

APPLICATIONS OF PROLATE SPHEROIDAL FUNCTION THEORY
TO GEOPHYSICAL DATA PROCESSING

by

JOSE ANTONIO RODRIGUEZ

B.A.(Physics), Cornell University, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Geophysics and Astronomy

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1987

©José Antonio Rodríguez, 1987

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Geophysics and Astronomy

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 5/4/87

ABSTRACT

Because signals that are simultaneously concentrated in time and frequency are common in geophysics, it is desirable to develop a set of basis functions with these properties to perform various types of data processing. The Prolate Spheroidal Functions (PSF's) form a complete orthonormal set, and the low-order PSF's span the space of functions which are simultaneously concentrated in time and frequency. In this study, the PSF's are utilized in three different data processing problems: spectrum estimation, signal-to-noise ratio enhancement, and wavelet estimation. All three problems are related by functions which are approximately time- and bandlimited: data tapers for spectrum estimates, time- and bandlimited signals, and seismic wavelets can all be expressed as linear combinations of the low-order PSF's.

Some of the results obtained by applying the PSF's to solving these problems are encouraging. In the problems of spectrum estimation and wavelet estimation in particular, the PSF's seem to extract much of the information present in the data. The application of PSF's to solving problems in geophysical data processing should be the focus of further research in the future.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
ACKNOWLEDGMENTS	viii
DEDICATION	ix
GENERAL INTRODUCTION	2
CHAPTER I The Prolate Spheroidal Functions	6
1. The Basic Equations	7
2. Properties of the PSF's	10
3. The 'Uncertainty Principle'	12
4. Discrete Time	13
CHAPTER II Spectrum Estimation	23
1. Conventional Estimators	24
A. The Periodogram	24
B. Autoregressive Estimates	25
2. The Intuitive Approach	27
3. Data-Adaptive Weights	30
4. Harmonic Analysis	34
A. Tests for single, isolated lines	34
B. Double line tests	37
5. Applications to Stationary Signals	39
6. Applications to Nonstationary Signals	56
7. Conclusion	67
CHAPTER III SNR Enhancement for Bandlimited Signals	69
1. The Problem in One Dimension	70
2. The Problem in Two Dimensions	77
3. Limitations and Conclusions	84
CHAPTER IV Seismic Wavelet Estimation	88
1. The Convolutional Model	88
2. The Prolate Spheroidal Wavelet	90
3. Example 1 - A Land Wavelet	92
4. Example 2 - A Marine Wavelet	93
5. Further Suggestions and Conclusions	100
CONCLUSION	101

REFERENCES	105
APPENDIX A	108
APPENDIX B	110
APPENDIX C	113

LIST OF TABLES

1. Eigenvalues for low-order DPSS's	21
---	----

LIST OF FIGURES

1.1 Low-order DPSS's	17
1.2 Low-order DPSWF's	18
1.3 Low-order DPSS's	19
1.4 Low-order DPSWF's	20
2.1 Effect of smoothing a spectrum estimate	26
2.2 A 'sinc' function	26
2.3 APS Spectrum, Example (A.1)	41
2.4 Periodogram, Example (A.1)	41
2.5 AR Spectra, Example (A.1)	42
2.6 APS Spectrum, Example (A.2)	45
2.7 Periodogram, Example (A.2)	45
2.8 AR Spectra, Example (A.2)	46
2.9 APS Spectrum and F-test, Example (B)	48
2.10 Periodogram, Example (B)	48
2.11 AR Spectra, Example(B)	49
2.12 APS Spectrum and F-test, Example (B)	51
2.13 Periodogram, Example (B)	51
2.14 AR Spectra, Example(B)	52
2.15 F-test, Example (C)	54
2.16 Periodogram, Example(C)	54
2.17 AR Spectra, Example(C)	55
2.18 Periodogram, changing frequencies	61
2.19 Adaptive AR, changing frequencies	61
2.20 PSS, changing frequencies	61
2.21 Periodogram, changing frequencies	62
2.22 Adaptive AR, changing frequencies	62
2.23 PSS, changing frequencies	62
2.24 Periodogram, changing frequencies	63
2.25 Adaptive AR, changing frequencies	63
2.26 PSS, changing frequencies	63
2.27 Love wave seismogram	64
2.28 Power spectrum of seismogram	64
2.29 Periodogram, Love waves	66
2.30 Adaptive AR. Love waves	66
2.31 PSS, Love waves	66
3.1 Low-order bandlimited functions	73
3.2 Frequency domain of 3.1	74

3.3 Low-order bandlimited functions	75
3.4 Frequency domain of 3.3	76
3.5 Ricker wavelet plus noise	78
3.6 Reconstructions of Ricker wavelet	79
3.7 Reconstructions of Ricker wavelet	80
3.8 Reconstructions of Ricker wavelet	81
3.9 Frequency-wavenumber space of seismic signals	82
3.10 2-D contaminated data	85
3.11 2-D data after filtering	85
4.1 Seismogram and reflectivity, land wavelet	94
4.2 Misfit curve and estimated wavelet	95
4.3 Misfit curve and estimated wavelet	96
4.4 Seismogram and reflectivity, marine wavelet	97
4.5 Misfit curve and estimated wavelet	98
4.6 Misfit curve and estimated wavelet	99

ACKNOWLEDGEMENTS

I wish to thank my advisor, Prof. Tad Ulrych, for his input and his patience during my time at U.B.C., and Colin Walker for all his support.

I also wish to thank Drs. David J. Thomson and Craig Lindberg for clarifying many things, and especially Dr. Alan Chave for providing me with an initial version of the spectrum estimation program used in this thesis.

Finally I wish to express my deepest thanks and my love to Minerva Camarena, who always came to my aid when I most needed it. *Gracias.*

This work was funded by NSERC grant A1804.

DEDICATORIA

A mis padres, a quienes les debo todo.

*Juventud, divino tesoro,
¡ ya te vas para no volver!
Cuando quiero llorar, no lloro...,
y a veces lloro sin querer.*

—Rubén Darío, ‘ Canción de Otoño en Primavera’

GENERAL INTRODUCTION

In many different aspects of signal processing one will encounter functions whose energy is highly concentrated simultaneously in both time and frequency domains. Such functions have been studied for some time: *Blackman and Tukey* (1958) noticed that certain time domain window functions have all their energy concentrated in some time interval and a large fraction in some frequency band. *Bartlett* (1950) had investigated window functions earlier, and *Parzen* (1961) also looked at this problem.

It was not until the early 1960's that a group of engineers in Bell Labs, New Jersey (D. Slepian, H.J. Landau, and H.O. Pollak) began a systematic approach to discover simultaneously time- and bandlimited functions that arose analytically rather than empirically. What they found was that the functions that achieve maximum simultaneous concentration of energy in the time and frequency domains had been studied previously in an entirely different context. *Flammer* (1953), *Stratton, et al.* (1956), *Meixner and Schafke* (1954), and *Morse and Feshbach* (1953) all present methods to solve the wave equation in prolate spheroidal coordinates, and *Slepian and Pollak* (1961) found that, curiously enough, the functions they were searching for satisfied the same equations as the angular solutions to the wave equation in prolate spheroidal coordinates. Since that

discovery, several books and publications have treated these functions in detail [*Landau and Pollak* (1961), *Landau and Pollak* (1962), *Slepian* (1964), *Abramowitz and Stegun* (1965), *Papoulis* (1977), *Slepian* (1978)].

This work began as a study of a method of spectrum estimation proposed by *Thomson* (1982) that utilizes the prolate spheroidal functions as data tapers. However, since timelimited and bandlimited functions appear in many different problems in data processing, the study of prolate spheroidal functions is of much interest. As the wide variety of applications of bandlimited function theory to problems in data processing became apparent, it seemed appropriate to direct this thesis toward the exploration of different applications. If the coverage of some of the applications seems limited only to preliminary research, it is because the objective of this thesis is mainly to point out the possibilities of using prolate spheroidal functions in the treatment of certain problems without going into lengthy discussions of computational problems.

As to the organization of this thesis, each chapter contains the treatment of a separate problem, and so each deserves a separate introduction. Chapter I deals with the fundamental mathematical properties of the prolate spheroidal functions. These functions are defined by an integral equation and form a complete orthonormal set. These properties, along with the bandlimited nature of these functions, are exploited in the later chapters.

As mentioned earlier, this thesis began as an investigation of a new method of spectrum estimation, and this is covered in Chapter II. Thomson's spectrum estimation method is tested on several different data sets, and the examples are subdivided into stationary and nonstationary data. Perhaps the most promising feature of this 'prolate spheroidal' method of spectrum estimation is that it can be modified to produce an

effective technique of harmonic analysis. What makes this harmonic analysis technique special is that it yields confidence levels for the amplitude and frequency estimates. This can be useful when trying to distinguish between peaks in a spectrum estimate that are produced by noise and peaks from harmonic components in the data. The examples in Chapter II are chosen to test Thomson's method and to compare it to some conventional techniques of spectrum estimation.

Chapter III deals with a technique to reduce additive noise that contaminates bandlimited signals. If one knows the frequency band that a signal exists in, then an expansion of the data in prolate spheroidal functions concentrated in the same band can be used as an alternative noise reduction scheme to regular bandpass filtering. The advantage of this approach is that it separates the data into components with spectra concentrated in and outside of the signal band. Fourier basis functions, on the other hand, do not distinguish between noise and signal inside the band of interest.

The final problem treated in this thesis, presented in Chapter IV, is seismic wavelet estimation. Seismic wavelets are observed to be finite in time and highly concentrated in frequency, so they can be well reproduced by linear combinations of prolate spheroidal functions. An attractive feature of the wavelet estimation method presented in Chapter IV is that it is computationally quite efficient, and it provides a criterion of how well the estimated wavelet and reflectivity approximate the true ones.

*Hay algo denso, unido, sentado en el fondo,
repitiendo su número, su señal idéntica.*

—Pablo Neruda, ‘Unidad’

CHAPTER I

The Prolate Spheroidal Functions

It is well known that a function that is finite in time cannot be bandlimited in frequency. It is of interest, however, to find how much it is possible to concentrate a time-limited function inside a certain frequency band, that is, to find the time-limited function that is optimally concentrated in the frequency domain. This important question arises in problems of spectrum estimation, filter design, and in general considerations of bandlimited functions. Some of these problems will be discussed in future chapters, while the present chapter contains some elementary theory that will become necessary later.

This chapter presents aspects of time- and band-limited functions which are relevant to spectrum estimation and filter design. General considerations and full mathematical details may be found in a series of excellent papers published in the Bell System Technical Journal [*Slepian and Pollak* (1961), *Landau and Pollak* (1961), *Landau and Pollak* (1962), *Slepian* (1978)]. Several applications of functions which are concentrated in both the time and frequency domains are suggested in these papers, but some of these are

useful almost exclusively to electrical engineers. Take, for instance, a problem in data transmission: to prevent a pulse from changing shape due to dispersion during transmission through a waveguide, the pulse must be concentrated in a narrow frequency band. The optimal pulse, therefore, should be finite in time and highly concentrated in the frequency domain. In the field of geophysics, however, applications of these functions also immediately come to mind. Seismic wavelets are observed to be finite in time and essentially bandlimited. Bandpass filters are concentrated in time and have an impulse response that is bandlimited.

1. The Basic Equations

A function of time $h(t)$ and its frequency-domain equivalent, $H(f)$, are related through the Fourier transform:

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{i2\pi ft} df \quad (1.1)$$

and

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi ft} dt. \quad (1.2)$$

Because of Parseval's theorem, the energy of this function may be expressed in the time or frequency domain:

$$E_{tot} = \int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |H(f)|^2 df. \quad (1.3)$$

If $h(t)$ is timelimited, that is, vanishes outside $|t| > T/2$, ($T > 0$), then the Fourier transform (1.2) becomes

$$H(f) = \int_{-T/2}^{T/2} h(t) e^{-i2\pi ft} dt. \quad (1.4)$$

To find the timelimited functions that have the largest possible fraction of their total energy inside some band $|f| \leq W$, ($W > 0$), it is necessary to find the maximum value of the ratio

$$\lambda = \frac{\int_{-W}^W H(f) H^*(f) df}{\int_{-\infty}^{\infty} H(f) H^*(f) df} \quad (1.5)$$

where $*$ indicates the complex conjugate. Using (1.3) and (1.4) the expression for λ becomes

$$\lambda = \frac{\int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' h(t) h^*(t') \int_{-W}^W df e^{i2\pi f(t'-t)}}{\int_{-T/2}^{T/2} h(t) h^*(t) dt}, \quad (1.6)$$

and using the fact that the Fourier transform of a "boxcar" function is a "sinc" function, i.e.,

$$\int_{-W}^W e^{i2\pi f(t'-t)} df = \frac{\sin 2\pi W(t'-t)}{\pi(t'-t)}, \quad (1.7)$$

we obtain

$$\lambda \int_{-T/2}^{T/2} h(t) h^*(t) dt = \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' h(t) h^*(t') \frac{\sin 2\pi W(t'-t)}{\pi(t'-t)}. \quad (1.8)$$

A solution to (1.8) may be found by solving the somewhat simpler equation

$$\lambda h(t) = \int_{-T/2}^{T/2} h(t') \frac{\sin 2\pi W(t' - t)}{\pi(t' - t)} dt'. \quad (1.9)$$

Note that since equation (1.8) may be obtained by multiplying (1.9) by $h^*(t)$, taking complex conjugates, and integrating over t , solutions of (1.9) are also solutions of (1.8). Letting $t = xT/2$, $t' = x'T/2$, and $\psi(x) = h(xT/2)$, (1.9) is normalized to

$$\lambda \psi(x) = \int_{-1}^1 \psi(x') \frac{\sin c(x' - x)}{\pi(x' - x)} dx', \quad (1.10)$$

where $c = \pi TW$. This is an eigenvalue-eigenfunction problem, and a separate family of solutions exists for each value of the parameter c .

The solutions of (1.10) also happen to be the angular solutions of the wave equation in prolate spheroidal coordinates, and so are known as the prolate spheroidal functions (PSF's). The PSF's also satisfy a differential equation [see *Slepian and Pollak* (1961), and *Abramowitz and Stegun* (1965)], but this aspect will not be pursued in this thesis.

The PSF's shall be denoted by $\phi_k(c; t)$, and their corresponding eigenvalues by $\lambda_k(c)$. (The renaming of symbols from $\psi(t)$ to $\phi_k(c; t)$ is done to show explicitly the dependence on k and c .) The eigenvalues are arranged so that

$$1 > \lambda_0(c) > \lambda_1(c) > \cdots > \lambda_n(c) > \cdots > 0, \quad 0 \leq n < \infty,$$

and therefore the function that is finite in $|t| \leq T/2$ whose spectrum is most concentrated in $|f| \leq W$ is $\phi_0(c; t)$.

The Fourier transforms of these functions shall be denoted by $\Phi_k(c; f)$:

$$\Phi_k(c; f) = \int_{-T/2}^{T/2} \phi_k(c; t) e^{-i2\pi ft} dt. \quad (1.11)$$

2. Properties of the PSF's

Since the PSF's are eigenfunctions of an integral equation with a symmetric, positive-definite kernel, they exhibit the following properties:

Orthonormality:

Within a constant factor, to each eigenvalue $\lambda_k(c)$ there corresponds only one eigenfunction $\phi_k(c; t)$. If this constant is chosen so that $\int_{-\infty}^{\infty} |\phi_k(c; t)|^2 dt = 1$, then

$$\int_{-\infty}^{\infty} \phi_i(c; t) \phi_j^*(c; t) dt = \int_{-T/2}^{T/2} \phi_i(c; t) \phi_j^*(c; t) dt = \delta_{ij}. \quad (1.12)$$

where δ_{ij} is the Kroenecker delta.

Completeness:

The eigenfunctions span the space of timelimited functions, so that any ~~timelimited~~ function $f(t)$ can be expanded as a linear combination of these eigenfunctions:

$$f(t) = \sum_{n=0}^{\infty} \alpha_n \phi_n(c; t) \quad (1.13)$$

with

$$\alpha_n = \int_{-\infty}^{\infty} f(t) \phi_n(c; t) dt = \int_{-T/2}^{T/2} f(t) \phi_n(c; t) dt. \quad (1.14)$$

The frequency-domain PSF's, $\Phi_k(c; f)$, exhibit a curious *double orthogonality*: they are orthonormal over $(-\infty, \infty)$ and orthogonal over $[-W, W]$. We prove this using (1.11):

$$\begin{aligned} \int_{-\infty}^{\infty} \Phi_i(c; f) \Phi_j^*(c; f) df &= \int_{-\infty}^{\infty} df \int_{-T/2}^{T/2} \phi_i(c; t) e^{-i2\pi ft} dt \int_{-T/2}^{T/2} \phi_j^*(c; t') e^{i2\pi ft'} dt' \\ &= \int_{-T/2}^{T/2} \phi_i(c; t) dt \int_{-T/2}^{T/2} \phi_j^*(c; t') dt' \int_{-\infty}^{\infty} e^{i2\pi f(t'-t)} df \\ &= \int_{-T/2}^{T/2} \phi_i(c; t) dt \int_{-T/2}^{T/2} \phi_j^*(c; t') \delta(t' - t) dt' \\ &= \int_{-T/2}^{T/2} \phi_i(c; t) \phi_j^*(c; t) dt \\ &= \delta_{ij}. \end{aligned} \quad (1.15)$$

Similarly, because of (1.11), (1.9), and (1.7),

$$\begin{aligned}
 \int_{-W}^W \Phi_i(c; f) \Phi_j^*(c; f) df &= \int_{-T/2}^{T/2} \phi_i(c; t) dt \int_{-T/2}^{T/2} \phi_j^*(c; t') \frac{\sin 2\pi W(t' - t)}{\pi(t' - t)} dt' \\
 &= \lambda_j \int_{-T/2}^{T/2} \phi_i(c; t) \phi_j^*(c; t) dt \\
 &= \lambda_j \delta_{ij}
 \end{aligned} \tag{1.16}$$

3. The 'Uncertainty Principle'

As has been shown so far, the timelimited function with the largest fractional energy inside the band $|f| \leq W$ is $\phi_0(c; t)$, and the fractional measure of the concentration is $\lambda_0(c)$. A similar argument would show that the greatest concentration in time achievable by a *bandlimited* function is also $\lambda_0(c)$. *Landau and Pollak* (1962) attempt to answer the question: what degree of *simultaneous* concentration in both time and frequency domains is possible? The results of this 'Fourier Uncertainty Principle' follow without proof.

If

$$\alpha^2 = \frac{\int_{-T/2}^{T/2} |h(t)|^2 dt}{\int_{-\infty}^{\infty} |h(t)|^2 dt}$$

and

$$\beta^2 = \frac{\int_{-w}^w |H(f)|^2 df}{\int_{-\infty}^{\infty} |H(f)|^2 df}$$

then it may be shown that

$$\cos^{-1} \alpha + \cos^{-1} \beta \geq \cos^{-1} \sqrt{\lambda_0(c)}$$

where α and β are positive and less than or equal to 1. The reader is referred to *Landau and Pollak* (1962) for a detailed proof of this theorem.

Although this ‘uncertainty principle’ is not directly relevant to the rest of this study, it is included here because of its great theoretical importance. It states that there is a limit to the degree of simultaneous concentration in time and frequency achievable, so that no function can be arbitrarily concentrated in both domains.

4. Discrete Time

Since geophysical data are in general digitized and put in the form of a discrete time series, the discrete-time formulation of the theory expressed so far is particularly important. The equations and the notation used are entirely analogous to the continuous-time case.

A function h_n has a Fourier transform $H(f)$:

$$H(f) = \sum_{n=-\infty}^{\infty} h_n e^{-i2\pi n f} \quad (1.17)$$

where the sampling interval is taken to be $\Delta t = 1$ for simplicity. Also,

$$h_n = \int_{-1/2}^{1/2} H(f) e^{i2\pi n f} df. \quad (1.18)$$

As before,

$$\lambda = \frac{\int_{-W}^W |H(f)|^2 df}{\int_{-1/2}^{1/2} |H(f)|^2 df}.$$

When $h_n = 0$ for $n < -N$ and $n > N$, and since

$$\int_{-W}^W e^{i2\pi f m} df = \frac{\sin 2\pi W m}{\pi m},$$

λ becomes

$$\lambda = \frac{\sum_{m=-N}^N \sum_{n=-N}^N \frac{\sin 2\pi W(m-n)}{\pi(m-n)} h_n h_m^*}{\sum_{n=-N}^N |h_n|^2}.$$

A solution is given by

$$\lambda h_n = \sum_{m=-N}^N \frac{\sin 2\pi W(m-n)}{\pi(m-n)} h_m.$$

This is a Toeplitz matrix eigenvalue problem with solutions $\phi_n^k(LW)$, ($L = 2N + 1$), corresponding to the L different eigenvalues $\lambda_k(LW)$, arranged as in the continuous case,

$$1 > \lambda_0(LW) > \lambda_1(LW) > \dots > \lambda_L(LW) > 0.$$

The eigenvectors $\phi_n^k(LW)$ are referred to as the discrete prolate spheroidal sequences (DPSS's), and their Fourier transforms $\Phi_k(LW; f)$ (not to be confused with the continuous time frequency-domain functions $\Phi_k(c; f)$) are known as the discrete prolate spheroidal wave functions (DPSWF's):

$$\Phi_k(LW; f) = \sum_{n=-N}^N \phi_n^k e^{-i2\pi f n}.$$

Notice that the DPSS's and the DPSWF's obey the same properties of orthogonality, completeness, and double orthogonality that were examined before for the PSF's. The DPSWF's, furthermore, satisfy another integral equation: since

$$\lambda_k \phi_n^k(NW) = \sum_{m=-N}^N \frac{\sin 2\pi W(m-n)}{\pi(m-n)} \phi_m^k(NW), \quad (1.19)$$

Fourier transforming both sides, and since the Fourier transform of

$$x_n = \begin{cases} 1, & \text{if } |n| \leq N; \\ 0, & \text{otherwise.} \end{cases}$$

is

$$\frac{\sin[2\pi f(N + \frac{1}{2})]}{\sin \pi f},$$

we obtain

$$\lambda_k \Phi_k(LW; f) = \int_{-W}^W \frac{\sin[2\pi(N + \frac{1}{2})(f' - f)]}{\sin[\pi(f' - f)]} \Phi_k(LW; f') df'. \quad (1.20)$$

(The kernel function in this equation is known as a ‘Dirichlet kernel’, and arises by Fourier transforming a discretized ‘boxcar’. See Appendix A for details.)

