\pi N)^{-\frac{3}{2}} \sum_{n \in S_r} u^{(N)}(n + N_j) x(n) \exp\{i(n - f_r)(\lambda_0 + \omega'_k)\}
$$

with

$$
f_r = -N_M + r(a_N + b_N)
$$

and $\sum_j^{(r)}$ is over those $j$ such that $(n-N_j) \in S_r$ for some $n$, $1 \leq n \leq N$ and $S_r \subset S$. $\zeta_N(T)$ contains all the other components of $\zeta_N$. Over the set $S_r$ the $u^{(N)}(n + N_j) = u_N(n/N + N_j/N)$ differ from $u(v_j - v_M + (1 + v_M - v_M)r/c_N)$ by $\epsilon$ where $\epsilon$ is arbitrarily small if $N$ is sufficiently large. From this it follows that we may replace (2.6) by

$$
\eta_{N,r} = \sum_k \alpha_k^* \bar{\eta}_{N,r,k} \tag{2.7}
$$

with

$$
a_{r,k}^* = \alpha_k^* \sum_j^{(r)} A_j \exp\{iN\omega'_k + if_r(\lambda_0 + \omega'_k)\} u(v_j - v_M + (1 + v_M - v_M)r/c_N)
$$

and

$$
\bar{\eta}_{N,r,k} = (2\pi N)^{-\frac{3}{2}} \sum_{S_r} x(n) \exp\{i(n - f_r)(\lambda_0 + \omega'_k)\}
$$
The point of (2.7) is that the $\eta_{N,r,k}$ have a distribution independent of $r$. Before going on to discuss the distribution of $\sum \eta_{N,r}$ we first show that $\zeta_N(T)$ is negligible.

Consider

$$
\zeta_N(k)(T) = \sum_{-M}^{M} A_j \exp\{iN\omega_j^k\} \sum_r \chi_{r,j}(2\pi N)^{-\frac{1}{2}} \sum_{n \in T_r} u_{N}(n+N_j) x(n) \exp\{i\theta_k^j\}
$$

where $\chi_{r,j}$ is unity if $T_r \subset \{n| -N_j < n \leq N-N_j\}$ and is zero otherwise.

Then

$$
\zeta_N(T) = \sum_k \alpha_k^* \zeta_N(k)(T)
$$

and we observe that

$$
\delta \left\{ \zeta_N(k)(T) \zeta_N(k)(T)^* \right\} = (2\pi N)^{-1} \int_{-\pi}^{\pi} K_N(k)(\theta) f(\theta+\omega_0) K_N(k)(\theta)^* d\theta
$$

where

$$
K_N(k)(\theta) = \sum_{-M}^{M} A_j \exp\{iN\omega_j^k\} \sum_r \chi_{r,j} \sum_{n \in T_r} u_{N}(n+N_j) \exp\{-i\theta+\omega_j^k\}
$$

Thus

$$
\left\| \delta \left\{ \zeta_N(k)(T) \zeta_N(k)(T)^* \right\} \right\| \leq (2\pi N)^{-1} \sup_{\theta} \|f(\theta)\| \left( \sum_{-M}^{M} |A_j| \right)^2 \sum_r \chi_{r,j} \sum_{n \in T_r} \left( u_{N}(n+N_j) \right)^2
$$

$$
\leq K_M \left\{ \frac{(c_M+1)b_N}{N} + (4M+1) \frac{a_N}{N} \right\}
$$

where $K_M$ is a constant. Since the above converges to zero we thus need only prove the sufficiency of (2.2) for $\sum \eta_{N,r}$.

Let $F_{N,r+1}(z)$ be the distribution function of the complex random variable $\eta_{N,r}$. This is equivalent to the joint distribution of a pair of real random variables but it is simpler to work in complex terms. Neglecting quantities which converge to zero we wish to show that, as $N \to \infty$ and all $z$,
where the star in this context denotes convolution. Once this is established it is sufficient to prove the central limit theorem treating the \( \eta_{N,r} \) as independent. The proof that (2.8) holds hardly differs from that given in Rosenblatt (1956) and also Hannan (1970), Chapter 4, but it is included here for the sake of completeness. We now consider events of the form

\[ E(j,m_j) = \{ m_j \tilde{c}_N^{-2} < \eta_{N,j} \leq (m_j+1) \tilde{c}_N^{-2} \} \]

Then

\[
\sum' P(\bigcap_j E(j,m_j)) \leq P(\sum_j \eta_{N,j} \leq z) \leq \sum'' P(\bigcap_j E(j,m_j)) \quad (2.9)
\]

where \( \sum' \) is over all different sets of \( \tilde{c}_N \) integers \( m_j \) such that

\[
\sum_{j=0}^{\tilde{c}_N^{-1}} (m_j+1)/\tilde{c}_N^2 \leq z
\]

and \( \sum'' \) is over all such sets for which

\[
\sum_{j=0}^{\tilde{c}_N^{-1}} m_j/\tilde{c}_N^2 \leq z.
\]

It is observed that the events \( \bigcap_j E(j,m_j) \) with the integers \( m_j \) varying arbitrarily constitute a subdivision of the whole \( \tilde{c}_N \)-dimensional space into non-intersecting elementary \( \tilde{c}_N \)-cells each having volume \( \tilde{c}_N^{-2} \).

Thus each realisation of \( \eta_{N,j} \) must lie in some \( E(j,m_j) \) and hence, for these particular \( m_j \), \( \sum_j \eta_{N,j} \leq z \) implies \( \sum_j m_j/\tilde{c}_N^2 \leq z \) and so the right-hand inequality of (2.9) holds. The event on the left-hand side implies \( \sum_j \eta_{N,j} \leq \sum (m_j+1)/\tilde{c}_N^2 \) which in turn implies \( \sum_j \eta_{N,j} \leq z \) and so (2.9) holds in its entirety. Now

\[
F_{N,1} * F_{N,2} * ... * F_{N,\tilde{c}_N^{-1}} (z - \tilde{c}_N^{-1}) \leq \sum_j P(E(j,m_j))
\]
and
\[ \sum_{j} \prod_{i} P(E(j,m_{j})) \leq F_{N,1} \cdots F_{N,N} (z + c_{N}^{-1}) \]
since, considering the first inequality, \( \sum \eta_{N,j} \leq z - c_{N}^{-1} \) implies that, for some set of \( m_{j} \) and events \( E(j,m_{j}) \), \( \sum (m_{j}+1)/c_{N}^{2} \leq z \) whereas, with reference to the last inequality, \( \sum \eta_{N,j} \leq \sum (m_{j}+1)/c_{N}^{2} \) implies \( \sum \eta_{N,j} \leq z + c_{N}^{-1} \). It will be shown that the right-hand side of (2.9) may be replaced by \( \sum_{j} \prod_{i} P(E(j,m_{j})) \) and the left-hand side of (2.9) may be similarly replaced by \( \sum_{j} \prod_{i} P(E(j,m_{j})) \).

By a repeated application of the uniform mixing condition we readily establish that
\[ |P(\bigcap_{j} E(j,m_{j})) - \prod_{j} P(E(j,m_{j}))| \leq c_{N} \psi(b_{N}) \]
since the sets \( E(j,m_{j}) \) involve the \( x(n) \) for \( n \in S_{j} \) and these latter are separated from one another by at least \( b_{N} \) time points. Now,
\[ P(\max_{j} |\eta_{N,j}| > \kappa \epsilon^{-1/2} \sqrt{c_{N}}) < \epsilon \]
since
\[ \delta(\max_{j} |\eta_{N,j}|)^{2} \leq \delta(\sum_{j} |\eta_{N,j}|)^{2} \leq c_{N} \kappa^{2} \]
for some suitable \( \kappa \). Thus, breaking up the sample space into the part for which
\[ \max_{j} |\eta_{N,j}| \leq \kappa \epsilon^{-1/2} \sqrt{c_{N}} \]
and its complement we see that
\[ |\sum_{j} P(\bigcap_{j} E(j,m_{j})) - \sum_{j} \prod_{i} P(E(j,m_{j}))| \leq c_{N} (2c_{N}^{5/2} \kappa \epsilon^{-1/2}) c_{N} \psi(b_{N}) + \epsilon \] (2.10)
where \( \Sigma \) can be interpreted as either \( \Sigma' \) or \( \Sigma'' \). If we choose \( c_{N} \) so that
it approaches infinity at a rate slower than \((- \log \psi(b_N))^{\frac{1}{2}}\) it is observed that

\[-(\log \psi(b_N) + \log \tilde{c}_N + (5\tilde{c}_N/2) \log \tilde{c}_N + \tilde{c}_N \log(2\kappa e^{-\frac{1}{2}}))\]

is greater than

\[c_N^2 \left(1 - c_N^{-2}(\log \tilde{c}_N + (5\tilde{c}_N/2) \log \tilde{c}_N + \tilde{c}_N \log(2\kappa e^{-\frac{1}{2}}))\right)\]

and so the right-hand side of (2.10) can be made arbitrarily small for \(c_N\) sufficiently large. Taking \(b_N = \sqrt{N}, a_N = [N + N_M - N_M]/c_N - \sqrt{N}\) and, without loss of generality, \(\psi(x) > (1 + |x|)^{-1}\), we see that \(b_N \to -\infty\),

\[a_N \geq \frac{[N + N_M - N_M]}{\log(1 + \sqrt{N})^{\frac{1}{2}}} - \sqrt{N} - \infty\]

\[a_N/b_N \geq \frac{[N + N_M - N_M]}{N \log(1 + \sqrt{N})^{\frac{1}{2}}} - 1 - \infty\]

and so sequences \(a_N, b_N, c_N\) can be found to satisfy all the imposed requirements. Moreover, we have now established (2.8) and may proceed to the last part of the proof assuming the \(\eta_{N,r}\) to be independent.

From Gnedenko and Kolmogorov (1954, p.103) it is sufficient for asymptotic normality that, for every \(\epsilon > 0\),

\[\lim_{N \to \infty} \sum_{r} \int_{|z| > \epsilon} |z|^2 dF_{N,r}(z) = 0 .\]

However

\[|\eta_{N,r}|^2 \leq (\sum_{k} a_{r,k}^* a_{r,k})(\sum_{N,r,k} \eta_{N,r,k}^*) (\sum_{N,r,k} \eta_{N,r,k})\]

and so

\[\sum_{r} \int_{|z| > \epsilon} |z|^2 dF_{N,r}(z) \leq \sum_{r} (\sum_{k} a_{r,k}^* a_{r,k}) \int_{x > \epsilon} xdG_N(x)\]

where \(G_N(x)\) is the distribution function of \(\sum_{k} \|\eta_{N,r,k}\|^2\) which does not depend on \(r\) and
\[ 0 < \inf \left\{ \frac{c^2}{\sum_{r,k} \| a_{r,k} \|^2} \right\} . \]

Since
\[ \sum_{r,k} a_{r,k} a_{r,k} \leq c^2 (\sum_{k} \| a_k \|^2) (\sum_{j} \| A_j \|^2), \quad c = \sup_x |u(x)| \]
and, moreover,
\[ c_n^{-1} \sum_{r,k} a_{r,k} a_{r,k} \]
is also bounded by a finite constant it is sufficient that, for every \( 0 > \delta \)
\[ \lim_{N \to \infty} \int_{x > \delta} x^d G_N(x) = 0. \quad (2.11) \]

Now, defining
\[ \tilde{\eta}_{N,r} = (2\pi N)^{-\frac{1}{2}} \sum_{S, r} x(n) \exp\{i(n-f_n) \lambda_0\} \]
it follows that
\[ c_n^{-\frac{1}{2}} (\tilde{\eta}_{N,r} - \tilde{\eta}_{N,r}) \]
converges in mean square to zero since
\[ d (\| c_n^{-\frac{1}{2}} (\tilde{\eta}_{N,r} - \tilde{\eta}_{N,r}) \|^2) \leq (\sup_\theta \| f(\theta) \|) c_n N^{-1} \sum_{n=0}^{a_n} \| e^{-i \omega k} - 1 \|^2 \]
and this latter expression converges to zero. Thus (2.11) may be replaced by
\[ \lim_{N \to \infty} \int_{z > \delta} z^d H_N(z) = 0 \]
where \( H_N(z) \) is the distribution function of \( \| \tilde{\eta}_{N,r} \|^2 \). Noting that \( N/c_n \)
is asymptotically equivalent to \( a_n \), the number of terms in the sum defining \( \tilde{\eta}_{N,r} \), we obtain the condition
\[ \lim_{N \to \infty} \int_{z > \delta} z^d H_N(z) = 0 \quad (2.12) \]
where $H_N(z)$ is the distribution function of

$$ \|(2\pi a_N)^{-\frac{1}{2}} \sum_{S_r} x(n) e^{in\lambda_0}\|^2 $$

and (2.12) is implied by (2.2) so that the proof of the theorem has now been established.

Remarks

The condition (2.2) is implied by conditions on higher moments of the $x(n)$. Thus we might require only that $x(n)$ be stationary to the fourth order (rather than strictly stationary) and that the fourth cumulant between $x_i(m)$, $x_j(m+n)$, $x_k(m+p)$, $x_l(m+q)$, which we might call $\kappa_{ijkl}(n,p,q)$, satisfy

$$ \sum_{n,p,q=0}^{\infty} |\kappa_{ijkl}(n,p,q)| < \infty. $$

In this case it is possible to show that

$$ \sum_r \delta(\eta_N, r \eta^4) $$

converges to zero as $N$ tends to infinity and so, with reference to the last section of the proof of Theorem 2.1, a Liapounov condition holds and thus the conclusion of Theorem 2.1 is established.

It will no doubt also be possible, following for example Ibragimov and Linnik (1971), Chapter 18, to replace (2.2) by conditions relating to the rate at which (1.2.6) decreases to zero.

We now examine (2.1). Let us define the matrix

$$ \Psi_u(v) = \Phi_u(v) \Lambda_u(v) $$

by requiring the typical elements of $\Phi_u(v)$ and $\Lambda_u(v)$ to be given for $-\frac{1}{2}m' < j,k \leq \frac{1}{2}m'$ by
\[
\{\Phi_{u(v)}\}_{j,k} = \int_{-\min(o,v)}^{1-\max(o,v)} u(\theta) u(\theta+v) e^{i2\pi a(j-k)\theta} d\theta
\]

and

\[
\{\Lambda_{u(v)}\}_{j,k} = \begin{cases} 
\exp[-i2\pi ajv] & (j = k) \\
0 & (j \neq k)
\end{cases}
\]

Then a typical sub-matrix of (2.1) is given by

\[
(2\pi)^{-1} \int_{-\infty}^{\infty} \{q(\theta-2\pi aj) h(\theta)\} f(\lambda_0) \{q(\theta-2\pi ak) h(\theta)\}^* d\theta
\]

and, noting that

\[
q(\theta-2\pi aj) h(\theta) = \int_{-1}^{1} e^{-i y(\theta-2\pi aj)} \int_{-\min(o,y)}^{1-\max(o,y)} u(x+y) e^{i2\pi ajx} dA(x) dy
\]

we may apply Plancherel's Theorem to obtain the matrix \( \Gamma_u \) in place of (2.1) where

\[
\Gamma_u = \int_{0}^{1} \int_{0}^{1} \psi_u(y-x) \otimes dA(x) f(\lambda_0) dA(y)^* \quad \text{.} \quad (2.14)
\]
3.1 Introduction

In §1.3 an example is given of a narrow band situation where, if phase is varying rapidly, the usual estimates of coherence given by (1.3.2) can be badly biased downwards. We shall, in this chapter, confine ourselves to this problem and shall examine the simple case where the vector of observations consists of two components \( x_1(n) \) and \( x_2(n) \) with spectral density matrix given by

\[
\begin{bmatrix}
    f_1(\lambda) & \sigma(\lambda) \left[ f_1(\lambda) f_2(\lambda) \right]^1 e^{-i\theta(\lambda)} \\
    \sigma(\lambda) \left[ f_1(\lambda) f_2(\lambda) \right]^1 e^{i\theta(\lambda)} & f_2(\lambda)
\end{bmatrix}
\]

We omit the subscripts on \( \sigma(\lambda) \) and \( \theta(\lambda) \) since they are redundant in this two-dimensional representation.

Given that we know that \( \theta(\lambda) \) is varying rapidly near the chosen frequency \( \lambda_0 \) we must select some appropriate functional relationship that adequately reflects this particular variation. A first case that arises quite naturally is that where we express \( \theta(\lambda) \) in the form of a truncated Taylor series so that, near \( \lambda_0 \),

\[
\theta(\lambda) = \theta(\lambda_0) + (\lambda - \lambda_0) \theta'(\lambda_0)
\]

(1.1)

and here \( \theta'(\lambda_0) \) is the derivative of \( \theta(\lambda) \) with respect to frequency evaluated at \( \lambda_0 \). Hence we are making the hypothesis that the phase change is near to linear over the frequencies involved. The quantity \( \theta'(\lambda_0) \) is called the group delay at \( \lambda_0 \) and seems to be a primary parameter of interest in a cross-spectral analysis. Consider, for example,
the situation where \((x_1(n), x_2(n))'\) has cross-spectral density such that over the narrow band of frequencies about \(\lambda_0\) the phase is effectively constant. If, over the narrow band, the second component were now to occur with lag \(\ell\) relative to the first component then the phase of the observed process near \(\lambda_0\) will be, approximately, \(\vartheta(\lambda_0) + \ell \lambda\). In this case the group delay, when estimated, would be near to \(\ell\). Thus we can interpret the group delay as the delay to which one component has been subjected to, relative to the other, over the band of frequencies near \(\lambda_0\). One can envisage situations where a pair of recorders are receiving a narrow band signal, together with noise, from some distant source. With reference to the diagram where \(R_1\) and \(R_2\) denote the two recorders and \(V\) represents the speed of propagation of the signal we see that the measurements at \(R_1\) will lag those at \(R_2\) by \(d \cos \phi / V\). Thus the group delay, when estimated, will give the direction of the signal provided we know both \(d\) and \(V\). In this last example the phase variation will, of course, be expected, but it may also be true that, unbeknown to the experimenter, one series has been lagged or rephased relative to the other due to unforeseen circumstances. Phase changes such as these can occur in many possible ways: a recording apparatus may incorporate a defect which rephases the incoming signal or there may be, perhaps, some inconsistency in the medium through which the vector signal travels whose net result is to make the propagation path of one component series longer than that of the other. In any event the complexity of the real world is such that this type of situation occurs only too frequently. Thus, although the problem is motivated in terms of the
need to estimate coherence we see that this will involve the estimation of group delay and it is this quantity which is the more interesting physically.

If we take the group delay to be of order $N$ it is evident that we may then apply Theorem 2.1 since, near $\lambda_0$, the spectral density of the observed process can be described as

$$h^{(N)}(\lambda) f(\lambda_0) h^{(N)}(\lambda)^*$$

where

$$h^{(N)}(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & \exp[i(\lambda-\lambda_0) \theta'(\lambda_0)] \end{bmatrix}$$

and $\theta'(\lambda_0)/N$ will be required to converge, as $N$ increases, to some value $\nu$, $|\nu| < 1$. Thus, in the sense of Theorem 2.1, we require $h^{(N)}(\theta/N + \lambda_0)$ to converge to

$$h(\theta) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\nu\theta} \end{bmatrix}.$$

We are effectively taking $\theta'(\lambda_0) = N\nu$ and are setting out to estimate $\nu$ and the other four spectral parameters via Theorem 2.1. This theorem is an appropriate choice since it retains the actual variation present across the band in its asymptotic form whereas Theorem 1.5, for instance, cannot. We are concerned therefore with the $m$ frequencies of the form $\theta_k$, for example, and the situation where (1.1) provides the discernible variation required by which we infer that

$$\frac{2\pi m}{N} \theta'(\lambda_0) \approx 2\pi m \nu$$

is not small. If we were not considering the estimation of coherence when group delay is large we would not, of course, have needed to phrase the problem in the way described in this introduction since,
clearly, the estimation of group delay can be considered whether it is large or not. However it seems likely that the estimation procedure described in this chapter will be a valid one even when \( \theta'(\lambda_0) \) is not large.

Selecting an appropriate form of Theorem 2.1 we now maximise this likelihood with respect to the various parameters. The exact distribution theory in relation to Theorem 2.1 is difficult to determine and indeed the likelihood equations are highly nonlinear. We thus proceed to find the asymptotic distribution of these statistics as \( m \) increases, assuming the asymptotic (with \( N \)) form of the likelihood for the \( w_j(\theta_k) \) to be the true likelihood. It will be evident that the resulting distribution, for \( m \) sufficiently large and \( N \) sufficiently large to a degree dependent on \( m \), will be an arbitrarily good approximation to the true distribution of the statistics introduced. Of course the method gives no precise indication of the size of \( N \) needed, or the way \( m \) and \( N \) are to be related. A problem arises here due to the fact that the asymptotic theory is developed in terms of \( v \) rather than \( \theta'(\lambda_0) \). Since the variance of the maximum likelihood estimate for \( v \) turns out to be \( O(m^{-3}) \) this yields a variance of \( O(N^2/m^3) \) for the corresponding estimate for the group delay. One would like this to approach zero as \( N \) becomes large if confidence intervals involving this latter quantity are, asymptotically, to decrease indefinitely, but, as yet, this has not been shown. However, in practice, such information will be of limited use since only one value of \( m \) and one value of \( N \) are experienced. In the final analysis it will, as always, be experience and familiarity with both the techniques used and the situation under study that will determine what values of \( m \) and \( N \) yield a good approximation to the actual distribution of the \( w_j(\theta_k) \). To that end the numerical studies in §3.5 should give some guide-lines. We note that if we now
consider the grid $\theta_k$ with reference to this situation where $m$ is also being increased, at a slower rate than $N$, then the variation in $\theta(\lambda)$ across the band will be consistent with an insignificant variation across a band of width $\pi/N$. Thus, for $m$ large, the $\theta_k$ may effectively be replaced by the more usual $\omega_k(\lambda)$. 

We now describe the particular distribution arising from Theorem 2.1 that is to be studied. This will be given for the finite Fourier transform case rather than the Cooley-Tukey technique since the additional complexity of the latter only serves to confuse the situation and make the results a little less perspicuous. It would, anyway, be a relatively simple matter to extend the results given to include this situation, but this has not been done here. In fact, throughout the remainder of this thesis, we shall consider only the case of the finite Fourier transform unless otherwise specified. In the situation considered $h(\lambda) f(\lambda_0) h(\lambda)^*$ is given by

$$
\begin{bmatrix}
  f_1 & \sigma \sqrt{f_1 f_2} \ e^{-i(\theta + v \lambda)} \\
  \sigma \sqrt{f_1 f_2} \ e^{i(\theta + v \lambda)} & f_2
\end{bmatrix}
$$

where the argument variable $\lambda_0$ has been omitted since this is held fixed. Now, with reference to Theorem 2.1, we may describe the form of the covariance matrix for the asymptotic distribution of the $w_j(\theta_k)$, $j = 1, 2, -\frac{1}{2}m < k \leq \frac{1}{2}m$. Taking $m = 2n+1$ for convenience we put

$$
\Gamma = (2\pi)^{-1} \int_{-\infty}^{\infty} h(\lambda) f(\lambda_0) h(\lambda)^* \otimes Q(\lambda) \ d\lambda
$$

and note that this is just

$$
\begin{bmatrix}
  f_1 I_m & \sigma \sqrt{f_1 f_2} \ e^{-i \theta} \Phi(v) \\
  \sigma \sqrt{f_1 f_2} \ e^{i \theta} \Phi(v)^* & f_2 I_m
\end{bmatrix}
$$

where $\Phi(v) = \Phi(v) \Lambda(v)$, $|v| < 1$. This follows from (2.2.14) and we
note that these latter two factors that constitute $T(v)$ have typical elements that can be described as
\[ \phi_{jk}(v) = \int_{-\min(0,v)}^{1-\max(0,v)} e^{i2\pi(j-k)\theta} d\theta \]  
(1.2)
and
\[ \Lambda_{jk}(v) = \begin{cases} \exp(-i2\pi kv) & (j = k) \\ 0 & (j \neq k) \end{cases} \]  
(1.3)
for $-n \leq j, k \leq n$. Thus if $w$ now denotes the vector having $w_j(\theta_k)$ as the $((j-l)m + k + n + 1)^{th}$ component then its asymptotic distribution is given by (1.2.5) with $A = \Gamma$ and can be written as
\[ \frac{1}{\pi^{2m} \det(\Gamma)} \exp(-w^* \Gamma^{-1} w). \]  
(1.4)
We have altered the previous definition of $w$ for the sake of presentation and trust that no confusion will result. Hence (1.4) is the initial likelihood on which the estimation procedure will be based.

Before going on to describe this procedure we point out that the approach is related to, but distinct from, that of Jones (1969). This author commences by regarding the $w_j(\omega_k(\lambda_0))$, $j = 1, 2$, $-\frac{1}{2}m < k \leq \lceil \frac{1}{2}m \rceil$, as having an asymptotic distribution given by Corollary 1.1 of Theorem 1.5. Thus he is essentially considering the standard situation where the spectral density may be regarded as effectively constant over the band of frequencies in question. He then considers this band to be composed of sub-bands, each containing a certain specified number of frequencies, and, over each sub-band, he forms the sum of products matrix given by
\[
\begin{bmatrix}
\sum_k w_1(\omega_k(\lambda_0))^2 & \sum_k w_1(\omega_k(\lambda_0)) w_2(\omega_k(\lambda_0)) \\
\sum_k w_2(\omega_k(\lambda_0)) w_1(\omega_k(\lambda_0)) & \sum_k |w_2(\omega_k(\lambda_0))|^2
\end{bmatrix}
\]  
(1.5)
where the summation is over the frequencies of the sub-band in question. The entries of this matrix have a complex Wishart distribution, in so far as the original distribution pertains, and at this stage the random variable representing the argument of the complex quantity

\[ \sum_{k} w_1(\omega_k(\lambda_0)) \overline{w_2(\omega_k(\lambda_0))} \]

is integrated out to yield a marginal distribution involving only those random variables given by the moduli of the distinct elements in (1.5). With reference to (1.3.2) it is evident that the random variable that has been integrated out is just the usual sample estimate of the phase evaluated over the frequencies of the sub-band considered. These marginal distributions are now combined to form a likelihood which is then maximised with respect to the two spectra and the coherence. Thus this procedure gives an alternative to the standard method for estimating coherence given by (1.3.2). The rationale behind this approach appears to be that of providing an estimation procedure that is relatively insensitive to phase change in that should phase variation now occur across the band of frequencies considered then, by judiciously selecting the various sub-bands, better estimates of coherence will result. However this does not seem at all obvious. Indeed, in the light of Theorem 2.1, it seems unlikely that the original independence assumption will still hold when phase is varying across the band. Moreover, this method gives no indication of the underlying phase structure involved. It is noted, however, that if the variation across the band was in fact due to the group delay and this latter was small then, considering the matrix \( \Gamma \), it is evident that the smaller \( v \) the more closely \( \Psi(v) \) approaches \( \Lambda(v) \) and the more independent the \( w(\omega_k(\lambda_0)) \) will become. In this sense Jones' results may hold in the case where \( v \), or rather \( mv \), is small.
3.2 The Maximum Likelihood Equations and Some Fundamental Lemmas

We shall put

\[ A(\sigma^2, v) = \left(I_m - \sigma^2 \phi(v)\right)^{-1}, \]

\[ B(v) = \frac{d}{dv} \{\phi(v)\}, \]

\[ C(\sigma^2, v) = \frac{d}{dv} \{\psi(v)\} + \sigma^2 B(v) A(\sigma^2, v) \psi(v) \]

and

\[ w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \]

where the vector \( w_1 \) contains the m quantities \( w_1(\theta_k) \) and similarly for \( w_2 \). With these considerations the logarithm of the likelihood function (1.4) can be written as

\[
L = -2m \log \pi - m \log f_1 - m \log f_2 + \log \det \{A(\sigma^2, v)\} - f_1^{-1} w_1^* A(\sigma^2, v) w_1
- \frac{f_2^{-1}}{2} w_2^* A'(\sigma^2, v) w_2 + \sigma f_1 f_2^{-1} \Re\{e^{-i\theta} w_1^* A(\sigma^2, v) \psi(v) w_2\}. \tag{2.1}
\]

The partial derivatives of \( L \) with respect to the various parameters are given by

\[
\frac{\partial L}{\partial f_1} = -\frac{m}{f_1} + \frac{1}{f_1 f_2} w_1^* A(\sigma^2, v) w_1 - \frac{\sigma}{f_1(f_1 f_2)^{\frac{1}{2}}} \Re\{e^{-i\theta} w_1^* A(\sigma^2, v) \psi(v) w_2\}, \tag{2.2}
\]

\[
\frac{\partial L}{\partial f_2} = -\frac{m}{f_2} + \frac{1}{f_2 f_1} w_2^* A'(\sigma^2, v) w_2 - \frac{\sigma}{f_2(f_1 f_2)^{\frac{1}{2}}} \Re\{e^{-i\theta} w_1^* A(\sigma^2, v) \psi(v) w_2\}, \tag{2.3}
\]

\[
\frac{\partial L}{\partial \theta} = 2\sigma(f_1 f_2)^{-\frac{1}{2}} f_1 e^{-i\theta} w_1^* A(\sigma^2, v) \psi(v) w_2, \tag{2.4}
\]

\[
\frac{\partial L}{\partial \sigma^2} = 2\sigma \text{ tr}[A(\sigma^2, v) \phi(v)] - \frac{2\sigma}{f_1} w_1^* A^2(\sigma^2, v) \phi(v) w_1
- \frac{2\sigma}{f_2} w_2^* \phi A'(\sigma^2, v) w_2 - (f_1 f_2)^{-\frac{1}{2}} 2\Re\{e^{-i\theta} w_1^* A(\sigma^2, v)(I - 2A(\sigma^2, v)) \psi(v) w_2\}. \tag{2.5}
\]
\[ \frac{\partial L}{\partial \nu} = \sigma^2 \text{tr}[A(\sigma^2, \nu) B(\nu)] - \frac{\sigma^2}{f_1} w_1^* A(\sigma^2, \nu) B(\nu) A(\sigma^2, \nu) w_1 \]
\[ - \frac{\sigma^2}{f_2} w_2^* A'(\sigma^2, \nu) B'(\nu) A'(\sigma^2, \nu) w_2 \]
\[ + \sigma(f_1 f_2)^{-\frac{1}{2}} 2\Re\{e^{-i\theta} w_1^* A(\sigma^2, \nu) C(\sigma^2, \nu) w_2\} \] (2.6)

where by tr(A) we mean the trace of the matrix A. Setting these to zero it is readily seen that \( \hat{\nu} \) and \( \hat{\sigma} \), the maximum likelihood estimates of \( \nu \) and \( \sigma \), make det(I) a minimum where \( f_1 \) and \( f_2 \) are now functions of \( \sigma^2 \), \( \nu \) and the data. Also, the likelihood equations (2.2), (2.3) and (2.4) imply that

\[ f_1^{-1} w_1^* A(\sigma^2, \nu) w_1 = f_2^{-1} w_2^* A'(\sigma^2, \nu) w_2 \] (2.7)

\[ f_1^{-1} w_1^* A(\sigma^2, \nu) w_1 = m + \sigma(f_1 f_2)^{-\frac{1}{2}} |w_1^* A(\sigma^2, \nu) \Psi(\nu) w_2| \] (2.8)

\[ \theta = \arctan\{f[w_1^* A(\sigma^2, \nu) \Psi(\nu) w_2]/R[w_1^* A(\sigma^2, \nu) \Psi(\nu) w_2]\}. \] (2.9)

The asymptotic properties of the solutions of equations (2.5) to (2.9) as \( m \) increases are not self evident; since the situation considered cannot be dealt with by conventional maximum likelihood theorems. To establish these properties the following lemmas are proved.