Equations (1.19) and (1.20) will be the basis of most of the work remaining in this study.

The DPSS’s may be computed by solving equation (1.19) directly, but this is inconvenient for large values of L . *Thomson* (1981) gives a procedure for the computation of the DPSS’s for large values of L that is outlined in Appendix B. Also in Appendix B is a Toeplitz matrix formula for calculating the DPSS’s with an even number of points, and the computational limitations of all these procedures. Figure 1.1 shows $\phi_k(LW)$ for $k = 1$ to 4 and a time-bandwidth product of 2.0 (the time-bandwidth product is given by $\Delta t LW$. In this case $\Delta t = 1$, $L = 50$, $W = .08$.) Figure 1.2 shows $|\Phi_k(LW; f)|^2$ for the DPSS’s in Figure 1.1. Figs. 1.3 and 1.4 are the same, but for $\Delta t LW = 4.0$. The eigenvalues corresponding to these DPSS’s are listed on Table 1. Note that larger eigenvalues correspond to larger fractional energy concentration inside the main lobe, and so the sidelobes increase with decreasing eigenvalue.

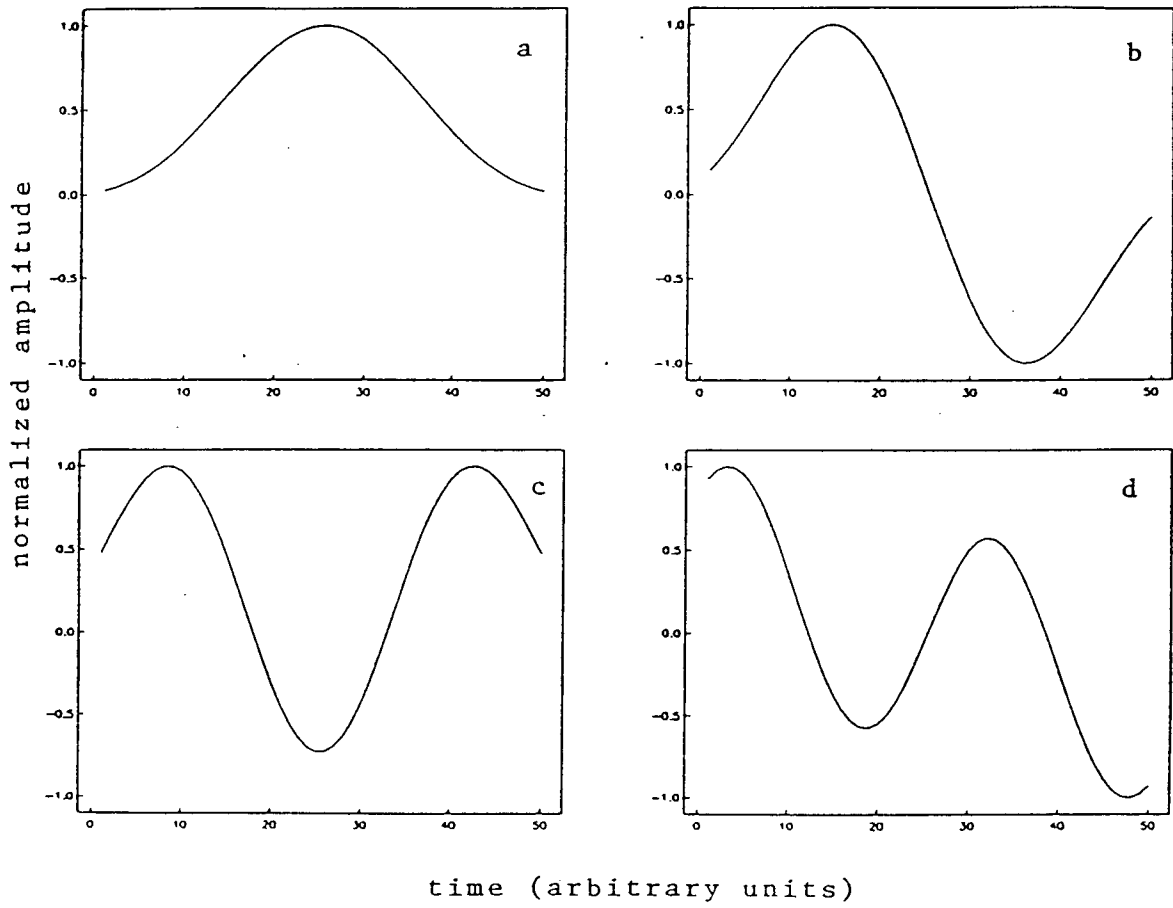


Figure 1.1 (a) Zero-order DPSS for time-bandwidth product 2.0. (b) First-order DPSS. (c) Second-order DPSS. (d) Third-order DPSS.

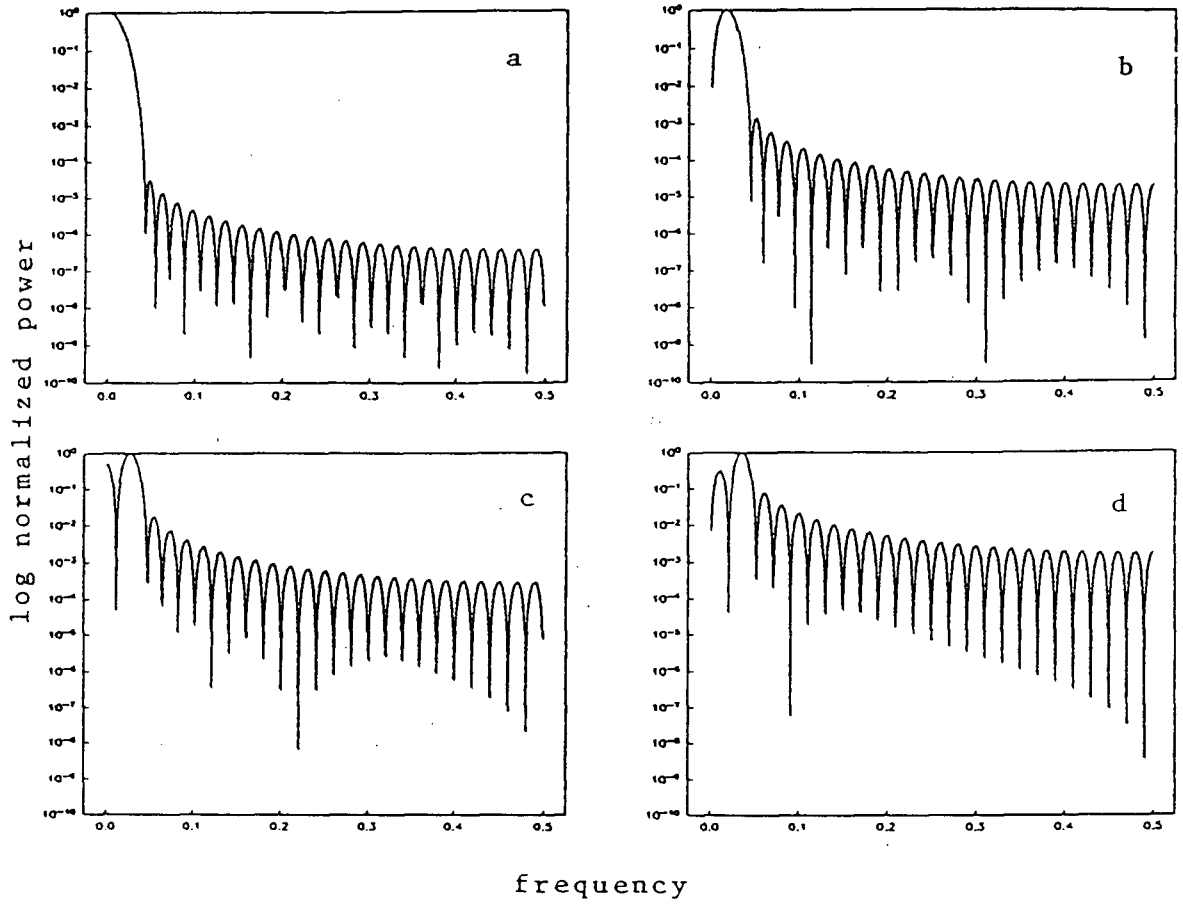


Figure 1.2 $|\Phi_k(LW; f)|^2$ for the DPSS's in Figure 1.1. Note log scale on the vertical axis.

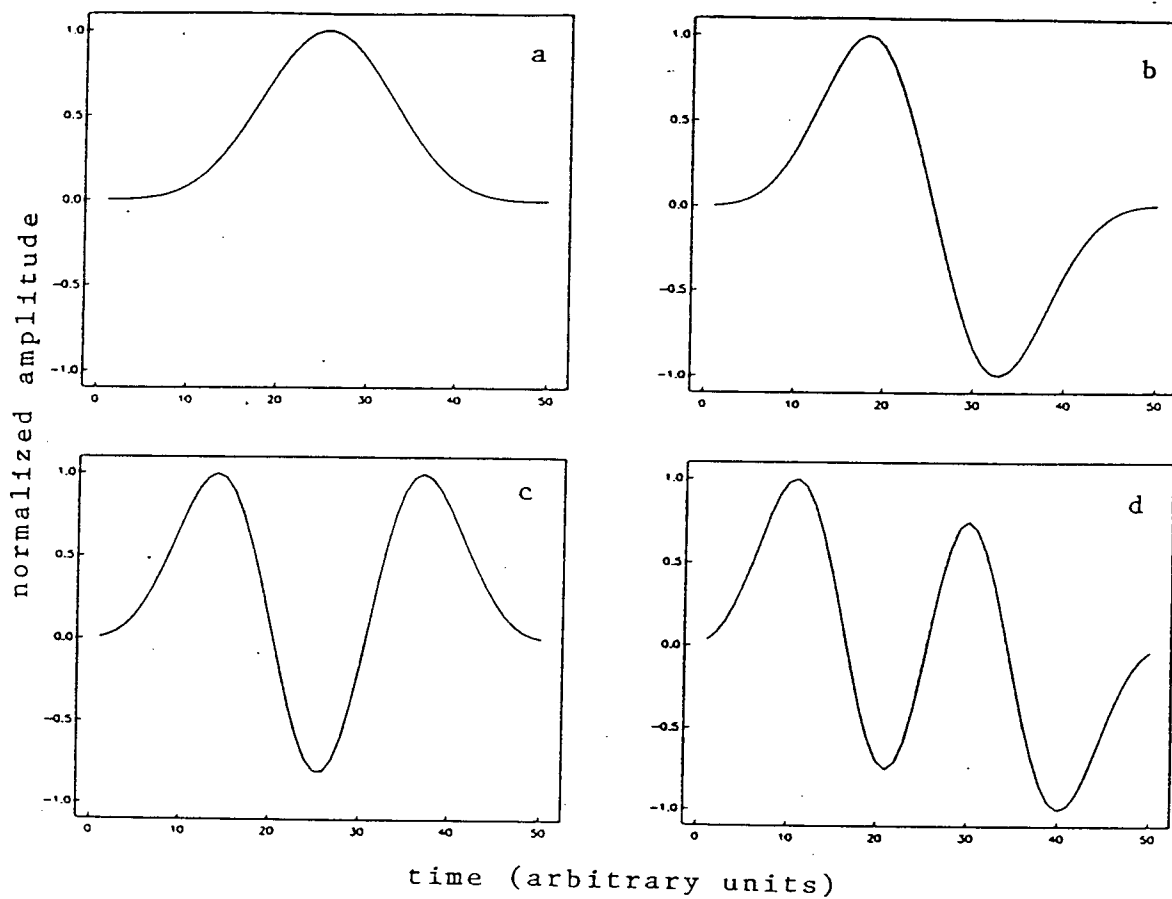


Figure 1.3 (a) Zero-order DPSS for time-bandwidth product 4.0. (b) First order DPSS. (c) Second order DPSS. (d) Third order DPSS.

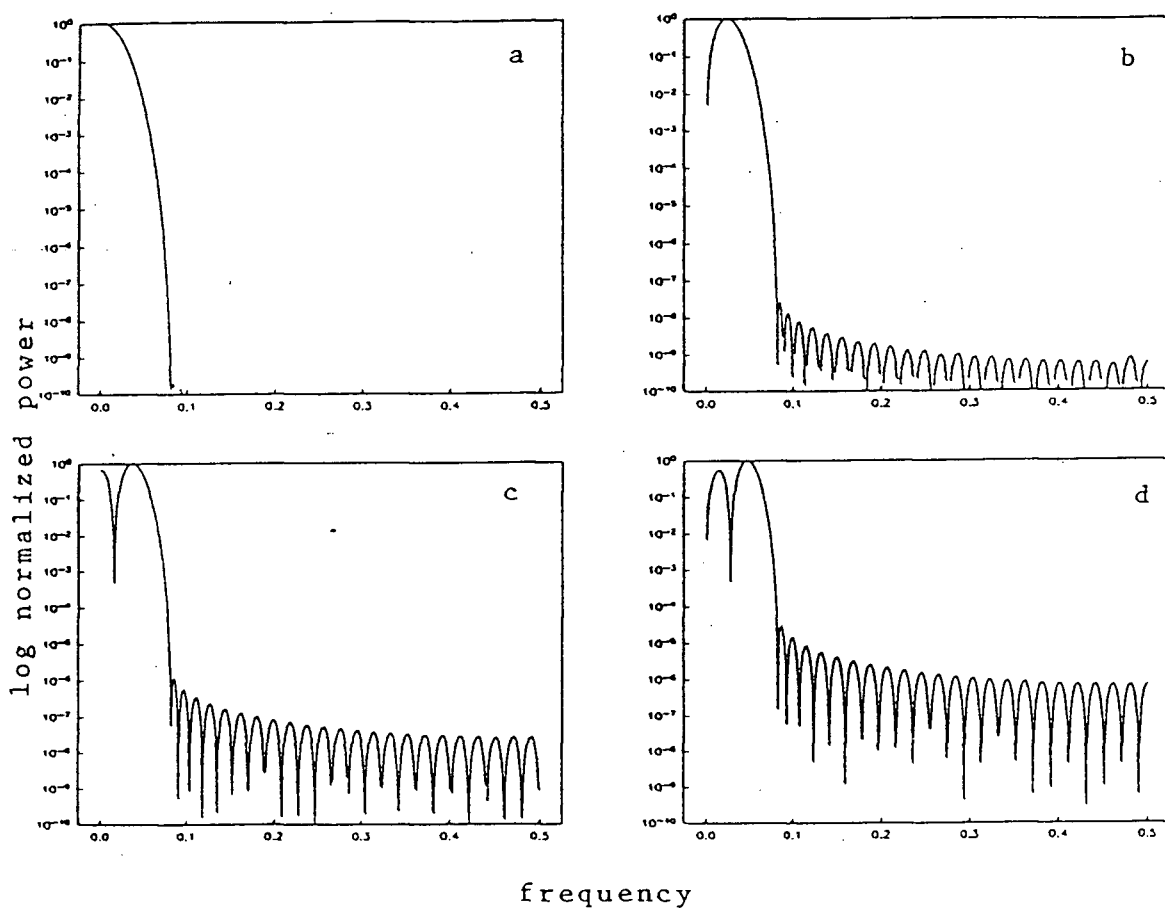


Figure 1.4 $|\Phi_k(LW; f)|^2$ for the DPSS's in Figure 1.3. Note log scale on the vertical axis.

k	$1 - \lambda_k, (TW = 2)$	$1 - \lambda_k, (TW = 4)$
0	0.5571e-04	0.2309e-09
1	0.2399e-02	0.2279e-07
2	0.4031e-01	0.1041e-05
3	0.2778	0.2905e-04
4	0.7257	0.5469e-03
5	0.9573	0.7159e-02
6	0.9966	0.6201e-01
7	0.9998	0.2998
8	0.9999	0.7019

Table 1 - Eigenvalues for $TW = 2$ and $TW = 4$, $k = 0$ to 8. $1 - \lambda_k$ is shown to illustrate how close to unity these eigenvalues come. Note that the eigenvalues decrease with increasing k , and that the largest eigenvalues are very close to 1.

Una tarde se entusiasmaron los muchachos con la estera voladora que pasó veloz al nivel de la ventana del laboratorio llevando al gitano conductor y a varios niños de la aldea que hacían alegres saludos con la mano, y José Arcadio Buendía ni siquiera la miró. “Déjenles que sueñen. —dijo— Nosotros volaremos mejor que ellos con recursos más científicos que ese miserable sobrecamas.”

—Gabriel García Márquez, Cien Años de Soledad

CHAPTER II

Spectrum Estimation

The bewildering variety of spectrum estimation techniques that exist today [illustrated in *Kay and Marple* (1981)] serves to point out the importance of the problem. Spectrum estimation methods may be divided into two types: *parametric* and *nonparametric*. Parametric methods assume that the data fit some sort of model [i.e. Autoregressive (AR), Moving Average (MA), or Autoregressive- Moving Average (ARMA) models] and set out to find the parameters of the model. Nonparametric methods, on the other hand, do not require that the data fit any presupposed model. It is the purpose of this chapter to discuss briefly some commonly used spectrum estimation methods, and compare them to a new nonparametric method proposed by *Thomson* (1982). For a thorough comparison of spectrum estimation methods in use prior to 1982, the reader is referred to *Kay and Marple* (1981).

1. Conventional Estimators

Two of the most common methods of performing spectrum analysis, well covered in *Jenkins and Watts* (1968) and *Priestley* (1981), are the periodogram and autoregressive techniques. This section outlines some of the advantages and disadvantages of using these estimators without going into mathematical details.

A. The Periodogram.-

This is the most commonly used spectrum estimate today, primarily because it is very simple to generate. Computation only involves taking an FFT of the data, then squaring it. This computational simplicity must be paid for, however, by certain disadvantages, which are:

- (i) The periodogram is *inconsistent*, that is, its variance does not decrease with increasing sample size.
- (ii) The variance of the periodogram estimate is roughly proportional to the power, so the variance properties are usually worse in the important parts of the spectrum.
- (iii) Truncation of the data in the time domain is equivalent to convolution with a 'sinc' function in the frequency domain, whose large sidelobes cause power leakage, which results in biasing the estimate.
- (iv) Attempts to reduce the variance of the periodogram are usually carried out by 'smoothing' the estimate by convolution with some window function, but this reduces the resolution and may introduce bias. For example, in Fig. 2.1, if the true peak was at position 'o', frequency f_o , convolution with a boxcar would displace the peak to a biased position 'x', frequency f_x .

- (v) Attempts to reduce bias by ‘tapering’ the data before transforming usually result in decreased resolution.

Problems (i) and (ii) arise because Fourier transforms and Fourier series (and thus the periodogram) are designed to analyze deterministic signals, and are not meant to deal with the randomness that is inherent in some processes. Problem (v) arises because tapering the data is equivalent to convolving with some other function in the frequency domain, and this function will invariably have a central lobe that is wider than that of a ‘sinc’ function, and thus will smear the fine structure in the spectrum.

Usually, tapering is done to reduce bias and smoothing is done to reduce variance, and the two problems are treated separately. Bias and variance are related, however, and may be taken into account by one single spectrum estimation method, as shall be observed later in section 2.

B. Autoregressive Estimates.-

There are three kinds of autoregressive estimates: (a) the *Yule-Walker* approach, (b) the *Burg* or *Maximum-Entropy* approach, and (c) the *Forward-Backward* or *Least Squares* approach. All of these are examined in *Kay and Marple* (1981). Autoregressive estimates, particularly the Maximum-Entropy and Least Squares estimates, have generally good resolution properties for closely spaced line components, which is why they are often used in the search for binary stars, speech processing, etc. However, there are other problems:

- (i) The data are being modelled as an autoregressive process, so a guess must be made as to the order of the process. A wrong guess of the order will either smear over details in the spectrum or introduce spurious peaks.

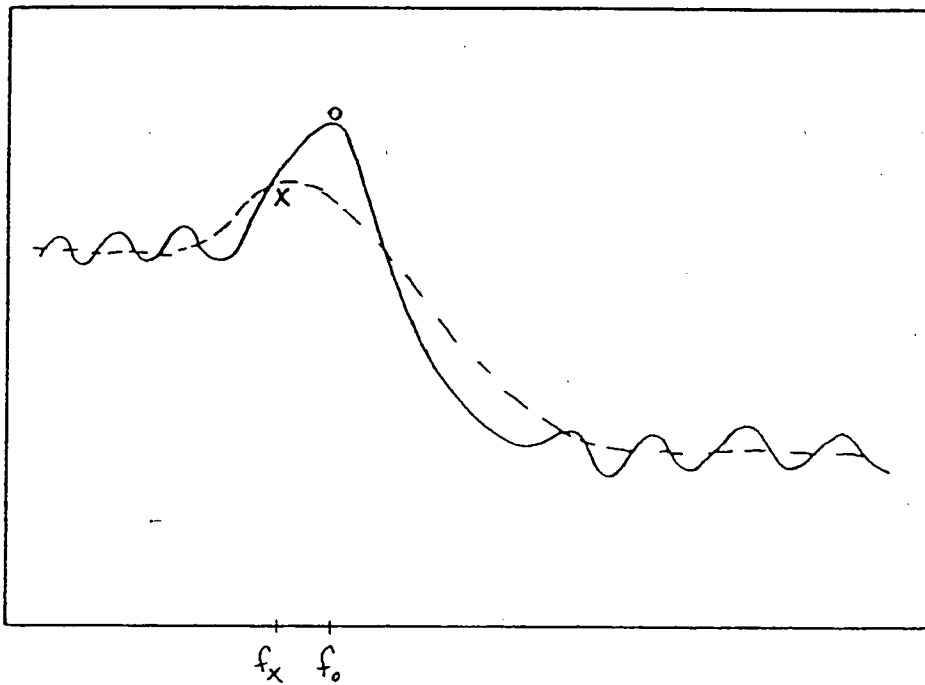


Figure 2.1 - Smoothing a spectrum estimate will move a peak from its true position 'o' to a biased position 'x'.