**Lemma 3.1**

Let the matrices \( H_j, j = 1, \ldots, q \), be given by

\[ H_j = \int_0^1 g_j(\theta) U(\theta) U(\theta)^* d\theta \] (2.10)

where \( U(\theta) \) is an \( m \)-dimensional vector with typical element \( \exp(i2\pi j\theta) \), \(-\frac{1}{2}m < j \leq \lceil \frac{1}{2}m \rceil \) and the \( g_j(\theta) \) are non constant functions given by

\[ g_j(\theta) = \sum_{k=0}^{r} g_k^{(j)} \chi(\theta_{E_k}), \quad \max_k |g_k^{(j)}| < \infty , \]

with
\[ \chi_{\theta}(E_k) = \begin{cases} 1 & (\theta \in E_k) \\ 0 & (\text{otherwise}) \end{cases} \]

and the intervals \( E_k \) are disjoint with

\[ E_k = \{ \theta | \alpha_k < \theta \leq \alpha_{k+1} \}, \quad \bigcup_{k=0}^{r} E_k = (0,1] . \]

Then

\[ m^{-1} \text{tr} \left\{ \prod_{j=1}^{q} H_j - H \right\} = O(m^{-1} \log m) \]

where

\[ H = \int_0^1 g(\theta) U(\theta) U(\theta)^* \, d\theta \]

and

\[ g(\theta) = \sum_{k=0}^{r} \left\{ \prod_{j=1}^{q} g_k^{(j)} \right\} \chi_{\theta}(E_k) \]

\[ = \prod_{j=1}^{q} g_j(\theta) . \]

**Proof**

The matrix \( H_j \) can be described as

\[ H_j = \sum_{k=0}^{r} g_k^{(j)} \Phi(E_k) \]

where

\[ \Phi(E_k) = \int_0^1 \chi_{\theta}(E_k) U(\theta) U(\theta)^* \, d\theta . \]

Hence

\[ m^{-1} \text{tr} \left\{ \prod_{j=1}^{q} H_j - H \right\} = \sum_{k=0}^{r} \left( g_k^{(1)} \ldots g_k^{(q)} \right) m^{-1} \text{tr} \left\{ \Phi(E_k) - \Phi(E_k) \right\} \]

\[ + \sum_{k_1, \ldots, k_q}^{\prime} \left( g_{k_1}^{(1)} \ldots g_{k_q}^{(q)} \right) m^{-1} \text{tr} \left\{ \Phi(E_{k_1}) \ldots \Phi(E_{k_q}) \right\} - (2.11) \]
and $\sum'$ is over all integers $k_i$, $0 \leq k_i \leq r$, except those for which 
$k_i = k$, $i = 1, \ldots, q$, and $k$ is an integer that lies between 0 and $r$.

We shall only consider the case where $m = 2n+1$ since the case 
for $m$ even is established in an identical manner. Now

$$m^{-1} \text{tr} \left\{ \Phi(E_j) \Phi(E_k) \right\} = \int_0^1 \int_0^1 \chi_{\Theta}(E_j) \chi_{\Phi}(E_k) \frac{\sin^2 m\pi(\Theta-\Phi)}{m \sin^2(\Theta-\Phi)} d\Theta d\Phi$$

$$= \int_0^1 \frac{\sin^2 m\pi\Theta}{m \sin^2(\Theta)} \int_0^1 \chi_{\Theta+\Phi}(E_j) \chi_{\Phi}(E_k) d\Phi d\Theta$$

$$= (2\pi)^{-1} \int_0^{2\pi} f_{jk}(\Theta) L_m(\Theta) d\Theta \quad (2.12)$$

where $L_m(\Theta)$ is given by (2.1.1) and the functions $\chi_{\Phi}(E_j)$ are defined 
outside $(0,1]$ by periodicity. It is noted that (2.12) represents the 
Césaro sum evaluated at the zero frequency of the function

$$f_{jk}(\Theta) = (2\pi)^{-1} \int_0^{2\pi} \chi_{(\Theta+\Phi)/(2\pi)}(E_j) \chi_{/(2\pi)}(E_k) d\Phi .$$

In the case $j > k$, $\alpha_{j+1} - \alpha_j \leq \alpha_{k+1} - \alpha_k$ this function may be described as

$$f_{jk}(\Theta) = \begin{cases} 
0 & \{ 0 \leq \Theta \leq 2\pi(\alpha_j - \alpha_{k+1}) \} , \\
\alpha_{k+1} - \alpha_j + \frac{\Theta}{2\pi} & \{ 2\pi(\alpha_j - \alpha_{k+1}) \leq \Theta \leq 2\pi(\alpha_{j+1} - \alpha_k) \} , \\
\alpha_{j+1} - \alpha_j & \{ 2\pi(\alpha_{j+1} - \alpha_k) \leq \Theta \leq 2\pi(\alpha_{j+1} - \alpha_k) \} , \\
\alpha_{j+1} - \alpha_k - \frac{\Theta}{2\pi} & \{ 2\pi(\alpha_j - \alpha_k) \leq \Theta \leq 2\pi(\alpha_{j+1} - \alpha_k) \} , \\
0 & \{ 2\pi(\alpha_j - \alpha_k) \leq \Theta \leq 2\pi \} , 
\end{cases}$$

and hence is continuous and satisfies a Lipschitz condition of order 
one, which is to say

$$\sup |f_{jk}(\lambda_1) - f_{jk}(\lambda_2)| = O(\delta)$$
whenever

$$|\lambda_1 - \lambda_2| < \delta \pmod{2\pi}.$$  

The other possible forms for $f_{jk}(\theta)$ yield similar results and so, from standard Fourier theory (Zygmund, 1959, p. 91)

$$(2\pi)^{-1} \int_0^{2\pi} f_{jk}(\theta) L_m(\theta) \, d\theta = f_{jk}(0) + O(m^{-1} \log m)$$

and it is observed that $f_{jk}(0)$ is zero for $j \neq k$ and for $j = k$ is given by $m^{-1} \text{tr} \{\Phi(E_j)\} = \alpha_{j+1} - \alpha_j$.

Considering the Rayleigh quotient

$$z^* \Phi(E_k) z \overline{z^* z} = \int_0^1 \chi_{\Phi(E_k)} z^* U(\theta) U(\theta)^* z \, d\theta$$

it can be seen that the eigenvalues of $\Phi(E_k)$ are bounded above by unity and below by zero. Thus, with reference to the first summation in (2.11)

$$|m^{-1} \text{tr} \Phi^q(E_k) - \Phi(E_k)| = m^{-1} \sum_{j=1}^{q-1} \text{tr} \Phi^j(E_k) - \Phi^{j+1}(E_k)$$

$$\leq (q-1) m^{-1} \text{tr} \Phi(E_k) - \Phi^2(E_k)$$

$$= O(m^{-1} \log m).$$

We note that the eigenvalues of any arbitrary product of $\Phi$ matrices are bounded above by unity and so, turning to the second term of (2.11), it is evident that it will be sufficient to consider

$$m^{-1} \text{tr} \Phi(E_j) \Phi(E_k) \Phi(E_\ell) \ldots, \quad j \neq k, \ k \neq \ell \quad (2.13)$$

for, if this is not so, we may replace $\Phi^2(E_j)$ by $\Phi(E_j)$ since
\[ m^{-1} \text{tr}([\Phi(E_j) - \Phi^2(E_j)]^{\frac{1}{2}} \Phi(E_j) \ldots [\Phi(E_j) - \Phi^2(E_j)]^{\frac{1}{2}}) \leq m^{-1} \text{tr}(\Phi(E_j) - \Phi^2(E_j)) = O(m^{-1} \log m). \]

Hence this process can be continued until either a result is attained or a situation of the form (2.13) eventuates. Now, assuming (2.13) to be the relevant form and denoting the product of the matrices other than \( \Phi(E_j), \Phi(E_k) \) and \( \Phi(E_{\ell}) \) by \( \Phi \) it can be seen that

\[ m^{-1} \text{tr}(\Phi(E_j) \Phi(E_k) \Phi(E_{\ell}) \Phi) \]

\[ \leq [m^{-1} \text{tr}(\Phi(E_j) \Phi(E_k))]^{\frac{1}{2}} [m^{-1} \text{tr}(\Phi^3(E_k) \Phi(E_{\ell}) \Phi(E_j) \Phi^3(E_k))]^{\frac{1}{2}} \]

\[ \leq [m^{-1} \text{tr}(\Phi(E_j) \Phi(E_k))]^{\frac{1}{2}} [m^{-1} \text{tr}(\Phi(E_k) \Phi(E_{\ell}))]^\frac{1}{2} \]

since

\[ \frac{z^* \Phi^3(E_k) \Phi^2(E_{\ell}) \Phi(E_j) \Phi^3(E_k) \Phi(E_{\ell}) \Phi^3(E_k) z}{z^* z} \leq \frac{z^* \Phi^3(E_k) \Phi^2(E_{\ell}) \Phi^3(E_k) \Phi(E_{\ell}) \Phi^3(E_k) z}{z^* z} \]

and so the lemma has now been established.

**Corollary 3.1**

Let \( \Phi(v) \) be given by (1.2) for \( |v| \leq 1 \). Then

\[ |m^{-1} \text{tr}(\Phi^q(v)) - (1 - |v|)| = O(m^{-1} \log m) \]

uniformly in \( v \).

**Proof**

Since

\[ \Phi(v) = \int_{-\min(0,v)}^{1-\max(0,v)} U(\theta) U(\theta)^* d\theta \]

the result follows by application of Lemma 3.1.
Lemma 3.2

Let \( H \) be an Hermitian matrix of the form given by (2.10) with

\[
H = \int_0^1 g(\theta)\, U(\theta)\, U(\theta)^*\, d\theta ,
\]

\[
g(\theta) = \sum_{j=0}^{r} g_j \chi_{\theta}(E_j)
\]

and \( \max_j |g_j| = c < \infty \). Then, if \( R(z) = \sum_{k=1}^{\infty} a_k z^k \) and \( \sum_{k=1}^{\infty} |a_k| c^k < \infty \),

\[
m^{-1} \text{tr} \{ R(H) - H_R \} = O(m^{-1} \log m)
\]

where

\[
H_R = \int_0^1 R[g(\theta)]\, U(\theta)\, U(\theta)^*\, d\theta .
\]

Proof

The eigenvalues of \( H^k \) and

\[
\int_0^1 g^k(\theta)\, U(\theta)\, U(\theta)^*\, d\theta
\]

are bounded in modulus by \( c^k \). Thus, for arbitrary \( \varepsilon > 0 \), there exists an integer \( M \) such that \( N > M \)

\[
\left| \sum_{k=N}^{\infty} a_k m^{-1} \text{tr} \{ H^k - \int_0^1 g^k(\theta)\, U(\theta)\, U(\theta)^*\, d\theta \} \right| \leq 2 \sum_{k=N}^{\infty} |a_k| c^k < \varepsilon .
\]

The fact that the finite sum

\[
\sum_{k=1}^{N-1} a_k m^{-1} \text{tr} \{ H^k - \int_0^1 g^k(\theta)\, U(\theta)\, U(\theta)^*\, d\theta \} = O(m^{-1} \log m)
\]

follows from Lemma 3.1 and so completes the proof.

Corollary 3.2

Let \( R(z) = \sum_{k=1}^{\infty} a_k(z^k - z) \), \( \sum_{k=1}^{\infty} |a_k| < \infty \) and let \( \Phi(v) \) be given by (1.2) for \( |v| \leq 1 \). Then

\[
|m^{-1} \text{tr} \{ R[\Phi(v)] \}| = O(m^{-1} \log m)
\]
uniformly in \( v \).

As an example of the use of the previous lemmas and in particular Corollary 3.2 we observe that for \( \sigma ^2 \leq 1 - \delta \), \( \delta > 0 \),

\[
m^{-1} \log \det(A(\sigma ^2, v)) = \sum_{k=1}^{\infty} \frac{2k}{k} m^{-1} \text{tr}(\Phi ^{2k}(v)) = -(1 - |v|) \log(1 - \sigma ^2) + O(m^{-1} \log m)
\]

and this result is uniform in \( v \) and \( \sigma ^2 \). It is noted that matrices of the type given by (2.10) are examples of Toeplitz matrices. (See Grenander and Szegö (1958).)

We conclude this section with the statement and proof of a certain inequality.

**Lemma 3.3**

Let

\[
x_m = m^{-1} \sum_{j=1}^{m} \lambda_{m,j} |\xi_j|^2 / \left( \prod_{j=1}^{m} \lambda_{m,j} \right)^{1/m}
\]

where the \( \xi_j \) are independent complex random variables with zero means and unit variances. Then

\[
\lim x_m \geq 1 \quad \text{a.s.}
\]

**Proof**

Now

\[
P(x_m < 1) = \int_{x_m < 1} \ldots \int x_m < 1 \exp \left( - \sum_{j=1}^{m} |\xi_j|^2 \right) \prod_{j=1}^{m} d\xi_j
\]

where by \( \prod_{j=1}^{m} d\xi_j \) we mean the differential form of the real and imaginary parts of the \( \xi_j \). It is easily verified that this represents an integration over an ellipsoid having the same volume as the sphere \( m^{-1} \sum_{j=1}^{m} |\xi_j|^2 < 1 \). Since the density function being integrated is spherically symmetrical and monotonically decreasing with the distance
from the origin, the integral is maximised for the spherical region. Thus

\[ P(x_m < 1) \leq P(m^{-1} \sum |\xi_j|^2 < 1) \]

and by Fatou's lemma we have

\[ P(\lim x_m < 1) \leq \lim P(x_m < 1) \leq \lim P(m^{-1} \sum |\xi_j|^2 < 1) . \]

However \( m^{-1} \sum |\xi_j|^2 \) converges to unity almost surely and hence

\[ P(\lim x_m \geq 1) = 1 \]

and the result is established.

3.3 A Strong Law of Large Numbers

It will now be convenient to indicate the row vector \((f_1, f_2, \theta, \sigma, \nu)\) by \(\tau^1\) and we put \(\tau' = (\mu', \nu')\) where \(\mu' = (f'_1, f'_2, \theta)\) and \(\nu' = (\sigma, \nu)\). The vector \(\hat{\tau}_m\) will denote the value of \(\tau\) which maximises (2.1) and so \(\hat{\tau}'_m = (\hat{\mu}'_m, \hat{\nu}'_m)\) with \(\hat{\mu}'_m = (\hat{\mu}'_1, m', \hat{\mu}'_2, m', \hat{\mu}'_m)\) and \(\hat{\nu}'_m = (\hat{\nu}'_m, \hat{\nu}'_m)\). In the theorems that follow we shall use a zero subscript for the true parameter set. The case where \(\sigma_0^2 = 1\) is not of interest since then \(\theta_0\) and \(\nu_0\) can be measured exactly in so far as the limiting distribution is concerned and this justifies the requirement \(\sigma_0^2 < 1\) which is introduced below. The requirements \(|\nu_0| < 1, f_1 > 0\) and \(f_2 > 0\) are justified similarly.

Theorem 3.1

Let the \(w_j(\theta_k)\) be generated by a complex normal process with probability density prescribed by (1.4) for \(\tau = \tau_0\) with \(\sigma_0^2 < 1\), \(|\nu_0| < 1, f_1 > 0\) and \(f_2 > 0\). Then, as \(m\) approaches infinity, \(\hat{\tau}_m\) converges almost surely to \(\tau_0\). Moreover, \(m(\hat{\nu}_m - \nu_0)\) converges almost surely to zero.
Proof

Throughout this proof the phrase "almost sure" is omitted when speaking of almost sure convergence. Writing $A(V)$ in place of $A(\sigma^2, \nu)$ we put

$$x_1(V) = m^{-1} w_1 A(V) w_1,$$
$$x_2(V) = m^{-1} w_2 A'(V) w_2,$$

and

$$y(V) = \sigma m^{-1} w_1 A(V) \bar{y}(v) w_2.$$

Then, for any $V$, the vector $\hat{m}(V)$ maximising (2.1) is given by

$$\hat{f}_{1,m}(V) = x_1(V) - \{x_1(V)/x_2(V)\}^{1/2} |y(V)|$$
$$\hat{f}_{2,m}(V) = x_2(V) - \{x_2(V)/x_1(V)\}^{1/2} |y(V)|$$

with $\theta_m(V)$ given by (2.9). Substituting these into $L$, dividing by $m$ and omitting a constant, we obtain

$$Q_m(V) = m^{-1} \log \det[A(V)] - 2 \log[m_x(V) m_y(V)] - |y(V)|.$$

We shall examine first the case where $\sigma^2_m$ is bounded away from unity and so we consider the above for values of $\sigma^2$ such that $\sigma^2 \leq 1 - \delta$ for some small $\delta > 0$. With reference to Corollary 3.2, it is noted that the first term converges uniformly in $V$ to $-(1 - |v|) \log(1 - \sigma^2)$. Now, for $j = 1, 2$, $x_j(V)$ can be written as

$$m^{-1} w_j [A(V) - I_m - \sigma^2/(1 - \sigma^2)] \Phi(v) w_j$$
$$+ m^{-1} w_j w_j + \sigma^2/(1 - \sigma^2) m^{-1} \Phi(v) w_j \tag{3.1}$$

and, in so far as the distribution (1.4) pertains, the first term of the above can be written as

$$(2m)^{-1} f_{j0} \sum_{m,j} \lambda_{m,j} \xi_j \tag{3.2}$$
where the $\lambda_{m,j}$ are the eigenvalues of $[A(v) - I_m - \sigma^2/(1 - \sigma^2) \phi(v)]$ and the $\xi_j$ are independent chi-square random variables with two degrees of freedom. Hence, using Schwartz's inequality and the lemmas it is evident that (3.2) converges uniformly to zero. The second term of (3.1) converges to $f_j \left( 1 + \sigma^2(1 - |v|)/(1 - \sigma^2) \right)$, for $j = 1, 2$, and since $w_j^* \phi(v) w_j$ is monotonically increasing over $[-1, 0]$ and monotonically decreasing over $[0, 1]$ the uniform convergence follows. Turning to $y(v)$ we observe that

$$|m^{-1} w_1^* [A(v) \phi(v) - (1 - \sigma^2)^{-1} \phi(v)] \Delta(v) w_2|$$

is dominated by

$$|m^{-1} w_1^* [A(v) \phi(v) - (1 - \sigma^2)^{-1} \phi(v)]^2 w_1|^{\frac{1}{2}} (m^{-1} w_2^* w_2)^{\frac{1}{2}}$$

which also converges uniformly to zero using the lemmas. Thus, in place of $y(v)$, we need only consider

$$[\sigma/(1 - \sigma^2)] z(v) = [\sigma/(1 - \sigma^2)] m^{-1} w_1^* \psi(v) w_2.$$

We put

$$\eta_m(v) = z(v) - \delta(z(v))$$

and let

$$\alpha(\ell) = \ell/m^{1+\epsilon}, \quad \epsilon > 0 ; \quad \ell = 0, \pm 1, \ldots, \pm \lfloor m^{1+\epsilon} \rfloor.$$

Now

$$\sup_{|v| \leq 1} |\eta_m(v) - \eta_m(\ell(v))|$$

(3.3)

converges uniformly to zero where $\ell(v)$ is the nearest $\alpha(\ell)$ to $v$. For example
\[ m^{-1} \psi(v) w_2 - m^{-1} \psi(\ell(v)) w_2 = m^{-1} \psi(\psi(v) - \psi(\ell(v)) \Lambda(v) w_2 \]
\[ \quad + m^{-1} \psi(\ell(v)) (\Lambda(v) - \Lambda(\ell(v))) w_2 \]  
(3.4)

and this is bounded above by
\[ [m^{-1} \psi(\psi(v) - \psi(\ell(v))) w_1^2]^{\frac{1}{2}} [m^{-1} w_2 w_2]^{\frac{1}{2}} \]
\[ + [m^{-1} \psi(\ell(v)) w_1^2]^{\frac{1}{2}} [m^{-1} \psi(\Lambda(v) - \Lambda(\ell(v)))^2 [\Lambda(v) - \Lambda(\ell(v))] w_2^2]^{\frac{1}{2}}. \]

However the maximum eigenvalue of \( \psi(\psi(v) - \psi(\ell(v))) \) is \( O(m^{-\epsilon}) \) and so the first term of the above is dominated by
\[ (m^{-1} \psi w_1^2 w_2^2 O(m^{-\epsilon}) \]

which converges to zero. Examining the second term we see that this is not greater than
\[ (m^{-1} \psi w_1^2) \frac{1}{2} (m^{-1} w_2 w_2^2) \frac{1}{2} O(m^{-\epsilon}) \]

and thus (3.4) converges uniformly to zero. The difference of the expectations of \( z(v) \) and \( z(\ell(v)) \) converges uniformly to zero in a similar manner and so (3.3) converges uniformly to zero as required.

Now \( m^{-1} \{ w^* H w - \text{tr}(H^0) \} \) has characteristic function
\[ \exp \{ \sum_{k=2}^{\infty} (i\theta/m)^k \frac{\text{tr}((H^0)^k)}{k} \}
\]
where \( H \) is any arbitrary Hermitian matrix. Hence it can be seen that \( \eta_m(v) \) and \( \eta_m(\ell(v)) \) have \( k^{th} \) cumulants of order \( m^{-k+1} \). Thus
\[ P(\sup_{\ell} \eta_m(\alpha(\ell)) > \delta) \leq \sum_{\ell} P(\eta_m(\alpha(\ell)) > \delta^6) = O(m^{-2+\epsilon}) \]
and by the Borel-Cantelli lemma $\sup_{\ell} \eta_{m}(\alpha(\ell))$, and therefore $\sup \eta_{m}(v)$, converges to zero. Noting that

$$||y(v)|| - [\sigma/(1 - \sigma^2)] ||z(v)|| \leq ||y(v)|| - [\sigma/(1 - \sigma^2)] ||z(v)|| + \sigma ||z(v)|| - ||\delta(z(v))||/(1 - \sigma^2)$$

$$\leq ||y(v)|| - [\sigma/(1 - \sigma^2)] ||z(v)|| + \sigma ||\eta_{m}(v)||/(1 - \sigma^2)$$

it is evident that

$$Q_{m}(v) + (1 - |v|) \log(1 - \sigma^2) + \log(f_{10}f_{20})$$

$$+ 2 \log[1 + \sigma^2/(1 - \sigma^2)]/(1 - \sigma^2) - [\sigma/(1 - \sigma^2)] ||\delta(z(v))|| (3.5)$$

converges uniformly to zero since the arguments of the respective logarithms are bounded away from zero. Now the expectation of $z(v)$ is given by

$$i_{0} \int f_{10}f_{20} e^{-i\theta} e^{-m^{-1} \text{tr}(\Psi(v) \Psi(v_0)^*)}$$

and this differs from

$$\sigma_{0}^{10}f_{10}f_{20} e^{-i\theta} e^{-m^{-1} \text{tr}(\Psi(v) \Psi(v_0)^*)}$$

by a term which converges uniformly to zero where

$$\chi_\theta(v) = \begin{cases} 1 & (-\min(0,v) < \theta \leq 1 - \max(0,v)) \\ 0 & \text{(otherwise)} \end{cases}$$

This latter follows from the fact that

$$m^{-1} \text{tr}(\Psi(v) \Psi(v_0)^*) = m^{-1} \text{tr}(\Phi(v_0) \Phi(v) \Lambda(v - v_0))$$

$$= m^{-1} \text{tr}[[\Phi(v_0) \Phi(v) - \int_{0}^{1} \chi_\theta(v) \chi_\theta(v_0) U(\theta) U(\theta)^* d\theta] \Lambda(v - v_0)]$$

$$+ [\sin\pi(v - v_0)/(m \sin\pi(v - v_0))] \int_{0}^{1} \chi_\theta(v) \chi_\theta(v_0) d\theta$$
and the first term on the right-hand side of the above is dominated by
the square root of

\[ m^{-1} \text{tr} [ (\Phi(v_0) \Phi(v) - \int_0^1 \chi_{\Theta}(v) \chi_{\Theta}(v_0) U(\Theta) U(\Theta)^* d\Theta)^2] \]

which converges uniformly to zero by Lemma 3.1. Putting

\[ Q(v) = \lim_{m} Q_m(v) \]

it follows that

\[ Q(v) = \begin{cases} 
(1 + |v_0|) \log(1 - \sigma^2) - 2 \log(1 - \sigma^2 |v_0| - \sigma v_0 (1 - |v_0|)) - \log(f_{10} f_{20}) & (v = v_0) \\
(1 + |v|) \log(1 - \sigma^2) - 2 \log(1 - \sigma^2 |v|) - \log(f_{10} f_{20}) & (v \neq v_0)
\end{cases} \]

where the convergence is uniform for $|v - v_0| \geq \epsilon \pmod{1}$, $\epsilon > 0$, since $\delta(z(v))$ converges in this fashion.

Let us assume that $\hat{v}_m$ does not converge to $v_0$ where $\hat{v}_m$ maximises $Q_m(v)$. Let $\hat{v}_m(j)$ be a subsequence converging to $v_1$ with $v_1 \neq v_0 \pmod{1}$. Then

\[ Q(v_1) = \lim_{j \to \infty} [Q_m(j) (\hat{v}_m(j))] \geq \lim_{j \to \infty} [Q_m(j)(v_0)] = Q(v_0). \]

However $Q(v) < Q(v_0)$, $v \neq v_0$, so that a contradiction is reached and hence $\hat{v}_m$ must converge to either $v_0$ or $v_0 - v_0 / |v_0|$. If we now assume that $\hat{v}_m(j)$ converges to $v_1$ with $v_1 = v_0 - v_0 / |v_0|$ it is apparent from (3.5) that we need only consider

\[ \lim_{j \to \infty} |\delta(z(\hat{v}_m(j)))|. \]

With reference to (3.6) it will be sufficient to look at
\[
\lim_{j \to -\infty} \left| \frac{\sin m\pi (\hat{v}_m(j) - v_0)}{m \sin (\hat{v}_m(j) - v_0)} \int_0^1 \chi_\theta (\hat{v}_m(j)) \chi_\theta (v_0) \, d\theta \right| \\
\leq \lim_{j \to -\infty} \left\{ \int_0^1 \chi_\theta (\hat{v}_m(j)) \chi_\theta (v_0) \, d\theta \right\}
\]

and this latter term converges to zero. Thus \(Q_m(j)(\hat{v}_m(j))\) converges to \(Q(v_1) < Q(v_0)\) yielding a contradiction as before and so \(\hat{v}_m\) converges to \(v_0\). Let us now take \(\hat{v}_m(j)\) as converging to \(v_1\) with \(v_1 = v_0\), but \(\sigma_1^2 \neq \sigma_0^2\). Considering (3.5) once more and observing that

\[
Q_m(j)(\hat{v}_m(j)) \geq Q_m(j)(v_1)
\]

we must have

\[
\lim_{j \to -\infty} |\delta(z(\hat{v}_m(j)))| \geq \lim_{j \to -\infty} |\delta(z(v_0))|.
\]

This is equivalent, by (3.6), to the requirement that

\[
\lim_{j \to -\infty} \left| \frac{\sin m\pi (\hat{v}_m(j) - v_0)}{m \sin (\hat{v}_m(j) - v_0)} \right| \geq 1
\]

and since \((\hat{v}_m(j) - v_0)\) converges to zero it follows that if the above relation is to hold then \(m(\hat{v}_m(j) - v_0)\) must also converge to zero. This in turn implies that

\[
|\delta(z(\hat{v}_m(j)))| - |\delta(z(v_0))|
\]

converges to zero and so \(Q_m(j)(\hat{v}_m(j))\) converges to \(Q(v_1)\). Since \(Q(v_1) < Q(v_0)\) this again implies a contradiction and thus \(\hat{v}_m\) converges to \(v_0\). A repetition of the proof just given shows that \(m(\hat{v}_m - v_0)\) converges to zero which implies that \(z(\hat{v}_m)\) converges to its limit at \(v_0\) and hence \(\hat{v}_m\) converges to \(\mu_0\).

The case where \(\hat{\sigma}_m^2\) is not strictly bounded below one is more complicated since now it may be true that \(\lim \hat{\sigma}_m^2 = 1\) and so a
subsequence of values of \( \hat{\tau}_m \) is now considered along which \( \hat{\sigma}_m^2 \) approaches unity. We commence by noting that maximising (1.4) is equivalent to minimising

\[
L_m = \exp(g_m x_m)/g_m \tag{3.7}
\]

where

\[
x_m = (2m)^{-1} \sum_k \Gamma_k^{-1} w / g_m
\]

and

\[
g_m = (\det(\Gamma_0^{-1}))^{2m}.
\]

It is observed that \( x_m \) may be described as

\[
(2m)^{-1} \sum_k \Gamma_k^{-1} w / ([f_{10} + f_{20})(\det(\Omega_0^{-1}))^{2m}]
\]

where \( \Omega = (f_{1} + f_{2})^{-1} \Gamma \) and thus

\[
x_m = (2m)^{-1} \sum_{j=1}^{2m} \lambda_{m,j} |\xi_j|^2/(\sum_{j=1}^{2m} \lambda_{m,j})^{2m} \frac{1}{2m}
\]

with the \( \lambda_{m,j} \) denoting the eigenvalues of \( \Omega_0^{\frac{1}{2}} \Omega^{-1} \Omega_0^{\frac{1}{2}} \) and the \( \xi_j \) are independent complex random variables having zero means and unit variances. However

\[
e^{C_0}/x, \quad c \geq 0, \quad x \geq 0,
\]

has a minimum at \( x = c^{-1} \) and so

\[
L_m \geq e x_m.
\]

Also, when \( \Gamma = \Gamma_0 \), \( L_m \) converges to \( e \). Hence, if it can be shown that \( \lim x_m \) is strictly greater than unity along this subsequence where \( \hat{\sigma}_m^2 \) tends to one, then a contradiction will have been established as before.

To simplify the notation the subscript \( m \) will be used to index this subsequence of \( \hat{\tau}_m \) in question. In this context it is noted that some
of the $\lambda_{m,j}$ may approach infinity as $m$ increases and so we are led to partition the $\lambda_{m,j}$ into two sets according as $\lambda_{m,j}^{-1} < d$ or $\lambda_{m,j}^{-1} \geq d$ for some $d$. Let there be $m'$ of the $\lambda_{m,j}$ in the first set with geometric mean $g'_m$ and $m'' (=2m-m') \lambda_{m,j}$ in the second set with geometric mean $g''_m$. Thus

$$x_m = \frac{p \, x'_m + q \, x''_m}{(g'_m)^p \, (g''_m)^q}$$

where

$$p_m = m'/(2m) = 1 - q_m$$

and

$$x'_m = (m')^{-1} \sum_{m,j} \lambda_{m,j} |\xi_j|^2, \quad x''_m = (m'')^{-1} \sum_{m,j} \lambda_{m,j} |\xi_j|^2$$

with the former summation over the last $m'$ $\lambda_{m,j}$ and the latter summation over the remaining $\lambda_{m,j}$. Since, for arbitrary two-dimensional vector $z_1$ and arbitrary $m$-dimensional vector $z_2$,

$$(z_1 \otimes z_2)^* \Omega (z_1 \otimes z_2)$$

$$= (f_1 + f_2)^{-1} (2\pi)^{-1} \int_{-\infty}^{\infty} z_1^* h(\lambda) f(0) h(\lambda)^* z_1 z_2^* Q(\lambda) z_2 d\lambda$$

$$\leq z_1^* z_1 (2\pi)^{-1} \int_{-\infty}^{\infty} z_2^* Q(\lambda) z_2 d\lambda$$

$$= (z_1 \otimes z_2)^* (z_1 \otimes z_2)$$

it is clear that the maximum eigenvalue of $\Omega$ is less than one and, similarly, the minimum eigenvalue of $\Omega$ is greater than zero. Thus there exists a constant $\delta_0 > 0$ such that

$$\delta_0 \leq \lambda_{m,j} < \infty, \quad 0 \leq \lambda_{m,j}^{-1} \leq \delta_0^{-1}.$$ 

We shall call $\lambda_m$ and $s_m^2$ the mean and variance respectively of the $\lambda_{m,j}$ and so
\[ \tilde{s}_m^2 = (2m)^{-1} \text{tr}((\Omega - \tilde{\lambda}_m \Omega_0^2)|\tilde{\tau}_m^2) \]

\[ \geq (2m)^{-1} \text{tr}((\Omega - \tilde{\lambda}_m \Omega_0^2)|\tilde{\tau}_m^2) \]

since the maximum eigenvalue of \( \Omega_0 \) is less than one. However, using the lemmas of §3.2 and the associated convergence proofs, it can be seen that \((2m)^{-1} \text{tr}((\Omega - \tilde{\lambda}_m \Omega_0^2)|\tilde{\tau}_m^2)\) differs from

\[ \frac{1}{2}(\tilde{f}_1 - \tilde{\lambda}_m \tilde{f}_1 10)^2 + \frac{1}{2}(\tilde{f}_2 - \tilde{\lambda}_m \tilde{f}_2 20)^2 + \sigma^2 \tilde{\tau}_1 \tilde{\tau}_2 (1 - |v|) \]

\[ + \tilde{\lambda}_m^2 \sigma_0^2 \tilde{\tau}_1 \tilde{\tau}_2 10^2 (1 - |v_0|) \]

\[ - 2\tilde{\lambda}_m \sigma_0 (\tilde{f}_1 \tilde{f}_2 10 \tilde{f}_2 20)^{\frac{1}{2}} \cos(\theta - \theta_0)(1 - |v_0|) \delta_{v_0} (v) \]

by a term which converges to zero. Here \( \delta_{v_0} (v) \) is unity if \( v = v_0 \) and is zero otherwise whereas

\[ \tilde{f}_j = f_j/(f_1 + f_2), \quad j = 1,2, \]

with the \( \tilde{f}_{j0} \) defined similarly. We wish to show that

\[ \lim[(2m)^{-1} \text{tr}((\Omega - \tilde{\lambda}_m \Omega_0^2)|\tilde{\tau}_m^2)] \geq b_\varepsilon > 0 \] (3.9)

for values of \( \tau \) such that \( |\sigma - \sigma_0| > \varepsilon \) where we have imposed this last condition because we are concerned with the case where \( \tilde{\lambda}_m^2 \) tends to one.