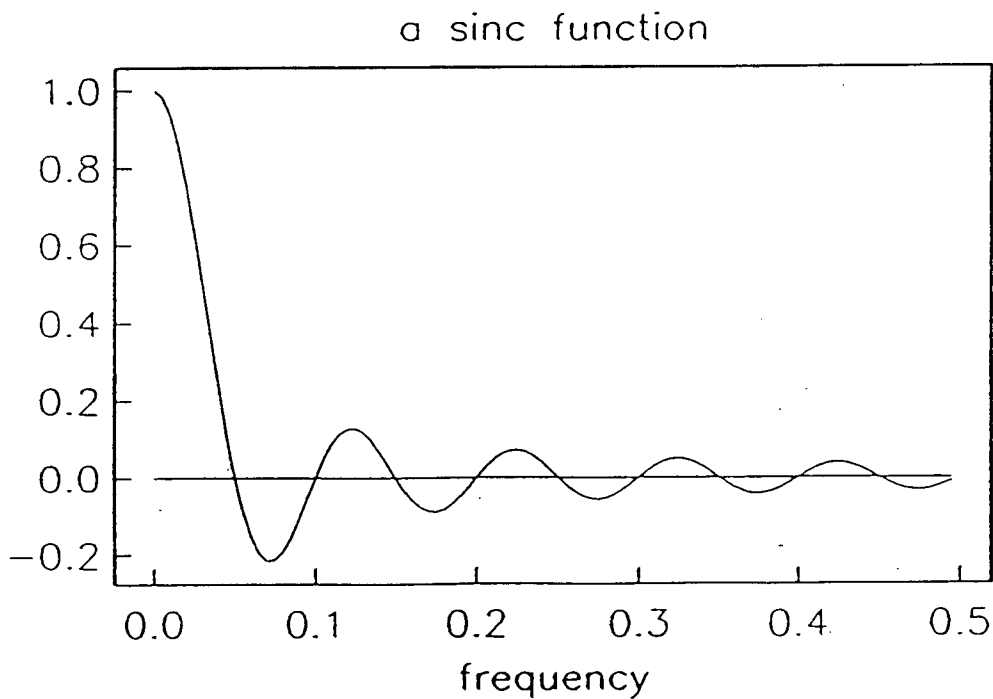


Figure 2.2 - The large sidelobes of the 'sinc' function cause power leakage, which may result in bias.

- (ii) Although the resolution properties for closely spaced line components are good, autoregressive estimates are not efficient at estimating ‘mixed’ spectra, like a continuous spectrum with sharp edges (see section 5.A.2 of this chapter).
- (iii) If the physical process that produces the signal is not well understood, there is a danger that if the data are produced by an MA or an ARMA process, the spectrum estimated by an AR model could differ substantially from the true one.

Because there are problems with most estimates, particularly when the data set is short, efforts to discover new spectrum estimation techniques continue. A 1982 paper by David J. Thomson presented the use of prolate spheroidal functions in an original nonparametric method of estimating spectra. This chapter outlines the core of Thomson’s method. The theory is derived from basic principles, and section 2 should give the reader an intuitive grasp of how this spectrum estimator works. As shall be demonstrated in the examples, this is indeed a good, useful estimator, and should be the focus of much interest in the coming years.

2. The Intuitive Approach

Consider some data x_n , $-\infty < n < \infty$, and its true Fourier transform $X(f)$:

$$X(f) = \sum_{n=-\infty}^{\infty} x_n e^{-i2\pi f n},$$

(the sampling interval is assumed to be unity). The true spectrum will be given by $|X(f)|^2$, but it must be noted that this is unobtainable, since one can never hope to have all the values of x_n , and must consider some sample of the data in the interval

$-N \leq n \leq N$. The assumption that the data are zero outside of this interval is equivalent to multiplication by a discretized ‘boxcar’ in the time domain, and therefore to convolution with a Dirichlet kernel in the frequency domain. As observed in the preceding section, this approach leads to severe problems in the final spectrum estimate.

Ideally, the true data should be convolved in the frequency domain with a δ -function, but again this is not possible, since it would be necessary to have infinite information in the time domain. We must settle, therefore, for approximations to a δ -function, and the Dirichlet kernel is a poor approximation. A good approximation must have its energy concentrated in some narrow band, and little energy elsewhere. These are criteria which describe the low-order DPSWF’s. Problem (iv) in the previous section may be circumvented by choosing the bandwidth of the central peak to be the same or narrower than that of a ‘sinc’ function.

What this suggests so far is using $\phi_n^0(LW)$ as a data taper, which would, in effect, convolve the true Fourier transform with $\Phi_0(LW; f)$. However, because this estimate is still inconsistent, this approach is only useful for routine data processing, not for crucial estimates where the final result must have good statistical properties. As observed previously, tapering the data is an effort to reduce the bias of the periodogram, but the variance problems are not solved. A solution to this problem may be achieved by utilizing the higher- order DPSWF’s.

The k^{th} -order *eigencoefficient* $y_k(f)$ is defined to be the Fourier transform of the finite data multiplied by the k^{th} -order DPSS:

$$y_k(f) = \sum_{n=-N}^N \phi_n^k(LW) x_n e^{-i2\pi f n}, \quad (2.1)$$

(LW is assumed to be chosen previously), and the k^{th} -order *eigenspectrum* is defined to be the square of the eigencoefficient:

$$\hat{S}_k(f) = |y_k(f)|^2. \quad (2.2)$$

Although each individual eigenspectrum is, by itself, an inconsistent estimate of the spectrum, the variance properties desired may be achieved by taking an average. The variance of the estimate of the mean of a process is reduced by taking more samples of the process, or, equivalently in this case, by averaging out more eigenspectra. Now, with a bigger data sample (bigger L) it is possible to use more eigenspectra in the average, thus tending toward consistency. A spectrum estimate achieved by these means is called an *Average Prolate Spheroidal spectrum (APS spectrum)* and is given by equation (2.3):

$$\bar{S}(f) = \frac{\sum_{k=0}^{K-1} |d_k(f)|^2 \hat{S}_k(f)}{\sum_{k=0}^{K-1} |d_k(f)|^2}. \quad (2.3)$$

The frequency-dependent weights $d_k(f)$ and the total number of eigenspectra K are chosen to reduce the bias of the final estimate. The mathematics necessary to obtain the $d_k(f)$ are somewhat involved, and the derivation is given in the next section.

The choice of K is rather simple: from Table 1, the eigenvalues of the DPSS's decrease rapidly for a given LW . The eigenvectors associated with these eigenvalues, furthermore, do not satisfy the criterion for a good data taper, namely, their Fourier transform is not concentrated inside the band $[-W, W]$. There is little to gain, therefore,

in using the higher-order DPSS's, and as a rule of thumb an upper limit to the value of K may be set at $K = 2\Delta t LW$.

3. Data-Adaptive Weights

To obtain the weights $d_k(f)$ it is necessary to look at the basic principles of spectrum estimation, even though these seldom appear in the geophysical literature. The reader is referred to *Priestley* (1981), section 4.11 for thorough coverage of this subject.

Because stochastic processes are not periodic, they cannot be represented by a finite Fourier series, and since stationary processes have infinite energy, they cannot be represented by a Fourier integral either [*Priestley* (1981), p.206]. The preferred spectral representation for a stationary stochastic process is the Cramér spectral representation:

$$x_n = \int_{-1/2}^{1/2} e^{i2\pi f n} dZ(f). \quad (2.4)$$

Note that, in some situations, $dZ(f) = X(f)df$, where $X(f)$ is the Fourier transform of x_n . In general, however, this derivative may not exist (see Appendix C for details). The $dZ(f)$ have the following properties for a zero-mean, stationary stochastic process:

$$E[dZ] = 0 \quad (2.5)$$

and

$$E[|dZ|^2] = S(f)df \quad (2.6)$$

where $S(f)$ is the spectrum.

The eigencoefficients $y_k(f)$, then, may be taken to be the convolution of the k^{th} -order DPSWF with dZ :

$$y_k(f) = \int_{-1/2}^{1/2} \Phi_k(LW; f - f') dZ(f). \quad (2.7)$$

The choice of $d_k(f)$ should minimize the effects of the sidelobes of $\Phi_k(LW; f)$. If the DPSWF's had no sidelobes, and existed only inside $[-W, W]$, then the convolution would be

$$z_k(f) = \frac{1}{\sqrt{\lambda_k(LW)}} \int_{-W}^W \Phi_k(LW; f - f') dZ(f), \quad (2.8)$$

where the normalization factor $1/\sqrt{\lambda_k(LW)}$ is introduced so that $E[|z_k(f)|^2] = S(f)$.

To minimize the effects of the sidelobes, therefore, we minimize

$$\epsilon_k = |z_k(f) - d_k(f)y_k(f)|^2, \quad (2.9)$$

where, utilizing (2.7) and (2.8),

$$\epsilon_k = \left| \frac{1}{\sqrt{\lambda_k(LW)}} \int_{-W}^W \Phi_k(LW; f - f') dZ(f) - d_k(f) \int_{-1/2}^{1/2} \Phi_k(LW; f - f') dZ(f) \right|^2,$$

and, splitting regions of integration,

$$\epsilon_k = \left| \left(\frac{1}{\sqrt{\lambda_k(LW)}} - d_k(f) \right) \int_{-W}^W \Phi_k(LW; f - f') dZ(f) - d_k(f) \int_{-1/2}^{1/2} \Phi_k(LW; f - f') dZ(f) \right|^2,$$

where the ‘cut’ integral $\int_{-1/2}^{1/2}$ implies integration over the entire region except $[-W, W]$.

Upon expanding, the cross-terms vanish, so that

$$\begin{aligned} \epsilon_k = & \left| \left(\frac{1}{\sqrt{\lambda_k(LW)}} - d_k(f) \right) \int_{-W}^W \Phi_k(LW; f - f') dZ(f) \right|^2 \\ & + \left| d_k(f) \int_{-1/2}^{1/2} \Phi_k(LW; f - f') dZ(f) \right|^2. \end{aligned} \quad (2.10)$$

Since $E[|z_k(f)|^2] = S(f)$, we obtain

$$E \left[\left| \int_{-W}^W \Phi_k(LW; f - f') dZ(f) \right|^2 \right] = \lambda_k(LW) S(f). \quad (2.11)$$

The expectation value of the ‘cut’ integral can also be given by a simple expression: note that the energy of $\Phi_k(LW; f)$ outside of $[-W, W]$ is $1 - \lambda_k(LW)$. Also, for a Gaussian variable, $E[S(f)] = \sigma^2$, where σ^2 is the process variance. With this information, the

expectation value of the 'cut' integral becomes

$$E \left[\left| \int_{-1/2}^{1/2} \Phi_k(LW; f - f') dZ(f) \right|^2 \right] = \sigma^2(1 - \lambda_k(LW)). \quad (2.12)$$

$E[\epsilon_k]$ now becomes

$$E[\epsilon_k] = \left| \frac{1}{\sqrt{\lambda_k(LW)}} - d_k(f) \right|^2 \lambda_k(LW) S(f) + |d_k(f)|^2 \sigma^2(1 - \lambda_k(LW)). \quad (2.13)$$

Minimizing this with respect to $d_k(f)$,

$$\left(\frac{-1}{\sqrt{\lambda_k(LW)}} + d_k(f) \right) \lambda_k(LW) S(f) + d_k(f) \sigma^2(1 - \lambda_k(LW)) = 0$$

and

$$d_k(f) = \frac{\sqrt{\lambda_k(LW)} S(f)}{\lambda_k(LW) S(f) + \sigma^2(1 - \lambda_k(LW))}. \quad (2.14)$$

Now an estimate $\bar{S}(f)$ of the spectrum may be obtained by equation (2.3).

Since the expression for the weights includes the unknown $S(f)$, $\bar{S}(f)$ must be obtained recursively. It is easy to show that $\bar{S}(f)$ may be found by solving

$$\sum_{k=0}^{K-1} \frac{\lambda_k(LW) (\bar{S}(f) - \hat{S}_k(f))}{[\lambda_k(LW) \bar{S}(f) - \sigma^2(1 - \lambda_k(LW))]} = 0. \quad (2.15)$$

4. Harmonic Analysis

Even though they are often computed with the same algorithms, harmonic analysis and spectrum estimation are two different problems. The typical harmonic analysis problem is to find line components hidden in noise, not to produce a smooth power spectrum. Still, both problems are related, and the prolate spheroidal approach to spectrum estimation can be modified to include a way of performing harmonic analysis. Note that, as the data will be modelled as a sum of line components in the frequency domain, Thomson's harmonic analysis is a parametric method, whereas the spectrum estimation method is nonparametric.

A. Tests for single, isolated lines:

Equation (2.5) no longer holds in the presence of line components in the data. If there is a line component of complex amplitude μ at frequency f' , then

$$E[dZ(f)] = \mu\delta(f - f'), \quad (2.16)$$

and the k^{th} -order eigencoefficient $y_k(f)$ has an expected value

$$E[y_k(f)] = \mu\Phi_k(LW; f - f'). \quad (2.17)$$

Using the values of K eigencoefficients at f' , it is possible to obtain an estimate of μ by complex least squares [Miller (1973)], by minimizing

$$\sum_{k=0}^{K-1} |y_k(f') - \mu\Phi_k(LW; 0)|^2 \quad (2.18)$$

with respect to μ , so that the estimated μ , $\hat{\mu}_p$, is

$$\hat{\mu}_p = \frac{\sum_{k=0}^{K-1} y_k(f') \Phi_k(LW; 0)}{\sum_{k=0}^{K-1} |\Phi_k(LW; 0)|^2}. \quad (2.19)$$

(Note that $\hat{\mu}_p$ is a function of frequency, $\hat{\mu}_p(f')$.) Thomson calls this 'point regression', thus the subscript 'p'.

To perform harmonic analysis, $\hat{\mu}_p(f)$ is calculated at every frequency, and at every frequency point one tests the hypothesis that there exists a line component with the estimated amplitude at that particular frequency. This test is done by comparing the estimated power in the line to the estimated power in the background. If we assume the probability distributions of the random variables $y_k(f)$ and $\hat{\mu}_p(f)$ to be complex Gaussian, then their squares will be χ^2_2 -distributed. Since there are K eigencoefficients, testing the hypothesis mentioned above involves computing an F-statistic with $2K - 2$ and 2 degrees of freedom [Hoel (1984)]:

$$F_p(f) = \frac{(K - 1) \sum_{k=0}^{K-1} |\hat{\mu}_p(f) \Phi_k(LW; 0)|^2}{\sum_{k=0}^{K-1} |y_k(f) - \hat{\mu}_p(f) \Phi_k(LW; 0)|^2}, \quad (2.20)$$

and any statistical package (IMSL, for example) will convert this quantity into the probability of there being a line component $\hat{\mu}_p(f)$ at each frequency.

To use information from a wider band, the amplitudes can be calculated by finding the $\hat{\mu}(f) \delta(f - f')$ which, when convolved with $\Phi_k(LW; f)$ will give the best fit to $y_k(f)$

in the band $[f' - W, f' + W]$, i.e. minimizing

$$\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} |y_k(f) - \mu(f') \Phi_k(LW; f - f')|^2 df$$

with respect to $\mu(f')$ at every frequency point f' . The new estimate of the amplitude for this 'integral regression' procedure is denoted by $\hat{\mu}_I(f')$:

$$\hat{\mu}_I(f') = \frac{\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} y_k(f) \Phi_k^*(LW; f - f') df}{\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} |\Phi_k(LW; f - f')|^2 df},$$

and, because of (1.16),

$$\hat{\mu}_I(f') = \frac{\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} y_k(f) \Phi_k^*(LW; f - f') df}{\sum_{k=0}^{K-1} \lambda_k(LW)}. \quad (2.21)$$

The ratio

$$F_I(f') = \frac{\text{estimated power due to the line in } [f' - W, f' + W]}{\text{estimated power in the background in } [f' - W, f' + W]}$$

now becomes

$$F_I(f') = \frac{\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} |\hat{\mu}_I(f') \Phi_k(LW; f - f')|^2}{\sum_{k=0}^{K-1} \int_{f'-W}^{f'+W} |y_k(f) - \hat{\mu}_I(f') \Phi_k(LW; f - f')|^2 df}. \quad (2.22)$$

Lindberg (1986) has shown that $F_I(f)$ does not follow an F-distribution, but probability values may be obtained empirically by taking several samples of a white noise process, and observing how often $F_I(f)$ exceeds different levels. If $p(f')$ is the probability of line component $\hat{\mu}_I(f')$ being present, then

$$p(f') = 1 - \frac{\# \text{ of times } F_I(f) \text{ exceeded } F_I(f')}{\text{total } \# \text{ of frequency points used}}. \quad (2.23)$$

B. Double-line tests:

It should be noted that line-line interactions have been neglected in the discussion up to now. If two lines are separated by $\Delta f \leq 2W$, strong interactions will occur between them, since the power will leak from one line into the other in the spectrum estimate because of the bandwidth of the DPSWF's. These closely spaced double lines will appear as wide peaks in $F_p(f)$, usually more than 95% significant over several frequency points. These peaks should be searched for the presence of double lines inside the width of the peak, so a different test is necessary: if there are two closely spaced line components at frequencies f_1 and f_2 , ($f_1 < f_2$), with respective amplitudes μ_1 and

μ_2 , then

$$E[y_k(f)] = \mu_1 \Phi_k(LW; f - f_1) + \mu_2 \Phi_k(LW; f - f_2). \quad (2.24)$$

To find estimates of μ_1 and μ_2 , we minimize

$$\sum_{k=0}^{K+1} \int_{f_1-W}^{f_2+W} |y_k(f) - \mu_1 \Phi_k(LW; f - f_1) - \mu_2 \Phi_k(LW; f - f_2)|^2 df$$

with respect to μ_1 and μ_2 . This gives rise to a matrix equation of the form

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (2.25)$$

where

$$a_{nm} = \sum_{k=0}^{K-1} \int_{f_1-W}^{f_2+W} \Phi_k(LW; f - f_m) \Phi_k^*(LW; f - f_n) df \quad (2.26)$$

and

$$b_n = \sum_{k=0}^{K-1} \int_{f_1-W}^{f_2+W} y_k(f) \Phi_k^*(LW; f - f_n) df. \quad (2.27)$$

With the new estimates $\hat{\mu}_1$ and $\hat{\mu}_2$, we need to find the value of the ratio

$$F_d(f_1, f_2) = \frac{\sum_{k=0}^{K-1} \int_{f_1-W}^{f_2+W} |\hat{\mu}_1 \Phi_k(LW; f - f_1) + \hat{\mu}_2 \Phi_k(LW; f - f_2)|^2 df}{\sum_{k=0}^{K-1} \int_{f_1-W}^{f_2+W} |y_k(f) - \hat{\mu}_1 \Phi_k(LW; f - f_1) - \hat{\mu}_2 \Phi_k(LW; f - f_2)|^2 df}. \quad (2.28)$$

To find double lines, therefore, one must compute $\hat{\mu}_1$, $\hat{\mu}_2$, and $F_d(f_1, f_2)$ for every possible combination of f_1 and f_2 inside the wide, significant peak in $F_p(f)$. Again, to assign probability values to this double-line statistic, one can obtain an empirical probability distribution function as suggested before for the integral regression statistic.

5. Applications to Stationary Signals

Perhaps the most important use of spectrum estimation techniques comes in the analysis of stationary signals. In fact, no study of a new spectrum estimation method would be complete without some mention of the results obtained by applying the method to a stationary process. This section contains a comparison of the performance of the prolate spheroidal method with the conventional estimators mentioned in the earlier sections. The two examples in (A) are designed to test the smooth 'APS spectrum' technique, whereas examples (B) and (C) examine the 'Harmonic Analysis' part of the prolate spheroidal method.

Example (A.1) - Red Noise

This example consists of 100 points of 'red' noise, that is, noise whose spectrum is dominant in the lower frequencies. The data were created by a process suggested by Oliver Jensen (personal communication) that works as follows:

A time series x_n with long-range periodicities is desired. Each x_n is created by

averaging out some random numbers a_{nm} , $1 \leq m \leq M$,

$$x_n = \frac{1}{M} \sum_{m=1}^{M-1} a_{nm}.$$

Starting at $n = 0$, the a_{nm} 's are changed every m^{th} time, so that a new value of a_{n1} is computed by some random number generator for every n , a new value of a_{n2} is computed every other n , a new a_{n3} every third time, and so on. Note that this procedure builds in long-range periodicities into x_n while still keeping some degree of randomness.