Now the sum of the last three terms of (3.8) is always positive and on examination of the first two terms of this expression we see that these will only approach zero if \( \tilde{f}_1 \) equals \( \tilde{f}_{10} \) and \( \tilde{\lambda}_m \) approaches one. For the case \( v \neq v_0 \) this latter implies that (3.8) is greater than

\[ \tilde{\lambda}_m^2 \sigma_0^2 \tilde{\tau}_1 \tilde{\tau}_2 10^2 (1 - |v_0|) \] whereas for \( v = v_0 \) it implies that (3.8) is not less than \( \tilde{f}_{10} \tilde{f}_{20} (1 - |v_0|)(\sigma - \tilde{\lambda}_m \sigma_0)^2 \). These two terms will be strictly positive since \( |\sigma - \sigma_0| > \varepsilon \) and so (3.9) has been established.

Hence, considering \( \tilde{\tau}_m^2 \), we now have \( \lim \tilde{s}_m^2 \geq b_\varepsilon > 0 \).
Returning to the statistic $x_m$ we examine two cases according as $\lim g_m = \infty$ or $\lim g_m < \infty$. Assuming the former we consider a subsequence along which $g_m$ approaches infinity. (The subscript $m$ will again be used to index this subsequence of a subsequence and also such subsequences mentioned later.) In this situation we may choose $d$ so that $0 < \lim p_m < 1$. This is always possible since, otherwise, as $d$ moves from 0 to $\delta_0^{-1}$, $\lim p_m$ is zero until some point $d'$ and is unity thereafter. The fact that there exists such a $d'$ follows from the observation that $\lim p_m$ is greater than zero for $d$ near to $\delta_0^{-1}$ because, assuming the contrary case and a subsequence of values along which $p_m$ converges to zero,

$$\tilde{s}_m^2 = (2m)^{-1} \sum_{\lambda_m, j}^{\lambda_m-2} \tilde{\lambda}_m^2 \leq p_m d^2 + q_m \delta_0^{-2} - q_m^2 d^2$$

can be made arbitrarily small contradicting the strictly positive nature of this variance. Thus almost all of the eigenvalues must lie in some arbitrarily small neighbourhood of $d'$ which again establishes a contradiction since $\lim \tilde{s}_m^2 > b > 0$. We now have a $d$ such that $0 < \lim p_m < 1$ and so a subsequence can be chosen such that $0 < \lim p_m < 1$. However, by Lemma 3.3,

$$x_m \geq \frac{p_m g_m' + q_m g_m''}{(g_m')^p m (g_m'')^q m}$$

and this, in turn, is greater than $p_m (g_m' / g_m'')^{-1} p_m$. Hence, along this subsequence, $x_m$ increases indefinitely contradicting the essential supposition that $\tilde{\sigma}_m^2$ approaches unity.

The second case where $\lim g_m < \infty$ is now considered. Given these circumstances we can always choose $d$ small enough so that $\lim p_m$ is arbitrarily small because, in the contrary case, $\lim g_m$ diverges to
infinity contradicting the original supposition. This latter follows because, for all arbitrarily small $d$,

$$g_m = (g'_m)^p_m (g''_m)^q_m \geq (d^{-1})^p_m q_m$$

and since $\lim p_m > 0$, $g_m$ increases indefinitely as required. Let us now denote by $\bar{\lambda}_m''$ and $s''_m$ the mean and variance respectively of those $\lambda_{m,j}^{-1} \geq d$. Since

$$\tilde{s}''_m = (2m)^{-1} \sum \lambda_{m,j}^{-2} - \bar{\lambda}_m''^2$$

$$\leq p_m d^2 + q_m \bar{s}''_m + q_m \bar{\lambda}_m'' - q_m^2 \bar{s}''_m$$

it is evident that

$$\tilde{s}''_m \geq \tilde{s}''_m - p_m d^2 - p_m \delta^{-2} .$$

With reference to the above it can be seen that for $d$ small enough $\lim \tilde{s}''_m \geq b''_c > 0$ and so, letting $\bar{\lambda}_m''$ and $s''_m$ represent the mean and variance of the $\lambda_{m,j} \leq d^{-1}$,

$$s''_m = (m'')^{-1} \sum (\lambda_{m,j} - \bar{\lambda}_m'')^2$$

$$\geq \delta^4 0 (m'')^{-1} \sum (\lambda_{m,j}^{-1} - \bar{\lambda}_m'')^2$$

$$\geq \delta^4 \times s''_m$$

which implies that $\lim s''_m \geq \delta^4 0 b''_c > 0$. Noting that $x''$ has bounded eigenvalues we see that this quantity converges to $\bar{\lambda}_m''$. However

$$\log(1 + x) \leq x - \frac{1}{2b} x^2 , \quad -1 \leq x \leq b^{-1} , \quad b > 1 ,$$

so that

$$(m'')^{-1} \sum \log(\lambda_{m,j} / \bar{\lambda}_m'') \leq (m'')^{-1} \sum \left\{ (\lambda_{m,j} / \bar{\lambda}_m'' - 1) - \frac{d}{\lambda_m''} (\lambda_{m,j} / \bar{\lambda}_m'' - 1)^2 \right\}$$

$$= - \frac{d}{\lambda_m''} s''_m$$
and hence
\[ \frac{\tilde{\lambda}''}{g''_m} \geq \exp\left[ d_0 \frac{\sigma^2}{2 \tilde{\lambda}''_m} \right]. \]

Now
\[ \lim \left( \frac{\tilde{\lambda}''}{g''_m} \right) \geq \exp\left[ d_0 \frac{\sigma^2}{2} \right] = 1 + \alpha, \quad \alpha > 0. \]

Considering the choice of \( d \), we observe that we can also take \( d \) small enough so that \( \lim q_m > 0 \) since, considering a subsequence along which \( q_m \) converges to zero,
\[ s_m^2 \leq p_m d_0^2 + q_m \delta_0^{-2} \]
may be made arbitrarily small contradicting the fact that \( \lim s_m^2 \) is bounded away from \( b_c > 0 \). Thus, for \( m \) large enough,
\[ x_m \geq \frac{p_m g_m' + q_m g''_m (1 + \alpha)}{(g'_m)^p (g''_m)^q} \]
where this follows from Lemma 3.3 and the development given above. Hence
\[ x_m \geq 1 + \alpha \delta_0 q_m / g_m \]
and so \( \lim x_m \) is strictly greater than unity as required. The above argument has shown that \( \tilde{\sigma}^2_m \) is bounded away from unity and that the original assumption concerning \( \sigma^2 \) may be imposed without cost. This completes the proof.

We note that the type of proof constructed for this case where \( \lim \tilde{\sigma}^2_m = 1 \) could have been used to establish Theorem 3.1 in its entirety. Indeed, this method will be adopted in the proof of an analogous theorem in Chapter 4. However, the approach given in Theorem 3.1 is instructive in that the limit function is evaluated. The latter is not required for the alternative method.
3.4 A Central Limit Theorem

The asymptotic normality of the estimates is established by the following theorem.

**Theorem 3.2**

Under the conditions of Theorem 3.1, for $\sigma_0^2 > 0$, $D_m (\hat{\tau}_m - \tau_0)$ is asymptotically normal with zero mean vector and covariance matrix $I_0^{-1}$ where

$$I_0^{-1} = \begin{bmatrix}
1 - \sigma_0^2 |v_0| & \sigma_0 f_{10}/(1 - \sigma_0^2 v_0^2) & 1 - |v_0| & 0 \\
1 - \sigma_0^2 |v_0| & \sigma_0 f_{20}/(1 - \sigma_0^2 v_0^2) & 0 & 0 \\
1 - \sigma_0^2 |v_0| & 0 & 0 & 0 \\
2(1 - |v_0|) & 0 & 0 & \frac{3(1 - \sigma_0^2)}{2\pi^2 \sigma_0^2 (1 - |v_0|)}
\end{bmatrix}$$

and $D_m$ is diagonal with $m^{\frac{1}{2}}$ in the first four places in the main diagonal and $m^{\frac{3}{2}}$ in the fifth place.

**Proof**

Letting $\partial L/\partial \tau$ denote the vector whose $i^{th}$ component is $\partial L/\partial \tau_i$ we have, expanding $\partial L/\partial \tau$ in a Taylor series about $\hat{\tau}_m$, that

$$D_m^{-1} \frac{\partial L}{\partial \tau} \bigg|_{\tau_0} = G(\tau_m) D_m (\hat{\tau}_m - \tau_0) ,$$

where $\tau_m$ is a random variable such that $|\tau_m - \tau_0| \leq |\hat{\tau}_m - \tau_0|$ and

$$G_{jk}(\tau) = -D_m^{-1} \frac{\partial^2 L}{\partial \tau_j \partial \tau_k} D_m^{-1} (j, k = 1, \ldots, 5).$$

The existence of $\tau_m$ is established as in Lemma 3 of Jennrich (1969).
The elements $G_{jk}(\tau_m)$, $j,k = 1, \ldots, 4$, converge almost surely to the respective $I_{0,jk}$ since $\tau_m$ converges almost surely to $\tau_0$. This result follows from the lemmas and corollaries given in §3.2 and to illustrate we note that

$$G_{11}(\tau_m) = \left[ \frac{2}{f_1} m^{-1} \sigma_1 \hat{A}(v) \psi_1 - \frac{1}{f_1^2} \right] \frac{3}{2} \frac{R}{f_1^2(f_1 f_2)} \left[ e^{-i\theta} \sigma_1^{-1} \hat{A}(v) \hat{\psi}(v) \psi_2 \right] \left| \tau_m \right|$$

in the notation of Theorem 3.1. However it can be seen that the various convergence proofs given in this previous result coupled with the fact that $\tau_m$ converges almost surely to $\tau_0$ yield the required result. To show that the remaining elements of $G(\tau_m)$ converge to their limits it will be sufficient to consider $G_{55}(\tau_m)$ and $G_{35}(\tau_m)$. The latter can be written as

$$-2 \left[ \sigma(f_1 f_2)^{-1/2} \left\{ e^{-i\theta} \sigma_1^{-2} \hat{A}(v) \hat{C}(\sigma^2, v) \psi_2 \right\} \right] \left| \tau_m \right|$$

and this differs from

$$-2 \left[ \sigma_0 \left( f_1 f_2 \right)^{1/2} \left( f_1 f_2 \right)^{1/2} \right] \left\{ e^{i(\theta, -\theta)} \sigma_1^{-2} \hat{A}(v) \hat{C}(\sigma^2, v) \hat{\psi}(v_0) \right\} \left| \tau_m \right|$$

by a term which converges almost surely to zero. Now

$$\frac{d}{dv} [\Phi(v)] = -\text{sign}(v) \hat{U}(v) \hat{U}^\prime(v), \quad \frac{d}{dv} [\psi(v)] = -\{\text{sign}(v) \hat{U}(v) \hat{U}^\prime + i2\pi \hat{\psi}(v) D\}$$

where $U(\theta)$ is as for Lemma 3.1, $U$ denotes $U(0)$ and $D$ is diagonal with $j$th entry $j - n - 1$. Hence
\[ |m^{-2} \text{tr}[A(V) C(\sigma^2, \nu) \psi(\nu_0)^*]| \]
\[ \leq \sigma^2 |m^{-2} \text{tr}[A(V) B(\nu) A(V) \psi(\nu) \psi(\nu_0)^*]| \]
\[ + |m^{-2} \text{tr}[A(V) \overline{U(\nu)} U'(\nu_0)^*]| + 2\pi |m^{-2} \text{tr}[A(V) \psi(\nu) D \psi(\nu_0)^*]| \]
\[ \leq \sigma^2 |m^{-2} U'(\nu) \phi(\nu) A(V) \psi(\nu) \psi(\nu_0)^* A(V) \overline{U(\nu)}| \]
\[ + \sigma^2 |m^{-2} U'(\nu) A(V) \psi(\nu) \psi(\nu_0)^* A(V) \phi(\nu) \overline{U(\nu)}| \]
\[ + |m^{-2} U'(\nu_0)^* A(V) \overline{U(\nu)}| + 2\pi |m^{-2} \text{tr}[A(V) \psi(\nu) D \psi(\nu_0)^*]| \]
\[ \leq 2\sigma^2 m^{-1} ||A(V)||^2 + m^{-1} ||A(V)|| \]
\[ + 2\pi |m^{-2} \text{tr}[\phi(\nu_0) A(V) \phi(\nu) \Lambda(\nu - \nu_0) D]| \]

and since the eigenvalues of \(A(V_m)\) are bounded it is evident that we need only consider the third term of the above evaluated at \(\tau_m\).

However this latter is bounded above by
\[ 2\pi |m^{-2} \text{tr}[(\phi(\nu_0) A(V) \phi(\nu - \nu_0) D)] + |m^{-2} \text{tr}[\phi(\nu - \nu_0) D]| \]

where
\[ \phi = (1 - \sigma^2) m^{-1} \int_0^1 \chi_\theta(v_m) \chi_\theta(\nu_0) U(\theta) U(\theta)^* d\theta \]

and the \(\chi_\theta(\nu)\) are as for (3.6). The first term converges to zero using Schwartz's inequality and Lemma 3.1 whereas the second term is given by
\[ \pi(1 - \sigma^2) m^{-1} \left[\left(\frac{m \cos \theta \sin \theta - \sin \theta \cos \theta}{m^2 \sin^2 \theta}\right)\right] \int_0^1 \chi_\theta(v_m) \chi_\theta(\nu_0) d\theta \]

which also converges to zero since \(\tau_m\) converges almost surely to \(\tau_0\) and \(m(v_m - \nu_0)\) converges almost surely to zero. Thus (4.1) converges almost surely to zero as required.
Turning now to $G_{55}(\tau_m)$ we note that

$$\frac{d^2}{dv^2} \psi(v) = i2\pi \text{sign}(v) \left[ D \overline{U(v)} U' + \overline{U(v)} U' D \right] - 4\pi^2 \psi(v) D^2$$

(4.4)

and using similar arguments to those given for the proof that (4.1) converges almost surely to zero we see that $G_{55}(\tau_m)$ is asymptotically equivalent to

$$8\pi^2 \left( \sigma_0 \left( f_1 f_2 \right)^{\frac{1}{2}} \right)^{\frac{1}{2}} \left[ e^{i(\theta_0 - \theta)} m^{-3} \text{tr}(A(v) \psi(v) D^2 \psi(v_0)^*) \right] \right|_{\tau_m} \left( \frac{m^2 + 1}{3} \sin \pi \theta \right) \left( \frac{m \cos \pi \theta \sin 2\pi \theta}{m^3 \sin 3\pi \theta} - 2 \sin \pi \theta \right) \right|_{\overline{v}_m - v_0} \cdot$$

With reference to (4.3) it is observed that

$$m^{-3} \text{tr}(A(\overline{v}_m) \psi(\overline{v}_m) D^2 \psi(v_0)^*) = m^{-3} \text{tr}(\Phi(v_0) A(\overline{v}_m) \Phi(\overline{v}_m) A(\overline{v}_m - v_0) D^2)$$

differs from

$$\left\{ (1 - \sigma_m^2)^{-1} \int_0^1 \chi_0(v_m) \chi_0(v_0) \, d\theta \right\} \left\{ m^{-3} \sum_{-n}^{n} j^2 \exp(-i2\pi j(v_m - v_0)) \right\}$$

by a term which converges to zero and, moreover, the second bracketed expression can be written as

$$\frac{1}{4} \left[ (m^2 + 1) \sin \pi \theta \right. \left. + \frac{m \cos \pi \theta \sin 2\pi \theta}{m^3 \sin 3\pi \theta} - 2 \sin \pi \theta \right] \left. \right|_{\overline{v}_m - v_0} \cdot$$

Since $m(v_m - v_0)$ converges to zero this latter term converges to 1/12 and so (4.5) converges to $2\pi^2 \sigma_0^2 (1 - |v_0|)/(3(1 - \sigma_0^2))$. Thus $G_{55}(\tau_m)$ converges almost surely to $I_{0,55}^*$.

All that remains to be proved is that $D_m^{-1} [\partial L / \partial \tau] |_{\tau_0}$ is asymptotically normally distributed with zero mean vector and covariance matrix $I_0$. Examining the arbitrary linear combination

$$z_m = \sum_{i=1}^{5} \alpha_i \left[ D_m^{-1} \frac{\partial L}{\partial \tau} \right]_i |_{\tau_0}$$

it is evident that this may be described as
\[
\sum_{i=1}^{5} \alpha_i D^{-1}_{m,i} \left\{ \sum_{i=1}^{\infty} \frac{r^{-1}}{\sigma_{\tau_i}} \frac{\partial r}{\partial \tau_i} \right\} \left( r^{-1} \frac{\partial r}{\partial \tau_i} \right) \right|_{\tau_0} \\
= \frac{1}{2} \sum_{i=1}^{2m} \mu_{m,i} (\xi_i - 2)
\]
where the \( \xi_i \) are distributed as independent chi-square random variables with two degrees of freedom and the \( \mu_{m,i} \) are the eigenvalues of

\[
T_{\alpha} = \sum_{i=1}^{5} \alpha_i D^{-1}_{m,i} \left\{ \sum_{i=1}^{\infty} \frac{r^{-1}}{\sigma_{\tau_i}} \frac{\partial r}{\partial \tau_i} \right\} \left( r^{-1} \frac{\partial r}{\partial \tau_i} \right) \right|_{\tau_0} 
\]

The limiting mean and variance of \( z_m \) are zero and \( \sigma_z^2 \) respectively where

\[
\sigma_z^2 = \lim_{m \to \infty} \left\{ \sum_{i=1}^{2m} \mu_{m,i} \right\} = \lim_{m \to \infty} \left\{ \sum_{i,j=1}^{m} \alpha_i \alpha_j D_{m,i}^{-1} D_{m,j}^{-1} \text{tr} \left( r^{-1} \frac{\partial r}{\partial \tau_i} \right) \right\} \left( r^{-1} \frac{\partial r}{\partial \tau_j} \right) \right|_{\tau_0} \\
= \sum_{i,j} \alpha_i \alpha_j I_{0,i,j}
\]

Now \( z_m / \sigma_z \) has characteristic function

\[
\exp \left[ - \sum_{j=1}^{2m} \log \left( 1 - \frac{i \mu_{m,j} \theta}{\sigma_z} \right) + \frac{i \mu_{m,j} \theta}{\sigma_z} \right] \]
and since \( \| T_{\alpha} \| \) is \( O(\sqrt{m}) \) the logarithm term may be expanded out to yield

\[
\exp \left[ - \frac{1}{2} \sum_{j=1}^{2m} \frac{\mu_{m,j} \theta^2}{\sigma_z^2} + O(m^{-\frac{1}{2}}) \right]
\]
which converges to \( \exp(-\theta^2/2) \). This completes the proof of Theorem 3.2.

We shall now, for convenience, omit the subscript \( m \) on \( \hat{\tau}_m \).

Let us examine the elements of \( I_0^{-1} \) in relation to the simple example given in §3.1 where two recorders are receiving the same signal together with noise from some distant source. With reference to the diagram we have endeavoured to represent the finite length of the observed vector time series by the solid lines \( R_1 B_1 \) and \( R_2 B_2 \) and so
the length of the line $R_1 A_1$ gives the group delay. Now, considering $I_0^{-1}$, it can be seen that the smaller $|v_0|$ the smaller the variances of $\hat{\theta}, \hat{\sigma}$ and $\hat{v}$. This accords with one's intuition in that the cross-spectral information will be largely determined from the common elements of the data implied by $A_1 B_1, R_2 A_2$ and these will be maximised when the group delay is minimised. Similarly, with the variances of $\hat{f}_1$ and $\hat{f}_2$ which increase as $|v_0|$ becomes smaller, one can reason that the knowledge of one series is augmented given $\sigma_0^2, \theta_0$ and $v_0$ as knowledge of these implies partial knowledge of extra data in the particular series. In other words, when we estimate the spectrum of the first series for example, it seems that as well as the information implied by $R_1 B_1$ we would also have, perhaps, some additional information borrowed from $A_2 B_2$. It is also observed that the variance of $\hat{\theta}$ is directly proportional to the variance of $\hat{v}$. This latter follows by considering (1.1) which can be interpreted as a simple regression of phase on frequency over the $m$ frequencies $\theta_k$ in question. Thus, using the standard expressions for the sample variances of the regression coefficients, we might infer that

$$\text{var}(\hat{\theta}) = \bar{\sigma}^2/m,$$

$$\text{var}(\hat{\theta}') \approx N^2 \text{var}(\hat{v}) = \bar{\sigma}^2/ \left( \frac{n}{\sum (2\pi j/N)^2} \right)$$

where $\bar{\sigma}^2$ represents the variance of the regression. However the ratio $m \text{var}(\hat{\theta})/(m^3 \text{var} \hat{v})$ is now given by

$$m^{-3} \frac{n}{\sum (2\pi j)^2}$$

which converges to $\pi^2/3$ as might be expected.
With regard to this simple model under study it can be seen that the group delay is also given by $d \cos \phi / V$ where $d$ represents the length of the line $R_1 R_2$, $\phi$ is as given, and $V$ is the velocity of the signal. Thus $v$ has the form $\alpha \cos \phi / V$ where $\alpha$ is a constant which, in the sample situation, will be given by $d/N$. It is evident that if we wish to determine either the velocity or azimuth of the signal from $v$ then one of these two former quantities must be known. If the azimuth is known then the variance corresponding to $\hat{\hat{V}}$ will be directly proportional to

$$
(1 - |\alpha \cos \phi / V|)^{-1} \left( \frac{\partial v}{\partial \phi} \right)^{-2} = V^2 (1 - |\alpha \cos \phi / V|)^{-1} (\alpha \cos \phi)^{-2}.
$$

If the velocity is known then the variance corresponding to $\hat{\hat{\phi}}$ will be directly proportional to

$$
(1 - |\alpha \cos \phi / V|)^{-1} \left( \frac{\partial v}{\partial \phi} \right)^{-2} = V^2 (1 - |\alpha \cos \phi / V|)^{-1} (\alpha \sin \phi)^{-2}.
$$

Hence, estimation will be better, generally speaking, if the speed of propagation is low. It can also be seen that optimum azimuth resolution is attained when the line of recorders is parallel to the wavefront whereas velocity determination is best achieved when the line of recorders is parallel to the direction of propagation. This agrees with the practical findings in other fields such as seismology where linear arrays of recorders are used to collect and study data.

As a final note to this section we reiterate that Theorems 3.1 and 3.2 give an asymptotic approximation to the distribution of the statistic $\hat{\phi}$ for $m$ sufficiently large and $N$ sufficiently large, the latter to an extent dependent on $m$. In this sense and subject to the remarks made in §3.1 the results hold under the very general conditions of Theorem 2.1.
3.5 The Numerical Procedure and Some Numerical Examples

For \( m \) small equations (2.5) to (2.9) can be solved numerically without too much trouble, but for large \( m \) the process becomes rather ponderous due to the inversion of \( (I_m - \sigma^2 \Phi^2(v)) \), an \( m \times m \) Hermitian matrix. Concentrating on this latter, most relevant, case, we shall now consider approximations to the maximum likelihood equations which will result in a simpler numerical procedure.

Considering (2.2), for example, we have

\[
m^{-1}[\partial L/\partial f_1] = -f_1^{-1} + f_1^{-2} m^{-1} w_1^* A(v) w_1
- \sigma f_1^{-1}(f_1f_2)^{-\frac{1}{2}} \mathcal{R}[e^{-i\theta} m^{-1} w_1^* A(v) \Psi(v) w_2]
= -f_1^{-1} + f_1^{-2} [m^{-1} w_1^* w_1 + (\sigma/(1 - \sigma^2)) m^{-1} w_1^* \Phi(v) w_1]
- (\sigma/(1 - \sigma^2)) f_1^{-1}(f_1f_2)^{-\frac{1}{2}} \mathcal{R}[e^{-i\theta} m^{-1} w_1^* \Psi(v) w_2]
+ (2m)^{-1} \sum_{j=1}^{2m} \alpha_j z_j \quad (5.1)
\]

where the \( z_j \) are chi-square random variables with two degrees of freedom and it is evident from Lemma 3.1 that \( m^{-1} \sum \alpha_j^2 = O(m^{-1} \log m) \).

Thus the last factor on the right-hand side of the above converges almost surely to zero. The other partial derivatives admit similar results and we are led to consider the resulting approximate likelihood equations which can be arranged to yield

\[
f_1/f_2 = [(1 - \sigma^2) w_1^* w_1 + \sigma^2 w_1^* \Phi(v) w_1]/[(1 - \sigma^2) w_2^* w_2
+ \sigma^2 w_2^* \Phi'(v) w_2] \quad (5.2)
\]

\[
f_1 = [w_1^* w_1 + \sigma^2 f_1^* f_2 (I_m - \Phi'(v)) w_2/f_2]/[m(1 + \sigma^2 |v|)] \quad (5.3)
\]

\[
\theta = \arctan(\frac{f_1^* w_2}{\mathcal{R}[w_1^* \Psi(v) w_2]}) \quad (5.4)
\]
\[ \sigma = (f_1 f_2)^{-\frac{3}{2}} \left| w_1^* \Psi(v) \right| \left| w_2 \right| / \left[ m(1 + |v|) - f_1^{-1} w_1^* (I_m - \Phi(v)) w_1 \right. \]
\[ \left. - f_2^{-1} w_2^* (I_m - \Phi'(v)) w_2 \right] \]  
\[ \Theta = \arctan(\Re[w_1^* \Psi(v) D w_2] / \Im[w_1^* \Psi(v) D w_2]) \]

where \( D \) is as for (4.2).

It is clear that (5.4) to (5.6) yield two nonlinear equations which must be solved for \( \hat{v} \) and \( \hat{\sigma} \) where now \( \hat{\cdot} \) refers to the solution of these approximate equations rather than the actual maximum likelihood estimates. However it is not immediately obvious as to how one obtains a first estimate of \( \hat{v} \) since, for example, the argument of \( w_2(\theta_k) w_1(\theta_k) \) can only be measured mod(2\pi). One method which can be applied is to plot the \( w_2(\theta_k) w_1(\theta_k) \) in polar coordinates, linking successive end points by straight lines. The number of times that the resulting line circles the origin will be, approximately, \( m\hat{v} \). One can, of course, look at the graph of the argument of \( w_2(\theta_k) w_1(\theta_k) \) in the hope of making some subjective attempt at estimating \( \hat{v} \), but this method and the one mentioned before will only give good results when \( m \) and \( \sigma_0^2 \) are large and \( v_0 \) is of moderate size. Once an initial estimate of \( \hat{v} \) has been found an initial estimate of \( \hat{\sigma} \) could be obtained from (5.5) with \( \hat{f}_1 \) and \( \hat{f}_2 \) estimated by \( m^{-1} w_1^* w_1 \) and \( m^{-1} w_2^* w_2 \) respectively.

Another obvious but more reliable procedure is to scan 
\[-(1 - |v|) \log(1 - \sigma^2) \]
for the value of \( v \) that maximises this expression where \( \sigma^2 \) is given as a function of \( v \) by equation (5.5) with \( \hat{f}_1 \) and \( \hat{f}_2 \) estimated, as before, by \( m^{-1} w_1^* w_1 \) and \( m^{-1} w_2^* w_2 \).

Perhaps the best way of obtaining initial estimates arises when one considers a certain approximation to the matrix \( \Phi(v) \) and related implications. In order to effectively introduce this approximation we shall prove the following lemma.
Lemma 3.4

Let $H$ be an Hermitian matrix of the form given by (2.10) with

$$H = \int_0^1 g(\theta) \ U(\theta) \ U(\theta)^* \ d\theta$$

where

$$g(\theta) = \sum_{k=0}^r g_k \chi_\theta(E_k), \quad \max_k |g_k| < \infty,$$

and the $\chi_\theta(E_k)$ are defined outside $(0,1]$ by periodicity. Then

$$m^{-1} \text{tr}[(H - \bar{H})^2] = O(m^{-1} \log m)$$

where

$$\bar{H} = TGT^* \quad (5.7)$$

and the matrices $T$ and $G$ have typical elements defined for

$$-\frac{1}{2}m < j,k \leq [\frac{1}{2}m]$$

by

$$G_{jk} = \begin{cases} 
  g(j/m) & (j = k) \\
  0 & (j \neq k)
\end{cases}$$

and

$$T_{jk} = m^{-\frac{1}{2}} e^{-i2\pi jk/m}.$$

Proof

We consider only the case where $m = 2n+1$ since the case $m = 2n$ is dealt with in a similar fashion. Now

$$m^{-1} \text{tr}[(H - \bar{H})^2] = m^{-1} \text{tr}(H^2) + m^{-1} \text{tr}(\bar{H}^2) - 2m^{-1} \text{tr}(HH)$$

$$= \int_0^1 g^2(\theta) \ d\theta + m^{-1} \sum_{j=-n}^n g^2(j/m) - 2m^{-1} \text{tr}(HH) + O(m^{-1} \log m)$$

and this follows from Lemma 3.1 and the fact that $T$ is unitary. Noting that

$$\bar{H} = m^{-1} \sum_{j=-n}^n g(j/m) \ U(j/m) \ U(j/m)^*$$
we may replace $m^{-1} \text{tr}(HH)$ by

$$m^{-1} \sum_{-n}^{n} g(j/m) \int_{0}^{1} g(\theta) \frac{\sin^{2} n \pi (\theta - j/m)}{m \sin^{2} n \pi (\theta - j/m)} d\theta = \int_{0}^{1} \left( m^{-1} \sum_{-n}^{n} g(\theta + j/m) g(j/m) \right) \frac{\sin^{2} n \pi \theta}{m \sin^{2} n \pi \theta} d\theta$$

since $g$ is periodic with period unity. This latter can be written as

$$\int_{0}^{1} \epsilon_{n}(\theta) \frac{\sin^{2} n \pi \theta}{m \sin^{2} n \pi \theta} d\theta + \int_{0}^{1} \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} g(\theta + \phi) g(\phi) d\phi \right) \frac{\sin^{2} n \pi \theta}{m \sin^{2} n \pi \theta} d\theta \quad (5.8)$$

where

$$\epsilon_{n}(\theta) = m^{-1} \sum_{-n}^{n} g(\theta + j/m) g(j/m) - \int_{-\frac{1}{2}}^{\frac{1}{2}} g(\theta + \phi) g(\phi) d\phi .$$

However

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} g(\theta + \phi) g(\phi) d\phi = \int_{0}^{1} g(\theta + \phi) g(\phi) d\phi = \sum_{j, k} g_{j} g_{k} \int_{0}^{1} \chi_{\theta + \phi}(E_{j}) \chi_{\phi}(E_{k}) d\phi$$

and from the development given in Lemma 3.1 it is evident that the last term of (5.8) can be expressed as

$$\int_{0}^{1} g^{2}(\theta) d\theta + O(m^{-1} \log m) .$$

Since

$$\left| \int_{0}^{1} \epsilon_{n}(\theta) \frac{\sin^{2} n \pi \theta}{m \sin^{2} n \pi \theta} d\theta \right| \leq \sup_{\theta} |\epsilon_{n}(\theta)| = O(m^{-1})$$

and

$$m^{-1} \sum_{-n}^{n} g^{2}(j/m) - \int_{0}^{1} g^{2}(\theta) d\theta = O(m^{-1})$$

the result of the lemma follows.

We note that we could, using this lemma, reinterpret the likelihood equations in terms of matrices of the type given by (5.7) since the error so incurred would be $O(m^{-1} \log m)$. However the principal asset of
this approximation is that it yields a ready interpretation of the
maximum likelihood procedure and also gives rise to an excellent method
for determining first estimates of \( \hat{v} \) and \( \hat{\sigma} \). Consider, for example, the
expression \( |w_1^* \psi(v) w_2| \). This converges almost surely to
\( (1 - |v_0|^2) \sigma \sqrt{\frac{\pi}{10}} \) for \( v = v_0 \) and to zero otherwise. We are now led
to approximate \( \Phi(v) \) in this expression by \( T J_m^* T \) where \( J_m \) is diagonal
with typical entry \( \chi_{j/m}(v), -n \leq j \leq n \), and \( \chi_\theta(v) \) is given by (3.6).
Thus \( |w_1^* \psi(v) w_2| \) is replaced by \( |w_2^* T J_m^* \Lambda(v) w_2| \) since,
using Lemma 3.4, the difference between these two quantities converges
almost surely to zero. However

\[
|w_1^* T J_m^* \Lambda(v) w_2| = |m^{-1} \sum_{-n}^{n} \chi_{t/m}(v) \xi_1(t) \xi_2(t + mv)| \quad (5.9)
\]

where

\[
\xi_j(t) = m^{-\frac{1}{2}} \sum_{-n}^{n} w_j(\theta_k) \exp(-i2\pi k t/m), \quad j = 1, 2.
\]

Since \( \xi_j(t) \) and \( \chi_{t/m}(v) \) are periodic in \( t \) with period \( m \) we may write
(5.9) as

\[
m^{-1} \left| \sum_{t=1}^{m} \chi_{t/m}(v) \xi_1(t) \xi_2(t + mv) \right| = \begin{cases} 
| \sum_{t=1}^{m(1-v)} \xi_1(t) \xi_2(t + mv) | & (v \geq 0) \\
| \sum_{t=-[mv]}^{m} \xi_1(t) \xi_2(t + mv) | & (v \leq 0).
\end{cases}
\]

Hence, for \( v \geq 0 \) and taking the usual estimates of \( \widehat{\theta}_1, \widehat{\theta}_2 \) as mentioned
before, this yields an estimate for \( \hat{\sigma} \) which can be described as

\[
\hat{\sigma} = \frac{| \sum_{t=1}^{m(1-v)} \xi_1(t) \xi_2(t + mv) |}{(1 - \hat{v}) \left( \sum_{-n}^{n} |w_1(\theta_k)|^2 \right) \left( \sum_{-n}^{n} |w_2(\theta_k)|^2 \right)^{\frac{1}{2}}} \quad (5.10)
\]
where \( \hat{v} \) is chosen to maximise \( \hat{o} \). This indicates that our transformed time series has been transformed back into the time domain, lagged, and then the lagging or rephasing has been optimised in the manner given above. This procedure would seem, intuitively speaking, to lie at the heart of the matter and one would expect that these estimates will be very close to the maximum likelihood estimates. In any event it provides us with excellent first estimates which, moreover, lend insight into the apparent course of the maximum likelihood procedure.

Before proceeding to describe some simulations we note that the method used to obtain the solution of the system of equations (5.2) to (5.6) was a modified Newton-Raphson procedure. Thus, representing (5.4) to (5.6) by \( h(v) = 0 \) and \( g(\sigma, v) = 0 \), we choose to iterate successively as

\[
\begin{align*}
\hat{v}_{i+1} &= \hat{v}_i - h(\hat{v}_i)/[\partial h/\partial v]_{\hat{v}_i} \\
\hat{\sigma}_{i+1} &= \hat{\sigma}_i - g(\hat{\sigma}_i, \hat{v}_i)/[\partial g/\partial \sigma]_{\hat{\sigma}_i, \hat{v}_i}
\end{align*}
\]

(5.11)

where \( \hat{v}_i \) and \( \hat{\sigma}_i \) denote the \( i \)th iterates of \( \hat{v} \) and \( \hat{\sigma} \) respectively. The results quoted in this section have all been obtained by this method.

We first describe a simulation where the group delay is large and, indeed, much larger than would usually be expected in most practical situations. The choice of simulation was stimulated by a desire to create a narrow band signal of high relative intensity at a known frequency as would be expected in the case of the determination of the direction of propagation of a signal with a high signal to noise ratio. A white noise time sequence with variance \( \sigma^2 \) and mean zero was generated and then filtered, using a least squares approximation, to give a series \( s(n) \) with spectrum
The vector process considered was

\[ x(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} s(n) + \varepsilon_1(n) \\ s(n-\ell) + \varepsilon_2(n) \end{bmatrix}, \quad (5.12) \]

where \( \ell \) is an integer lag and \( \varepsilon_1(n), \varepsilon_2(n) \) and \( s(n) \) are incoherent with the components \( \varepsilon_1(n) \) and \( \varepsilon_2(n) \) having zero means, common variance \( \sigma_2^2 \) and uniform spectrum. Thus the spectral density matrix of the above process is

\[ f(\lambda) = \begin{bmatrix} g(\lambda) + \frac{\sigma_2^2}{(2\pi)} & g(\lambda) e^{-i\ell\lambda} \\ g(\lambda) e^{i\ell\lambda} & g(\lambda) + \frac{\sigma_2^2}{(2\pi)} \end{bmatrix}. \quad (5.13) \]

In the example considered the values of the various parameters were \( N = 2048, m = 25, \ell = 512, \sigma_1^2 = 10\pi, \sigma_2^2 = \pi, \lambda_0 = \pi/4 \) and \( \delta = 25\pi/2048 \). Hence, with regard to the model, \( \ell/N \) corresponds to \( v_0 \) and the narrow band in question is centred at \( \pi/4 \). The various spectral values at \( \lambda_0 \) are given by \( f_{10} = 5.5, f_{20} = 5.5, \theta_0 = 0.0, \sigma_0 = 0.909 \) and \( v_0 = 0.25 \).

It is observed in this lag situation that \( \theta_0 \) will in fact be \( \check{\theta}(\lambda_0) + \ell\lambda_0 \), where \( \check{\theta}(\lambda) \) refers to the phase inherent in the process before the lagging took place. If one wanted an estimate of \( \check{\theta}(\lambda_0) \) rather than \( \theta_0 \) then this would have to be disentangled from the resulting estimate of \( \theta_0 \). However, in the results quoted, the phase estimated is \( \theta_0 \), in keeping with the formulation of the group delay model. The first estimate procedures mentioned before yielded varying results with the two graphical methods giving first estimates for \( \hat{v} \) of 0.1 and 0.3 respectively. The former is not very satisfactory, due in part to the large value chosen for \( v_0 \), but the latter is acceptable.

The scanning procedure took about 65 seconds of computer time and gave
first estimates for $\hat{v}$ and $\hat{c}$ of 0.24 and 0.78 respectively. The method involving the transformed transforms took 2 seconds and first estimates of 0.24 and 0.94 for $\hat{v}$ and $\hat{c}$ respectively were realised. This latter is clearly the first estimate procedure that should be adopted and in the simulations described later this was the method used. The main routine converged rapidly taking approximately 4.5 seconds per iteration. Five iterations were required to obtain accuracy to three decimal places and resulted in the following estimates together with the appropriate confidence intervals obtained using $I^{-1}_0$ evaluated at the true parameter point. With the confidence level set at 95% the results were

$$\hat{f}_1 = 5.10, \quad 3.09 \leq f_{10} \leq 7.11;$$
$$\hat{f}_2 = 5.11, \quad 3.10 \leq f_{20} \leq 7.12;$$
$$\hat{\theta} = 0.11, \quad -0.04 \leq \theta_{0} \leq 0.26;$$
$$\hat{\sigma} = 0.93, \quad 0.88 \leq \sigma_{0} \leq 0.98;$$
$$\hat{v} = 0.25, \quad 0.244 \leq v_{0} \leq 0.256.$$

It is interesting to observe that the usual estimate of coherence gave a value of $\hat{\sigma} = 0.20$ which is considerably less than the true value of 0.91.

We now turn to some examples where the values chosen for the group delay are of a more moderate size. We shall consider two models and shall examine the estimates arising from these models for different values of $m$ and $N$. It is hoped, in this way, to give some idea of the values of $m$ and $N$ that might be needed in order that Theorem 3.2 should adequately describe the distribution of the $w_j(\theta_k)$. Now the models studied were identical to that given by (5.12) and (5.13) except that the filtering was omitted for reasons of computational efficiency.
This makes no essential difference to the simulations, but, in practice, if one knew that the variation was broad band in nature then one should use a more appropriate technique other than that described here. (See Hannan and Robinson (1971).) Thus (5.13) can be written as

\[
f(\lambda) = (2\pi)^{-1} \begin{bmatrix}
\sigma_1^2 + \sigma_2^2 & \sigma_1^2 e^{-i\lambda} \\
\sigma_1^2 e^{i\lambda} & \sigma_1^2 + \sigma_2^2
\end{bmatrix}
\]

and in the models considered \( \sigma_1^2 \) and \( \sigma_2^2 \) were once again set at \( 10\pi \) and \( \pi \) respectively. The lags \( \ell \) were given for these simulations by \( N_\ell \) and two values of \( v_0 \) were examined. These were \( v_0 = 0.0625 \) and \( v_0 = 0.015625 \) where the former is moderately large, but the latter is small. The other spectral values are now given by \( f_0 = 5.5, \)

\( f_2 = 5.5, \theta_0 = N_\ell v_0 \lambda_0 \) and \( \sigma_0 = 0.909 \). The phase \( \theta_0 \) is zero in all cases except for the one given in Table 9 where \( v_0 = 0.015625, N = 128 \) and so \( \theta_0 = \pi \). The transformed transform method was used to determine first acceptable estimates in all the cases considered and this provided very acceptable results. The figures relating to computational times given in the previous simulation where \( v_0 = 0.25 \) apply here, although these improved significantly as \( m \) and \( N \) became smaller to the extent that in the case \( m = 9, N = 128 \) a typical result took 0.6 seconds per iteration and 4 iterations to obtain the required 3 decimal figure accuracy.

Each of the displayed tables refers to one set of 20 replications based on independent realisations of the particular model concerned. The values of \( N, m, \lambda_0 \) and \( v_0 \) are given for each table and the \( x \) and \( s \) rows denote the sample means and standard deviations of the estimates obtained. The \( \sigma_x \) row gives the theoretical standard deviation and the \( z, t \) and \( \chi^2 \) rows are calculated from the other rows by the formulae

\[
z = \frac{|\bar{x} - \mu|}{\sigma_x/\sqrt{20}}, \quad t = \frac{|\bar{x} - \mu|}{s_x/\sqrt{20}}
\]
and

\[ \chi^2 = 19 \frac{s^2}{\sigma^2_x} \]  

where \( \mu \) denotes the appropriate true parameter value. The quantity \( z \), for example, should, in terms of Theorem 3.2, be a realisation of a distribution that is approximately the standard normal distribution. The quantities \( t \) and \( \chi^2 \) admit similar interpretations in terms of the \( t \) distribution and the chi-square distribution and so, for a 5\% level of significance, the critical values of \( z \), \( t \) and \( \chi^2 \) are

\[
\begin{align*}
Z_{0.025} &= 1.96, & t_{0.025} &= 2.093, \\
\chi^2_{0.025} &= 32.852, & \chi^2_{0.975} &= 8.907,
\end{align*}
\]

(5.15)

where the \( t \) and \( \chi^2 \) values are based on 19 degrees of freedom.

On examination of the results it can be seen that estimation is more difficult in the case of large \( \nu \) as was expected. (See the discussion at the end of §3.4.) For \( \nu = 0.015625 \) the results seem to be quite reasonable although the \( \chi^2 \) criterion becomes a little erratic in Tables 7 and 9 indicating significant departures of the sample standard deviations from their postulated values. For \( N = 2048 \) and \( N = 1024 \) the spectra estimates are slightly better in the case of the larger \( \nu \), but, otherwise, this is not necessarily true. Considering \( \nu = 0.0625 \) it is evident that the sample estimates of the standard deviations differ from their theoretical values, significantly on the whole, from about Table 6 onwards. If one subsequently assumes that the sample standard deviations are more representative of the population standard deviations and uses the \( t \) criterion, then the results are good except for those of the \( \hat{\sigma} \) estimates. There appears to still be a downwards bias in this statistic for the case where \( \nu = 0.0625 \). This could be due to residual phase variation that has not been accounted
(i) $N = 2048, m = 25, \lambda_0 = \pi/8$.

**TABLE 1**

<table>
<thead>
<tr>
<th>$v_0 = 0.015625$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>5.702</td>
<td>5.829</td>
<td>0.016</td>
<td>0.907</td>
<td>0.0152</td>
</tr>
<tr>
<td>$s_x$</td>
<td>1.233</td>
<td>1.131</td>
<td>0.073</td>
<td>0.025</td>
<td>0.0013</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>1.094</td>
<td>1.094</td>
<td>0.065</td>
<td>0.025</td>
<td>0.0014</td>
</tr>
<tr>
<td>$z$</td>
<td>0.826</td>
<td>1.345</td>
<td>1.095</td>
<td>0.364</td>
<td>1.2414</td>
</tr>
<tr>
<td>$t$</td>
<td>0.733</td>
<td>1.301</td>
<td>0.977</td>
<td>0.355</td>
<td>1.3847</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>24.131</td>
<td>20.295</td>
<td>23.845</td>
<td>20.027</td>
<td>15.2939</td>
</tr>
</tbody>
</table>

**TABLE 2**

<table>
<thead>
<tr>
<th>$v_0 = 0.0625$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>5.418</td>
<td>5.233</td>
<td>-0.013</td>
<td>0.889</td>
<td>0.0631</td>
</tr>
<tr>
<td>$s_x$</td>
<td>0.894</td>
<td>1.040</td>
<td>0.076</td>
<td>0.028</td>
<td>0.0015</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>1.077</td>
<td>1.077</td>
<td>0.067</td>
<td>0.025</td>
<td>0.0015</td>
</tr>
<tr>
<td>$z$</td>
<td>0.341</td>
<td>0.111</td>
<td>0.866</td>
<td>3.560</td>
<td>1.8000</td>
</tr>
<tr>
<td>$t$</td>
<td>0.411</td>
<td>1.148</td>
<td>0.763</td>
<td>3.179</td>
<td>1.8000</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>13.079</td>
<td>17.712</td>
<td>24.456</td>
<td>23.574</td>
<td>19.0127</td>
</tr>
</tbody>
</table>

(ii) $N = 1024, m = 21, \lambda_0 = \pi/4$.

**TABLE 3**

<table>
<thead>
<tr>
<th>$v_0 = 0.015625$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>5.642</td>
<td>5.672</td>
<td>-0.025</td>
<td>0.906</td>
<td>0.0159</td>
</tr>
<tr>
<td>$s_x$</td>
<td>1.058</td>
<td>1.141</td>
<td>0.073</td>
<td>0.028</td>
<td>0.0021</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>1.194</td>
<td>1.194</td>
<td>0.071</td>
<td>0.027</td>
<td>0.0019</td>
</tr>
<tr>
<td>$z$</td>
<td>0.532</td>
<td>0.644</td>
<td>1.556</td>
<td>0.500</td>
<td>0.5858</td>
</tr>
<tr>
<td>$t$</td>
<td>0.600</td>
<td>0.674</td>
<td>1.516</td>
<td>0.486</td>
<td>0.5214</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>14.927</td>
<td>17.369</td>
<td>20.022</td>
<td>20.154</td>
<td>23.9904</td>
</tr>
</tbody>
</table>

**TABLE 4**

<table>
<thead>
<tr>
<th>$v_0 = 0.0625$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>5.503</td>
<td>5.467</td>
<td>0.018</td>
<td>0.893</td>
<td>0.0631</td>
</tr>
<tr>
<td>$s_x$</td>
<td>1.263</td>
<td>0.864</td>
<td>0.082</td>
<td>0.034</td>
<td>0.0028</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>1.176</td>
<td>1.176</td>
<td>0.073</td>
<td>0.027</td>
<td>0.0019</td>
</tr>
<tr>
<td>$z$</td>
<td>0.011</td>
<td>0.126</td>
<td>1.114</td>
<td>2.653</td>
<td>1.4687</td>
</tr>
<tr>
<td>$t$</td>
<td>0.011</td>
<td>0.171</td>
<td>0.991</td>
<td>2.098</td>
<td>1.0206</td>
</tr>
</tbody>
</table>
(iii) \( N = 512, m = 17, \lambda_0 = \pi/2 \).

### TABLE 5

<table>
<thead>
<tr>
<th>( v_0 = 0.015625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.634</td>
<td>5.652</td>
<td>0.018</td>
<td>0.917</td>
<td>0.0155</td>
</tr>
<tr>
<td>( s_x )</td>
<td>1.538</td>
<td>1.175</td>
<td>0.083</td>
<td>0.025</td>
<td>0.0020</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.327</td>
<td>1.327</td>
<td>0.079</td>
<td>0.030</td>
<td>0.0026</td>
</tr>
<tr>
<td>( z )</td>
<td>0.452</td>
<td>0.512</td>
<td>0.994</td>
<td>1.201</td>
<td>0.2557</td>
</tr>
<tr>
<td>( t )</td>
<td>0.390</td>
<td>0.578</td>
<td>0.951</td>
<td>1.405</td>
<td>0.3241</td>
</tr>
</tbody>
</table>

### TABLE 6

<table>
<thead>
<tr>
<th>( v_0 = 0.0625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.926</td>
<td>6.187</td>
<td>0.021</td>
<td>0.894</td>
<td>0.0622</td>
</tr>
<tr>
<td>( s_x )</td>
<td>1.692</td>
<td>1.570</td>
<td>0.099</td>
<td>0.039</td>
<td>0.0037</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.307</td>
<td>1.307</td>
<td>0.081</td>
<td>0.030</td>
<td>0.0026</td>
</tr>
<tr>
<td>( z )</td>
<td>1.458</td>
<td>2.351</td>
<td>1.185</td>
<td>2.238</td>
<td>0.4994</td>
</tr>
<tr>
<td>( t )</td>
<td>1.126</td>
<td>1.957</td>
<td>0.974</td>
<td>1.737</td>
<td>0.3556</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>31.832</td>
<td>27.416</td>
<td>28.125</td>
<td>31.565</td>
<td>37.4778</td>
</tr>
</tbody>
</table>

(iv) \( N = 256, m = 13, \lambda_0 = \pi/2 \).

### TABLE 7

<table>
<thead>
<tr>
<th>( v_0 = 0.015625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.666</td>
<td>5.961</td>
<td>0.005</td>
<td>0.899</td>
<td>0.0153</td>
</tr>
<tr>
<td>( s_x )</td>
<td>1.596</td>
<td>1.508</td>
<td>0.109</td>
<td>0.044</td>
<td>0.0052</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.518</td>
<td>1.518</td>
<td>0.091</td>
<td>0.034</td>
<td>0.0038</td>
</tr>
<tr>
<td>( z )</td>
<td>0.489</td>
<td>1.358</td>
<td>0.253</td>
<td>1.312</td>
<td>0.4308</td>
</tr>
<tr>
<td>( t )</td>
<td>0.465</td>
<td>1.367</td>
<td>0.210</td>
<td>1.008</td>
<td>0.3173</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>21.001</td>
<td>18.754</td>
<td>27.501</td>
<td>32.201</td>
<td>35.0255</td>
</tr>
</tbody>
</table>

### TABLE 8

<table>
<thead>
<tr>
<th>( v_0 = 0.0625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.081</td>
<td>5.183</td>
<td>-0.008</td>
<td>0.874</td>
<td>0.0638</td>
</tr>
<tr>
<td>( s_x )</td>
<td>1.797</td>
<td>1.484</td>
<td>0.107</td>
<td>0.052</td>
<td>0.0052</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.494</td>
<td>1.494</td>
<td>0.093</td>
<td>0.034</td>
<td>0.0039</td>
</tr>
<tr>
<td>( z )</td>
<td>1.254</td>
<td>0.949</td>
<td>0.394</td>
<td>4.565</td>
<td>1.4653</td>
</tr>
<tr>
<td>( t )</td>
<td>1.043</td>
<td>0.955</td>
<td>0.343</td>
<td>3.031</td>
<td>1.1156</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>27.485</td>
<td>18.745</td>
<td>25.094</td>
<td>43.093</td>
<td>32.7841</td>
</tr>
</tbody>
</table>
(v) \( N = 128, \ m = 9, \ \lambda_0 = \pi/2. \)

### TABLE 9

<table>
<thead>
<tr>
<th>( v_0 = 0.015625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.930</td>
<td>5.669</td>
<td>3.194</td>
<td>0.908</td>
<td>0.0139</td>
</tr>
<tr>
<td>( s_x )</td>
<td>2.449</td>
<td>1.926</td>
<td>0.162</td>
<td>0.044</td>
<td>0.0063</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.824</td>
<td>1.824</td>
<td>0.109</td>
<td>0.041</td>
<td>0.0067</td>
</tr>
<tr>
<td>( z )</td>
<td>1.054</td>
<td>0.414</td>
<td>2.133</td>
<td>0.148</td>
<td>1.1621</td>
</tr>
<tr>
<td>( t )</td>
<td>0.785</td>
<td>0.392</td>
<td>1.432</td>
<td>0.138</td>
<td>1.2278</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>34.284</td>
<td>21.193</td>
<td>42.127</td>
<td>22.084</td>
<td>17.0206</td>
</tr>
</tbody>
</table>

### TABLE 10

<table>
<thead>
<tr>
<th>( v_0 = 0.0625 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( \theta )</th>
<th>( \sigma )</th>
<th>( v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} )</td>
<td>5.119</td>
<td>5.321</td>
<td>-0.021</td>
<td>0.877</td>
<td>0.0591</td>
</tr>
<tr>
<td>( s_x )</td>
<td>1.486</td>
<td>1.550</td>
<td>0.210</td>
<td>0.041</td>
<td>0.0180</td>
</tr>
<tr>
<td>( \sigma_x )</td>
<td>1.796</td>
<td>1.796</td>
<td>0.112</td>
<td>0.041</td>
<td>0.0068</td>
</tr>
<tr>
<td>( z )</td>
<td>0.950</td>
<td>0.446</td>
<td>0.834</td>
<td>3.478</td>
<td>2.2212</td>
</tr>
<tr>
<td>( t )</td>
<td>1.148</td>
<td>0.517</td>
<td>0.443</td>
<td>3.466</td>
<td>0.8424</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>13.004</td>
<td>14.148</td>
<td>67.224</td>
<td>19.144</td>
<td>132.0850</td>
</tr>
</tbody>
</table>

### TABLE 11

<table>
<thead>
<tr>
<th></th>
<th>( f_1 )</th>
<th>( f_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x}_1 )</td>
<td>5.936</td>
<td>5.667</td>
</tr>
<tr>
<td>( \bar{x}_2 )</td>
<td>5.100</td>
<td>5.279</td>
</tr>
</tbody>
</table>
for by the estimated group delay. If the group delay was not small then one might expect that a small error in the estimation of $\hat{v}$ might lead to such significant residual variation especially since the chance of such an error occurring is greater the larger the value of $v_0$. In any event the results for $\hat{\theta}$, $\hat{\varphi}$ and $\hat{v}$ seem, generally speaking, to be worse in the case of large group delay. From the statistic given by (5.10) it would appear, intuitively speaking, that the larger $m\hat{v}$ the greater the loss of sample cross-spectral information. Thus for small values of $m$ the loss represented by $m\hat{v}$ could be quite critical. That is, a loss of $6^{1/2}$ from 25 statistics may not amount to much, but a loss of $6^{1/2}$ from 13 or less statistics could be crucial. One suspects that values of $N$ greater than 512 or, perhaps, 256 will be large enough to ensure approximately normal behaviour of the $\omega_j(\pi_k)$. Thus the critical factor seems to be the balance between $m$ and $m\hat{v}$ as mentioned before.

In Table 11 the sample means of the usual estimates of $\hat{\mathbf{f}}_1$, $\hat{\mathbf{f}}_2$, $\hat{\theta}$ and $\hat{\varphi}$ are given for the simulations that form the basis of Tables 9 and 10. Here the $\bar{x}_1$ row applies to the results given in Table 9 and the $\bar{x}_2$ row to those of Table 10. The spectra estimates do not differ to any degree, but the coherence estimate is biased downwards in both cases and this is particularly bad in relation to the case of the larger group delay. Also, in this latter case, the usual estimates of phase fluctuated to a far greater extent that those obtained from the actual maximum likelihood procedure.

These results appear to support the asymptotic theory developed especially when one considers that $N < 500$ is rather rare in most narrow band spectral analyses and $N = 128$ is an extreme case. Moreover, $v_0 = 0.0625$ represents a lag that is larger than would normally be expected in practice.
4.1 Introduction

An estimation problem is now examined where a scalar signal is received together with lagged and attenuated forms of that signal. Let us commence by considering the scalar time series \( s(t) \) and echo times \( t_j, \ j = 1, \ldots, r \), so that the observed process has the form

\[
x(t) = \sum_{j=0}^{r} a_j s(t - t_j) + e(t), \quad a_0 = 1, \ t_0 = 0.
\]

The \( a_j \) are the amplitude factors representing the degree to which the original signal has been attenuated at the \( j \)th echo whereas \( e(t) \) and \( s(t) \) are incoherent, with \( e(t) \) denoting the noise inherent in the system. If \( s(t) \) and \( e(t) \) are stationary and have absolutely continuous spectral distribution functions with spectra \( f_s(\lambda) \) and \( f_e(\lambda) \) respectively, then the spectral density of the \( x(t) \) process is given by

\[
f_x(\lambda) = \left| \sum_{j=0}^{r} a_j e^{it_j \lambda} \right|^2 f_s(\lambda) + f_e(\lambda).
\]

Moreover, it will be assumed that the \( a_j, t_j \) are frequency dependent so that now (1.2) becomes

\[
f_x(\lambda) = \left| \sum_{j=0}^{r} a_j(\lambda) e^{it_j(\lambda) \lambda} \right|^2 f_s(\lambda) + f_e(\lambda).
\]

This latter state of affairs will arise if the medium through which the signal is transmitted is dispersive so that the speed of propagation of the signal is frequency dependent. In this connection the estimation of the \( a_j \) is of some importance since it will give some knowledge of that particular medium. If, for example, the signal is the result of some distant earthquake which, on arrival at the recording apparatus,
consists of the original wavetrain plus echoes caused by reflections from various layers within the earth's crust, then the $t_j$ may give information about the boundaries to these various layers whereas the $a_j$ might relate to the nature of the particular layer involved. Thus (1.3) represents a model of some interest. The problem where the $a_j(\lambda)$, $t_j(\lambda)$ are independent of frequency is not specifically discussed here since this type of formulation might, perhaps, be more adequately treated by some appropriate broad band theory rather than the methods developed in this thesis. It is noted that problems of this latter type are considered in Bogert, Healy and Tukey (1963) and, in particular, Hannan and Robinson (1971).

Now, if we are to make any effective measurement of the $a_j(\lambda)$, $t_j(\lambda)$ then it is evident that we must be able to say something a priori about the various component functions that make up (1.3). The assumptions that we shall require will accord with the basic requirements of Theorem 2.1 in that the $a_j(\lambda)$, $t_j(\lambda)$, $f_s(\lambda)$ and $f_e(\lambda)$ will be taken to be sufficiently smooth over a narrow band of frequencies centred at $\lambda_0$ and the $t_j(\lambda)$ to be sufficiently large over this band for $\sum_{j=0}^{\infty} a_j(\lambda_0) \exp\{it_j(\lambda_0) \lambda\}^2$ to be discernible. If such an assumption is not made then it may be impossible to determine the lag structure $t_j(\lambda_0)$ from the observed spectral density over the narrow band in question.

The series $x(t)$ given by (1.1) is now assumed to have been sampled at unit time intervals with the sampling interval chosen small enough for aliasing effects to be ignored. Furthermore, the $a_j(\lambda)$, $j \neq 0$, in (1.3) will now be regarded as complex functions. Thus, apart from the lag function $t_j(\lambda)$, allowance is also made for some rephasing of the signal components. We mean, by the latter, that even if the components at two frequencies were adjusted relative to each other to
allow for the different lengths of the paths travelled and the
different speeds of propagation, it would be natural to expect that
there would still be phase differences introduced due to distortion in
the echoing process. Under these conditions the spectral density of
the x(t) process near \( \lambda_0 \) may be described as

\[
\left| \sum_{j=0}^{r} a_j(\lambda_0) \exp(it_j(\lambda_0) - \lambda) \right|^2 f_s(\lambda_0) + f_e(\lambda_0) .
\]  

(1.4)

Omitting the argument variable \( \lambda_0 \) since this is held fixed, (1.4) may be
reparameterised to yield

\[
\sigma^2 \left\{ 1 + \sum_{j \neq k} \rho_j \rho_k e^{i(t_j - t_k)(\lambda_0 - \lambda)} \right\}
\]

(1.5)

where

\[
\sigma^2 = \left\{ \sum_{j=0}^{r} |a_j|^2 \right\} f_s + f_e ,
\]

\[
\rho_j = a_j \exp(it_j(\lambda_0) f_s^2/\sigma , \ j = 0, 1, \ldots, r ,
\]

with \( \rho_0 \) real and positive and \( \sum_{j=0}^{r} |\rho_j|^2 < 1 \). To account for the
requirement that the \( t_j \) are so large as to cause discernible variation
in (1.5) over the band we shall assume that the \( t_j \) are of order \( N \) and
shall put

\[
\lim_{N \to \infty} \frac{t_j}{N} = v_j , \quad |v_j| < 1 , \quad j = 1, \ldots, r .
\]

The spectrum (1.5) is now in the form required by Theorem 2.1 in that,

\[
f_x(\lambda_0 + \lambda/N) = \sigma^2 |h(N)(\lambda_0 + \lambda/N)|^2
\]

with

\[
|h(N)(\lambda)|^2 = 1 + \sum_{j \neq k} \rho_j \rho_k e^{i(t_j - t_k)(\lambda - \lambda_0)}
\]

(0)
and \( h^{(N)}(\lambda_0 + \lambda/N) \) is required to converge to \( h(\lambda) \) in the appropriate manner with

\[
|h(\lambda)|^2 = 1 + \sum_{j \neq k} \rho_j \rho_k e^{i(v_j - v_k)\lambda}.
\]

Thus, we have required the process to depend on \( N \) in such a way as to maintain a pattern of echo times \( t_j = Nv_j \). This pattern will now be present in the limiting distribution given by the appropriate form of Theorem 2.1. The quantity \( \rho_j \) has been treated as being independent of \( N \) which may sound a little contradictory since it involves \( t_j \).