Results of calculating the different spectrum estimates on this data set are shown in Figs. 2.3-2.5. As expected, the periodogram has severe variance problems. Though it is still possible to observe that the power drops off with increasing frequency, it is much easier to see this in the AR estimates (order 5 was used) and in the APS spectrum. Note further that the drastic reduction in variance from the periodogram to the APS spectrum was achieved without smoothing. The time-bandwidth product of 4.0 chosen to produce this estimate would smooth over the details in the spectrum more than the periodogram's time-bandwidth product of 1.0, but for spectra as smooth as this, the tradeoff between variance and resolution is practically meaningless, since there are no details to resolve.

Example (A.2) - Bandlimited Noise

This example is designed to highlight the differences between techniques more than the previous one. Using the same procedure as before, 1000 points of red noise were computed, then bandpass filtered between frequencies 0.1 and 0.3. After filtering, a 100-point sample was taken. A spectrum estimate from this sample, therefore, should

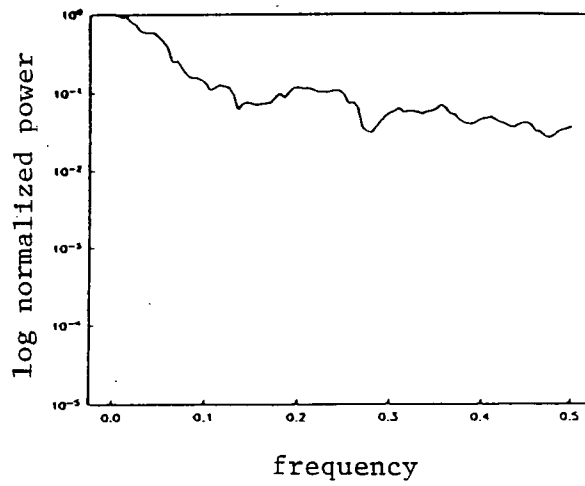


Figure 2.3 - The APS spectrum estimate for Example (A.1).

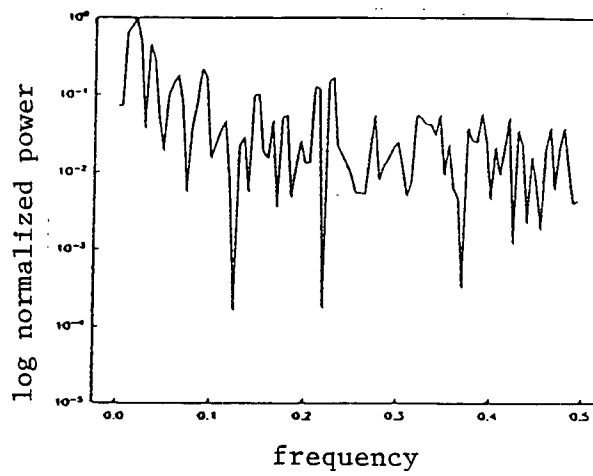


Figure 2.4 - The periodogram spectrum estimate for Example (A.1).

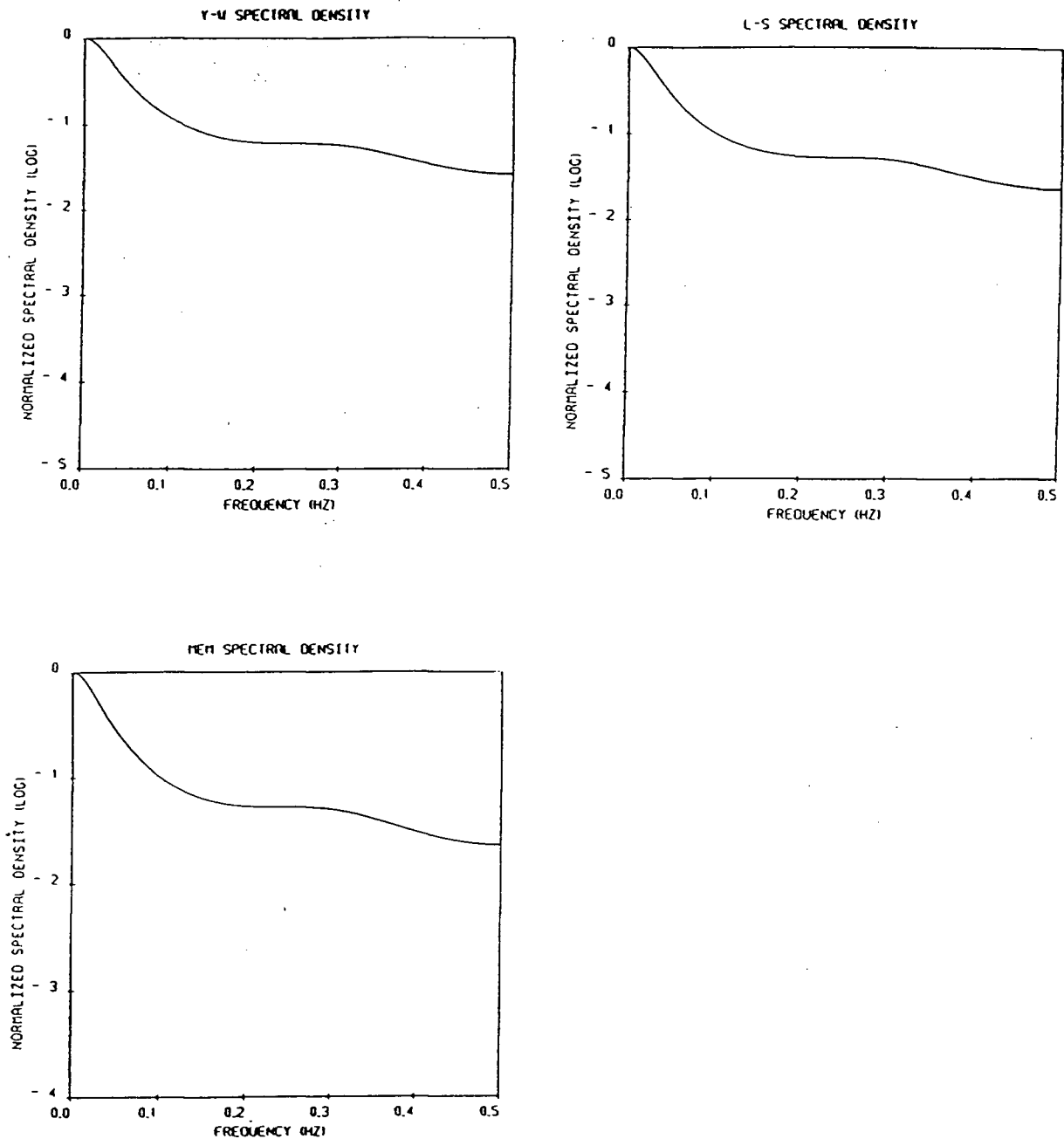


Figure 2.5 - (a) The Yule-Walker estimate for Example (A.1). (b) The Maximum-Entropy estimate. (c) The Least Squares estimate.

drop off smoothly between frequencies 0.1 and 0.3, and be zero outside this band. Such a spectrum with smooth components and sharp edges is said to be ‘mixed’.

Different spectrum estimates for this process are shown in Figs. 2.6-2.8. The periodogram gives very large oscillations inside the signal band, and rather large sidelobes outside this band. As a matter of fact, the sidelobes off the low-frequency end look larger than the signal around $f = 0.20$. This sort of behavior is precisely the reason why other estimates are needed.

The autoregressive estimates, in particular the Maximum Entropy and the Least-Squares techniques, all reproduced the band edges very well for model order 20, but introduced spurious peaks inside the band. Modelling the data with lower orders causes the spurious peaks to disappear, but then the band edges are not reproduced as well. This illustrates the problem of choosing a model order for an autoregressive estimate. The FPE criterion [Akaike (1970)] reaches a minimum at order 45, but the resultant spectrum for that order is excessively ‘spikey’ inside the band of interest. Choosing a model order is an important consideration when dealing with AR estimates, so the choice must be made with care.

The APS spectrum estimate does manage to identify the smoothness of the process for $0.1 \leq f \leq 0.3$, and the power does drop off sharply outside this band. The band edges, however, are not found accurately. From Fig. 2.6, one might think that the edges are at $f \simeq 0.06$ and $f \simeq 0.34$. However, as the time-bandwidth product chosen was 4.0, an overshoot of 0.04 is expected to appear on a sharp drop such as this for a data set 100 points long. One last disagreeable feature of this APS spectrum is the spurious noise at the low-frequency end. This is probably caused by the sidelobes of the higher

order eigenspectra, and can be removed by choosing $TW = 5.0$. Such choice, however, increases the overshoot at the band edges.

To some degree, all estimates exhibit some tradeoff between resolution and variance. The periodogram can be smoothed to reduce variance at the expense of resolution, AR estimates may be obtained by modelling as high or low-order processes, and the APS spectrum may be computed with different time-bandwidth products, depending on the degree of smoothness desired. Still, this does not mean that all techniques are equally good, just that they should be applied to the proper problem. Although the APS spectrum may not be the final solution to all problems of estimating smooth spectra, it is a good way of obtaining a low-variance estimate of a spectrum without making any assumptions about the data *a priori*.

Example (B) - Closely Spaced Spectral Lines

As Example (A) was chosen to test the performance of Thomson's algorithm for smooth spectra, Example (B) tests the high-resolution capabilities of the harmonic analysis algorithm. Since spectral lines are produced by a wide variety of physical processes, and the frequencies and amplitudes usually carry information about the process itself, it is interesting to see how well different algorithms resolve closely spaced lines. The limit for resolving such components is usually taken to be the 'Rayleigh limit', or $\Delta f_R = 1/T$, where T is the total time of observation. It is a popular belief that the periodogram will not resolve lines spaced closer than this. One purpose of this example is to demonstrate that this Rayleigh limit is only useful as a coarse measure of resolution capacity, since line frequency estimates depend on the initial phase of the signals.

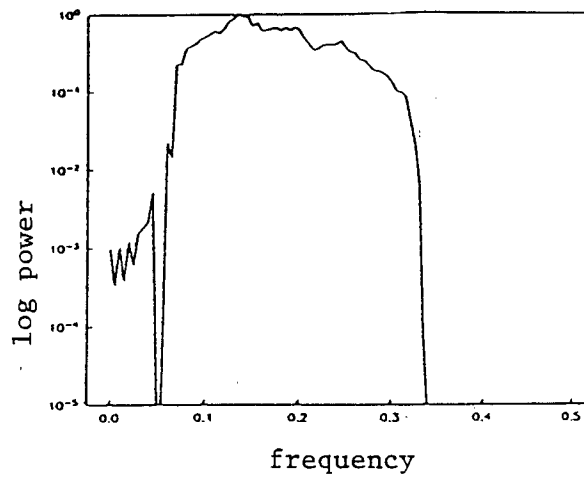


Figure 2.6 - APS spectrum for Example (A.2).

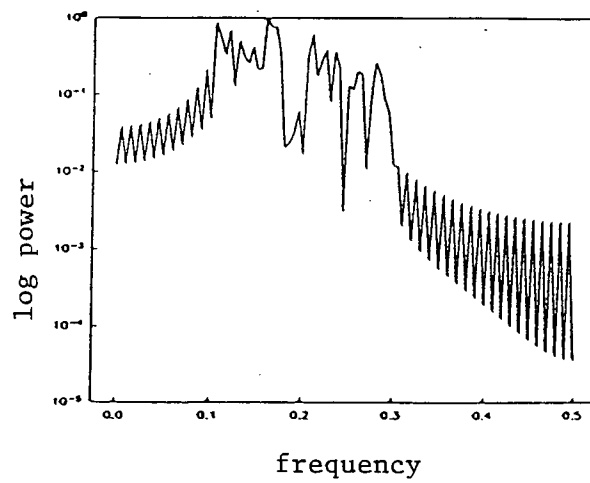


Figure 2.7 - Periodogram for Example (A.2).

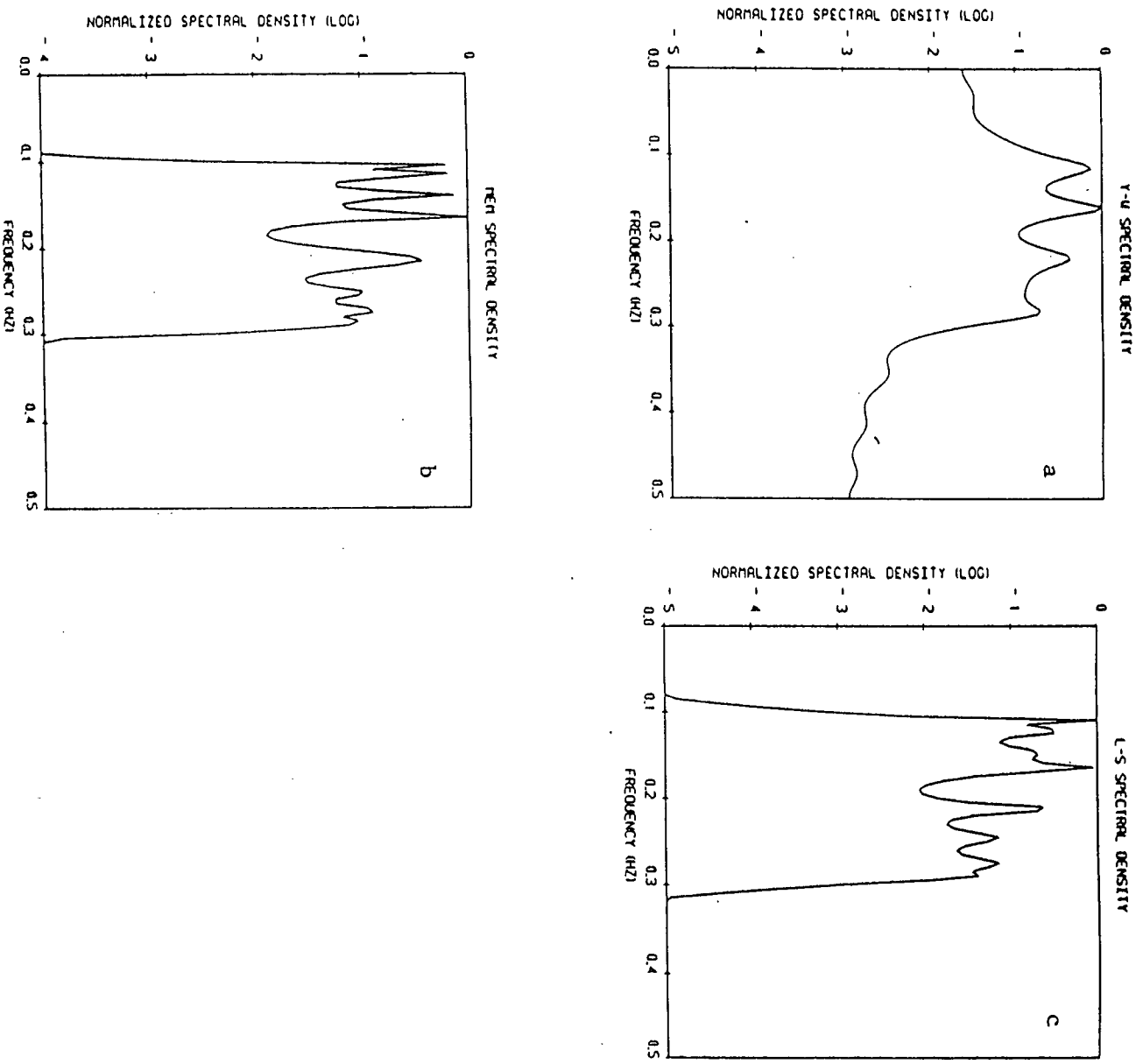


Figure 2.8 - (a) Yule-Walker spectrum for Example (A.2). (b) Maximum-Entropy spectrum (c) Least Squares spectrum.

The synthetic data used were

$$x_n = \cos 2\pi f_1 n + \sin 2\pi f_2 n + s_n,$$

with $f_1 = 0.195$, $f_2 = 0.210$, and s_n a white noise process with standard deviation $\sigma_s = 0.1$. One hundred points of these data were computed, and two thirty-point samples were examined in detail. For a thirty-point sample, the Rayleigh resolution limit is $\Delta f_R = 0.033$, so the two sinusoids in this example are over two times closer than the Rayleigh limit. The first sample consisted of points 24-53, and its spectral estimates are plotted in Figs. 2.9-2.11.

For this particular initial phase, the periodogram has no problem in recognizing the presence of two lines, although the periodogram peaks occur at the wrong frequencies, $f_1 = 0.185$ and $f_2 = 0.220$. For the AR spectra, an order of 10 was used, and all three techniques recognized the two peaks. Furthermore, all three gave more accurate positions for the lines than the periodogram.

The procedure to obtain Fig. 2.9a is rather complicated. First, a choice must be made for the time-bandwidth product. If a small value is chosen, only a few DPSS's may be utilized as data tapers, and the F-test performs better when there are several eigencoefficients to work with. Experience shows that, for signal-to-noise ratios such as the one in this example, $3.5 \leq TW \leq 5.5$ gives good results, and so TW is usually chosen to be 4.0, and this automatically sets the upper limit of K to 8. Eigencoefficients are then computed and an F-test is performed at every frequency for 2 and 14 degrees of freedom (the F-test is plotted in Fig. 2.9b). Two peaks in the F-test are significant above the 99% level (this is given by $F = 6.51$), reaching maxima at $f = 0.185$ and $f = 0.220$, the

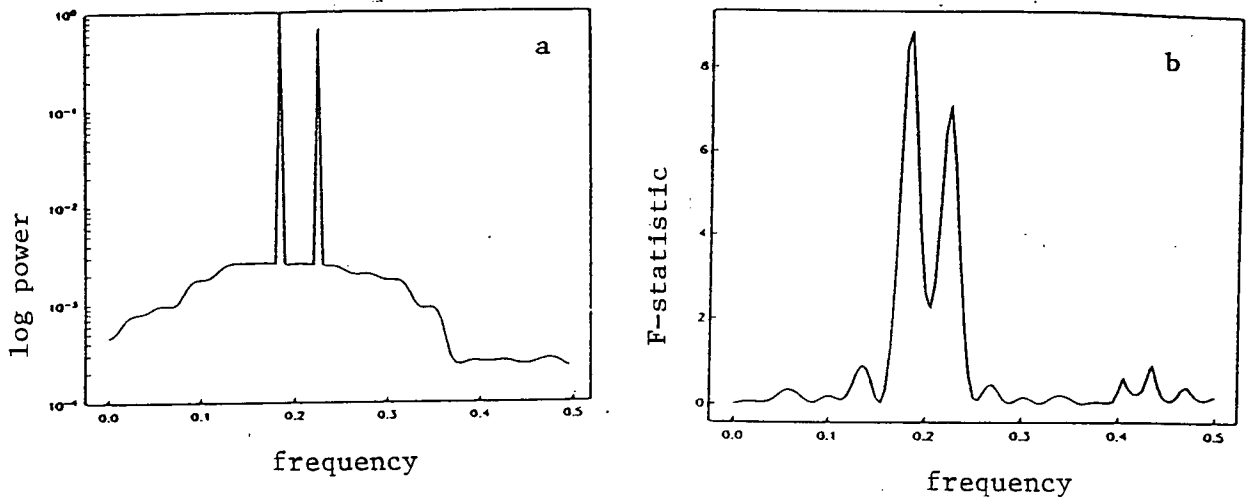


Figure 2.9 - (a) APS spectrum for Example (B), first sample. (b) F-statistic for the same data set.

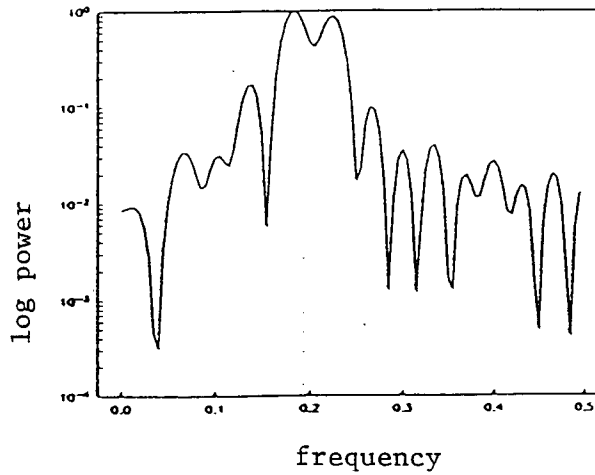


Figure 2.10 - Periodogram for Example (B), first sample.