However it must be remembered that only one value of \( N \) is experienced and, with reference to Chapter 2, the sequence of data sets indexed by \( N \) is constructed in such a way that an asymptotic approximation is found to the distribution of the statistics \( w(\lambda) \) considered.

We now proceed to prove results relating to the estimation of \( \sigma^2 \) and the \( \rho_j \)\( v_j \) using an analogous procedure to that developed in §3.3 and §3.4. As before, we shall be concerned with the \( m \) frequencies of the form \( \theta_k = \lambda_0 + 2\pi k/N, -\frac{1}{2}m < k \leq \frac{1}{2}m \), so that the band considered has width \( 2\pi m/N \). The various comments with regard to the relative sizes of \( m \) and \( N \) given in §3.1 can also be seen to apply in this context of estimating echo times and need not be repeated here.

Denoting \( w \) as the vector having \( w(\theta_k) \) in the \([\lfloor \frac{1}{2}(m-1) \rfloor + k + 1]^{th}\) place, \( -\frac{1}{2}m < k \leq \frac{1}{2}m \), it is evident from Theorem 2.1 and (2.2.14) that this vector has the asymptotic distribution function

\[
\pi^{-m} \det(F^{-1}) \exp(-w^* \Gamma^{-1} w)
\]

with

\[
\Gamma = \sigma^2 \{ I_m + \sum_{j \neq k} \rho_j \rho_k \Psi(v_k - v_j) \}.
\]

The matrix \( \Psi(v) \) is as for Chapter 3 and we choose to call \( \Gamma/\sigma^2 \) the
We take $0 < v_1 < v_2 < \ldots < v_r < 1$ without any essential restriction and seek to estimate $\rho_j$, $j = 0, 1, \ldots, r$; $v_j$, $j = 1, \ldots, r$ and $\sigma^2$.

One further point needs to be discussed. As yet we have not examined the possibility that there may be alternative sets of the $\rho_j$, $v_j$ and $\sigma^2$ that yield the same likelihood given by (1.6). We now consider this problem and commence by showing that, for $m > 2r(r + 1)$, $\Gamma$ uniquely determines the expression

$$
\phi(\theta) = \sigma^2 \left( 1 + \sum_{j}^{r} \rho_j \rho_k e^{i(v_j - v_k)\theta} \right).
$$

Thus if $\phi_1(\theta)$ and $\phi_2(\theta)$ are two such functions giving rise to the same $\Gamma$ then we require

$$
(2\pi)^{-1} \int_{-\infty}^{\infty} (\phi_1(\theta) - \phi_2(\theta)) Q(\theta) \, d\theta
$$

to be the null matrix. The above matrix is of the form

$$
\sum_{s} \tilde{c}_s \tilde{Q}(w_s),
$$

where $s \leq r(r + 1)$, $\tilde{c}_s = -\tilde{c}_s$, $w_j = -w_k$ and $w_j \neq w_k$ for $j \neq k$. Using the formulae (3.1.2) and (3.1.3) and taking, for example, $m = 2n+1$, it can be seen that when the various elements of (1.9) are set to zero the following system of equations results:

$$
\sum_{s} u_s \sin 2\pi j w_s = 0 , \quad \sum_{s} u_s (1 - w_s)(1 - \cos 2\pi j w_s) = 0 ,
$$

$$
\sum_{s} v_s (1 - \cos 2\pi j w_s) = 0 , \quad \sum_{s} v_s (1 - w_s) \sin 2\pi j w_s = 0 ,
$$

$$
\sum_{s} u_s (1 - w_s) = -ic_0
$$

where $c_s = u_s + iv_s$, $s \neq 0$, and $1 \leq j \leq n$. Now, considering the equations for $v_s$, we note that the determinant of the matrix having
typical entry \((1 - \cos 2\pi jw_k), \ 1 \leq j, k \leq t\), can be written as
\[
\prod_{1 \leq j < k \leq t} (\cos 2\pi w_k - \cos 2\pi w_j) = 0
\]
apart from a constant multiplicative factor. However this can only be
zero if \(w_k + w_j = 1\) for some \(j, k, j \neq k\), since \(0 < w < 1\) for
\(0 \leq \ell \leq n\). Let us suppose that there are \(\tau\) pairs of such \(w_j\) given by
the relations
\[
w_{p_j} + w_{q_j} = 1; \quad p_j, q_j \neq 0, \quad 1 \leq j \leq \tau.
\]
Then
\[
\sum_{1 \leq \ell \leq s} v_{\ell}(1 - \cos 2\pi jw_{\ell}) = \sum_{1 \leq \ell \leq s-\tau} \tilde{v}_{\ell}(1 - \cos 2\pi j\tilde{w}_{\ell}) \quad (1.10)
\]
where the \(\tilde{w}_{\ell}\) contain some of the original \(v_{\ell}\) and there are \(\tau\) \(\tilde{v}_{\ell}\) which
are of the type \(v_{p_j} + v_{q_j}\). The \(\tilde{w}_{\ell}\) are now such that no relation of the
form \(\tilde{w}_j + \tilde{w}_k = 1\) can hold and so \((1.10)\) implies that the \(\tilde{v}_{\ell}\) are
identically zero. Thus the \(v_{\ell}\) are all zero except for the \(v_{p_j}, v_{q_j}\) and
so
\[
\sum_{1 \leq \ell \leq s} v_{\ell}(1 - w_{\ell}) \sin 2\pi jw_{\ell} = \sum_{k=1}^{\tau} v_{p_k}(1 - w_{p_k}) \sin 2\pi jw_{p_k}
\]
\[
+ \sum_{k=1}^{\tau} v_{q_k}(1 - w_{q_k}) \sin 2\pi jw_{q_k}
\]
\[
= \sum_{k=1}^{\tau} v_{p_k} \sin 2\pi jw = 0. \quad (1.11)
\]
However the matrix having typical entry \(\sin 2\pi jw_k, \ 1 \leq j, k \leq t\), has a
determinant given by
\[
\prod_{1 \leq j < k \leq t} (\sin 2\pi w_k - \sin 2\pi w_j) = 1
\]
apart from some constant multiplicative factor. Since the \(w_{p_j}\) of
(1.11) are such that no relation of the type \( w_j + w_k = 1, \ j \neq k, \) can hold it is evident that the remaining \( v_{p,j} \) must be zero. The \( u_j \) may be treated in a similar fashion and so the required result has been established for \( n \geq s \) or, equivalently, \( m > 2(r+1) \). This last condition seems to be too strong since one would expect that the requirement \( m > r(r+1) \) should be enough. Indeed, in the case \( r = 1 \), it is not hard to see that \( m > 2 \) will ensure that \( \Gamma \) uniquely specifies \( \phi(\theta) \).

If, considering the original equations, we can show that the determinant of the partitioned matrix

\[
\begin{bmatrix}
A \\
B
\end{bmatrix}
\]

where

\[
A_{jk} = \sin 2\pi j w_k, \quad B_{jk} = (1 - w_k)(1 - \cos 2\pi j w_k),
\]

\(1 \leq j \leq t, \quad 1 \leq k \leq 2t,\)

is non-zero, then the \( u_j \)'s will all be zero if \( 2n \geq s \). If the analogous statement held for the \( v_j \)'s then we would have the required result; that is \( m > r(r+1) \). This latter supposition looks to be almost certainly true, but has not, as yet, been shown. However \( m > 2r(r+1) \), although not optimal, at least serves to provide an upper bound of \( 2r(r+1) \) to the actual result.

We now consider the function \( \phi(\theta) \) given by (1.8) with regard to the uniqueness of its specification in terms of the \( \rho_j, v_j \) and \( \sigma^2 \).

When the case \( r = 1 \) is considered, for example, there are two sets of \( \rho_j \) that give the same \( \phi(\theta) \) since, putting \( \rho_1 = |\rho_1| e^{i\theta} \),

\[
\sigma^2 \left\{ 1 + \rho_0 \rho_1 e^{-i\theta} + \rho_1 \rho_0 e^{i\theta} \right\}
\]

\[
= \sigma^2 \left\{ 1 + |\rho_1| \left( \rho_0 e^{-i\theta} \right) e^{-i\theta} + |\rho_1| \left( \rho_0 e^{-i\theta} \right) e^{i\theta} \right\}.
\]
In this situation a uniquely defined set of parameters is achieved by putting $\rho = \rho_0 \rho_1$, $|\rho| < 0.5$. For $r = 2$ there are three cases according as $v_1 < v_2 - v_1$, $v_1 > v_2 - v_1$ or $v_1 = v_2 - v_1$. In the first two cases where $v_1$, $v_2$ and $v_2 - v_1$ are all distinct the parameters may be taken to be $\sigma^2$, $v_1$, $v_2$, $\rho_0$, $\rho_1$ and $\rho_2$. It is noted that these two cases in question cannot be coalesced into the one situation where $v_1 \neq v_2 - v_1$ since

$$
\sigma^2 \left\{ 1 + \sum_{j \neq k \neq 0}^\infty \rho_j \rho_{j-k} e^{i(v_j - v_k)\theta} \right\} = \sigma^2 \left\{ 1 + \sum_{j \neq k \neq 0}^\infty \rho'_j \rho'_{j-k} e^{i(v'_j - v'_k)\theta} \right\}
$$

where

$$
v'_0 = v_0, \quad v'_1 = v_2 - v_1, \quad v'_2 = v_2,
$$

$$
\rho'_0 = |\rho_2|, \quad \rho'_1 = \rho_1 e^{i\theta_2}, \quad \rho'_2 = \rho_0 e^{i\theta_2}
$$

and $\theta_2$ is the argument of $\rho_2$. If $v_2 = 2v_1$ we may put

$$
\sigma^2 \left\{ 1 + \sum_{j \neq k \neq 0}^\infty \rho_j \rho_{j-k} e^{i v_1 (j-k)\theta} \right\} = \left| \sum_{j} \gamma_j e^{i v_1 \theta} \right|^2
$$

where $\sum \gamma_j z^j$ has all zeros on or outside the unit circle and now the $\gamma_j$ and $v_1$ may be taken as parameters. For $r = 3$ the situation becomes more complicated. Considering the tableau

<table>
<thead>
<tr>
<th></th>
<th>$v_0$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_0$</td>
<td>0</td>
<td>$v_1$</td>
<td>$v_2$</td>
<td>$v_3$</td>
</tr>
<tr>
<td>$v_1$</td>
<td>$-v_1$</td>
<td>0</td>
<td>$v_2-v_1$</td>
<td>$v_3-v_1$</td>
</tr>
<tr>
<td>$v_2$</td>
<td>$-v_2$</td>
<td>$-(v_2-v_1)$</td>
<td>0</td>
<td>$v_3-v_2$</td>
</tr>
<tr>
<td>$v_3$</td>
<td>$-v_3$</td>
<td>$-(v_3-v_1)$</td>
<td>$-(v_3-v_2)$</td>
<td>0</td>
</tr>
</tbody>
</table>

which indicates the various possibilities for the $v_j - v_k$ we observe that the elements of any sub-diagonal are less than the elements of the
next sub-diagonal above. Thus, considering the identification of the
\( v_j \), it is evident that if all the \( v_j - v_k \) are distinct we have only two
cases according as \( v_2 > v_3 - v_1 \) or \( v_2 < v_3 - v_1 \). In both of these
situations we may take the \( \sigma^2, \rho_j, v_j \) to be the parameters. In the
case where \( v_2 = v_3 - v_1 \) we must now consider the sub-cases where
\( v_1 > v_2 - v_1, v_1 < v_2 - v_1 \) and \( v_2 = 2v_1 \). In the first two a canonical
choice must be made on the basis of the \( \rho_j \)'s whereas in the latter
circumstance we have

\[
\sigma^2 \left\{ 1 + \sum_{j \neq k} \rho_j \rho_k \exp \left( iv_1(j-k)\theta \right) \right\} = \sum_{j=0}^3 \gamma_j \exp \left( iv_1\theta \right)^2
\]

and this is treated in the same way as the analogous case when \( r = 2 \).
It can be seen that the identification problem becomes more complex as
\( r \) increases.

Let us now examine the case of general \( r \). The situation
where \( f_e = 0 \) is considered first and so

\[
\phi(\theta) = \sigma^2 \left| \sum_{j=0}^{r} \rho_j \exp \left( iv_j\theta \right)^2 \right| = \left| \sum_{j=0}^{r} b_j \exp \left( iv_j\theta \right)^2 \right| \tag{1.12}
\]

where \( b_j = \sigma \rho_j, 0 \leq j \leq r, \text{ and } 0 = v_0 < v_1 < v_2 < \cdots < v_r < 1 \). We
wish to determine all possible factorisations of \( \phi(\theta) \) in the knowledge
that \( \phi(\theta) \) is known as a function of \( \theta \) and that such a representation
given by (1.12) exists. Ritt (1927) has shown that \( \phi(\theta) \) can be
decomposed uniquely as

\[
b^2 \prod_{j=0}^{r} |I_j|^2 \prod_{k=0}^{r} |S_k|^2.
\]

Here the \( I_j \) are irreducible components with

\[
I_j = \sum_{t=0}^{r_j} a_{j,t} \exp \left( iw_{j,t} \theta \right), \quad a_{j,0} = 1, \quad 0 = w_{j,0} < w_{j,1} < \cdots < w_{j,r_j} < 1,
\]

and these cannot be factored into factors of a like kind whereas
\[ S_k = \sum_{t=0}^{s_k} \bar{a}_{k,t} e^{i n_k, t \omega_k^\theta}, \quad \bar{a}_{k,0} = 1, \omega_k > 0, \]

with the \( n_{k, t} \) being integers such that \( 0 = n_{k, 0} < n_{k, 1} < \ldots < n_{k, s_k} \).

The \( \omega_k \) are irrationally related by which we mean that \( \omega_k / \omega_{\ell} \) is irrational for \( k \neq \ell \). Considering a typical \( I_j \) we observe that

\[
\tilde{I}_j = \frac{-i \omega_j, r_j \theta}{a_j, r_j} \sum_{t=0}^{\theta} a_{j, t} e^{i(n_j, r_j - \omega_j, t) \theta}
\]

where

\[
a_{j, t} = \frac{a_j, t}{\bar{a}_j, r_j}.
\]

Thus \( \tilde{I}_j \) is a factor of the type \( I_j \) and is indistinguishable from \( I_j \) since

\[
b_0^2 \prod_j |I_j|^2 = \tilde{b}_0^2 \prod_j |\tilde{I}_j|^2
\]

where \( \tilde{b}_0 = b_0 \prod_j |a_j, r_j| \). One of these two alternatives must be included in the \( h(\theta) \) of \( \phi(\theta) = |h(\theta)|^2 \) and hence some rule concerning the selection of one of these irreducible components is required. We could, for example, choose \( I_j \) if \( \omega_j, t > \omega_j, r_j - \omega_j, t' \) some \( t \), and if \( \omega_j, t = \omega_j, r_j - \omega_j, t \) for all \( t \) a choice would then be based on some appropriate criterion concerning the \( a_{j, t} \)'s. Turning to the \( S_k \) we consider the expression

\[
\tilde{\phi}(\theta) = \prod_k |S_k|^2.
\]

Now, putting \( z = i \theta \), we may find a unique representation

\[
\tilde{\phi}(\theta) = |\tilde{h}(i \theta)|^2
\]

where \( \tilde{h}(z) \) has no zeros in the left half plane. (See, for example, Hannan (1970) p.154.) This latter follows since we may choose each \( S_k \)
to have no zeros in the left half plane. Let us denote a typical $S_k$ by

$$S = \sum_{t=0}^{s} a_t e^{n_twz}.$$ 

This can be written as

$$\prod_{j} (1 - u_j e^{wz})$$

which has zeros at

$$z = w^{-1} \{- \log|u_j| + i(\arg(u_j) + 2\pi k)\}.$$ 

However $S$ has the form

$$\prod_{j} (1 - \overline{u}_j e^{-wz})$$

and this has zeros at

$$z = w^{-1} \{ \log|u_j| - i(\arg(u_j) + 2\pi k)\}.$$ 

It is evident that the components $(1 - u_j e^{wz})$ and $(1 - \overline{u}_j e^{-wz})$ are such that one has a zero which lies in the left half plane whereas the other does not. We now select the appropriate factors for $S_k$ and hence obtain the required result. Thus, for the case $f_e = 0$, the model can be parameterised in this rather complex fashion.

When $f_e \neq 0$ we can describe $\phi(\theta)$ as

$$\phi(\theta) = \left| \sum_{j=0}^{r} b_j e^{iv_j \theta} \right|^2 + f_e$$

where the $b_j$, $v_j$ are as for the case $f_e = 0$. However this problem is more difficult since now, for example, there is the possibility that

$$\left| \sum_{j=0}^{r} b_j e^{iv_j \theta} \right|^2 + f_e = \left| \sum_{j=0}^{r} c_j e^{iv_j \theta} \right|^2.$$ 

One might surmise that if $\sum_{j=0}^{r} b_j e^{iv_j \theta}$ was irreducible and some condition relating to the identifiability of the $v_j$'s was imposed (see the
following discussion) then \( \Phi(\theta) \) would be uniquely determined as a function of its parameters. This, however, has not been established and we must now resort to a slightly stronger set of conditions. We consider the tableau

\[
\begin{array}{cccccc}
 & v_0 & v_1 & \cdots & v_{r-2} & v_{r-1} & v_r \\
 v_0 & 0 & v_1 & \cdots & v_{r-2} & v_{r-1} & v_r \\
v_1 & -v_1 & 0 & \cdots & v_{r-1} - v_1 & v_r - v_1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\end{array}
\]

(1.13)

which, as for the case \( r = 3 \), gives the various possibilities for the \((v_j - v_k)\). The case of interest is that where the elements of this tableau are all distinct and this will be so if

\[ v_k - v_j \neq v_t - v_s, \quad 0 \leq j < k \leq r, \ 0 \leq s < t \leq r, \]

with the possibility \( s = j, t = k \) excluded. Examining the various possibilities in the above we are led to require that

\[ \sum n_j v_j \neq 0, \quad n_j = 0, \pm 1, \pm 2. \]  

(1.14)

Hence we are excluding a finite number of "surfaces" in the space of the \( v_j \)'s. However, with probability one a priori, no parameter point will occur on such a surface nor, for sufficiently accurate computation, will a point fall on it. Thus the restriction is of no consequence.

In any case, for fixed \( r \), the process can be exhaustively investigated in the manner given for \( r = 3 \). Assuming that (1.14) holds we consider the problem of identifying the \( v_j, 1 \leq j < r \), given that we know only the differences \( v_j - v_k, 0 \leq j, k \leq r \). It will be shown that there exist only two possible selections for \( v_j, 1 \leq j \leq r \), and these are
given by \( v_j, 1 \leq j \leq r \), and \( v_r - v_{r-j}, 1 \leq j \leq r \). Selecting \( v_r \) is clearly no problem but, considering (1.13), we see that there are two possibilities for the choice of \( v_{r-1} \). These are \( v_{r-1} \) and \( v_r - v_1 \). Denoting our choices for the \( v_j \) as \( v'_j, 1 \leq j \leq r \), we consider the selection \( v'_{r-1} = v_{r-1} \). This implies that \( v'_r - v'_1 = v_r - v_1 \) which requires \( v'_1 \) to be \( v_1 \). Since we now have \( v'_1, v'_{r-1}, v'_r \) we can identify \( v'_{r-1} - v'_1 = v_{r-1} - v_1 \) and so we must select \( v'_{r-2} \) from \( v_{r-2} \) and \( v_r - v_2 \). Let us proceed by induction and assume that we have \( v'_j, v'_{r-j}, 0 \leq j \leq t \) given by \( v_j, v_{r-j}, 0 \leq j \leq t \). Clearly, we now have to choose \( v'_{r-t-1} \) from \( v_{r-t-1} \) and \( v_r - v_{t+1} \). Suppose we take \( v'_{r-t-1} = v_r - v_{t+1} \) which implies that \( v'_{r-t-1} - v'_1 = v_r - v_{t+1} - v_1 \). However, on examination of the appropriate sub-diagonal, it can be seen that \( v_r - v_{t+1} - v_1 \) must now occur among \( v_{r-t-2}, v_{r-t-1} - v'_1, v_{r-1} - v_{t+1} \) and \( v_r - v_{t+2} \). This is impossible because of (1.14) and hence we must choose \( v'_{r-t-1} = v_{r-t-1} \). Thus the initial selection \( v'_{r-1} = v_{r-1} \) leads to the single possibility given by \( v_j, 1 \leq j \leq r \). The selection \( v'_{r-1} = v_r - v_1 \) yields a similar result except that now the resulting possibility is given by \( v_r - v_{r-j}, 1 \leq j \leq r \). It can be seen, therefore, that if we can distinguish between \( v_t \) and \( v_r - v_{r-t} \) for some \( t \) then we will have uniquely determined the \( v_j \). We note, as an aside, the well known fact that when two notes of different frequencies are sounded together the listener hears the two original notes together with the difference of these two frequencies. This latter is called the beat frequency. Thus an analogy can be drawn between the identification of beat frequencies and the choice of the \( v_j \) as described above. This completes the discussion on the identification of the model and, for the remaining sections of this chapter, we shall require (1.14) to hold together with \( v_1 > v_r - v_{r-1} \).
4.2 A Strong Law of Large Numbers

The vector of all parameters \( \tau \) is defined by \( \tau' = (\sigma^2, \nu') \)
where \( \nu' = (\rho', \nu') \) and \( \rho \) and \( \nu \) are the vectors having typical elements
\( \rho_j, \nu_j \) respectively. The logarithm of the likelihood given by (1.6) is
denoted by \( L \) and so

\[
L = -m \log \gamma - \log \det(\Gamma) - w^* \Gamma^{-1} w. \tag{2.1}
\]

The likelihood equations obtained from (2.1) by differentiating with
respect to the various parameters are complicated and do not lead to
any major simplification. The one exception to this last statement is
the maximising value for \( \sigma^2 \) which is evidently

\[
\hat{\sigma}^2 = m^{-1} w^* \hat{\Omega}^{-1} w \tag{2.2}
\]

where \( \hat{\Omega}_m \) is the maximum likelihood estimate of \( \Omega \).

We now proceed to find the asymptotic distribution of \( \hat{\tau}_m \),
which is the maximum likelihood estimate of \( \tau \), taking (1.6) to be the
true distribution of the \( w(\theta_k) \). A zero subscript is again used to
indicate the true parameter set and it is assumed that \( \rho_0^* \rho_0 < 1 \),
\( 0 < \nu_{10} < \nu_{20} < \ldots < \nu_{r0} < 1 \), with the model identified in the way
described in §4.1. It is noted that we have not notationally dis­
tinguished between the element of \( \rho \) represented by \( \rho_0 \) and the vector of
ture values \( \rho_0 \). However it will, in all cases, be evident as to which
of these definitions applies.

Theorem 4.1

Let the \( w(\theta_k) \) be generated by a complex normal process with
probability density given by (1.6) for \( \tau = \tau_0 \). Then \( \hat{\tau}_m \) converges
almost surely to \( \tau_0 \) as \( m \) tends to infinity. Moreover, \( m(\hat{\nu}_m - \nu_0) \)
converges almost surely to zero.
Proof

Throughout this proof convergence will be taken to mean almost sure convergence, but the qualifying phrase will be omitted.

With reference to (2.1) and (2.2) we are led to consider

\[-\log(m^{-1} w^* \Omega^{-1} w) - m^{-1} \log \det(\Omega)\]

and, in particular, adding the constant \(m^{-1} \log \det(\Gamma_0)\), we introduce

\[Q_m(\nu) = \log \sigma^2_0 + m^{-1} \log \det(\Omega_0^{-1}) - \log(m^{-1} w^* \Omega^{-1} w)\]

Considering first the case where \(\rho^* \rho \leq 1-\delta, \delta > 0\), it is observed that \(\Omega\) has eigenvalues which are uniformly bounded above by \((r + 1)\) and uniformly bounded below by \(\delta\). Indeed, with reference to §2.2 and §4.1,

\[
\Omega = (1 - \rho^* \rho) I_m + (2\pi)^{-1} \int_{-\infty}^{\infty} \left| \sum_{j=0}^{r} \rho_j e^{iv_j \theta} \right|^2 Q(\theta) \, d\theta
\]

and since \(z^* Q(\theta) z\) is not less than zero for all complex vectors \(z\) it is evident that

\[z^* \Omega z / z z \geq 1 - \rho^* \rho \geq \delta > 0\]

and so the minimum eigenvalue of \(\Omega\) is greater than \(\delta\). The bound for the largest eigenvalue follows in a similar manner since

\[
z^* \Omega z / z z = 1 - \rho^* \rho + (2\pi)^{-1} \int_{-\infty}^{\infty} \left| \sum_{j=0}^{r} |\rho_j|^2 \right|^2 \{z^* Q(\theta) z / z z \} \, d\theta
\]

\[\leq 1 - \rho^* \rho + (r + 1) \left( \sum_{j=0}^{r} |\rho_j|^2 \right) (2\pi)^{-1} \int_{-\infty}^{\infty} \{z^* Q(\theta) z / z z \} \, d\theta
\]

\[= 1 + r \rho^* \rho < r + 1\]

It can also be seen from the above that

\[A_m(\nu) = \sum_{j \neq k} \rho_j \rho_k \Psi(v_k - v_j)\]
has eigenvalues which lie between \(-\rho \rho\) and \(\rho \rho\). Now the matrix \(\Omega\) may be written as

\[(r + 1)[I_m - B_m(V)]\]

where

\[B_m(V) = (r + 1)^{-1} [r I_m - A_m(V)]\]

has eigenvalues that are uniformly bounded below by \(\delta - \delta/(r + 1)\) and uniformly bounded above by \(1 - \delta/(r + 1)\). Thus, for arbitrary \(\epsilon > 0\), there exists an \(M\) such that

\[\|\Omega^{-1} - (r + 1)^{-1} \sum_{j=0}^{M} B_m^j(V)\| < \epsilon\]

(2.3)

uniformly in \(V\).

We now wish to establish that

\[\eta_m(V) = m^{-1} w^* \Omega^{-1} w - \delta [m^{-1} w^* \Omega^{-1} w]\]

converges uniformly to zero and commence by noting that this quantity is bounded above by

\[|m^{-1} w^* (\Omega^{-1} - (r + 1)^{-1} \sum_{j=0}^{M} B_m^j(V)) w|\]

\[+ \sigma_0^2 m^{-1} \text{tr}[(\Omega^{-1} - (r + 1)^{-1} \sum_{j=0}^{M} B_m^j(V)) \Omega_0]\]

\[+ (r + 1)^{-1} \sum_{j=0}^{M} |m^{-1} w^* B_m^j(V) w - \sigma_0^2 m^{-1} \text{tr}(B_m^j(V) \Omega_0)|\].

The sum of the first two terms in the above can be made arbitrarily small, uniformly in \(V\), since it is not greater than

\[\epsilon [m^{-1} w^* w + \sigma_0^2 (m^{-1} \text{tr}(\Omega_0^2))]^{\frac{1}{2}}\]

However \(B_m^j(V)\) is just the sum of components of the type...
where the $s_i, t_i$ are integers and each component is multiplied by some appropriate constant factor. Hence it will be sufficient to show that the third term or, equivalently,

$$|m^{-1} w^* \left( \prod_{i} \Psi(v_{t_i} - v_{s_i}) \right) w - \frac{\sigma_0^2}{m} \operatorname{tr}[\left( \prod_{i} \Psi(v_{t_i} - v_{s_i}) \right) \Omega_0]| \quad (2.5)$$

converges uniformly to zero. Observing that

$$\{\Lambda(v_1) \phi(v_2)\}_{j,k} = \left\{ \int_{-\min(0,v_1)}^{1-\max(0,v_2)} e^{i2\pi(j-k)(\theta-v_1)} d\theta \right\} e^{-i2\pi kv_1}$$

where $\mathcal{F}$ is a matrix of the same type as the $H_j$ of Lemma 3.1, it can be seen that the $\Lambda$ matrices in (2.4) may be multiplied through to yield an expression of the form

$$\left[ \prod_{i} \rho_{s_i,t_i}^* \mathcal{F}_i \right] \Lambda(\tilde{v})$$

with the $\mathcal{F}_i$ representing the appropriate $\mathcal{F}$ matrices and

$$\tilde{v} = \sum_{i} (v_{t_i} - v_{s_i}).$$

Now (2.5) is dominated by

$$|m^{-1} w^* \left( \prod_{i} \mathcal{F}_i - \mathcal{F}_0 \right) \Lambda(\tilde{v}) w| + \frac{\sigma_0^2}{m} \operatorname{tr}[\left( \prod_{i} \mathcal{F}_i - \mathcal{F}_0 \right) \Lambda(\tilde{v}) \Omega_0]$$

$$+ |m^{-1} w^* \mathcal{F}_0 \Lambda(\tilde{v}) w - \frac{\sigma_0^2}{m} \operatorname{tr}[\mathcal{F}_0 \Lambda(\tilde{v}) \Omega_0]|$$

where $\mathcal{F}_0$ denotes the appropriate $H$ matrix in the sense of Lemma 3.1.

Thus, using Schwartz's inequality together with this lemma, we see, with reference to (3.3.1), that the first two terms of the above
expression converge uniformly in \( v \) to zero. The proof that the third term of the above converges uniformly in \( v \) to zero now follows in a completely analogous fashion to the result proved in relation to the \( \eta_m(v) \) of Theorem 3.1.

We may now replace \( \log[m^{-1} \omega^{-1}] \) in \( Q_m(v) \) by \( \log[\omega^2 m^{-1} \text{tr}(\Omega_0 \Omega^{-1})] \) since the argument of this last logarithm is uniformly bounded away from zero. The resulting expression will be called \( R_m(v) \) with

\[
R_m(v) = m^{-1} \text{tr} \left[ \log(\Omega_0 \Omega^{-1}) \right] - \log[m^{-1} \text{tr}(\Omega_0 \Omega^{-1})] \leq 0.
\]

Here the inequality follows from that which holds between the arithmetic and geometric means of the eigenvalues of \( \Omega_0 \Omega^{-1} \). We now proceed to show, for every \( v \neq v_0 \), that \( R_m(v) \) is strictly negative and that

\[
\limsup_{m \to \infty} R_m(v) \leq a_\varepsilon < 0. \tag{2.6}
\]

This rather indirect way of proceeding is adopted because the convergence of \( R_m(v) \) to its limit is both complex and non uniform. Indeed the limit function will have singularities, possibly at an infinite set of points, in the space over which \( v \) varies. The largest of the singularities is at the point \( v = v_0 \) and it is the presence of this which explains the rapid convergence of \( \hat{\varphi}_m \) to \( v_0 \). (See Theorem 4.2.) These singularities arise, for example, from the form of \( m^{-1} \text{tr}(\Omega_0 \Omega^{-1}) \) which, considering (2.3) and (2.4), may be written as a sum of components of the type

\[
m^{-1} \text{tr} \left[ \rho_{i_0} \rho_{k_0} \varphi(v_{k_0} - v_{j_0}) \Pi \left( \rho_i \rho_{t_i} \varphi(v_{t_i} - v_{s_i}) \right) \right]. \tag{2.7}
\]

Once again we may multiply the \( \Lambda \) matrices through and we can replace (2.7) by
where
\[ \tilde{\mathbf{v}}_0 = \mathbf{v}_{ko} - \mathbf{v}_{jo} + \sum \mathbf{t}_i \mathbf{s}_i \]
and \( \tilde{\mathbf{v}}_0 \) represents the appropriate \( \mathbf{H} \) matrix in the sense of Lemma 3.1.

However (2.8) can now be written as the product of two functions; the first is a continuous function of \( \rho, \rho_0, \mathbf{v} \) and \( \mathbf{v}_0 \) whereas the second is the function
\[ \sin m \pi \tilde{\mathbf{v}}_0 / (m \sin \pi \tilde{\mathbf{v}}_0) . \]

It is this latter that gives rise to the singularities since it will be unity if
\[ \tilde{\mathbf{v}}_0 = \mathbf{v}_{ko} - \mathbf{v}_{jo} + \sum \mathbf{t}_i \mathbf{s}_i = 0 \pmod{1} \]
and will converge to zero otherwise. The last condition can be described equivalently by
\[ \mathbf{v}_{jo} - \mathbf{v}_{ko} = \sum \sum n_{st} \mathbf{v}_t \mathbf{v}_s \pmod{1}, \quad (2.9) \]
where the \( n_{st} \) are integers, and thus there will be singular behaviour in the limit function at a parameter point with such values for the \( \mathbf{v}_j \).

In any case (2.6) ensures that \( \mathbf{v}_m \) converges to \( \mathbf{v}_0 \) for if that were not true then there would be a subsequence of \( m \) values, \( m(j) \) let us say, such that \( \mathbf{v}_{m(j)} \) converges to \( \mathbf{v}_1 \neq \mathbf{v}_0 \). If \( \| \mathbf{v}_1 - \mathbf{v}_0 \| > \varepsilon \) then
\[ \lim R_{m(j)}(\mathbf{v}_{m(j)}) \leq a_{\varepsilon} < 0. \]
However
\[ \lim R_{m(j)}(\mathbf{v}_{m(j)}) \geq \lim R_{m(j)}(\mathbf{v}_0) = 0, \]
and a contradiction is obtained, thus establishing the first part of the theorem.
Let \( \lambda_{m,j} \) be the \( j \)th eigenvalue of \( \Omega_0^{\frac{1}{2}} \Omega_0^{-1} \Omega_0^{\frac{1}{2}} \) with \( \lambda_{m,1} \leq \lambda_{m,2} \leq \cdots \leq \lambda_{m,m} \). Since \( \rho^* \leq 1-\delta \), \( \delta \geq 0 \), it is evident that there exist constants \( a \) and \( \delta_0 \) such that \( \lambda_{m,1} \geq \delta_0 > 0 \) and \( \lambda_{m,m} \leq a < \infty \). Here \( a \) is determined by \( \Omega \) as well as \( \Omega_0 \) but \( \delta_0 \) depends only on \( \Omega_0 \). Denoting the mean and variance of the \( \lambda_{m,j} \) by \( \overline{\lambda}_m \) and \( s_m^2 \) respectively we observe that

\[
 s_m^2 = m^{-1} \text{tr}((\Omega_0 - \overline{\lambda}_m I_m)^2) 
\]

\[
 \geq (r + 1)^{-2} m^{-1} \text{tr}((\Omega_0 - \overline{\lambda}_m \Omega)^2) 
\]

since the greatest eigenvalue of \( \Omega \) is less than or equal to \( r + 1 \).

We wish to prove that

\[
 \lim_{m \to \infty} s_m^2 \geq c_\varepsilon > 0, \quad \|v - v_0\| > \varepsilon .
\]

Now

\[
 \Omega_0 = I_m + \sum_{j \neq k=0}^{r} \rho_{jko} \overline{\Omega}(v_{ko} - v_{jo}) 
\]

(2.10)

can be written as

\[
 \sum_{-t}^{t} c_{j,o} \overline{\Omega}(w_{j,o}), \quad c_{0,0} = 1 ,
\]

(2.11)

where \( c_{j,o} = c_{-j,o}^* \), \( w_{j,o} = -w_{-j,o}^* \), and \( w_{j,o} \neq w_{k,o} \) for \( j \neq k \). From the discussion given in §4.1 concerning the identification of the model it can be seen that (2.11) uniquely determines the representation (2.10).

It is also observed that \( \Omega \) may be put in the form

\[
 \sum_{-s}^{s} c_{j} \overline{\Omega}(w_j), \quad c_0 = 1 ,
\]

(2.12)

where \( c_{j} = c_{-j} \), \( w_{j} = -w_{-j} \), and \( w_{j} \neq w_{k} \) for \( j \neq k \). However, from previous considerations we know that

\[
 m^{-1} \text{tr}(\overline{\Omega}(v_1)^* \overline{\Omega}(v_2)) = m^{-1} \text{tr}(\Phi(v_2)^* \Phi(v_1) \Lambda(v_1 - v_2))
\]
converges to zero for $v_1 \neq v_2$ and to $1 - |v_1|$ for $v_1 = v_2$ and so

$$m^{-1} \text{tr}[(\Omega_0 - \overline{\lambda}_m^2)^2]$$

approaches

$$\sum' |c_{j,0}|^2 (1 - |w_{j,0}|) + \sum'' |c_j|^2 (1 - |w_j|) + \sum''' |c_{j,0} - \overline{\lambda}_m \tilde{c}_j|^2 (1 - |w_{j,0}|).$$

Here $\sum'$ is over those $j$ such that $w_{j,0}$ is not equal to any $w_j$, $\sum''$ is over those $j$ such that $w_j$ is not equal to any $w_{j,0}$ and $\sum'''$ is over those $j$ where $w_j$ does equal some $w_{j,0}$ with $\tilde{c}_j$ representing the corresponding $c_j$ and $\tilde{c}_0 = c_0$. Now the first term will be strictly positive in all cases where there are some $w_{j,0}$ that do not equal any $w_j$. In the case where each $w_{j,0}$ equals a $w_j$ it can be seen that the sum of the last two terms will only be zero when $\overline{\lambda}_m$ approaches unity, $\sum''$ is null and $\tilde{c}_j$ equals $c_{j,0}$. However, since $\|v - v_0\| > \epsilon$ implies that $\tilde{c}_j \neq c_{j,0}$ each non zero $j$, this last event cannot happen. Thus there exists a constant $c_\epsilon$ such that

$$\lim s_m^2 > c_\epsilon > 0, \quad \|v - v_0\| > \epsilon.$$

Now

$$\log(1 + x) \leq x - (2b)^{-1} x^2, \quad -1 \leq x \leq b - 1, \quad b > 1,$$

so that

$$\log(\lambda_{m,j}/\overline{\lambda}_m) \leq (\lambda_{m,j}/\overline{\lambda}_m - 1) - \frac{\overline{\lambda}_m}{2a} (\lambda_{m,j}/\overline{\lambda}_m - 1)^2$$

and

$$m^{-1} \sum \log \lambda_{m,j} - \log \overline{\lambda}_m \leq \frac{s_m^2}{(2a \overline{\lambda}_m)}.$$

Since $\overline{\lambda}_m$ is uniformly bounded above by $a$, the right hand side of the last expression is bounded above by $-c_\epsilon/(2a^2)$ which we shall call $a_\epsilon$. Thus
\[
\lim \sup_m \{ m^{-1} \sum_{j} \log \lambda_{m,j} - \log \bar{\lambda}_m \} \leq a_\varepsilon < 0.
\]

However the bracketed expression is just \( R_m(\nu) \) and we have established what we wish.

The case where \( \rho^* \rho \) is not required to lie below \( 1-\delta \) is now examined and a subsequence of values of \( m \) is considered along which \( \hat{\rho}_m^* \hat{\rho}_m \) approaches unity. If it can be shown that along such a subsequence \( \lim Q_m(\hat{\nu}_m) \) is strictly negative then a contradiction will have been attained as before. We observe that \( Q_m(\nu) \) may be denoted as \(-\log x_m \) where

\[
x_m = m^{-1} \sum \lambda_{m,j} |\xi_j|^2/(\prod_j \lambda_{m,j})^{1/m}
\]

and the \( \xi_j \) are independent complex random variables having zero means and unit variances. However \( x_m \) has exactly the same form as the \( x_m \) given in the development for the case \( \lim \hat{\sigma}_m^2 = 1 \) of §3.3. Thus the proof that \( \lim x_m \) is strictly greater than unity will follow in an analogous fashion if the variance of the \( \lambda_{m,j}^{-1} \) can be shown to be strictly positive. The proof of this latter parallels that given for \( s_m^2 \) and hence the condition \( \rho^* \rho \leq 1-\delta, \delta > 0 \), can now be imposed without cost.

It remains only to prove that \( m(\hat{\nu}_m - \nu_0) \) converges to zero and \( \hat{\sigma}_m^2 \) converges to \( \sigma_0^2 \). We first observe that

\[
\lim Q_m(\hat{\nu}_m) \geq \lim Q_m(\nu_0) = 0.
\]

However

\[
\lim Q_m(\hat{\nu}_m) = \lim R_m(\hat{\nu}_m) \leq 0
\]

and so

\[
\lim Q_m(\hat{\nu}_m) = 0.
\]
It is evident from what has been proved before that this can only happen if

\[
\lim_{m \to \infty} m^{-1} \text{tr}\{(\Omega_0 - \frac{1}{m} \Omega(\hat{V}_m))^2\} = 0 \tag{2.13}
\]

where now we know that \(\hat{V}_m\) converges to \(V_0\). However, in the notation given by (2.11) and (2.12) and for \(\hat{V}_m\) close to \(V_0\) it is evident that

\[
m^{-1} \text{tr}\{(\Omega_0 - \frac{1}{m} \Omega(\hat{V}_m))^2\}
\]

\[
= m^{-1} \sum_{-t}^{t} \sum_{-t}^{t} c_{j,o} \hat{c}_k \text{tr}\{(\xi(w_j,o) - \frac{1}{m} \xi(\hat{w}_j))(\xi(w_k,o) - \frac{1}{m} \xi(\hat{w}_k))\}^{*}
\]

differs from

\[
\sum_{-t}^{t} |c_{j,o}|^2 (1 - |w_j,o|) \left\{1 + \frac{\sigma^2}{m} - 2\frac{\sin\pi(\hat{w}_j - w_j,o)}{m \sin\pi(\hat{w}_j - w_j,o)}\right\}
\]

by a term which can be made arbitrarily small as \(\hat{V}_m\) approaches \(V_0\).

Since this expression must converge to zero both the limit superior and limit inferior must go to zero and so it can be seen that \(\frac{1}{m}\) converges to unity and \(m(\hat{w}_j - w_j,o)\) converges to zero. This ensures that \(m(\hat{V}_m - V_0)\) converges to zero and also that \(\sigma^2\), which approaches \(\sigma^2_0\), converges to \(\sigma^2_0\). This completes the proof.

4.3 A Central Limit Theorem

The asymptotic normality of the estimates is established by the following theorem. Here we shall assume that \(\rho_j = |\rho_j| e^{i\theta_j}\), \(1 \leq j \leq r\), and that \(\tau\) is now a \((3r + 2)\)-dimensional vector of the form \((\sigma^2, \rho_0, |\rho|', \theta', v')'\) with \(|\rho|\) and \(\theta\) having typical elements \(|\rho_j|\) and \(\theta_j\) respectively.

Theorem 4.2

Under the conditions of Theorem 4.1 \(D_m(\hat{\tau}_m - \tau_0)\) is asymptotically normal with zero mean vector and covariance matrix \(I^{-1}\).
where the typical element of $I$ is given by

$$I_{jk} = \lim_{m} \left[ D_{m,j}^{-1} D_{m,k}^{-1} \left( \Gamma_{m,j}^{-1} \frac{\partial \Gamma_{m,j}}{\partial r_{j}} \Gamma_{m,k}^{-1} \frac{\partial \Gamma_{m,k}}{\partial r_{k}} \right) \right]_{0}$$

and $D_{m}$ is diagonal with $m^{\frac{3}{2}}$ in the first $2(r + 1)$ places on the main diagonal and $m^{\frac{3}{2}}$ in the remaining $r$ places.

**Proof**

As in Theorem 3.2 we let $\frac{\partial L}{\partial r}$ denote the vector whose $i^{th}$ component is $\frac{\partial L}{\partial r_{i}}$ where $L$ is given by (2.1). Expanding $\frac{\partial L}{\partial r}$ in a Taylor series about $\hat{r}_{m}$ we have

$$D_{m}^{-1} \frac{\partial L}{\partial r} \bigg|_{0} = G(\bar{r}_{m}) D_{m} (\hat{r}_{m} - r_{0})$$

where $\bar{r}_{m}$ is a random vector such that $\|\bar{r}_{m} - r_{0}\| \leq \|\hat{r}_{m} - r_{0}\|$ and the typical element of $G(\tau)$ is given by

$$G_{jk}(\tau) = -D_{m,j}^{-1} \frac{\partial^{2} L}{\partial r_{j} \partial r_{k}} D_{m,k}^{-1}.$$  

The existence of $\bar{r}_{m}$ is established as for Lemma 3 of Jennrich (1969).

To show that $G(\bar{r}_{m})$ converges to $I$ it will be sufficient to prove the result for a particular element of $G(\bar{r}_{m})$ since the proof that the remaining elements converge to their stated limits can be established in a similar fashion. Examining the diagonal element corresponding to $\tau_{i} = \nu_{i}$ we see that this is

$$-m^{-3} \text{tr} \left\{ \frac{\partial^{2} \Gamma_{m}^{-1}}{\partial \nu_{i} \partial \nu_{j}} (\Gamma_{m} - \nu \nu^{*}) \right\} \bigg|_{\bar{r}_{m}} - m^{-3} \text{tr} \left\{ \frac{\partial \Gamma_{m}^{-1}}{\partial \nu_{i}} \frac{\partial \Gamma_{m}}{\partial \nu_{j}} \right\} \bigg|_{\bar{r}_{m}}.$$  

The second term of (3.1) converges to the appropriate element of $I$ and to show this we consider

$$m^{-3} \text{tr} \left( \Omega_{m}^{-1} \frac{\partial \Omega_{m}}{\partial \nu_{i}} \Omega_{m}^{-1} \frac{\partial \Omega_{m}}{\partial \nu_{j}} \right) \bigg|_{\bar{r}_{m}} - m^{-3} \text{tr} \left( \Omega_{m}^{-1} \frac{\partial \Omega_{m}}{\partial \nu_{i}} \Omega_{m}^{-1} \frac{\partial \Omega_{m}}{\partial \nu_{j}} \right) \bigg|_{0}.$$  

(3.2)
Here

\[
\frac{\partial \Omega}{\partial v_1} = \sum_{j \neq 1}^{\infty} \left[ \rho_j \frac{\partial}{\partial v_1} (\mathbb{P}(v_1 - v_j)) + \rho_j^* \frac{\partial}{\partial v_1} (\mathbb{P}(v_1 - v_j)^*) \right]
\]

and, as indicated in the proof of Theorem 3.2,

\[
\frac{\partial \mathbb{P}(v_1 - v_j)}{\partial v_1} = -\text{sign}(v_1 - v_j) \frac{\partial}{\partial v_1} \mathbb{U}(v_1 - v_j) \mathbb{U}' - i2\pi \mathbb{P}(v_1 - v_j) \mathbb{D}
\]

with \( \mathbb{U}(\theta) \) as for Lemma 3.1, \( \mathbb{U} = \mathbb{U}(0) \), and \( \mathbb{D} \) is diagonal with \( j \)th entry \( j - \left[ \frac{1}{2}(m-1) \right] - 1 \). Since

\[
\left\| \frac{\partial \mathbb{P}(v_1 - v_j)}{\partial v_1} \right\| \leq \left\| \mathbb{U}(v_1 - v_j) \right\| \mathbb{U}' \right\| + 2\pi \left\| \mathbb{P}(v_1 - v_j) \right\| \mathbb{D} \|
\]

\[
\leq (\pi + 1) m
\]

it is clear that

\[
\left\| \frac{\partial \Omega}{\partial v_1} \right\| \leq \kappa m
\]

for some suitable constant \( \kappa \). To simplify the notation the subscript \( T \) will denote that the implied expression is to be evaluated at \( \tau_m \) and so \( \left[ \frac{\partial \Omega}{\partial v_1} \right]_T \) will be equivalent to \( \partial \Omega_T / \partial v_1 \). Now (3.2) is given by

\[
m^{-3} \text{tr} \left\{ (\Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} - \Omega_0^{-1} \frac{\partial \Omega_0}{\partial v_1}) \Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \right\}
\]

\[
+ m^{-3} \text{tr} \left\{ \left( \Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} - \Omega_0^{-1} \frac{\partial \Omega_0}{\partial v_1} \right) \Omega_0^{-1} \frac{\partial \Omega_0}{\partial v_1} \right\}
\]

and the first term of the above may, for example, be written as

\[
m^{-3} \text{tr} \left\{ (\Omega_0 - \Omega_T) \Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \right\}
\]

\[
+ m^{-3} \text{tr} \left\{ \left( \frac{\partial \Omega_T}{\partial v_1} - \frac{\partial \Omega_0}{\partial v_1} \right) \Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \Omega_0^{-1} \right\}. \tag{3.3}
\]

Considering the first expression of (3.3) we observe that this is dominated by
\[ m^{-2} \left( m^{-1} \text{tr}(\Omega_0 - \Omega_T)^2 \right)^{\frac{3}{2}} \|\Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \| \|\Omega_0^{-1} \frac{\partial \Omega_0}{\partial v_1} \| \leq \kappa^2 \|\Omega_T^{-1}\|^2 \|\Omega_0^{-1}\| \left( m^{-1} \text{tr}(\Omega_0 - \Omega_T)^2 \right)^{\frac{3}{2}} \]

and, since \( \hat{\beta}_m \beta_m < 1 \), \( \|\Omega_T^{-1}\|^2 \) is bounded above and so this latter term converges to zero by virtue of (2.13). The second term of (3.3) is dominated by

\[
\left\{ m^{-3} \text{tr} \left( \frac{\partial \Omega_T}{\partial v_1} - \frac{\partial \Omega_0}{\partial v_1} \right)^2 \right\}^{\frac{3}{2}} \|\Omega_T^{-1} \frac{\partial \Omega_T}{\partial v_1} \| \|\Omega_0^{-1} \frac{\partial \Omega_0}{\partial v_1} \|
\]

\[
\leq \kappa' \sum_{j \neq 1} \left\{ m^{-3} \text{tr} \left( \left( \rho_j \rho_T \frac{\partial \psi_T}{\partial v_1} (v_1 - v_j) - \rho_j \rho_{10} \frac{\partial \psi_0}{\partial v_1} (v_1 - v_j) \right) \right) \right\}^{\frac{3}{2}}
\]

\[
\leq \kappa' \sum_{j \neq 1} |\rho_j \rho_T| \left| \rho_{10} \right| \left( m^{-2} (U(v_1T - v_jT) - U(v_{10} - v_{j0})) \right) \left( m^{-3} \text{tr}(\psi(v_1T - v_jT) - \psi(v_{10} - v_{j0})) \right)^{\frac{3}{2}}
\]

\[
+ 2\pi \kappa' \sum_{j \neq 1} |\rho_j \rho_T| \left| \rho_{10} \right| \left( m^{-3} \text{tr}(\psi(v_1T - v_jT) - \psi(v_{10} - v_{j0})) \right)
\]

\[
\times D^2(\psi(v_1T - v_jT) - \psi(v_{10} - v_{j0})) \right)^{\frac{3}{2}}
\]

(3.4)

where

\[
\kappa' = 2\kappa \|\Omega_T^{-1}\| \|\Omega_0^{-1}\| .
\]

Now the second last term of (3.4) is bounded above by

\[
m^{-\frac{1}{2}} \left\{ 2\kappa' \sum_{j \neq 1} |\rho_j \rho_T| \left| \rho_{10} \right| \right\}
\]
whereas the last term is not greater than

\[
\pi \kappa' \sum_{j\neq 1}^r |\rho_j \rho_0| |\rho_1| \left[ m^{-1} \text{tr}\left( (\Psi(v_1 - v_j) - \Psi(v_0 - v_j))\right) \right] \times (\Psi(v_1 - v_j)^* - \Psi(v_0 - v_j)^*) \right]^{1/2}.
\]

Since \( \tau_m \) converges to \( \tau_0 \) and \( m(v_m - v_0) \) converges to zero it is clear that (3.4) converges to zero as required.

We now examine the first expression in (3.1) and note that this can be written as

\[
m^{-3} \text{tr}\left\{ \frac{\partial^2 \Gamma^{-1}}{\partial v_1^2} (\Gamma_0 - \Gamma) \right\}_{\tau_m} + \left[ m^{-3} w \frac{\partial^2 \Gamma^{-1}}{\partial v_1^2} w - m^{-3} \text{tr}\left( \Gamma_0 \frac{\partial^2 \Gamma^{-1}}{\partial v_1^2} \right) \right]_{\tau_m}
\]

where

\[
\frac{\partial^2 \Gamma^{-1}}{\partial v_1^2} = 2 \Gamma^{-1} \frac{\partial \Gamma}{\partial v_1} \Gamma^{-1} \frac{\partial \Gamma}{\partial v_1} \Gamma^{-1} - \Gamma^{-1} \frac{\partial^2 \Gamma}{\partial v_1^2} \Gamma^{-1}
\]

with

\[
\frac{\partial^2 \Gamma}{\partial v_1^2} = \sigma^2 \sum_{j \neq 1}^r \left[ \rho_j \rho_1 \frac{\partial^2 \Psi(v_1 - v_j)}{\partial v_1^2} + \rho_1 \rho_j \frac{\partial^2 \Psi(v_1 - v_j)}{\partial v_1^2} \right]
\]

and

\[
\frac{\partial^2 \Psi(v_1 - v_j)}{\partial v_1^2} = 12 \pi \text{ sign}(v_1 - v_j) \left[ D \frac{U(v_1 - v_j)}{U'(v_1 - v_j)} U' + \frac{U(v_1 - v_j)}{U'(v_1 - v_j)} U' D \right] - 4 \pi^2 \Psi(v_1 - v_j) D^2.
\]

Observing that

\[
\left\| \frac{\partial^2 \Psi(v_1 - v_j)}{\partial v_1^2} \right\| \leq 4 \pi \| D \| \left\| \frac{U(v_1 - v_j)}{U'(v_1 - v_j)} U' \right\| + 4 \pi^2 \left\| \Psi(v_1 - v_j) D^2 \right\|
\]

\[
\leq 2 \pi m^2 \pi + m^2 \pi^2
\]

it is evident that the first term of (3.5) is bounded above by
and this converges to zero as a result of (2.13). The last expression in (3.5) converges almost surely to zero and so we have established the first part of this proof.

It remains to be shown that $D_m^{-1} \frac{\partial L}{\partial \tau} |_{\tau_0}$ is asymptotically normal with zero mean vector and covariance matrix I. Putting $V_2 = m^{-\frac{1}{2}}$ where $\delta_i$ is one or three depending on the element chosen we consider the linear combination

$$z_m = \sum_{i=1}^{3r+2} \alpha_i \frac{-\delta_i/2}{m} \frac{\partial L}{\partial \tau_i} |_{\tau_0}$$

$$= \sum_{i=1}^{3r+2} \alpha_i \frac{-\delta_i/2}{m} \left\{ w^* \Gamma_0^{-1} \frac{\partial \Gamma_0}{\partial \tau_j} \Gamma_0^{-1} w - \text{tr} \left( \Gamma_0^{-1} \frac{\partial \Gamma_0}{\partial \tau_j} \right) \right\}$$

$$= \sum_{j=1}^{m} \mu_{m,j} (|\xi_j|^2 - 1)$$

where the $\xi_j$ are independent complex random variables each having zero mean and unit variance and the $\mu_{m,j}$ represent the eigenvalues of $T_\alpha = \sum_{i=1}^{3r+2} \alpha_i m^{-\frac{1}{2}} \frac{\partial \Gamma_0}{\partial \tau_i} \Gamma_0^{-\frac{1}{2}}$.

Since $2|\xi_j|^2$ is just a chi-square random variable with two degrees of freedom it is evident that the characteristic function of $z_m$ is just

$$\exp \left[ - \sum_{j=1}^{m} \left\{ \log(1 - i\mu_{m,j} t) + 1^{\mu_{m,j}} t \right\} \right].$$

Noting that $\|T_\alpha\|$ is $O(m^{-\frac{3}{2}})$ we may expand out the logarithmic term in the above and obtain

$$\exp \left[ - \frac{1}{2} \sum_{j=1}^{m} \mu_{m,j}^2 t^2 + O(m^{-\frac{3}{2}}) \right].$$

However
which converges to

\[ \sum_{i,j} \alpha_i \alpha_j I_{ij} \]

as required. Thus \( D_m \) is asymptotically normal with covariance matrix \( I \) which in turn implies that \( D_m (\tau_m - \tau_0) \) approaches a multi-variate normal distribution with zero mean vector and covariance matrix \( I^{-1} \). This completes the proof.

The elements of \( I \) do not admit to much simplification and indeed the only notable exceptions are the first diagonal element which is \( \sigma_0^{-4} \) and the elements whose normalising factor is \( m^2 \) and these have value zero. To see the latter we consider, for example,

\[ m^{-2} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \nu_i} \Omega^{-1} \frac{\partial \Omega}{\partial \theta_j} \right) \]

and note that this is just the sum of terms of the type

\[ m^{-2} \text{tr} \left[ \rho_j \rho_i (\text{sign}(v_i - v_j) \overline{U(v_i - v_j)} U' + i2\pi \Psi(v_i - v_j) D) \right] \]

\[ \times \prod_{i,j} \rho_{s_i t_i} \Psi(v_i - v_{s_i}) \]

where the \( s_i, t_i \) are integers and each of these terms given above is multiplied by some appropriate constant factor. With reference to the discussion resulting from (2.4) it is clear that we need only examine an expression such as

\[ m^{-2} \text{sign}(v_i - v_j) U' \Phi \Lambda(\tilde{v}) \overline{U(v_i - v_j)} + i2\pi m^{-2} \text{tr}(\Phi \Lambda(\tilde{v}) D) \quad (3.6) \]

where \( \Phi, \tilde{\Phi} \) denote the appropriate H matrices in the sense of Lemma 3.1 and

\[ \tilde{v} = \sum_{i} (v_{t_i} - v_{s_i}), \quad \nu = \tilde{v} + v_i - v_j. \]
The first term of (3.6) is $O(m^{-1})$ whereas the second term converges to zero for all $v$ since

$$m^{-2} \left\{ \frac{d}{d\theta} \left( \sum_j e^{i2\pi j\theta} \right) \right\}_v$$

is also $O(m^{-1})$.

On a less rigorous basis we observe that if the $v_j$ are so small that

$$\phi(v_k - v_j) = \int_{-\min(0, v_k - v_j)}^{1 - \max(0, v_k - v_j)} U(\theta) U(\theta)^* d\theta$$

then $\phi(v_k - v_j)$ may be approximated by $\Delta(v_k - v_j)$. In this situation $\Gamma$ is diagonal with typical entry

$$\sigma^2 \left( 1 + \sum_{j \neq k} \rho_j e^{i2\pi t(v_j - v_k)} \right) = \sigma^2 |h(2\pi t)|^2, \quad -\frac{1}{2} m < t \leq \frac{1}{2} m,$$

and so the matrix $I$ now has elements given by

$$I_{jk} = \lim_m \left[ D_m^{-1} D_m, j \sum_k \left\{ \frac{\sigma^2 |h(2\pi t)|^2}{\sigma^4 |h(2\pi t)|^4} \frac{\partial}{\partial t} (\sigma^2 |h(2\pi t)|^2) \right\} \right].$$

This throws a little more light on the structure of $I$ and hence $I^{-1}$.

The above information matrix would serve, no doubt, as a good approximation in this not unimportant case where the $v_j$ are all small yet still able to cause discernible variation in the spectral density over the narrow band in question.

4.4 Some Numerical Studies

The problem of obtaining the maximum likelihood estimates in a practical situation is now discussed with reference to certain
computer simulations. From the development given in §4.2 it can be seen that maximising (2.1) is equivalent to minimising

\[ \bar{L} = \log \sigma^2 + m^{-1} \log \det(\Omega) \]  

(4.1)

where

\[ \sigma^2 = m^{-1} w^* \Omega^{-1} w . \]  

(4.2)

Thus the equations that must be solved to determine the minimum of (4.1) are given by

\[ m^{-1} w^* \Omega^{-1} \frac{\partial \Omega}{\partial \tau_i} \Omega^{-1} w/\sigma^2 - m^{-1} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \tau_i} \right) = 0 \]  

(4.3)

for \( 2 \leq i \leq 3r+2 \). The optimising value for \( \sigma^2 \) will be given by (4.2) evaluated at the resulting \( \hat{\tau}_i \), \( 2 \leq i \leq 3r+2 \). We have, for convenience, omitted the subscript \( m \) on \( \hat{\tau}_i \). The principal method used to obtain the solution of (4.3) was the one of scoring as detailed in Rao (1965), Chapter 5. Hence we choose to iterate successively as

\[ \hat{\tau}^{(k+1)}_i = \hat{\tau}^{(k)}_i - \sum_j \left[ I_{ij} \left( \frac{\partial \bar{L}}{\partial \tau_j} \right) \right] \hat{\tau}^{(k)}_i , \quad 2 \leq i, j \leq 3r+2 , \]

where \( \hat{\tau}^{(k)}_i \) denotes the \( k \)th iterate of \( \hat{\tau}_i \) and

\[ \hat{I}_{ij} = m^{-1} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \tau_i} \Omega^{-1} \frac{\partial \Omega}{\partial \tau_j} \right) - \left[ m^{-1} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \tau_i} \right) \right] \left[ m^{-1} \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \tau_j} \right) \right] \]

This method differs from the standard Newton-Raphson procedure in that the matrix of second derivatives of \( \bar{L} \) which has typical element

\[ \frac{\partial^2 \bar{L}}{\partial \tau_i \partial \tau_j} , \quad 2 \leq i, j \leq 3r+2 , \]

is approximated by

\[ \delta \left[ \frac{\partial^2 \bar{L}}{\partial \tau_i \partial \tau_j} \right] \hat{\tau}^{(k)} \approx \hat{I}_{ij} \hat{\tau}^{(k)} \]

for \( \hat{\tau}^{(k)} \) sufficiently close to \( \tau_0 \). In this sense the procedure has the disadvantage that, if the data is sufficiently badly behaved, \( \hat{I}_{ij} \hat{\tau}^{(k)} \)
may differ significantly from the actual matrix of second derivatives even at \( \hat{\gamma}(k) = \tau_0 \). Moreover, the number of iterations required to obtain satisfactory convergence is often greater than the corresponding number for the Newton-Raphson procedure. However the main advantage of the scoring technique in this situation is the relative ease of computation of the matrix \( \tilde{I} \).

Regardless of the procedure chosen one must have some mechanism for determining a first estimate of \( \hat{\gamma} \). Considering Lemma 3.4 and, in particular, (3.5.9) and (3.5.10) it is observed that we may approximate \( m^{-1} w^* \Psi(v)^* w \) by \( m^{-1} w^* \Lambda(v)^* T_j m T^* w \) which can be written as

\[
m^{-1} \sum_t \frac{\xi(t)}{\xi(t + mv)}
\]

where the summation is over those \( t \) such that

\[
\max(1, -[mv]) \leq t \leq \min(m, [m(1-v)])
\]

and

\[
\xi(t) = m^{-\frac{1}{2}} \sum_{[m]} w(\theta_k) \exp(-i2\pi kt / m).
\]

(4.4)

In this situation \( m^{-1} w^* \Psi(v)^* w \) has expectation

\[
\sigma_0^2 m^{-1} \text{tr}(\Omega_0 \Psi(v)^*) = \sigma_0^2 m^{-1} \text{tr}(\Psi(v)^*)
\]

\[
+ \sigma_0^2 \sum_{j \neq k} \rho_{j \rightarrow k} m^{-1} \text{tr}(\Psi(v_{ko} - v_{jo}) \Psi(v)^*)
\]

(4.5)

and this converges to

\[
\begin{cases}
\sigma_0^2 & (v = 0) \\
\sigma_0^2 \rho_{j \rightarrow k} (1 - |v_{ko} - v_{jo}|) & (v = v_{ko} - v_{jo}, j \neq k, 0 \leq j, k \leq r) \\
0 & (\text{otherwise})
\end{cases}
\]
Thus, for $v \geq 0$, we are led to determine the $\frac{1}{2} r(r + 1) + 1$ principal maxima of the function

$$c(v) = (m(1 - v))^{-1} \sum_{t=1}^{\lfloor m(1-v) \rfloor} \xi(t) \xi(t + mv)$$

(4.6)

since $c(0)$ gives an estimate of $\hat{\sigma}^2$ and $c(\hat{\nu}_k - \hat{\nu}_j)/c(0)$, $k > j$, gives an estimate of $\hat{\sigma}_{jk}$ where $\hat{\nu}_k - \hat{\nu}_j$ denotes a point at which one of the maxima occurs. These resulting estimates can be disentangled by the identification process described in §4.1 to give the required estimates.