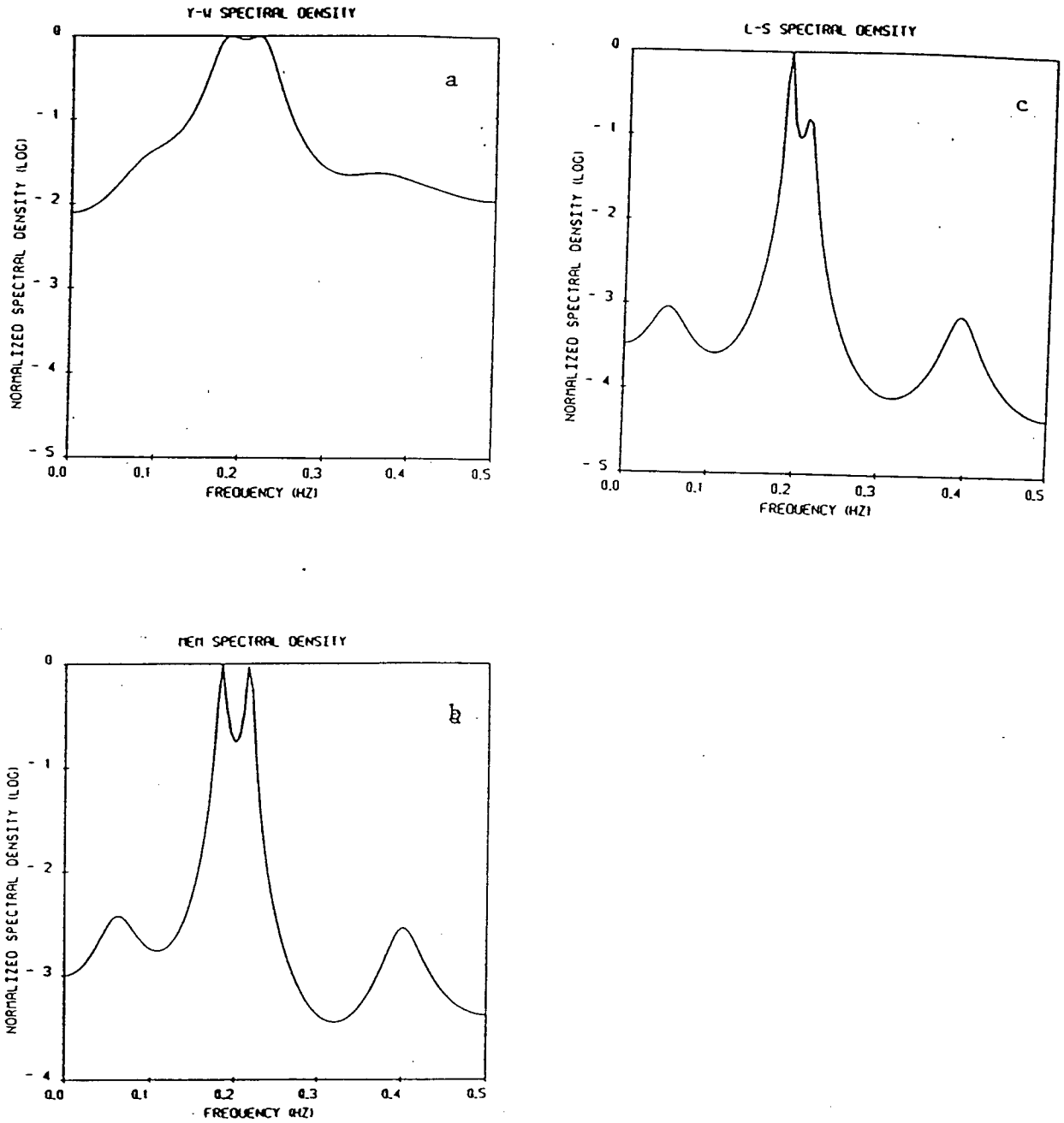


Figure 2.11 - (a) The Yule-Walker spectrum for Example (B), first sample. (b) Maximum-Entropy spectrum. (c) Least Squares spectrum.

same as the periodogram peaks. The two sinusoids at these frequencies and with their corresponding amplitudes are then removed from the data, and the average spectrum is computed for the residuals. The two lines at their estimated frequencies and amplitudes are then replaced, and the result of this is shown in Fig. 2.9a. For all this complicated procedure, however, Thomson's method shows no improvement over the periodogram in the position of the lines in this example, and the power of the line at $f = 0.220$ is estimated to be down by a factor of 3 relative to the other, whereas in fact they should have the same power.

The other data set gives reason to be more optimistic about this technique, however. These data consisted of points 58-87 of the same process, but the spectral estimates (Figs. 2.12-2.14) are very different from the previous ones. The periodogram failed to resolve the two frequencies completely, and it was necessary to go to order 17 in the AR estimates to distinguish the two lines. Even then, the Yule-Walker technique could not resolve the components. The single-line F-test in Thomson's method gave a highly significant single peak at $f = 0.205$ (Fig. 2.12b). The F-statistic value at this frequency is 283, so the probability of the F-statistic being this high without the presence of a line is less than 10^{-6} . The peak in the F-test is wide, however. It may not be noticeable from Fig. 2.12b, but this peak is significant above the 99% level over a band of width $W/2$, whereas peaks for single lines on other synthetic data were usually narrower than $W/4$. A double-line test was then done inside the width of this peak, and a very high double line probability was found. (At this point it should be stressed that the double-line probability distribution function was approximated by an F-distribution to avoid the great computational expense incurred in calculating high probability values for $F_d(f_1, f_2)$ on several different time-bandwidth products. Thomson himself, in *Thomson*

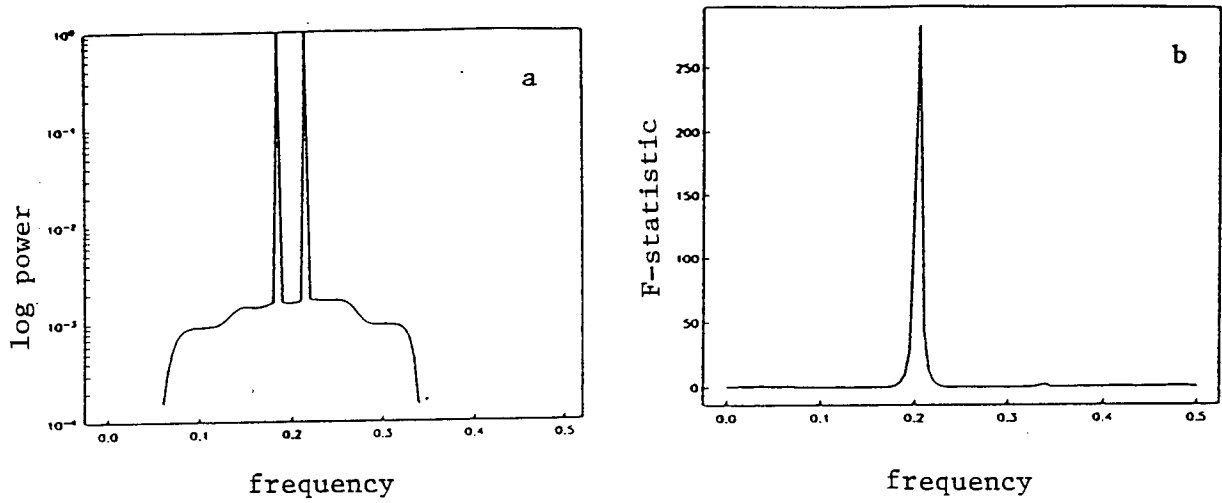


Figure 2.12 - (a) APS spectrum for Example (B), second sample. (b) F-statistic for the same data set.

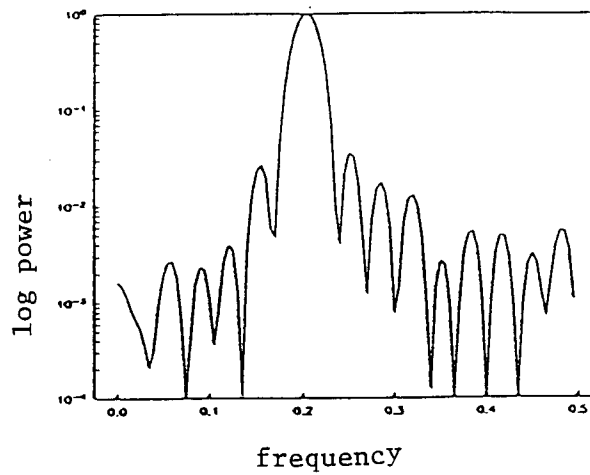


Figure 2.13 - Periodogram for Example (B), second sample.

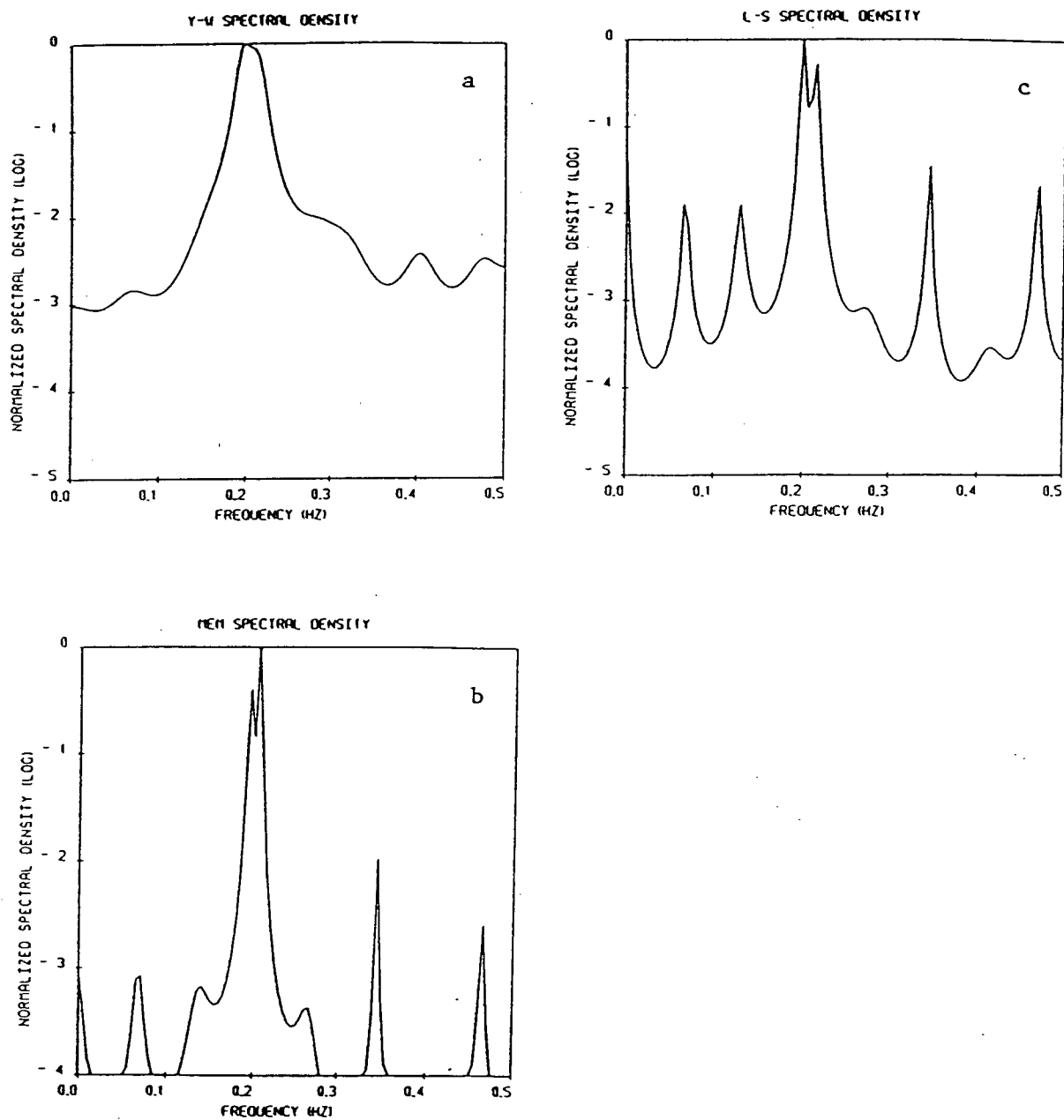


Figure 2.14 - (a) Yule-Walker spectrum for Example (B), second sample. (b) Maximum-Entropy spectrum. (c) Least Squares spectrum.

(1982), used this approximation when working on the Kay and Marple example.) The double line probability exceeded 99.9999% for $f_1 = 0.195$ and $f_2 = 0.210$, so a choice had to be made between a double and a single line. Again, experience shows that F-statistic peaks for single lines are usually narrow, and their double line probabilities are low (less than 95%). For these reasons, in cases when both single and double line probabilities are high, the double line is chosen preferentially. Double lines are chosen when the double line probability is higher than the single line probability (the most common case in the actual presence of two lines), and when both single and double line probabilities are high. Final results are shown in Fig. 2.12a.

Example (C) - Sinusoids in Noise

Although at first it may seem a disadvantage to have to set a probability threshold to choose spectral line components, it is useful to know what probability there is that a peak in a spectrum estimate is caused by a sinusoid in the data or by noise. Example (C) illustrates the usefulness of the F-test in discriminating between sinusoids and noise.

The data consist of 25 points of two sinusoids, one at $f = 0.17$ with amplitude 1.0, the other at $f = 0.30$ with amplitude 0.6. White noise with a standard deviation of 1.0 was added to the two sinusoids.

The periodogram (Fig. 2.16) shows peaks at roughly the correct frequencies, plus an extraneous peak near the Nyquist frequency and a possible peak near the low frequency end. All the AR estimates (order 7 was seen to be appropriate) yield roughly the same result (see fig. 2.17), so there is no doubt that the noise must look like rather large amplitude sinusoids at low and high frequencies. The F-test (Fig. 2.15) demonstrates that, although the power at some frequency may be large, this does not necessarily

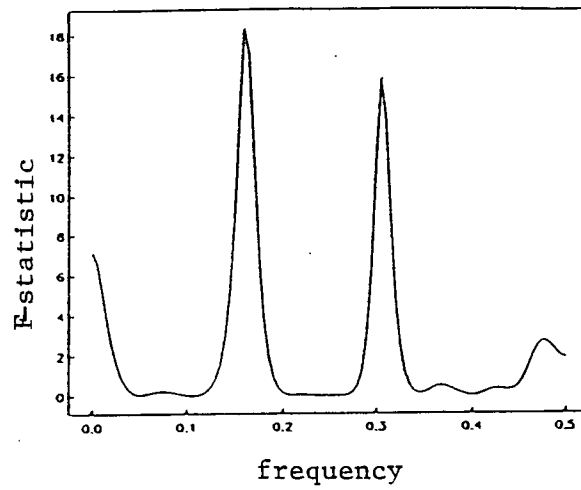


Figure 2.15 - Single-line F-statistic for Example (C).

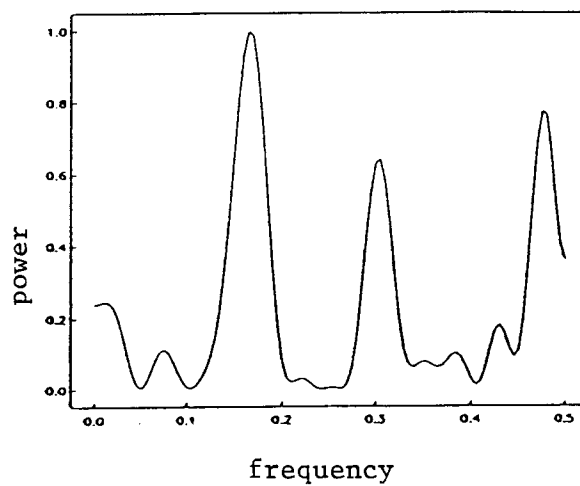


Figure 2.16 - Periodogram for Example (C).

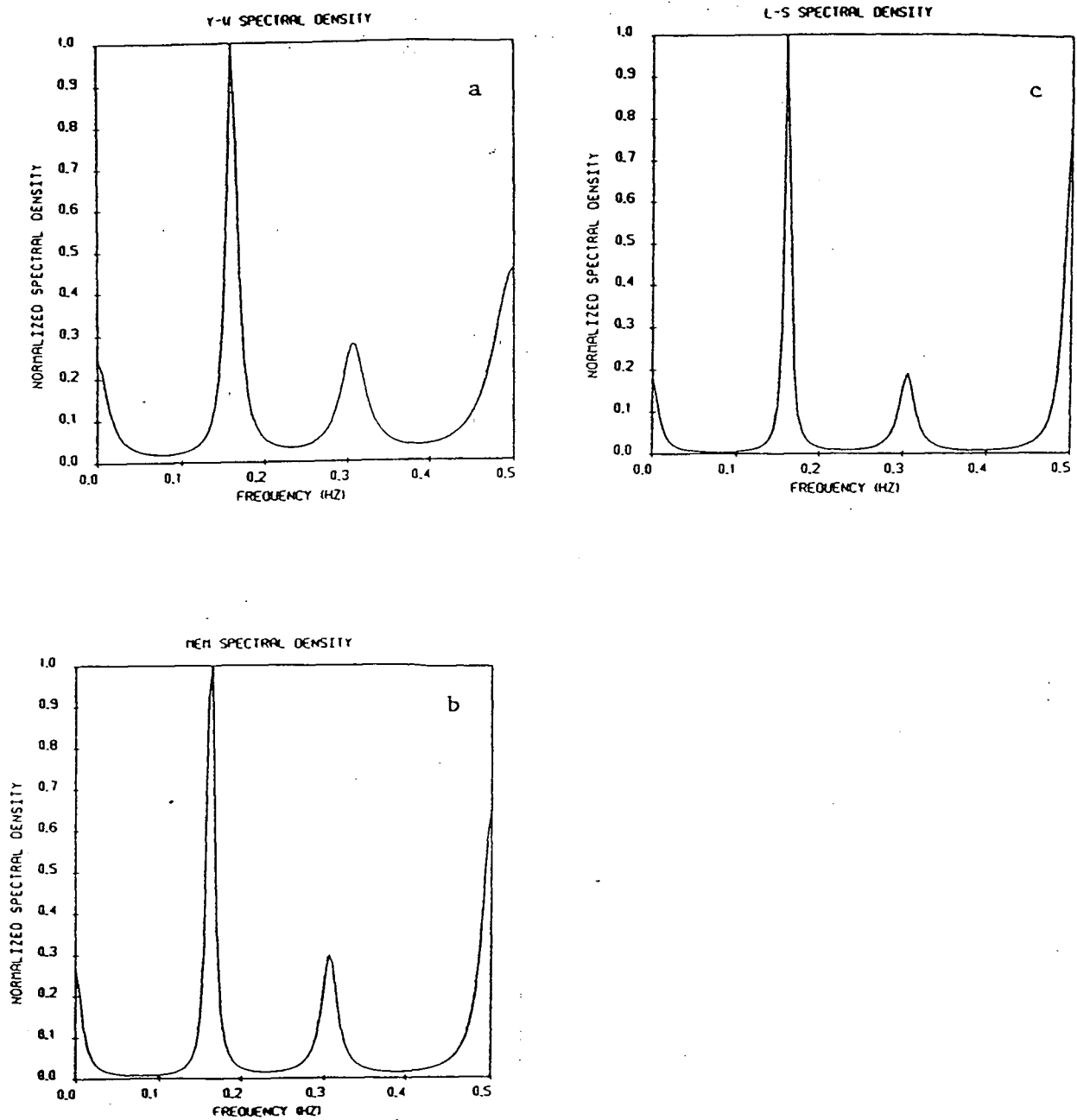


Figure 2.17 - (a) Yule-Walker spectrum for Example (C). (b) Maximum-Entropy spectrum. (c) Least Squares spectrum.

mean that it is due to a line component. In noisy data, large time-bandwidth products may prove disastrous when doing harmonic analysis, because they may leak power from the noise into the sinusoidal frequency. For this reason, a rather small $TW = 3.0$ was chosen to analyze these data. Also, the higher-order data tapers would be the first affected by the noise, since they are the least bandlimited, so $K = 5$ was used, instead of the upper limit $K = 6$ for the chosen time-bandwidth product. The F-test shows a peak of 18.4 at $f = 0.165$ for 99.9% probability, and a peak of 15.9 at $f = 0.035$ for 99.8% probability. The peak near DC has about 97% probability, and the peak near the Nyquist frequency is below 94%.

6. Applications to Nonstationary Signals

As a curve can be approximated by a series of straight-line segments, so a nonstationary time series can be represented as a collection of short, independent, stationary time series, each with its own independent spectrum. Observation of how these spectra evolve with time is what gives insight into the frequency domain of the nonstationary signal. The accuracy of this approximation depends on the length of each 'locally stationary' sample: the shorter the windows are, the better the approximation will be. With this approach, the problem of nonstationarity becomes one of resolution, since we need to draw as much information as possible from a very short, often noisy time series.

To find the spectrum as a function of time, then, we take a window of length W_l , 'slide' it across the data, and compute the spectrum of the data inside this window at each step. The window length should be chosen as short as possible without causing problems of resolution. Using the rough criterion of the Rayleigh limit of resolution, W_l

should be chosen so that $W_l \sim 1/f_l$, where f_l is the lowest frequency of interest. If the data have a broad dynamic range, they should be subdivided into different frequency ranges and analyzed separately for each.

The present section is an application of the prolate spheroidal harmonic analysis technique to analyze nonstationary data. Three examples are covered:

- (A) Amplitude variation with time (exponentially damped sinusoids)
- (B) Frequency variation with time.
- (C) Real Earthquake data

Examples (B) and (C) include a comparison of the prolate spheroidal method to the periodogram and to the adaptive AR algorithm suggested by *Griffiths and Prieto-Diaz* (1977). Note that the adaptive AR method is not based on sliding a window across the data.

Example (A) - Exponentially Damped Sinusoids

One of the more interesting types of nonstationary signals is one where the amplitude is decaying exponentially with time. Given such a case and as part of the preprocessing, we can multiply the data by $e^{\beta t}$ to make the signal closer to stationary. If only one decaying sinusoid is present in the data, β should be a first guess of the actual damping factor, and can be obtained by sliding a short window across the data and fitting an exponential decay curve to the decaying amplitudes obtained. In the presence of two or more sinusoids, β should be chosen to be somewhere between the different damping factors. The choice of β is not crucial, however, and in practice one can try different values of β and average out the results.

In this example, the data consisted of two exponentially damped sinusoids with different damping factors:

$$x(t) = e^{\alpha_1 t} \cos 2\pi f_1 t + e^{\alpha_2 t} \cos 2\pi f_2 t + s(t)$$

where $\alpha_1 = -.09$, $f_1 = .20$, $\alpha_2 = -.13$, $f_2 = .30$, and $s(t)$ was a white noise process with standard deviation $\sigma_s = .1$. The Rayleigh limit for these two sinusoids would be given by $\Delta f_R = 0.04$ if they were stationary, but if a locally stationary sample is taken to be one where the amplitude of a sinusoid changes by 10%, then the Rayleigh limit becomes about 0.1, close to the separation of the sinusoids in the example.

A preliminary run was performed with a 6-point window, and an average frequency of .208 was found. The best-fitting exponential decay curve fitted to the amplitudes was $e^{-.115t}$. After this, ten trials were performed on the data using window lengths ranging from 10 to 13 points, and multiplying the data by $e^{\beta t}$ with β ranging from .1 to .2. The average frequency and damping factor estimates after the ten trials were as follows:

$$f_1 = .194 \pm .0076,$$

$$\alpha_1 = -.099 \pm .0114,$$

$$f_2 = .294 \pm .0061,$$

$$\alpha_2 = -.126 \pm .0085,$$

$$f_3 = .499 \pm .0027,$$

$$\alpha_3 = -.006 \pm .0029.$$

The spurious frequency near $f = .5$ appeared only late in the data set, where the SNR was low, and showed up in only five of the ten trials. The sinusoid near $f = .3$ was usually seen only near the beginning of the data set where the SNR was high, and died away halfway through.

Note that the amplitude of a sinusoid at $f = .2$ and $\alpha = -.09$ will decay by roughly 10% after each cycle, making this example a sharply nonstationary signal. On longer data sets with lower damping constants and only one frequency of interest, errors of the order 1% can be obtained even under the adverse conditions of peak SNR=3 (throughout this study, SNR is defined as the ratio of the maximum signal amplitude to the standard deviation of the noise).

A possible drawback of using this method on damped sinusoids is that the estimates deteriorate at low frequencies and high damping constants, because a data set of length $l = 1/f$ is needed to estimate a frequency accurately, and thus estimation of low frequencies depends on long data sets. High damping factors, furthermore, make it difficult to choose a window in which the signal is 'locally stationary'.

Example (B) - Changing Frequencies

When the frequency of a signal changes with time, the 'goodness' of an estimator should be judged by how quickly and accurately it can adapt to these changes in frequency. Figs. 2.18-2.26 are a comparison of doing a periodogram on each window versus doing adaptive AR or Thomson's harmonic analysis. In all these figures, PGM stands for periodogram estimate, ADAR for adaptive AR estimate, PSS for prolate spheroidal estimate, WL is the window length, FL is the AR filter length, MU the adapting param-

eter μ , SS the number of points between successive WL-point windows, and SNR is the signal-to-noise ratio as defined above. The solid lines indicate the theoretical response.

The example chosen is the following:

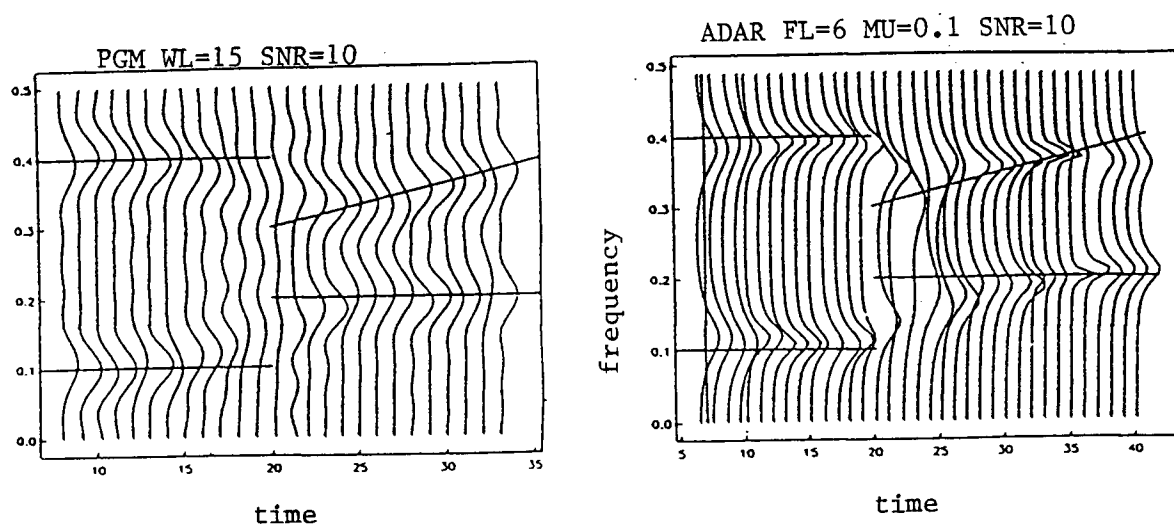
$$x(t) = \begin{cases} \cos 2\pi(.1t) + \cos 2\pi(.4t), & \text{if } t < 20; \\ \cos 2\pi(.2t) + \cos 2\pi g(t)t, & \text{if } t \geq 20; \end{cases}$$

where $g(t)$ is a linear function of time. Noise is added so that SNR is as shown on each figure.

It may be observed that the shorter window lengths and the larger values of μ give faster convergence, and that all three methods perform reasonably well. Both the periodogram and the adaptive AR techniques have the advantage that they may be computed about five times faster than the prolate spheroidal method, since the latter involves taking several FFT's at each window.

Example (C) - A Real Data Test

The data used in this example are from a magnitude 6.0 earthquake focussed under Vancouver island on December 16, 1957. The seismograms were recorded in Pasadena, California, and digitized at UBC in 1985 by John Cassidy. The digitized data were interpolated to yield equispaced time intervals of 1.0 second, and the DC component was filtered out. The seismogram used was the E-W component of ground motion, caused by Love waves. Because the waves travelled a long distance, the frequencies should be separated well enough to perform harmonic analysis on them. Seismograms from nearby earthquakes do not lend themselves to this type of analysis.



(left) Figure 2.18 - Periodogram applied to Example (B), window size 15, SNR=10. (right) Figure 2.19 - Adaptive AR algorithm applied to Example (B), order 5, $\mu = 0.1$.

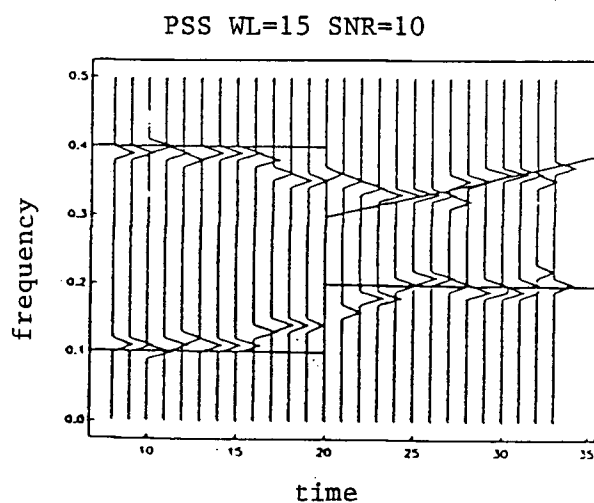
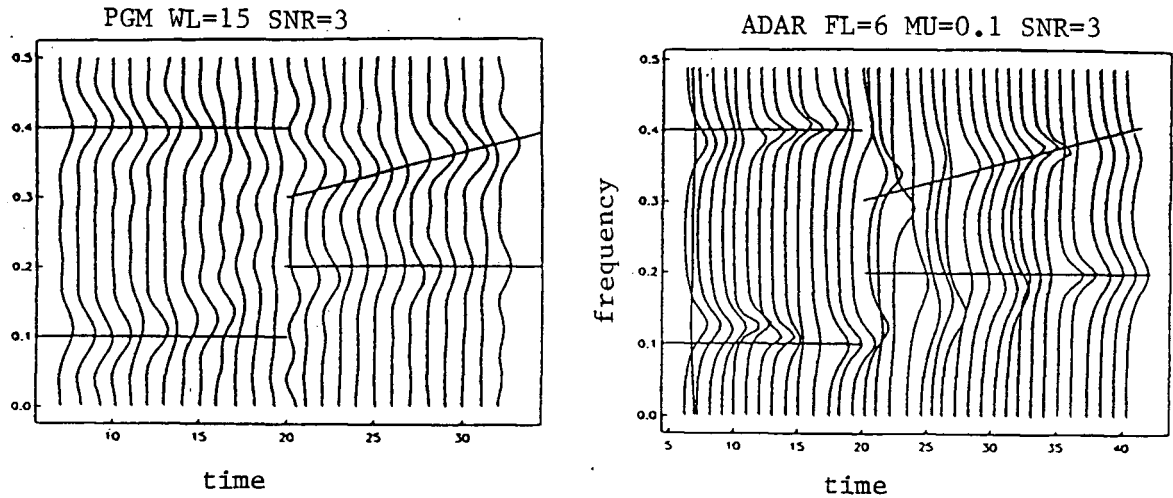


Figure 2.20 - Prolate spheroidal harmonic analysis applied to Example (B), window size 15, SNR=10.



(left) **Figure 2.21** - Periodogram applied to Example (B), window size 15, SNR=3. (right) **Figure 2.22** - Adaptive AR algorithm applied to Example (B), order 5, $\mu = 0.1$

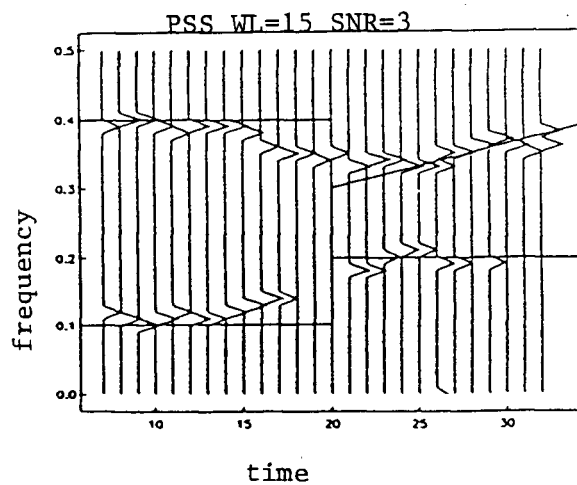
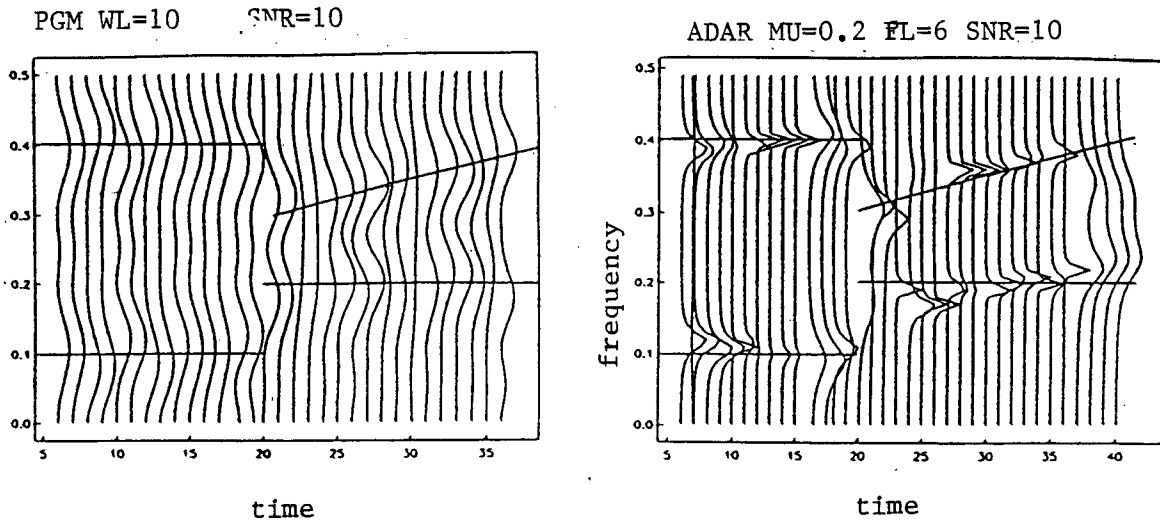


Figure 2.23 - Prolate spheroidal harmonic analysis applied to Example (B), window size 15, SNR=3.



(left) **Figure 2.24** - Periodogram applied to Example (B), window size 10, SNR=10. (right) **Figure 2.25** - Adaptive AR algorithm applied to Example (B), order 5, $\mu = 0.2$.

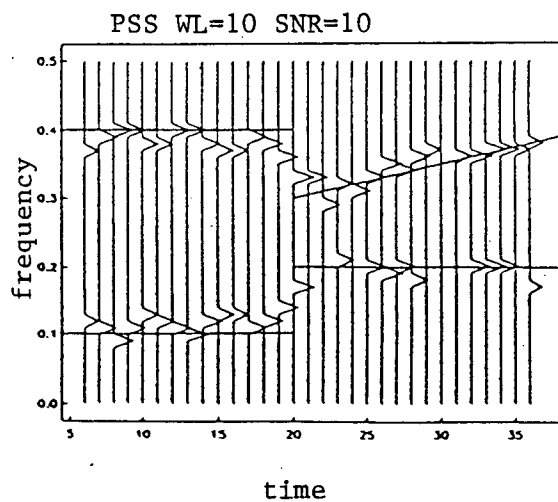


Figure 2.26 - Prolate spheroidal harmonic analysis applied to Example (B), window size 10, SNR=10.

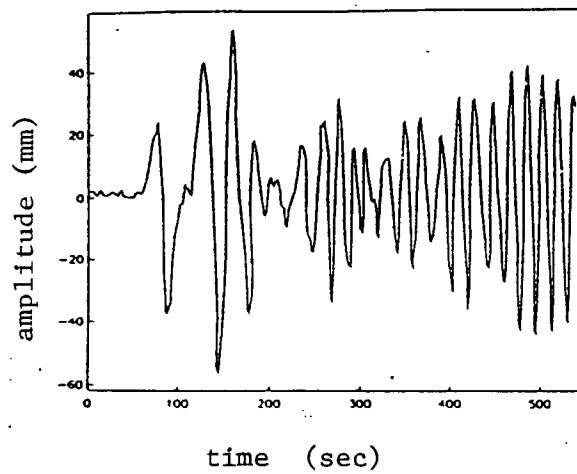


Figure 2.27 - Love wave seismogram for Example (C).

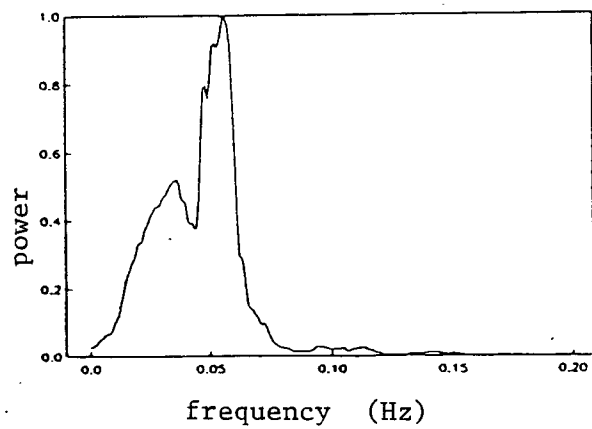


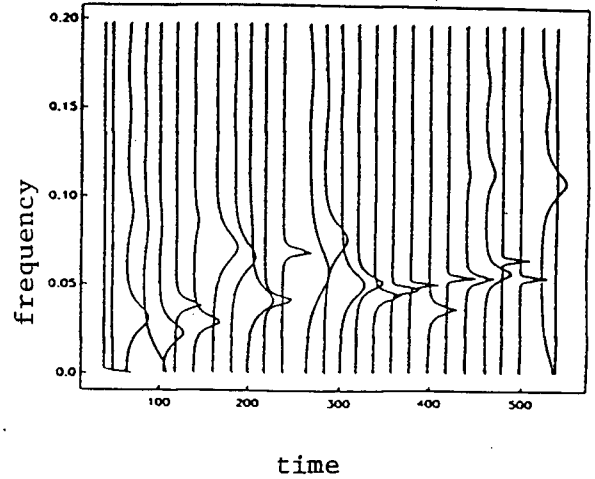
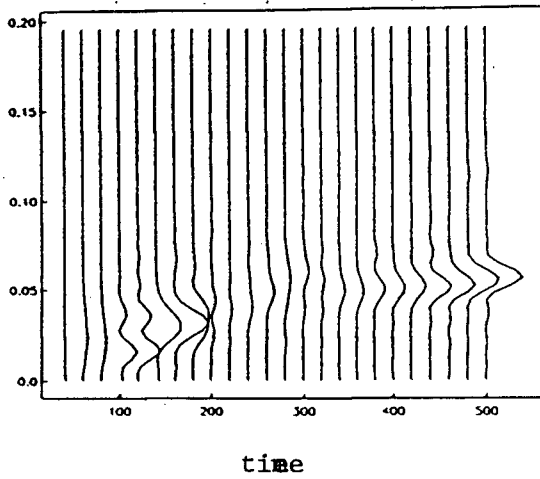
Figure 2.28 - Power spectrum of the Love wave seismogram.

Prior to examining the time dependence of the data, a power spectrum of the entire seismogram was computed to provide an idea of what frequency range was of interest, to allow a choice of an appropriate window length. The dynamic range of the signal is not too broad, so it is not necessary to use two windows. From Fig. 2.28, the Love wave seismogram has $f_l \sim .015$, so the window length was chosen to be around 70 seconds. What we expect to see is a slow rise in frequency as time progresses, because group velocity for surface waves increases with decreasing frequency [Garland (1979)]. Furthermore, above frequencies of .05Hz we expect to see noise caused by crustal scattering.

The Love wave seismogram yields some interesting results. A *double peak* can be seen between $t \sim 80s$ to $300s$ and $f \sim .01Hz$ to $.05Hz$ (see Fig. 2.31), indicating that there are two frequencies with the same group velocity, a fact confirmed by conventional processing methods [Hermann (1978)].

This may be because the path of the waves is partly in oceanic lithosphere and partly in continental lithosphere. The double peak in the last two traces may be due to noise, although there does seem to be some modulation in the data in the time domain (Fig. 2.27) around 500s. This observation may be a topic for further examination, and shall not be discussed here.

Note that the periodogram failed to detect some of the double peaks around 180s, whereas the double peak appeared on some of the other traces. This is because periodogram spectral estimates have a fairly strong phase dependence which is relieved somewhat by the harmonic analysis algorithm presented here. The adaptive AR algorithm also missed the double peak when modelled with orders between 5 and 25. Higher orders were not attempted. For this particular real data test, Thomson's harmonic



(left) **Figure 2.29** - Periodogram method applied to the Love wave seismogram. (right) **Figure 2.30** - Adaptive AR applied to the Love wave seismogram.

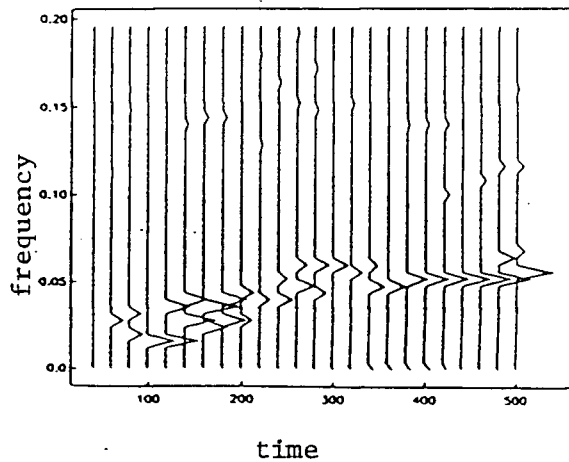


Figure 2.31 - Prolate spheroidal harmonic analysis applied to the Love wave seismogram.

analysis performed significantly better than a straight periodogram or the adaptive AR method, justifying to some extent its computational burden.

7. Conclusion

The prolate spheroidal technique of spectrum estimation is a useful method of reducing the variance involved in computing unsmoothed periodogram estimates. It has the further advantage that the data are not assumed to fit any sort of model, so there is no concern about incorrectly modelling the observations. The tradeoff between bias and resolution that affects this technique is also present in most methods of spectrum estimation. The harmonic analysis technique presented in this chapter enhances the resolution of spectral line components, and has the advantage that confidence levels are set to distinguish signal peaks from noise peaks.

Perhaps the most promising applications of the prolate spheroidal methods outlined in this chapter are to three different problems:

- (a) Estimating the spectrum of data whose true spectrum is known to be slowly varying, and using this as a guide to discriminate between different models of the data.
- (b) Discriminating between noise peaks in a spectrum and real sinusoidal components.
- (c) Finding the parameters of exponentially damped sinusoids in noise.

Both the methods for smooth spectra and harmonic analysis suffer from a large computational burden, and so rather than being used for routine data processing, they should be applied on occasions when accurate frequency-domain information is required. When only a rough idea of the spectrum is desired, the straight periodogram or the Fourier transform of the data windowed by the zeroth-order DPSS are an adequate alternative.

-¿Qué rumor es ése, Sancho?

-No sé, señor,—respondió él—. *Alguna cosa nueva debe ser, que las aventuras y desventuras nunca comienzan por poco.*

—Miguel de Cervantes Saavedra, *Don Quijote de la Mancha*

CHAPTER III

SNR Enhancement for Bandlimited Signals

One of the principal aims in signal processing is to eliminate noise from a corrupted version of some true signal. If the true signal is known to be bandlimited, and the band that it exists in can be found, then the prolate spheroidal functions (or a modified version of these) can be used to separate signal from noise. Consider the following two-dimensional problem: there is a signal $u(x, t)$ that is limited to the domain $-X/2 \leq x \leq X/2$ and $-T/2 \leq t \leq T/2$ and highly concentrated in the spectral domain in both dimensions, i.e. most of its energy is concentrated in some region of the $k_x - f$ plane. We record some corrupted version of the true signal, $g(x, t)$, such that

$$g(x, t) = u(x, t) + s(x, t), \quad (3.1)$$

where $s(x, t)$ is some noise function that is not limited in the spectral domain. We wish to recover $u(x, t)$.