We note that the approximation of $\phi(v)$ by $T^*_{m,m}$ given by Lemma 3.4 is essentially the same as replacing

$$\phi(v) = \int_{-\min(0,v)}^{1-\max(0,v)} U(\theta) U(\theta)^* d\theta$$

by the approximating sum

$$\min\left( m, \lfloor m(1-v) \rfloor \right) \sum_{t=\max(1,-[mv])} \left( U(t/m) U(t/m)^* \right)$$

Thus a modification of the above may be considered where $\phi(v)$ is replaced by

$$\min\left( M, \lfloor M(1-v) \rfloor \right) \sum_{t=\max(1,-[Mv])} \left( U(t/M) U(t/M)^* \right)$$

for some positive integer $M$. Now $m^{-1} \sum \overline{\psi}(v) \psi(w)$ may be approximated by

$$m^{-1} \sum \tilde{\xi}(t) \xi(t + Mv)$$

where the summation is over those $t$ such that

$$\max(1, -[Mv]) \leq t \leq \min(M, \lfloor M(1 - v) \rfloor)$$

and

$$\tilde{\xi}(t) = M^{-\frac{1}{2}} \sum_{-\left[ \frac{1}{2} (m-1) \right]}^{\left[ \frac{1}{2} m \right]} w(\theta_k) \exp(-i2\pi kt/M).$$

Hence, in place of $c(v)$, one might consider the various maxima of
\[
\tilde{c}(v) = (m(1-v))^{-1} \sum_{t=1}^{[M(1-v)]} \tilde{\xi}(t) \tilde{\xi}(t+Mv). \tag{4.8}
\]

The virtue of these first estimation methods is that one may use the fast Fourier transform procedure to effect the backwards transformation of the type (4.7) and then estimates can be obtained in the manner detailed above. This can be done for any particular \( M \) and so \( m^{-1} \Psi(v)^* w \) may be conveniently calculated for values of \( v \) of the form \( j/M \). The resulting \( \tilde{c}(j/M) \) would then be scanned for the appropriate maxima. The basic limiting factor on the size of \( M \) chosen is, in terms of computation, the calculation of the \( \tilde{c}(j/M) \) from the \( \tilde{\xi}(t) \).

However one would probably not need to consider all values of the type \( j/M \) since an experimenter would no doubt have some idea of the size of lag that he is hoping to model and, for example, might only need consider those \( j/M \) less than 0.25. We also note that since the transformation from the \( w(\varphi_k) \) to the \( \tilde{\xi}(t) \) is of rank \( m \), we are still essentially dealing with \( m \) statistics and, in this sense, the last method does not yield any significant improvement over the former method implied by (4.6). However, in terms of the minimisation problem, the transformation with \( M > m \) provides a useful method for evaluating \( m^{-1} \Psi(v)^* w \) at the points \( j/M \).

We illustrate the above with reference to the simple echo situation given by

\[
x(n) = s(n) + \alpha s(n - N') + e(n) \tag{4.9}
\]

where \( s(n) \) and \( e(n) \) are incoherent with zero means and the uniform spectra given by 5.0 and 0.5 respectively. The constant \( \alpha \) was chosen to be 0.4 and the record length consisted of 1024 time points with the band of frequencies in question centred at \( \pi/4 \). This latter contained 25 fundamental frequencies. Thus the spectrum of the \( x(n) \) series is
given by
\[ f(\lambda) = 5 |1 + 0.4 e^{iN'\lambda}|^2 + 0.5 \]

which implies, in terms of our model, that the appropriate \( \Gamma \) matrix is

given by
\[ \sigma_0^2 \left[ I_m + \rho_0 e^{i\theta_0} \bar{\psi}(v_0) + e^{-i\theta_0} \psi(v_0)^* \right] \]

where \( \sigma_0^2 = 6.3 \), \( \rho_0 = 0.318 \), \( \theta_0 = -N'\pi/4 \) and \( v_0 = N'/1024 \). This is consistent with the type of situation that we hope to model in that, over the narrow band, the echoed signal is of high intensity relative to the noise.

We first examine a case where \( v_0 = 0.125 \) which represents a moderately large lag of 128 time points. Hence, it can be seen that \( \theta_0 \) is zero since this parameter is measured mod(2\( \pi \)). The first estimation procedure represented by (4.4) and (4.6) was used and this took less than two seconds to give the following first estimates; namely
\( \hat{\sigma}^2 = 6.49 \), \( \hat{\rho} = 0.29 \), \( \hat{\theta} = -0.6 \) and \( \hat{v} = 0.12 \). Iterating on these values further refined them to yield the results listed below where the 95% confidence intervals quoted were calculated using \( \hat{\Lambda} \) evaluated at the true parameter point. Each iteration took approximately 10 seconds and 8 iterations were required to obtain accuracy to three decimal places. The final estimates and the various 95% confidence intervals were

\[ \hat{\sigma}^2 = 6.77 \ , \ 4.02 \leq \sigma_0^2 \leq 9.52 \ ; \]
\[ \hat{\rho} = 0.30 \ , \ 0.11 \leq \rho_0 \leq 0.49 \ ; \]
\[ \hat{\theta} = -0.59 \ , \ -1.40 \leq \theta_0 \leq 0.22 \ ; \]
\[ \hat{v} = 0.137 \ , \ 0.118 \leq v_0 \leq 0.157 \ . \]

We now consider the same model given by (4.9) and examine the two situations where \( v_0 = 0.0625 \) and \( v_0 = 0.03125 \). The former
represents a lag of moderate size whereas the latter is small. The parameter $\theta_0$ is zero in both of the cases under study and the results about to be quoted were based on 20 independent realisations of each of the models concerned. The first estimation method described in the previous example was used in the $v_0 = 0.0625$ situation and this gave good results. Again, the time taken to accomplish this was invariably less than two seconds. For the $v_0 = 0.03125$ case the procedure indicated by (4.7) and (4.8) was used with $M$ set at 256. However it was found that when $\tilde{c}(v)$, or equivalently $m^{-1} w^* \Psi(v)^* w$, was scanned for values of $v$ less than $1/m$, a problem arose in that it became difficult to distinguish the subsidiary peak due to $v_0 < 1/m$ from the main peak at $v = 0$. With reference to (4.5) it can be seen that the main peak has the form $c_0^2 (m(1-v))^{-1} \text{tr}[(\Psi(v))^*]$ and this will be the dominant term of $\tilde{c}(v)$ if $\rho_0$ is small compared to unity and $v < 1/m$. Hence, one way around this is to consider, in place of $\tilde{c}(v)$, the modified statistic

$$\tilde{c}(v) - \tilde{c}(0) (m(1-v))^{-1} \text{tr}[\Psi(v)^*]$$

since this effectively removes the peak at the origin. This method was adopted and it took approximately 5.5 seconds to scan this function over the values $j/256$, $1 \leq j \leq 32$. The resulting first estimates were reasonable in all cases. It is noted, by way of comparison, that it took nearly 24 seconds to evaluate the function

$$m^{-1} w^* \Psi(v)^* w - (m^{-1} w^* w) ((m(1-v))^{-1} \text{tr}(\Psi(v)^*))$$

at the points $j/100$, $1 \leq j \leq 12$. The times relating to the iteration procedure are the same as those quoted for the case where $v_0 = 0.125$. However, for $v_0 = 0.03125$, the scoring technique failed to converge in 7 out of the 20 occasions. This was due, no doubt, to the occurrence
of significant differences between $\bar{I}$ and the matrix of second derivatives. These seven cases were treated with an I.B.M. subroutine based on the minimisation method of Fletcher and Powell (1963). This procedure took, in each case, approximately two minutes to give the results to the required accuracy.

The displayed tables list the results obtained for each set of simulations. The rows in these tables have precisely the same interpretation as the rows of Tables 1 to 10. (See §3.5 and, in particular, (3.5.14).) The critical values for the $z$, $t$ and $\chi^2$ statistics are given by (3.5.15). In the case where $v_0 = 0.0625$ it is evident that the sample standard deviations, although erratic, do not differ significantly from their theoretical values. The $z$ and $t$ statistics for this table are again not significant except for those corresponding to $\hat{\sigma}$ which are just significant at the 5% level. In Table 13 the $\chi^2$ statistic clearly indicates that the standard deviations differ significantly from their postulated values. Taking the sample standard deviations as being more representative of the population standard deviations it can be seen, in this case, that the $t$ statistics are not significant except for the one corresponding to $\hat{\sigma}^2$. In general, the estimation procedure is considerably less effective in the case of the smaller $v_0$. If the procedure indicated by (4.6) is representative of the actual maximum likelihood procedure then one would expect poor estimation to result if $v_0$ is large and also if $v_0$ is quite small. The first extreme follows in that $\hat{\sigma}$, $\hat{\sigma}$ and $\hat{\nu}$ will be determined from $c(\hat{\nu})$ which summarises information from $[m(1 - \hat{\nu})]$ statistics whereas the latter extreme comes about because it will become increasingly difficult to disentangle $c(\hat{\nu})$ from $c(0)$ if $\hat{\nu}$ is small. Indeed, this latter was the reason that $\tilde{c}(v)$ had to be modified in the way described previously. These observations are also
<table>
<thead>
<tr>
<th>$v_0 = 0.0625$</th>
<th>$\sigma^2$</th>
<th>$\rho$</th>
<th>$\theta$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>6.751</td>
<td>0.362</td>
<td>0.143</td>
<td>0.0609</td>
</tr>
<tr>
<td>$s_x$</td>
<td>1.418</td>
<td>0.069</td>
<td>0.317</td>
<td>0.0089</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>1.508</td>
<td>0.095</td>
<td>0.357</td>
<td>0.0072</td>
</tr>
<tr>
<td>$z$</td>
<td>1.336</td>
<td>2.107</td>
<td>1.796</td>
<td>0.9709</td>
</tr>
<tr>
<td>$t$</td>
<td>1.421</td>
<td>2.916</td>
<td>2.021</td>
<td>0.7775</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>16.788</td>
<td>9.914</td>
<td>15.004</td>
<td>29.6291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$v_0 = 0.03125$</th>
<th>$\sigma^2$</th>
<th>$\rho$</th>
<th>$\theta$</th>
<th>$v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>7.639</td>
<td>0.350</td>
<td>-0.108</td>
<td>0.0530</td>
</tr>
<tr>
<td>$s_x$</td>
<td>2.183</td>
<td>0.082</td>
<td>0.541</td>
<td>0.0624</td>
</tr>
<tr>
<td>$\sigma_x$</td>
<td>5.730</td>
<td>0.611</td>
<td>0.380</td>
<td>0.0351</td>
</tr>
<tr>
<td>$z$</td>
<td>1.045</td>
<td>0.237</td>
<td>1.277</td>
<td>2.7661</td>
</tr>
<tr>
<td>$t$</td>
<td>2.743</td>
<td>1.771</td>
<td>0.896</td>
<td>1.5565</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>2.757</td>
<td>0.340</td>
<td>38.604</td>
<td>60.0080</td>
</tr>
</tbody>
</table>

Intuitively obvious in that if a time series of finite length is correlated with a lagged form of the same series then echo detection will be difficult if the signal and echo are far apart and will be virtually impossible if they are very close together. It appears then, from the results given, that for $v_0 < 1/m$ the estimation problem becomes very difficult. However, in the light of this difficulty, the $t$ criterion of Table 13 gives partially acceptable results.
5.1 Introduction

A vector generalisation of the type of model discussed in the preceding chapters is now examined. It is assumed as before that the situation is such that any spectral analysis must be confined to some narrow band of frequencies centred about a chosen frequency. Thus, if \( x(t) \) is the \( p \)-dimensional vector whose \( j \)th component is the measurement at time point \( t \) of the \( j \)th recorder, then

\[
x(t) = \sum_{j=0}^{r} A_j s(t - t_j) + e(t), \quad t_0 = 0,
\]

where the \( A_j \) are \( p \times q \) matrices and the \( q \)-dimensional vector \( s(t) \) represents the original signal. The noise inherent in the system is denoted by the \( p \)-dimensional vector \( e(t) \) which is assumed to be incoherent with \( s(t) \) and the \( t_j \) represent the appropriate lags. This type of model includes the particular circumstance where the vector signal is received by some spatial arrangement of recorders or antennae. This spatial arrangement could be three-dimensional but the more usual type of array would consist, perhaps, of a linear arrangement of recorders separated by certain specified distances or a two-dimensional array with a given coordinate structure. The study of such arrays has been extensively discussed by research workers in other fields such as seismology for example. (See Birtill and Whiteway (1965).) This has led to a number of methods of analysis of data originating from such arrays among which will be found the sum-squared response method and the delay-sum-correlate method. A detailed description of such methods is presented in Birtill and Whiteway (1965).
The procedures developed in this chapter will present a statistical approach to this problem based on the theory given in the preceding chapters. We shall call \( f_x(\lambda) \) the \( p \times p \) spectral density matrix of the \( x(t) \) series, \( f_s(\lambda) \) the \( q \times q \) spectral density matrix of the \( s(t) \) series and \( f_e(\lambda) \) the \( p \times p \) spectral density matrix of the \( e(t) \) series. As in §4.1 we assume that the \( A_j \) are complex and that the \( A_j \), \( t_j \) are frequency dependent. Moreover, the \( A_j(\lambda) \), \( t_j(\lambda) \) together with \( f_s(\lambda) \) and \( f_e(\lambda) \) are required to be smooth over the narrow band of \( m \) frequencies in question that are centred about the chosen frequency \( \lambda_0 \). Thus the spectral density matrix \( f_x(\lambda) \) may be described, near \( \lambda_0 \), as

\[
H(\lambda) f_s(\lambda_0) H(\lambda)^* + f_e(\lambda_0)
\]

where

\[
H(\lambda) = \sum_{j=0}^{r} A_j(\lambda_0) \exp[\imath t_j(\lambda_0) \lambda] .
\]

The \( t_j(\lambda_0) \) are now taken to be large enough to cause discernible variation in \( H(\lambda) \) across the band. These assumptions are consistent with those required to enable an accurate measurement to be made with an interferometer and would be accomplished with such a device by ensuring, for example, that the antennae were spread over a long enough distance so that a small change, say, in the direction of a signal will result in an appreciable alteration in the phase differentials at the antennae. Thus the problem is closely bound up with the making of measurements by means of interference effects.

Sampling the \( x(t) \) process at unit time intervals and assuming that aliasing effects can be ignored we now take the \( t_j(\lambda_0) \) to be of order \( N \), the number of time points, and shall require, as before, that

\[
\lim_{N \to \infty} \frac{t_j(\lambda_0)}{N} = v_j , \quad |v_j| < 1 , \quad j = 1, \ldots, r .
\]

The constant \( \exp[\imath t_j(\lambda_0) \lambda_0] \) is absorbed into \( A_j(\lambda_0) \) and, assuming that
the conditions of Theorem 2.1 are met, we see that the vector \( w \) which has \( w_j(\theta_k) \) as the \( ((j - 1)m + k + \lfloor \frac{(m - 1)}{2} \rfloor + 1) \) th component has an asymptotic distribution function given by

\[
\pi^{-mp} \det(\Gamma^{-1}) \exp(-w^* \Gamma^{-1} w)
\]

(1.4)

with

\[
\Gamma = \sum_{j,k=0}^{r} \left[ A_j(\lambda_0) f_s(\lambda_0) A_k(\lambda_0)^* \right] \otimes \Psi(v_k - v_j) + f(e(\lambda_0) \otimes I_m.
\]

(1.5)

The matrix \( \Gamma \) may be reparameterised to accord with the type of model postulated and the new parameters so chosen must be individually distinguishable in order that they may be correctly identified as in the case described in Chapter 4.

We proceed by the method of maximum likelihood and hence (1.4) is now maximised with respect to the parameters in \( \Gamma \) so that the maximum likelihood estimates of the true values of these parameters might be ascertained. The establishment of the statistical properties of these estimates parallels the development given in the proofs of Theorems 4.1 and 4.2. Indeed, the maximisation of (1.4) is clearly equivalent to minimising

\[
L_m = \{\exp((mp)^{-1} w^* \Gamma^{-1} w)\}/\{\det(\Gamma_0 \Gamma^{-1})\}^{mp}
\]

(1.6)

where \( L_m \) converges almost surely to \( e \) when \( \Gamma = \Gamma_0 \). Hence, if it can be shown that

\[
\lim L_m > e
\]

outside some arbitrarily small closed neighbourhood containing the vector of true parameter values then the result will follow from the development given in the proof of Theorem 4.1. However (1.6) has the same form as (3.3.7) and so the proof for this vector case can now be
established in precisely the same way as the proof for the scalar case. Thus the results of Theorems 4.1 and 4.2 may be carried over to include the vector situation.

As a final note to this introductory section we observe that since we have effectively allowed m to increase it is evident that the basic condition for the application of this theory is the requirement that

\[
\min_{1 \leq j \leq r} \{m t_j(\lambda_0)/N\}
\]

should cause discernible variation in the spectral density function over the narrow band of interest. This condition is once again of the type required in order to make an effective measurement with an interferometer. In the sections that follow we shall be concerned with certain special cases of this general vector model given by (1.1) with particular reference to the situation mentioned before where the various observations are received by an array of recorders or antennae.

5.2 The Estimation Problem

We consider now an array of \((r + 1)\) recorders which receives lagged and attenuated forms of some common scalar signal together with noise. Thus, in terms of the formulation given by (1.1) we are requiring that \(q\) be unity and

\[
x(t) = (s(t), a_1 s(t - t_1), \ldots, a_r s(t - t_r))' + \epsilon(t) \quad (2.1)
\]

with the \(a_j\) denoting the various attenuation coefficients. The basic assumptions described in §5.1 are once again applied here and, furthermore, we shall take \(\epsilon(t)\) as having components that are incoherent with one another. This latter implies that \(\hat{f}_e(\lambda)\) is diagonal with \(j^{th}\) entry \(\hat{f}_{e,j}(\lambda)\). Hence, putting
\[ b_j = a_j(\lambda_0) \exp(it_j(\lambda_0) \omega_0) \]

and

\[ h(\lambda) = (1, b_1 e^{it_1(\lambda)(\lambda-\lambda_0)}, \ldots, b_r e^{it_r(\lambda)(\lambda-\lambda_0)}), \]

(1.2) becomes

\[ h(\lambda) h(\lambda)^* f_s(\lambda_0) + f_e(\lambda_0). \]

Omitting the argument variable \( \lambda_0 \) since this is held fixed it is now evident that \( f \) has typical sub-matrix

\[ f_s b_j b_k \Psi(v_k - v_j) \]

for \( j \neq k \) and

\[ (f_s |b_j|^2 + f_e, j) I_m \]

for \( j = k \) where \( 0 \leq j, k \leq r \).

Let the figure described below depict a general two-dimensional array and let the velocity of the signal be \( V \) with the \( j \)th recorder located at the point \( P \) which has polar coordinates \( (r_j, \theta_j) \). Relating the various leads and lags to the arbitrary origin at \( 0 \) we observe that, relative to this point, the \( j \)th recorder has a lag of \( t_j = (r_j \cos(\phi - \theta_j))/V \).

For convenience, we now select the origin of the reference frame to be relocated at that particular recorder which has the minimum lag and relabel the others so that \( v_0 = 0 < v_1 < v_2 < \ldots < v_r < 1 \) is true.

Noting that \( \Psi(v_k - v_j) \) can be written as
\[ A(v_j)^* Q(v_j,v_k) A(v_k) \]

(2.3)

where the typical element of \( Q(v_j,v_k) \) is given by

\[
\int_0^1 q(v_j,v_k;\theta) e^{i2\pi(s-t)\theta} \, d\theta, \quad -\frac{1}{2m} < s, t \leq \frac{1}{2m},
\]

with

\[
q(v_j,v_k;\theta) = \begin{cases} 
0 & \min(1-v_j,1-v_k) \leq \theta \leq \max(1-v_j,1-v_k) \\
1 & \text{otherwise}
\end{cases}
\]

we can reparameterise \( T \) to yield

\[
\Gamma = A^* Q A.
\]

(2.4)

Here \( A \) is diagonal with \( A(v_j) \) as the \((j+1)\)th sub-matrix along the main diagonal and \( Q \) has typical sub-matrix

\[
\sqrt{f_j f_k} \sigma_{jk} \exp(i\theta_{jk}) Q(v_j,v_k)
\]

for \( j \neq k \) and

\[
f_j I_m
\]

for \( j = k \). Also, for \( 0 \leq j, k \leq r \),

\[
\sigma_{jk} = f_s |b_j| |b_k|/\sqrt{f_j f_k} = \sigma_{kj},
\]

\[
\theta_{jk} = \arctan(f(b_j b_k^*)/R(b_j b_k)) = -\theta_{kj}
\]

(2.5)

and

\[
f_j = f_s |b_j|^2 + f_{e,j}.
\]

As for Theorem 3.1, we shall require \( f_{j0} > 0, \sigma_{jk}, 0 < 1, 0 \leq j,k \leq r, \)

and \( v_{j0} < 1, 1 \leq j \leq r \), where the zero subscript indicates the true parameter value. Moreover we shall consider maximising the likelihood for values of \( \sigma_{jk}, 0 \leq j,k \leq r \), such that \( 0 \leq \sigma_{jk} \leq 1 - \delta \) for some small \( \delta > 0 \) since, as in Theorem 3.1, it can be shown that this requirement is unrestricted.
The logarithm of the likelihood given by (1.4) with $\Gamma$ as
indicated above can be written as

$$-\log \det(Q) - w^* \Lambda Q^{-1} \Lambda w$$ (2.6)

apart from a constant factor. Using similar methods to those used in
Chapters 4 and 5 we shall establish an asymptotically equivalent log-
likelihood to that given by (2.6) and, to that end, we now prove the
following generalisation of Lemma 3.1.

**Lemma 5.1**

Let the matrices $H_j$, $j = 1, \ldots, q$, be given by

$$H_j = \int_0^1 G_j(\theta) \otimes U(\theta) U(\theta)^* d\theta$$ (2.7)

where $U(\theta)$ is as for Lemma 3.1 and the $G_j(\theta)$ are non-constant $p \times p$
Hermitian matrix functions given by

$$G_j(\theta) = \sum_{k=0}^t G^{(j)}_k \chi_\theta(E_k), \quad \max_k \|G_k^{(j)}\| < \infty,$$

with

$$\chi_\theta(E_k) = \begin{cases} 1 & (\theta \in E_k) \\ 0 & (\text{otherwise}) \end{cases}$$

and the intervals $E_k$ are disjoint with

$$E_k = \{ \theta | \alpha_k < \theta \leq \alpha_{k+1} \}, \quad \bigcup_{k=0}^t E_k = (0,1].$$

Then

$$m^{-1} \text{tr} \left\{ \frac{1}{q} \sum_{j=1}^q H_j - H \right\} = O(m^{-1} \log m)$$

where

$$H = \int_0^1 G(\theta) \otimes U(\theta) U(\theta)^* d\theta$$
and

\[ G(\theta) = \sum_{k=0}^{t} \left\{ \prod_{j=1}^{q} c_{k}^{(j)} \right\} \chi_{\theta}(E_{k}) = \prod_{j=1}^{q} G_{j}(\theta) . \]

**Proof**

The matrix \( H_{j} \) can be described as

\[ H_{j} = \sum_{k=0}^{t} c_{k}^{(j)} \otimes \phi(E_{k}) \]

where

\[ \phi(E_{k}) = \int_{0}^{1} \chi_{\theta}(E_{k}) U(\theta) U(\theta)^{*} \, d\theta . \]

Hence

\[ m^{-1} \text{tr} \left\{ \prod_{j=1}^{q} H_{j} \right\} = \sum_{k_1, \ldots, k_q} \text{tr} \left( c_{k_1}^{(1)} \ldots c_{k_q}^{(q)} \right) m^{-1} \text{tr} \left( \phi(E_{k_1}) \ldots \phi(E_{k_q}) \right) = 0 \]

and the result of the lemma follows as a consequence of Lemma 3.1.

In a similar fashion we also establish a generalisation of Lemma 3.2.

**Lemma 5.2**

Let \( H \) be an Hermitian matrix of the form given by (2.7) with

\[ H = \int_{0}^{1} G(\theta) \otimes U(\theta) U(\theta)^{*} \, d\theta , \]

\[ G(\theta) = \sum_{j=0}^{t} G_{j} \chi_{\theta}(E_{j}) \]

and \( \max_{j} \|G_{j}\| = c < \infty \). Then, if \( R(z) = \sum_{k=1}^{\infty} a_{k} z^{k} \) and \( \sum_{k=1}^{\infty} |a_{k}| e^{k} < \infty \),

\[ m^{-1} \text{tr} [R(H) - H_{R}] = O(m^{-1} \log m) \]

where
Proof

This follows directly from Lemma 5.1 and the proof of Lemma 3.2.

We now observe that $Q$ can be written as

$$Q = \int_0^1 B(\theta) \otimes U(\theta) U(\theta)^* d\theta,$$

(2.8)

$$B(\theta) = \Delta P(\theta) \Delta$$

where $P(\theta)$ and $\Delta$ have typical elements given for $0 \leq j, k \leq r$ by

$$P_{jk}(\theta) = \begin{cases} 1 & (j = k) \\ \sigma_{jk} \exp(i\theta_{jk}) q(v_j, v_k; \theta) & (j \neq k) \end{cases}$$

and

$$\Delta_{jk} = \begin{cases} \sqrt{f_j} & (j = k) \\ 0 & (j \neq k) \end{cases}$$

However $P(\theta)$ can also be described equivalently as

$$\sum_{j=0}^{r} P \chi_\theta(E_j)$$

where

$$E_j = \{ \theta | 1 - v_{j+1} < \theta \leq 1 - v_j \}, \quad v_{r+1} = 1,$$

and $P_j$ has typical element

$$[P_j]_{s,t} = \begin{cases} 1 & (s = t) \\ \sigma_{jk} \exp(i\theta_{jk}) & (s, t \leq j; s, t > j); \ 0 \leq s, t \leq r \end{cases}$$

(2.10)
Hence $P(\theta)$ is of the form required by Lemmas 5.1 and 5.2. Now, the negative of the log-likelihood given by (2.6) may, after normalisation by $m$, be written as

$$m^{-1} \log \det(Q) + m^{-1} w^* Q^{-1} A w.$$  

However, from Lemma 5.2,

$$m^{-1} \log \det(Q) = m^{-1} \text{tr} \{ \log Q \}$$

converges uniformly to

$$\int_0^1 \log \det(B(\theta)) \, d\theta .$$

Moreover,

$$w^* A^* \left\{ Q^{-1} - \int_0^1 B^{-1}(\theta) \otimes U(\theta) \, U(\theta)^* \, d\theta \right\} A w$$

converges uniformly to zero, almost surely, and this follows from Lemmas 5.1 and 5.2 in an identical fashion to corresponding results given in the proof of Theorem 3.1. Thus we may replace the original likelihood by the asymptotically equivalent form

$$L = \int_0^1 \log \det(B(\theta)) \, d\theta + m^{-1} w^* A \left\{ \int_0^1 B^{-1}(\theta) \otimes U(\theta) \, U(\theta)^* \, d\theta \right\} A w .$$

Assuming that this reflects the true nature of the statistics $w_j(\theta_k)$, we are reduced to the minimisation of $L$ with respect to the parameters $f_j, \sigma_{jk}, \theta_{jk}; 0 \leq j, k \leq r$, and $v_j, 1 \leq j \leq r$.

It is noted that

$$m^{-1} w^* A \left\{ B^{-1}(\theta) \otimes U(\theta) \, U(\theta)^* \right\} A w$$

$$= m^{-1} \sum_{j,k=0}^r w_j^* A(v_j)^* B^{-1}(\theta) U(\theta) \, U(\theta)^* A(v_k) w_k$$

$$= \text{tr} \{ B^{-1}(\theta) \, C(\theta) \}$$
where \( w_j \) is the vector consisting of the \((j+1)\)th string of \(m\) elements of \(w\) and

\[
C_{jk}(\theta) = \xi_j(\theta + v_j) \xi_k(\theta + v_k)
\]

with

\[
\xi_j(\theta) = m^{-\frac{1}{2}} \sum_{k=-\frac{1}{2}(m-1)}^{\frac{1}{2}m} w_j(\theta_k) e^{-i2\pi k\theta}.
\]

Thus \( L \) can now be written as

\[
L = \int_0^1 \left[ \text{tr}(B^{-1}(\theta) C(\theta)) + \log \det(B(\theta)) \right] d\theta
\]

\[
= \sum_{j=0}^r \left[ \text{tr}(B_j^{-1} C_j) + (v_{j+1} - v_j) \log \det(B_j) \right]
\]

where

\[
C_j = \int_0^{1-v_j} C(\theta) d\theta, \quad B_j = \Delta P \Delta.
\]

When \( L \) is minimised with respect to \( \sigma_{jk}, \theta_{jk} \) and \( f_j \), the following equations result:

\[
\int_0^1 B_{jk}^{-1}(\theta) B_{kj}(\theta) d\theta = \int_0^1 \left[ B_{jk}^{-1}(\theta) C(\theta) B_{kj}^{-1}(\theta) \right] d\theta,
\]

\[0 \leq j, k \leq r.\]

However, noting that

\[
q(v_j, v_s; \theta) q(v_k, v_t; \theta) q(v_j, v_k; \theta) = q(v_j, v_s; \theta) q(v_s, v_t; \theta) q(v_t, v_k; \theta)
\]

and also that

\[
B^{-1}(\theta) = \sum_{j=0}^r B_j^{-1} \chi_\theta(E_j)
\]

implies

\[
B_{jk}^{-1}(\theta) = q(v_j, v_k; \theta) R_{jk}(\theta)
\]

for some matrix function \( R(\theta) \), it is evident that (2.13) may be rewritten as
\[ \int_0^1 \left( B^{-1}(\theta) \tilde{C}(\theta) B^{-1}(\theta) - B^{-1}(\theta) \right) \, d\theta = 0 \] (2.14)

where

\[ \tilde{C}_{jk}(\theta) = q(v_j, v_k; \theta) \tilde{C}_{jk}(\theta) . \]

With reference to (2.13) it implies that

\[ (r + 1)^{-1} \int_0^1 \text{tr}[B^{-1}(\theta) C(\theta)] \, d\theta = 1 \]

it can be seen that the maximum likelihood estimates of the \( v_j \), which we shall call \( \hat{v}_j \), are found by minimising

\[ \int_0^1 \log \det(B(\theta)) \, d\theta = \sum_{j=0}^{r} (v_{j+1} - v_j) \log \det(B_j) . \]

Here the \( \sigma_{jk} \), \( \theta_{jk} \), \( \hat{f}_j \) will be functions of the \( v_j \) found by solving the system of equations given by (2.14) and the maximum likelihood estimates \( \hat{C}_{jk} \), \( \hat{\theta}_{jk} \) and \( \hat{f}_j \) will be the values of these functions computed at the \( \hat{v}_j \).