Although the goal of this chapter is to treat the SNR enhancement problem in two dimensions, understanding of the one-dimensional problem is necessary prior to moving on to the multidimensional case. The 1-D problem follows in section 1.

1. The Problem in One Dimension

The recorded signal in the 1-D case is

$$g(t) = u(t) + s(t).$$

The most common problem occurs when $s(t)$ is Gaussian white noise, and thus has a spectrum that is relatively flat. If $u(t)$ is bandlimited, its spectrum will be very small outside of a certain band $a \leq |f| \leq b$, and then the energy of $g(t)$ outside this band can be considered to be due to the noise.

One way to get rid of the noise, then, is to take the Fourier transform of $g(t)$ and apply a frequency-domain filter $D(f)$, so that

$$\hat{U}(f) = D(f)G(f),$$

where $\hat{U}(f)$ is the estimated true signal in the f -domain, and $G(f)$ is the Fourier transform of $g(t)$. The choice of $D(f)$ is crucial for the final estimate of $u(t)$. There are two common approaches: (a) $D(f)$ is taken to be a 'boxcar' function, and (b) $D(f)$ is unity inside most of the band of interest, but drops off to zero smoothly outside this band. There are two serious problems with approach (a):

- (i) Truncation of the Fourier transform will introduce discontinuities of order at least σ_s (σ_s is the standard deviation of the noise), and this will cause Gibbs' oscillations in the time domain.
- (ii) This does nothing about the noise inside the signal band.

Approach (b) is an attempt to solve problem (i), but objection (ii) still stands.

The problem with this kind of filtering may be seen as a problem with the choice of basis functions: the data are expanded by Fourier basis functions, and then those components in the expansion with properties dissimilar to those known properties of the data are downweighted or thrown away. However, the Fourier basis functions make no distinction between signal and noise in the signal band.

A better solution is to expand $g(t)$ by some basis that separates the data into correlated and uncorrelated components (the Karhunen-Loève approach), because this separates noise from signal in the entire frequency domain. The solution presented in this chapter is to expand the data into components that are concentrated in the signal band and components that are not.

If the signal band is $0 \leq |f| \leq b$, then $g(t)$ may be expanded by the prolate spheroidal functions described in Chapter I, whereas if the lower limit is nonzero, then a modification must be made: to find the desired functions that are concentrated in some arbitrary band $a \leq |f| \leq b$, we find the large values of the ratio

$$\lambda = \frac{\int_{-b}^{-a} |H(f)|^2 df + \int_a^b |H(f)|^2 df}{\int_{-1/2}^{1/2} |H(f)|^2 df}.$$

Proceeding as in Chapter I, we obtain an integral equation

$$\lambda h(t) = \int_{-T/2}^{T/2} \left[\frac{\sin 2\pi b(t' - t)}{\pi(t' - t)} - \frac{\sin 2\pi a(t' - t)}{\pi(t' - t)} \right] h(t') dt'$$

normalized to

$$\lambda \psi(x) = \int_{-1}^1 \psi(x') \left[\frac{\sin c_b(x' - x)}{\pi(x' - x)} - \frac{\sin c_a(x' - x)}{\pi(x' - x)} \right] dx' \quad (3.3)$$

($c_a = \pi Ta$, $c_b = \pi Tb$), so that, again, the desired functions are eigenfunctions of a symmetric kernel, and they are denoted by $\phi_k(c_a, c_b; t)$. Note that the prolate spheroidal functions can be obtained by setting $a = 0$. Figure 3.1 contains these functions for $k = 0$ to 3 and $a = .01$, $b = .1$. Figure 3.2 contains the power spectrum of these functions. Figs. 3.3 and 3.4 are the same for $a = .1$, $b = .2$.

Now $g(t)$ may be expanded by this basis to give an estimate of $u(t)$,

$$\hat{u}(t) = \sum_{k=0}^{K-1} \alpha_k \phi_k(c_a, c_b; t).$$

For the case where $c_a = 0$, the value of K is given by *Landau and Pollak* (1962) to be $K = 2Tb$. The case where c_a is nonzero must be studied further, but as a rule of thumb only those eigenfunctions whose associated eigenvalue is $\lambda_k \geq .7$ should be used in the expansion, and empirically that results in $K = 2T(b - a)$. An example was designed to test this filtering technique.

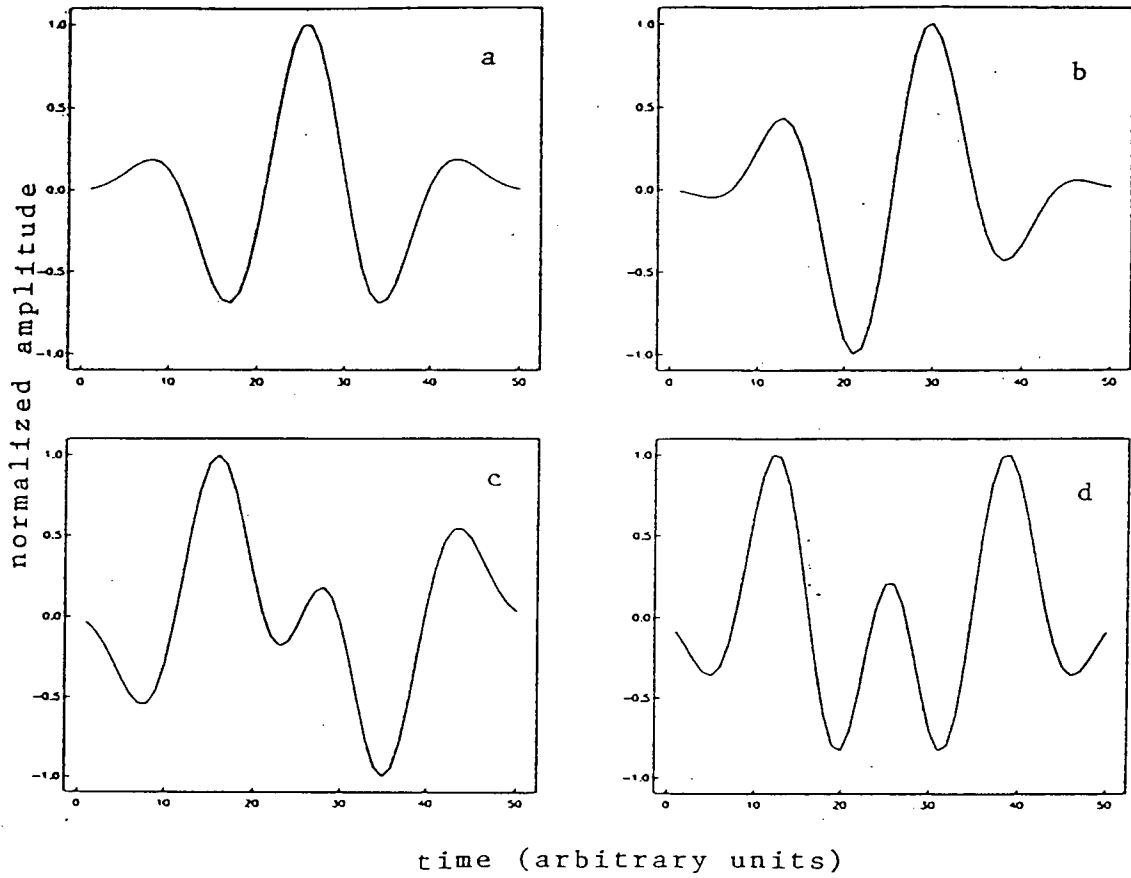


Figure 3.1 (a) Zero-order bandlimited function for $a=.01$, $b=.1$. (b) First-order bandlimited function. (c) Second-order bandlimited function. (d) Third-order bandlimited function.

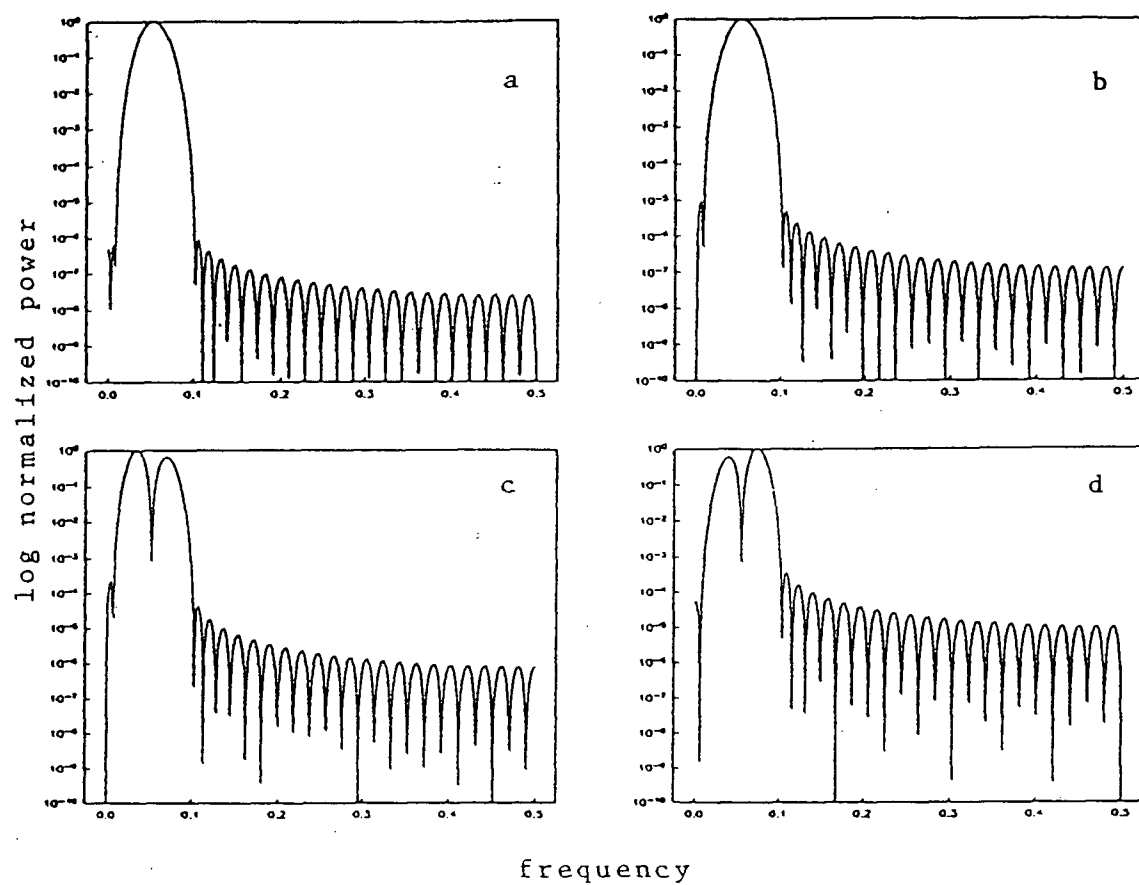


Figure 3.2 $|\Phi_k(La, Lb; f)|^2$ for the functions in Figure 3.1 (log scale on the vertical axis). Figs. a-d are the spectra of their corresponding time domain functions in Fig. 3.1.

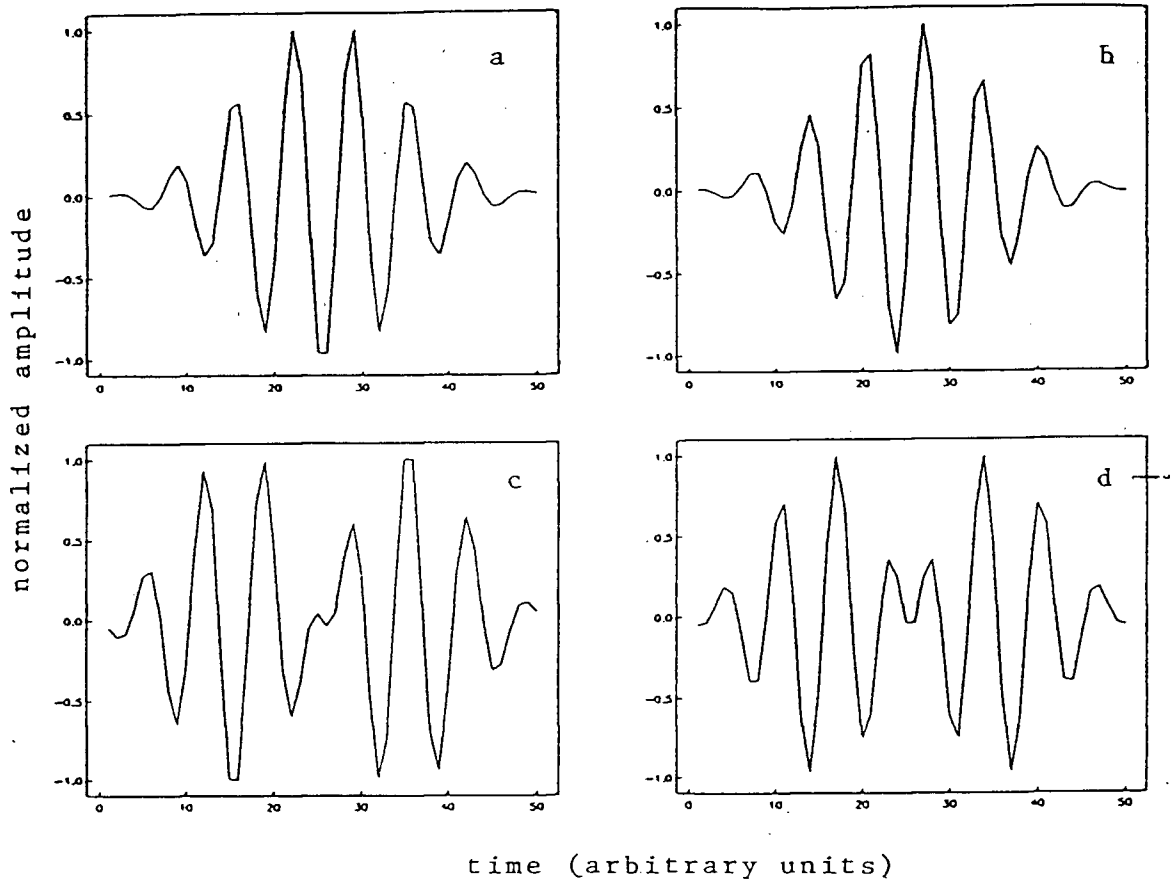


Figure 3.3 (a) Zero-order bandlimited function for $a=.1$, $b=.2$. (b) First-order bandlimited function. (c) Second-order bandlimited function. (d) Third-order bandlimited function.

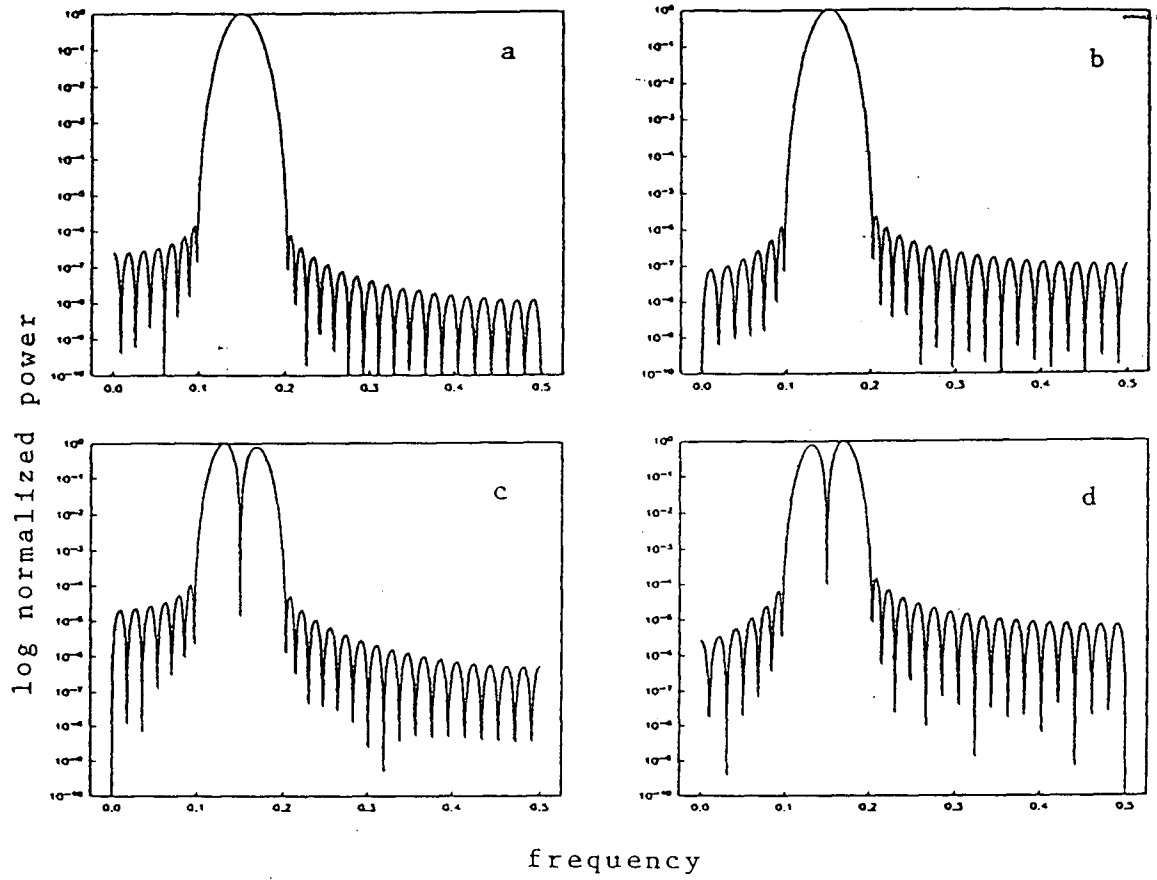


Figure 3.4 $|\Phi_k(La, Lb; f)|^2$ for the functions in Figure 3.3 (log scale on the vertical axis).

In the example, $u(t)$ was taken to be a Ricker wavelet, and three types of noise reduction schemes were attempted: (a) a 'boxcar' filter in the frequency domain, (b) a 'boxcar' tapered at the ends by a Gaussian drop-off curve (GDO for short), (c) an expansion as discussed in the previous section. Also, three different noise levels were added: $\sigma_s = .05, .20$, and $.30$. The three different data sets are compared to the original Ricker wavelet in Fig. 3.5, and the performance of the three techniques on each of the three data sets is given in Figs. 3.6-3.8. To determine what frequencies determined the signal band for all three techniques, a periodogram was taken to observe the band where the power of the signal was above noise level. The same band was used for all three methods, and the Gaussian decay constant that produced best results was the one used for the GDO filter in each case.

The performance of the three methods may be judged by finding the least-squares error (lse) between the true signal and the estimated signal. The lse for the expansion was smaller than the lse for either filter on all three data sets, and each time about 20% less than the lse of the GDO filter. Note how Gibbs' phenomenon becomes more of a problem on the 'box' filter as the noise level is increased, because the discontinuity introduced in the Fourier transform is larger.

2. The Problem in Two Dimensions

To extend this same procedure to two dimensions, we could try to find 2-D basis functions which are concentrated in some region of the $k_x - f$ plane, however, this may be a much harder problem. The reflection seismology experiment illustrates this very well: from *Sheriff and Geldart* (1983), reflection seismic signals usually have an

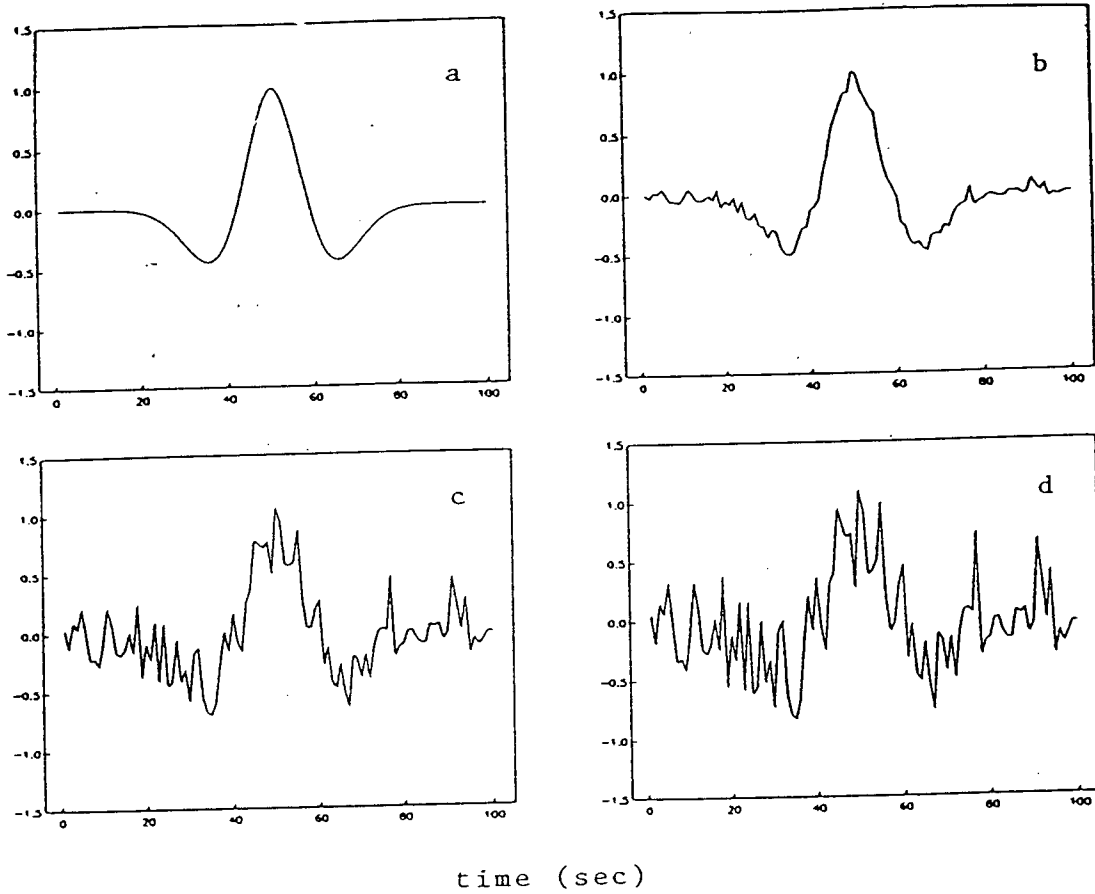


Figure 3.5 (a) The true data, a Ricker wavelet. (b) The true data plus noise, $\sigma_s = .05$. (c) The true data plus noise, $\sigma_s = .2$. (d) The true data plus noise, $\sigma_s = .3$.