Proceeding on a more heuristic level we observe that

\[ B^{-1}(\theta) = B^{-1} + \sum_{j=0}^{r-1} (B_{j}^{-1} - B^{-1}) \chi_{\theta}(E_j) \]

\[ = B^{-1} + \bar{B}^{-1}(\theta) \]

where

\[ B_{jk} = [B_r]_{jk} = \begin{cases} \hat{f}_{j} \sigma_{jk} \exp(i\theta_{jk}) & (j = k) \\ \sqrt{\hat{f}_{j} \hat{f}_{k}} \sigma_{jk} \exp(i\theta_{jk}) & (j \neq k) \end{cases} \] (2.15)

and thus (2.14) becomes

\[ \int_0^1 (B^{-1} + \bar{B}^{-1}(\theta))(\tilde{C}(\theta) - B(\theta))(B^{-1} + \bar{B}^{-1}(\theta)) \, d\theta = 0 . \]

Since \( \bar{B}^{-1}(\theta) \) is zero for \( \theta \) such that \( 0 < \theta \leq 1 - v \), it is apparent, if the \( v_j \) are all small, that \( B^{-1}(\theta) \) may be approximated by \( B^{-1} \) and so, neglecting the terms involving \( \bar{B}^{-1}(\theta) \), we obtain the system of equations
\[ \int_0^1 B(\theta) \, d\theta = \int_0^1 \bar{c}(\theta) \, d\theta. \]  \tag{2.16} 

Hence we can now form

\[ \hat{f}_j = \int_0^1 |\xi_j(\theta + v_j)|^2 \, d\theta = m^{-1} \sum_{t=-[\frac{1}{2}(m-1)]}^{[\frac{1}{2}m]} |w_j(t)|^2, \]

and \( \hat{\theta}_{jk}(v) \) is given by the argument of \( \int_0^1 \bar{c}_{jk}(\theta) \, d\theta \). As before one would now obtain the \( \hat{v}_j \) by minimising

\[ \int_0^1 \log \det(B(\theta)) \, d\theta = \log \det(B) + \sum_{j=0}^{r-1} (v_{j+1} - v_j) \log \det(B_j B_j^{-1}) \]

with the \( f_j, \sigma_{jk} \) and \( \theta_{jk} \) given by (2.17). In the spirit of the previous approximation for small \( v_j \) we might consider this latter procedure as being close to the minimisation of \( \det(B) \) where the entries in \( B \) are given once again by (2.17). Once the \( \hat{v}_j \) are determined, \( \hat{\theta}_{jk} \) and \( \hat{\theta}_{jk}(\hat{v}) \) and \( \hat{\theta}_{jk}(\hat{\theta}) \) respectively.

Another approximation may also be considered where the right-hand side of (2.16) is replaced by an approximating sum so that

\[ \int_0^1 q(v_j, v_k; \theta) C_{jk}(\theta) \, d\theta \approx m^{-1} \sum_{t} \xi_j(t, t+mv_j) \xi_k(t, t+mv_k) \]

with the summation over those \( t \) such that

\[ \min([m(1 - v_j)], [m(1 - v_k)]) < t \leq \max([m(1 - v_j)], [m(1 - v_k)]) \]

and \( \xi_j(t) = \xi_j(t/m) \). It is noted that the \( \xi_j(t) \) are just inverse transforms of the type discussed in §3.5 and §4.4. Of course, as in §4.4, we could have chosen an integer other than \( m \), but in view of the properties of the transformation using \( m \) which are described in §3.5 we have chosen this latter. In terms of these quantities the maximum
likelihood estimates given by (2.17) can now be approximated as

$$\hat{f}_j = m^{-1} \sum_t |\xi_j, t|^2 = m^{-1} \sum_t |w_j(t)|^2,$$

$$\hat{v}_{jk} = \frac{|\sum_t \xi_{j,t+m\hat{v}_j} \xi_{k,t+m\hat{v}_k}|}{(m - m|\hat{v}_j - \hat{v}_k|)\sqrt{\sum_t \hat{f}_j, \hat{f}_k}},$$  \hfill (2.18)

$$\hat{\vartheta}_{jk} = \arctan\left(\frac{\sum_t \xi_{j,t+m\hat{v}_j} \xi_{k,t+m\hat{v}_k}}{\sqrt{\sum_t \hat{f}_j, \hat{f}_k}}\right)$$

where the $\hat{v}_j$ are obtained by minimising $\det(B)$ with entries as above.

These estimates are intuitively appealing since they have a direct interpretation in terms of leads or lags of the original time series. Hence we have transformed back into the time domain, lagged or rephased, and then optimised the rephasing. A discussion of estimation methods similar to that given by (2.18) is presented in Hannan (1972).

So far only the one parameterisation given by (2.5) has been considered. Other schemes (cf. Hannan (1972)) could have been adopted in the light of the requirements that need to be imposed on the structure of the model. However, the resulting modifications to the general development given in this section are straightforward and will be omitted here.

5.3 An Optimality Criterion for Array Structures

Considering the original lags $t_j$ as depicted in the diagram given at the beginning of §5.2 we see that

$$t_j/N = r_j \cos(\phi - \theta_j)/(NV), \quad 1 \leq j \leq r,$$

where the $\theta_j$ and $r_j$ are known. Since we require

$$\lim_{N \to \infty} (t_j/N) = v_j, \quad 1 \leq j \leq r,$$
the \( v_j \) will now be parameterised as

\[
v_j = v r_j \cos(\phi - \theta_j), \quad 1 \leq j \leq r ,
\]

and so, in the sample situation, \( v \) will be given by \((NV)^{-1}\). Thus \( v \) is directly proportional to the wavenumber of the signal at this fixed angular frequency \( \lambda_0 \). It is \( v \) and \( \phi \) together with the other spectral parameters that are estimated by the methods proposed in §5.2.

We now proceed to evaluate the variance-covariance matrix of our estimates using the vector generalisation of Theorem 4.2 as discussed in §5.1. Letting \( t_j \) denote any general parameter other than \( v \) or \( \phi \) this states that

\[
m^\frac{1}{2} [(\widehat{t}_1, \ldots, \widehat{m\phi}, \widehat{mv})' - (t_1, \ldots, m\phi, mv)']
\]

is asymptotically normal with zero mean vector and covariance matrix \( \Gamma^{-1} \) where the typical element of \( \Gamma \) is given by

\[
I_{jk} = \lim_{m \to \infty} \left\{ m^{-1} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial t_j} \Gamma^{-1} \frac{\partial \Gamma}{\partial t_k} \right) \right\}
\]

with \( t_j \) being either \( mv \), \( m\phi \) or a \( t_j \) and the zero subscript denoting the true parameter set.

Using Lemmas 5.1 and 5.2 it is evident that

\[
\lim_{m \to \infty} \left\{ m^{-1} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial t_j} \Gamma^{-1} \frac{\partial \Gamma}{\partial t_k} \right) \right\} = \lim_{m \to \infty} \left\{ m^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial t_j} Q^{-1} \frac{\partial Q}{\partial t_k} \right) \right\}
\]

\[
= \int_0^1 \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial t_j} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial t_k} \right) \, d\theta . \quad (3.3)
\]

These various elements can now be deduced from the following equations:

\[
\text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \sigma_{jk}} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \sigma_{st}} \right)
\]

\[
= \sigma^{-1}_{jk} \sigma^{-1}_{st} 2 \Re \left( B^{-1}(\theta) B_{kj}(\theta) B^{-1}(\theta) B_{st}(\theta) + B^{-1}(\theta) B_{jk}(\theta) B_{ks}(\theta) \right)
\]

\[
= \sigma^{-1}_{jk} \sigma^{-1}_{st} 2 \Re \left( B^{-1}(\theta) B_{kj}(\theta) B^{-1}(\theta) B_{st}(\theta) \right) + B^{-1}(\theta) B_{jk}(\theta) B_{ks}(\theta) \right)
\]

\[
= \sigma^{-1}_{jk} \sigma^{-1}_{st} 2 \Re \left( B^{-1}(\theta) B_{kj}(\theta) B^{-1}(\theta) B_{st}(\theta) \right) + B^{-1}(\theta) B_{jk}(\theta) B_{ks}(\theta) \right)
\]

\[
= \sigma^{-1}_{jk} \sigma^{-1}_{st} 2 \Re \left( B^{-1}(\theta) B_{kj}(\theta) B^{-1}(\theta) B_{st}(\theta) \right) + B^{-1}(\theta) B_{jk}(\theta) B_{ks}(\theta) \right)
\]
\[ \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \theta_{jk}} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \theta_{st}} \right) \]

\[ = 2 \mathcal{L} \left( B^{-1}_{tk}(\theta) B_{kj}(\theta) B^{-1}_{js}(\theta) B_{st}(\theta) - B^{-1}_{tj}(\theta) B_{jk}(\theta) B^{-1}_{ks}(\theta) B_{st}(\theta) \right) \]

\[ \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_j} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_k} \right) = (2 f_j f_k)^{-1} \mathcal{L} \left( B^{-1}_{jk}(\theta) B_{kj}(\theta) + \delta_{jk} \right) \]

\[ \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \theta_{jk}} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial \theta_{st}} \right) \]

\[ = - \mathcal{L} \left( B^{-1}_{tk}(\theta) B_{kj}(\theta) B^{-1}_{js}(\theta) B_{st}(\theta) + B^{-1}_{tj}(\theta) B_{jk}(\theta) B^{-1}_{ks}(\theta) B_{st}(\theta) \right) \]

\[ \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_j} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_s} \right) = f_j^{-1} \mathcal{L} \left( B^{-1}_{jk}(\theta) B_{kj}(\theta) (\delta_{sk} + \delta_{sj}) \right) \]

\[ \text{tr} \left( B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_j} B^{-1}(\theta) \frac{\partial B(\theta)}{\partial f_s} \right) = f_s^{-1} \mathcal{L} \left( B^{-1}_{jk}(\theta) B_{kj}(\theta) (\delta_{sk} + \delta_{sj}) \right) \]

where \( \delta_{jk} \) takes the value unity for \( j = k \) and is zero otherwise. If the \( \nu_j \) are small and the approximation \( B(\theta) = B \) is made, the above realise the usual form of the information matrix when no lagging has occurred.

The elements corresponding to \( \hat{\nu}, \hat{\phi} \) are now considered.

Taking the diagonal element corresponding to \( \hat{\nu} \) for example we observe that

\[ m^{-3} \text{tr} \left( r^{-1} \frac{\partial \Gamma}{\partial \nu} r^{-1} \frac{\partial \Gamma}{\partial \nu} \right) = - m^{-3} \text{tr} \left( \frac{\partial \Gamma^{-1}}{\partial \nu} \frac{\partial \Gamma}{\partial \nu} \right) \]

\[ = - m^{-3} \text{tr} \left\{ \left( \frac{\partial \Lambda}{\partial \nu} \right)^* \Lambda^{-1} \Lambda + \Lambda^* \frac{\partial \Lambda^{-1}}{\partial \nu} \Lambda + \Lambda^* \Lambda Q^{-1} \frac{\partial \Lambda}{\partial \nu} \right\} \]

where \( \Lambda \) is the Lagrange multiplier. However
\[
\frac{\partial \Lambda}{\partial v} = -i2\pi TA
\]

where \( T = H \otimes D \) is diagonal with

\[
H_{jk} = \frac{\partial v_j}{\partial v} \delta_{jk}, \quad 0 \leq j, k \leq r,
\]

and

\[
D_{jk} = j^5 j^k, \quad -\frac{1}{2}m < j, k \leq \left\lfloor \frac{1}{2}m \right\rfloor.
\]

Hence

\[
m^{-3} \text{tr} \left( r^{-1} \frac{\partial \Gamma}{\partial v} r^{-1} \frac{\partial \Gamma}{\partial v} \right) = 8\pi^2 m^{-3} \text{tr}(T Q T^{-1} - T^2)
\]

\[
- m^{-3} \text{tr} \left( \frac{\partial Q}{\partial v} \frac{\partial Q}{\partial v} \right) - 4\pi i m^{-3} \text{tr}\left( \frac{\partial Q}{\partial v} (T Q^{-1} - Q^{-1} T) \right). \quad (3.4)
\]

Since

\[
Q = \sum_{j=0}^{r} B_j \otimes \int_{1-v_j}^{1} U(\theta) U(\theta)^* d\theta = \sum_{j=0}^{r} B_j \otimes \Phi(E_j)
\]

it can be seen that

\[
\frac{\partial Q}{\partial v} = - \sum_{j=1}^{r} \left( B_j - B_{j-1} \right) \frac{\partial v_j}{\partial v} \otimes U(1 - v_j) U(1 - v_j)^*
\]

and so the second term of (3.4) is a sum of terms of the type

\[
\left( \frac{\partial v_k}{\partial v} \frac{\partial v_t}{\partial v} \right) \text{tr} \left\{ B_j^{-1}(B_k - B_{k-1}) B_s^{-1}(B_t - B_{t-1}) \right\}
\]

\[
\times m^{-3} \text{tr} \left\{ \Phi(E_j) U(1 - v_k) U(1 - v_k)^* \Phi(E_s) U(1 - v_t) U(1 - v_t)^* \right\}.
\]

We have, in the above, taken \( Q^{-1} \) to be given by

\[
\sum_{j=0}^{r} B_j^{-1} \otimes \Phi(E_j) \quad (3.5)
\]

with the knowledge that, in so doing, we have introduced an error which converges uniformly to zero. However

\[
|U(1 - v_t)^* \Phi(E_j) U(1 - v_k)| \leq m
\]

and so the second term of (3.4) converges uniformly to zero. The third
term of (3.4) can be represented, in a similar fashion, as the sum of terms of the type

\[ (4\pi) \left[ \frac{\partial v_j}{\partial v} \text{tr} \left\{ (B_j - B_{j-1}) B_k^{-1} H \right\} m^{-3} \text{tr} \left\{ U(1 - v_j) U(1 - v_j)^* \Phi(E_k) D \right\} \right. \]

- \left. \frac{\partial v_j}{\partial v} \text{tr} \left\{ (B_j - B_{j-1}) H B_k^{-1} \right\} m^{-3} \text{tr} \left\{ U(1 - v_j) U(1 - v_j)^* D \Phi(E_k) \right\} \right].

In this case \(|U(1 - v_j)^* \Phi(E_k) D U(1 - v_j)|\) and \(|U(1 - v_j)^* D \Phi(E_k) U(1 - v_j)|\) are both bounded above by \(m^2\) and so the third term of (3.4) converges uniformly to zero. Thus we may now confine our attention to the first term of (3.4) which, with reference to (3.5), is asymptotically equivalent to

\[ 8\pi^2 \sum_{j,k} \text{tr}(P_j H P_k^{-1} H) m^{-3} \text{tr}(\Phi(E_j) D \Phi(E_k) D) - 8\pi^2 \text{tr}(H^2) m^{-3} \text{tr}(D^2). \]  

(3.6)

However

\[ m^{-3} \text{tr}(\Phi(E_j) D \Phi(E_k) D) = m^{-3} \text{tr}[D \Phi(E_j) [D \Phi(E_k) - \Phi(E_k) D]] \]

\[ + m^{-3} \text{tr}[D^2 \Phi(E_j) \Phi(E_k)] \]  

(3.7)

and the first term on the right-hand side of the above can be written as

\[ (i2\pi)^{-1} m^{-3} \text{tr}[D \Phi(E_j) [U(1 - v_k) U(1 - v_k)^* - U(1 - v_{k+1}) U(1 - v_{k+1})^*]] \]

which is bounded in modulus by

\[ \pi^{-1} m^{-2} \|D \Phi(E_j)\| \leq (2\pi)^{-1} m^{-1}. \]

The second term on the right-hand side of (3.7) differs from

\[ \delta_{jk} (v_{j+1} - v_j) [m^{-3} \text{tr}(D^2)] \]

by the term
which is bounded in modulus by
\[
[m^{-5} \text{tr}(D^4)]^{\frac{1}{2}} \left[ m^{-1} \text{tr}(\Phi(E_j) \Phi(E_k) - \delta_{jk} \Phi(E_j)) \right]^{\frac{1}{2}}
\]
and this latter converges uniformly to zero by Lemma 3.1. Now (3.6)
can be replaced by
\[
8\pi^2 \left[ m^{-3} \text{tr}(D^2) \right] \left\{ \int_0^1 \text{tr}(P(\theta) \ H \ P^{-1}(\theta) \ H - H^2) \, d\theta \right\}
\]
which converges, as \( m \) becomes large, to
\[
\frac{2\pi^2}{3} \left\{ \int_0^1 \text{tr}(P(\theta) \ H \ P^{-1}(\theta) \ H - H^2) \, d\theta \right\}.
\]
The proofs for the remaining elements of the information matrix are
established in the same way. Hence, putting \( G \) as the matrix with
\[
\frac{\partial}{\partial \phi} \delta_{jk}
\]
as the \((j,k)\)th element, these can now be inferred from the
following relations:
\[
\lim_{m \to \infty} \left\{ m^{-3} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial \nu} \Gamma^{-1} \frac{\partial \Gamma}{\partial \phi} \right) \right\} = \frac{2\pi^2}{3} \mathcal{R} \left\{ \int_0^1 \text{tr}(P(\theta) \ H \ P^{-1}(\theta) \ G - HG) \, d\theta \right\}
\]
\[
\lim_{m \to \infty} \left\{ m^{-3} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial \phi} \Gamma^{-1} \frac{\partial \Gamma}{\partial \phi} \right) \right\} = \frac{2\pi^2}{3} \left\{ \int_0^1 \text{tr}(P(\theta) \ G \ P^{-1}(\theta) \ G - G^2) \, d\theta \right\}
\]
\[
\lim_{m \to \infty} \left\{ m^{-2} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial \nu} \Gamma^{-1} \frac{\partial \Gamma}{\partial \nu} \right) \right\} = 0
\]
\[
\lim_{m \to \infty} \left\{ m^{-2} \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial \phi} \Gamma^{-1} \frac{\partial \Gamma}{\partial \phi} \right) \right\} = 0
\]
where \( \tau \) denotes a parameter other than \( \nu \) or \( \phi \).

We now examine the section of the information matrix corres­
ponding to \( \hat{\nu} \) and \( \hat{\phi} \) and note that this is independent of the rest of the
information matrix. Since
\[ \frac{\partial v_j}{\partial \phi} = v_j \sin(\theta_j - \phi), \]
\[ \frac{\partial v_j}{\partial \theta} = r_j \cos(\theta_j - \phi), \]

It can be seen that
\[ \mathcal{R} \left\{ \int_0^1 \text{tr}(P(\theta) H P^{-1}(\theta) G - HG) \, d\theta \right\} \]
\[ = \mathcal{R} \left\{ \sum_{j,k=0}^r \int_0^1 \left( P^{-1}_{jk}(\theta) \frac{\partial v_k}{\partial \phi} P^{-1}_{kj}(\theta) \frac{\partial v_j}{\partial \phi} - \delta_{jk} \frac{\partial v_k}{\partial \phi} \frac{\partial v_j}{\partial \phi} \right) \, d\theta \right\} \]
\[ = v x' A y \]

where the matrix A is real and symmetric and has typical element given by
\[ A_{jk} = \int_0^1 \left\{ \mathcal{R}\left( P^{-1}_{jk}(\theta) P_{kj}(\theta) - \delta_{jk} \right) \right\} \, d\theta = \int_0^1 A_{jk}(\theta) \, d\theta, \]
\[ 0 \leq j, k \leq r. \quad (3.8) \]

Here x and y are vectors with entries \[ r_j \cos(\theta_j - \phi) \] and \[ r_j \sin(\theta_j - \phi) \] respectively for \[ 0 \leq j \leq r. \] Thus it is evident that the sub-matrix in question can be described as
\[ \frac{2\pi^2}{3} \begin{bmatrix} x' A x & v x' A y \\ v x' A y & v^2 y' A y \end{bmatrix}. \quad (3.9) \]

Previously we referenced the recorders relative to axes chosen to have origin at that recorder representing the smallest lag, but it will be more convenient at this stage to consider new axes that have as origin a point, not necessarily a recorder, central to the array. With reference to the diagram we observe that

\[ x = r' \cos(\phi' - \phi) \mathbf{l} + \mathbf{x}, \quad y = r' \sin(\phi' - \phi) \mathbf{l} + \mathbf{y} \]
where \( \mathbf{x} \) and \( \mathbf{y} \) are vectors with \( r_j \cos(\theta_j - \phi) \) and \( r_j \sin(\theta_j - \phi) \) as their respective \( j \)th entries.

However

\[
[A]_j = \sum_{k=0}^{r} A_{jk} = 0
\]

and so

\[
x'Ay = \tilde{x}'A\tilde{y}
\]

with the other elements of (3.9) admitting similar results. Moreover,

\[
A_{jk} = \int_{0}^{1} [q(P^{-1}_{jk}(\theta - \omega) P_{kj}(\theta - \omega) - \delta_{jk})] d\theta
\]

since \( q(v_j, v_k, \theta) \) is periodic with period unity and hence, choosing \( \omega \) to be \( v_j \cos(\phi' - \phi) \), we see that the matrix \( A \) in (3.10) may be replaced by \( \tilde{A} \) where \( \tilde{A} \) is the same as \( A \) except that the \( v_j = v_{x_j} \) have now been replaced by \( v'_{j} = v_{x_j}' \). Thus this section of the information matrix is invariant under translation. Since \( v_j, x_j, \) and \( y_j \) are functions of \( (\theta_j - \phi) \) which is independent of the orientation of the axes, it is evident that (3.9) is also invariant under rotation. We may now, without any loss of generality, consider the \( r_j, \theta_j \) to be the polar coordinates of the \( j \)th recorder relative to any particular axes of choice. The origin of these latter could, for convenience, be a central point within the array structure.

We now consider the problem of choosing the \( r_j, \theta_j \) so as to minimise the variances of \( \hat{\nu} \) and \( \hat{\phi} \). This is, in a sense, rather artificial because the other spectral components could no doubt depend to some degree on the particular recorder configuration adopted. However we shall assume that the change in such components caused by an array change is slight in comparison to the variation introduced by the
various phasing effects at the individual antennae. The coordinates of the recorders will be chosen subject to some restriction and it is clear that by varying the restrictions imposed one would obtain different results. For example, if some prior distribution were assigned to such that the signal would almost always enter through the third quadrant then some asymmetrical recorder pattern might no doubt eventuate. The problem is basically that of making \( \frac{3}{2\pi^2} (X' A X)^{-1} \) as small as possible where \( X \) is a matrix with typical row \( (x^j, y^j) \), \( 0 \leq j \leq r \). Noting that

\[
A = A_r + v \sum_{j=1}^r x_j (A_{j-1} - A_j)
\]

where

\[
(A_{j})_{s,t} = A_{j} \left( \begin{array}{c} P^{-1}_j \times t, s - s \times t \end{array} \right)
\]

we shall now make the additional assumption, as mentioned at the end of §5.2, of taking the \( v^j = v x^j \) to be small enough so that \( A \) may be approximated by \( A_r \). Within the context of these approximations and qualifications we now examine \( \frac{3}{2\pi^2} (X' A_r X)^{-1} \) where \( A_r \) is independent of \( X \).

We choose to work in terms of the ordinates and abscissae of the recorders relative to the given reference axes. Putting \( s \) and \( t \) as the vectors whose typical elements are \( r_j \cos \phi_j \) and \( r_j \sin \phi_j \) respectively, it is evident that

\[
x = s \cos \phi + t \sin \phi
\]

and

\[
y = t \cos \phi - s \sin \phi
\]

Hence we have
\[ x'_r A_r x = s'_r A_r s \cos^2 \phi + t'_r A_r t \sin^2 \phi + 2s'_r A_r t \sin \phi \cos \phi , \]
\[ y'_r A_r y = s'_r A_r s \sin^2 \phi + t'_r A_r t \cos^2 \phi - 2s'_r A_r t \sin \phi \cos \phi , \]
\[ x'_r A_r y = s'_r A_r t \cos^2 \phi + (t'_r A_r t - s'_r A_r s) \sin \phi \cos \phi \]
and the determinant of \( X'_r A_r X \) is given by
\[ v^2(s'_r A_r s t'_r A_r t - (s'_r A_r t)^2) . \]

Let us now impose the restrictions
\[ s'_r A_r s = \alpha_1 > 0 , \quad t'_r A_r t = \alpha_2 > 0 , \quad (3.11) \]
which require the \( s, t \) to lie on ellipsoids. If one transformed to
new coordinates \( \xi = T s, \eta = T t \) where \( T \) is the orthogonal matrix that
diagonalises \( A_r \), then this requires that the average squared lengths of
the \( \eta \) and the \( \xi \), both weighted according to the appropriate eigen­
values of \( A_r \), should be \( \alpha_1 \text{ tr}(A_r) \) and \( \alpha_2 \text{ tr}(A_r) \) respectively. The
physical interpretation of these conditions will become more apparent
as we progress.

The optimality criterion used is the minimisation of the
generalised variance which is given by
\[ \left( \frac{3}{2\pi^2} \right)^2 \text{det}\{ (X'_r A_r X)^{-1} \} \]
\[ = \left( \frac{3}{2\pi^2} \right)^2 \left[ v^2 s'_r A_r s t'_r A_r t \left( 1 - \frac{(s'_r A_r t)^2}{s'_r A_r s t'_r A_r t} \right) \right]^{-1} . \]

It can be seen that a minimum is attained when \( s'_r A_r t = 0 \) yielding a
covariance matrix for \( \hat{\nu} \) and \( \hat{\phi} \) that can be described as
\[ \frac{3}{2\pi^2} \begin{bmatrix} \alpha_2^{-1} \sin^2 \phi + \alpha_1^{-1} \cos^2 \phi & v^{-1}(\alpha_1^{-1} - \alpha_2^{-1}) \sin \phi \cos \phi \\ v^{-1}(\alpha_1^{-1} - \alpha_2^{-1}) \sin \phi \cos \phi & v^{-2}(\alpha_2^{-1} \cos^2 \phi + \alpha_1^{-1} \sin^2 \phi) \end{bmatrix} \quad (3.12) \]
If $a_1 = a_2 = a$ then this becomes

$$\frac{3a^{-1}}{2\pi^2} \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

(3.13)

which is independent of the angle $\phi$. This is reasonable in that knowledge that the signal can arrive from any direction is consistent with requiring the variation of the ordinates and abscissae to be the same.

It is noted that if we had retained $A$ instead of $A_r$ the result would remain unchanged, but now (3.11) becomes rather difficult to interpret. Thus the essential requirement of an array design is the orthogonality of the vectors $s$, $t$ relative to the matrix $A_r$.

In order to obtain some insight into the above we shall now proceed on a heuristic basis and examine these conditions in terms of the actual structure of the model. Thus we shall choose to describe the elements of $A_r$ in terms of the attenuation coefficients $b_j$ since these latter form the basis for the original model given by (2.1).

Suppressing the subscript $r$ on $P$, we now consider $P$ to be given by

$$P_{jk} = \begin{cases} 1 & (j = k) \\ \rho_j \rho_k & (j \neq k) \end{cases}, \quad 0 \leq j, k \leq r,$$

where, with reference to (2.2) and (2.5),

$$\rho_j = b_j \sqrt{f_j} f_j.$$

Hence

$$P = E + \rho \rho^*,$$

$$P^{-1} = E^{-1} - \tilde{\alpha} E^{-1} \rho \rho^* E^{-1}$$

where $\rho$ is a vector with typical element $\rho_j$, $E$ is diagonal with $1 - |\rho_j|^2$ as the $(j + 1)^{th}$ diagonal entry and $\tilde{\alpha}$ is given by
We can now describe $A_r$ as

$$D((1 - \alpha) I_{r+1} - \alpha uu') D$$

where $D$ is diagonal and $u$ is a vector with

$$D_{jj} = u_j = \sqrt[2r]{|\rho_j|/(1 - |\rho_j|^2)^\frac{1}{2}}$$

and

$$\tilde{\alpha} = \left(1 + \sum_{j=0}^{r} v_j\right)^{-1}.$$

However

$$v_j = \frac{|\rho_j|^2}{1 - |\rho_j|^2} = \frac{|b_j|^2 f_s}{f_{e,j}}$$

is just the signal to noise ratio at the $j^{th}$ recorder. Now

$$s'' A_r t = (1 - \tilde{\alpha}) \left\{ \sum_{j=0}^{r} v_j s_j t_j \right\}$$

$$- \left( \sum_{j=0}^{r} v_j \right)^{-1} \left( \sum_{j=0}^{r} v_j s_j \right) \left( \sum_{j=0}^{r} v_j t_j \right)$$

with analogous formulae holding for $s' A_r s$ and $t' A_r t$. Hence, putting

$$r_{st} = s' A_r t / \{s' A_r s t' A_r t\}^{\frac{1}{2}},$$

an optimal configuration is achieved when $r_{st}$ is zero. This represents a correlation coefficient describing the degree of association between the ordinates and abscissae of the recorders each weighted according to the signal to noise ratio at that particular recorder. The conditions given by (3.11) can now be interpreted as the requirement that the average squared difference of the ordinates from their average value should be constant and similarly for the abscissae. This also implies that the average squared Euclidean distance between the recorders and the point corresponding to the average ordinate and average abscissa
should be constant. Thus the conditions imposed by (3.11) fix the spread of the array and hence the optimum array has been sought from the class of such arrays. We shall now confine our attention to the case where the two variances are the same so that $\alpha_1 = \alpha_2 = \alpha$ and certain special arrays (cf. Birtill and Whiteway (1965)) will be examined with regard to the optimality criterion proposed.

The family of linear cross arrays is considered and, with reference to the diagram, we select the axes to lie on the lines of the recorders as depicted. This implies that

$$s = (s_0, s_1, \ldots, s_q, 0, 0, \ldots, 0)' \quad \text{and} \quad t = (0, 0, \ldots, 0, t_{q+1}, t_{q+2}, \ldots, t_r)'$$

for some particular $q$. Thus $r_{st}$ will be zero when either $s$ or $t$ is zero where

$$s = \frac{\sum_{i=0}^{r} v_i s_i}{\sum_{i=0}^{r} v_i} \quad \text{and} \quad t \text{ is defined analogously. If } s = t = 0 \text{ and the attenuation coefficients are all the same then}$$

$$\frac{v_i}{\sum_{i=0}^{r} v_i} = (r + 1)^{-1}$$

and the array that results will be a symmetrical cross arrangement provided that the recorders are equidistant from one another on the axes. An example of the latter is indicated in the diagram. It is noted that the symmetrical cross is equivalent, with regard to the optimality criterion, to the array that results when one of the lines of recorders of the symmetrical cross is displaced some fixed distance
along its axis relative to the origin of the reference axes. These latter arrays include the class of T-shaped arrays. However, given that the array structure has been determined prior to the experiment, it is evident that the symmetrical cross has an advantage over the T-shaped array in that both \( s \) and \( t \) might be expected to be small for the former whereas only one of these quantities need be small in the latter. Hence, for the symmetrical cross \( r_{st} \) should be small, but this does not necessarily follow for the T-shaped array.

For a linear cross array shaped in the form of the letter L it is evident that this implies that both \( s \) and \( t \) are non-zero unless \( \bar{v}_i = 0 \) for all \( i \) and thus the L-shaped array cannot be optimal with regard to the criterion given above. If, in regard to this latter array, either \( s \) or \( t \) is small then it may be near to optimal.

No mention yet has been made as to the number of recorders on any one arm of the array. It would appear that whereas one could have more recorders on one arm than the other this could lead to sub-optimal estimation. By this we mean that if the \( v_i \) vary from one experimental situation to the other and the array is the same in all cases, then the variance of one of the sets of coordinates could be more prone to error than that of the other set and this might lead to a violation of the initial constraint that these variances should be the same. If this latter were to occur then the array would no longer be optimal.

Turning to other arrays we now assume that the origin of the array has been chosen so as to ensure that \( \bar{s} = \bar{t} = 0 \). Thus \( r_{st} \) is now proportional to \( r_{st} = \sum_{i=0}^{N} v_i s_i t_i \). However, when we come to examine circular arrays, triangular arrays, rectangular arrays and all other symmetrical arrays we see that \( r_{st} \) will be zero when the \( v_i \) assume some common value, but need not necessarily be so otherwise. Hence the symmetrical
cross array appears to be the most robust design in that it yields, within the context considered, minimum variances for \( \hat{v} \) and \( \hat{s} \) even when the individual signal to noise ratios are not the same.
Bibliography