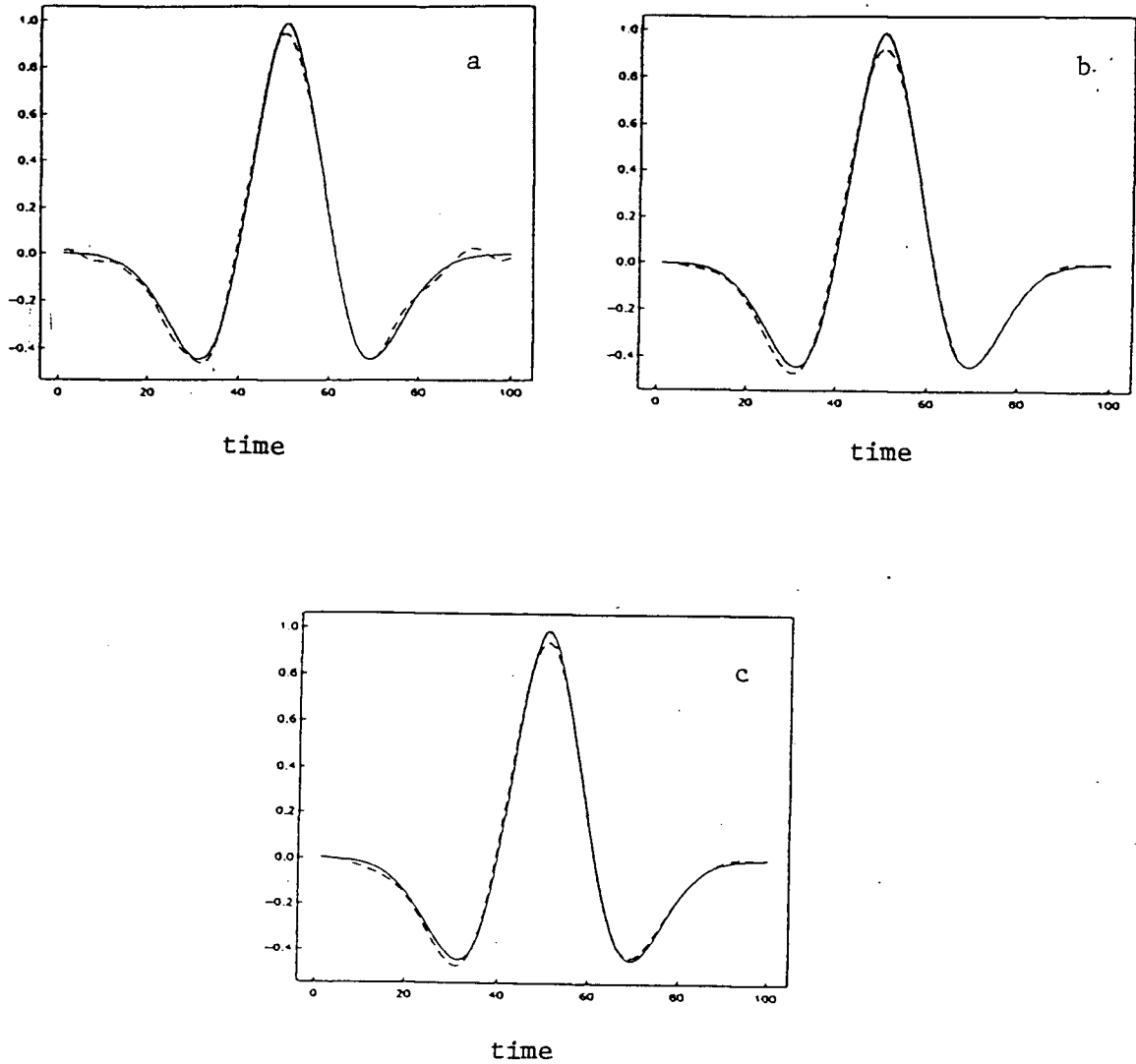


Figure 3.6 The true signal (solid line) and the reconstructions from three different methods (dashed lines) for $\sigma_s = .05$. (a) Using a 'boxcar' filter, $lse = .057$. (b) Using a GDO filter, $lse = .055$. (c) Using an expansion by bandlimited functions, $lse = .038$.

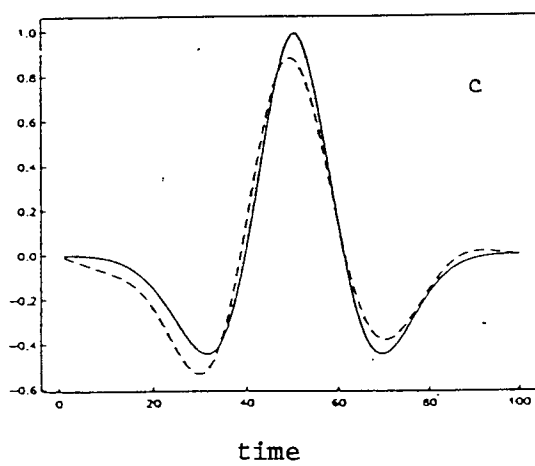
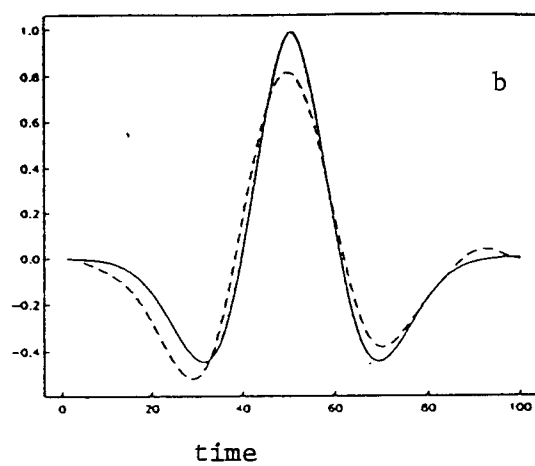
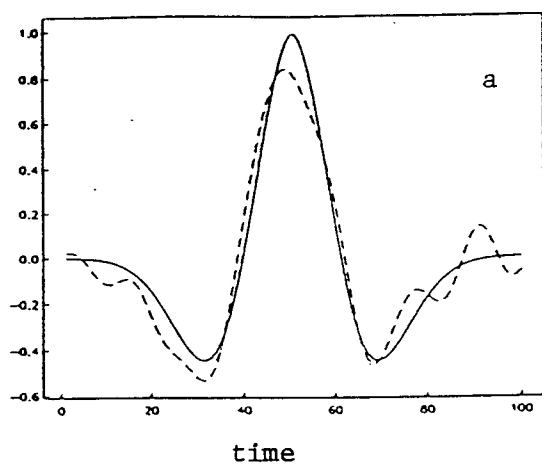


Figure 3.7 Same as Figure 3.6, except $\sigma_s = .2$. (a) $lse = .86$. (b) $lse = .68$.
(c) $lse = .48$.

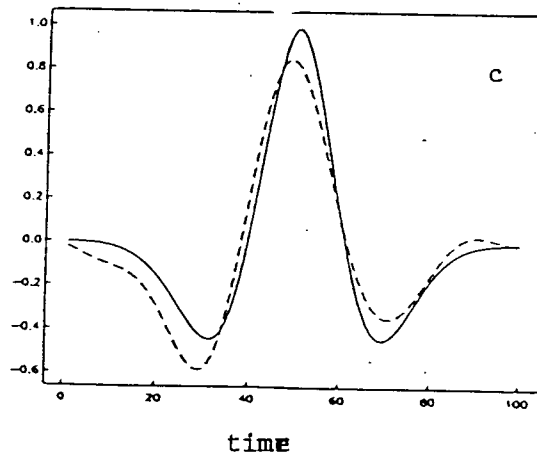
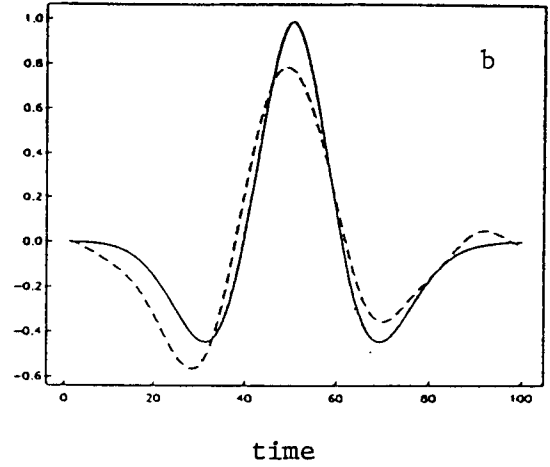
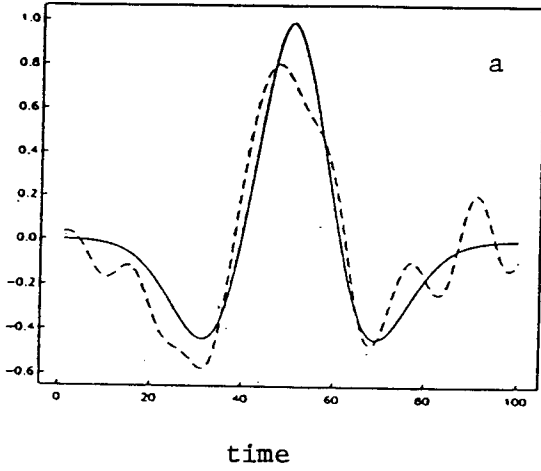


Figure 3.8 Same as Figures 3.6 and 3.7, except $\sigma_s = .3$. (a) $lse = 1.9$. (b) $lse = 1.2$. (c) $lse = 1.0$.

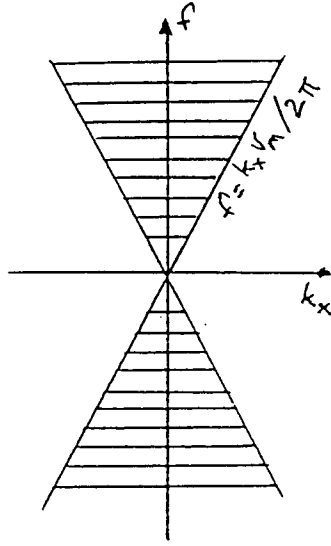


Figure 3.9 Seismic signals are confined to an ‘hourglass’ shaped region in the $k_x - f$ plane. Energy outside this region is probably noise.

apparent velocity in the x -direction $v_x = 2\pi f/k_x$ that is larger than some minimum, v_m .

Taking this into account, and since seismic wavelets are bandlimited, seismic sections are confined to an ‘hourglass’ shaped region in $k_x - f$ space (see Fig. 3.9). Attempts to find basis functions $\psi(x, t)$ which are confined to this region S in the $k_x - f$ plane prove very difficult, as one must solve an equation of the form

$$\lambda \psi(x, t) = \int_R \int K_S(x' - x, t' - t) \psi(x', t') dx' dt',$$

where R is the finite region in the $x - t$ plane for the desired functions, and K_S is a kernel that depends on the shape of the region S in the $k_x - f$ plane. In the one-dimensional case, to construct a set of 10-point basis functions, the kernel is a 10×10 matrix. Here, to compute a set of 10×10 basis functions, the kernel is a formidable $10 \times 10 \times 10 \times 10$ matrix! The enormity of this problem suggests that one find a better way of expanding in two dimensions.

An alternative is to find a space where the expansions may be done separately for each direction. The $\tau-p$ plane suggested in *Schultz and Claerbout* (1978) and *McMechan and Ottolini* (1980) is one such space. If p is defined to be $p = \Delta t / \Delta x = 1/v_x$, then the minimum x -apparent velocity v_m translates to a maximum value of p , p_m . Seismic data, therefore, are bandlimited in p , and the 2-D noise lies outside this p -band. In one dimension, to filter out high frequencies, an expansion may be done in the time domain by f -limited functions. Here, to filter out high values of p , an expansion can be done in the q -domain by p -limited functions, where q is defined to be the inverse Fourier dual of p :

$$G(p) = \int_{-Q/2}^{Q/2} g(q) e^{-i2\pi qp} dq \quad (3.4)$$

when $g(q)$ is zero outside the interval $-Q/2 \leq q \leq Q/2$. The same functions used to do f -filtering may be used to do p -filtering, and the parameters that define these functions are now p_m and Q . Q is chosen so that the basis functions in q -space are not aliased: $Q \leq 1/\Delta p$, where Δp is the p -sampling interval.

The 2-D filtering may be done in the following steps:

- (i) Slant stack the data.
- (ii) Compute the basis functions in q .
- (iii) Fourier transform the q -basis functions to obtain p -basis functions.
- (iv) Expand the data in p by these basis functions to filter out high values of p .
- (v) Inverse slant stack to get back to $x-t$ space.
- (vi) Expand the data in t by $\phi_k(c_a, c_b; t)$ to filter out high frequencies in each trace.

Step (iii) must be done carefully, so that the p -basis functions do not come out complex. To prevent this, the odd-ordered DPSS's in q must be multiplied by i , since they have odd symmetry. The expansion may be done in the p -domain because, as demonstrated in Chapter I, the prolate spheroidal functions retain orthogonality upon Fourier transformation. The expansion in p rather than q saves the computation time of doing a 2-D FFT on the $\tau - p$ data.

This procedure was tested by a synthetic example (Fig. 3.10) which consisted of three events: one horizontal, one dipping gently to the right, and one dipping steeply to the right. The purpose of the processing was to eliminate the steeply dipping event while leaving the other two intact. Results are shown in Figure 3.11. Note that the steep event has been eliminated almost completely, but the wavelet shape has been stretched. The stretching occurs because slant stacking and inverse slant stacking are both integration procedures, and they act as low-enhance filters. These effects are counteracted somewhat by applying a high-enhance filter after inverse slant stacking, but it is not easy to find just the right balance to cancel the integration effects.

3. Limitations and Conclusions

The computational limitations outlined in Appendix B prevent this from being an effective method for noise reduction, even in one dimension. Since there are problems computing the functions for large time-bandwidth products, the method outlined in this chapter fails to work for these cases. Other forms of computing the DPSS's, perhaps through some asymptotic formulae, are needed to calculate the low-order basis functions to do expansions in general [see *Slepian* (1978), and *Slepian* (1965)]. Another possibility

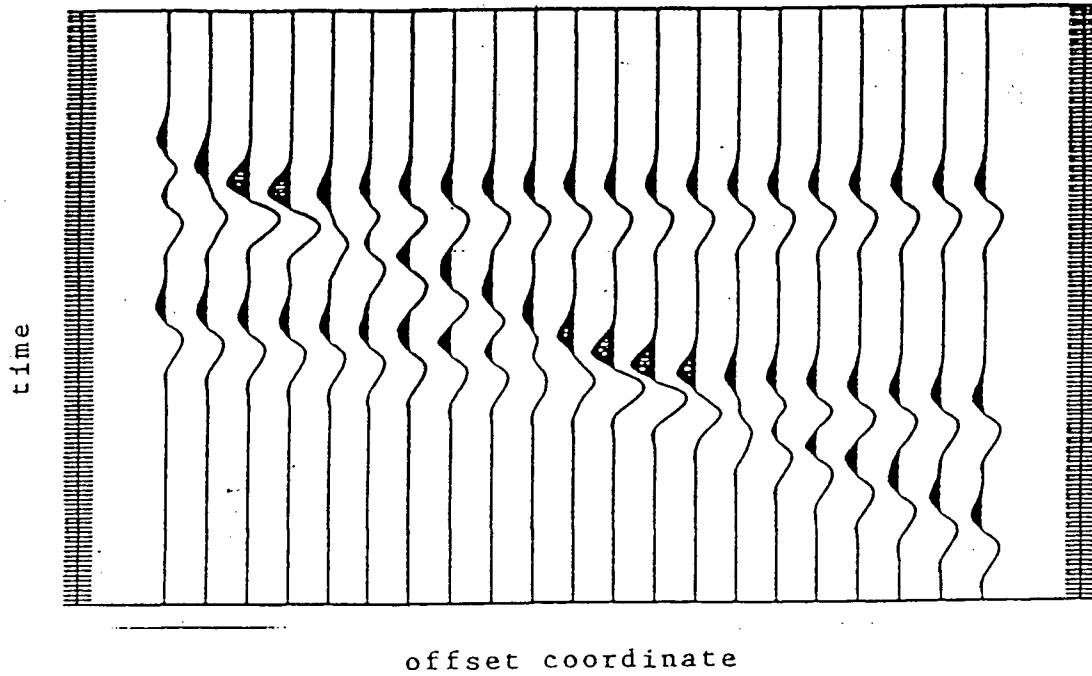


Figure 3.10 The original data in two dimensions. The steeply dipping event is considered to be noise and should be filtered out.

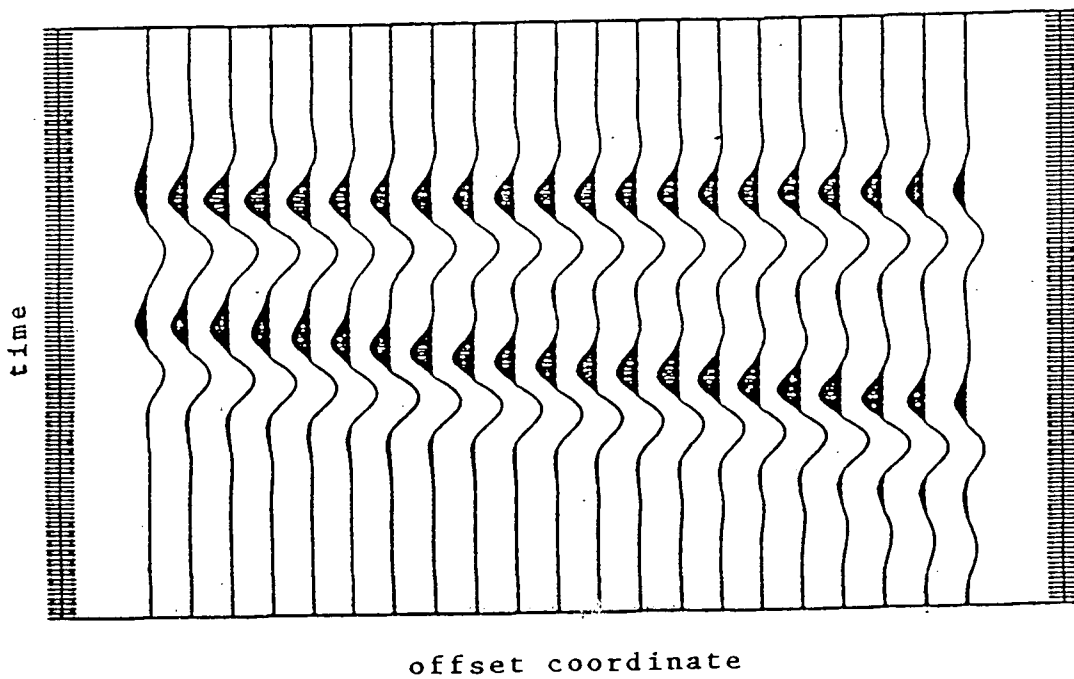


Figure 3.11 The data after filtering in p by prolate spheroidal expansion. The steeply dipping event has essentially disappeared.

to deal with large time-bandwidth products would be to take different windows of the data and expand each separately, but this may introduce discontinuities. A solution to this may be to slide a window across the data, and take the average value of all the expansions at each point, but this is computationally inefficient.

Filtering by bandlimited expansion looks encouraging at first, but the computational problems are difficult to overcome. On the other hand, the good performance of these expansions on signals with small time-bandwidth products suggests applications to wavelet estimation, because wavelets usually have small time-bandwidth products. This possibility is explored in Chapter IV.

El no trabajaba. El buscaba las botijas llenas de bombas doradas que hacen “plo-cosh!” cuando la reja las topa, y vomitan plata y oro, como el agua del charco cuando el sol comienza a ispiar detrás de lo del ductor Martínez, que son los llanos que topan al cielo.

—Salarrué, ‘La Botija’

CHAPTER IV

Seismic Wavelet Estimation

1. The Convolutional Model

Seismic data are usually modelled as the convolution of a reflectivity function $r(t)$ with a wavelet $w(t)$:

$$x(t) = \int_{\tau_1}^{\tau_2} w(\tau)r(t - \tau)d\tau. \quad (4.1)$$

The objective of seismic data processing is to extract from the seismogram $x(t)$ the geological information contained in $r(t)$. Without any knowledge of the wavelet or the reflectivity, the space of possible solutions to this problem is so large that there is no hope of knowing whether a calculated solution is anywhere close to the true one. In order to reduce the solution space, constraints are introduced and a norm is chosen such that the desired solution optimizes this norm. The final solution, then, is only as good as the assumptions made in imposing the constraints and choosing the norm.

Consider the following case: there is an acoustic log from a well near a seismic survey area, and this log yields a reflectivity sequence $r_{log}(t)$ which is close to the reflectivity $r_0(t)$ that produced some data $x_0(t)$. From this approximate reflectivity and the data we wish to recover the wavelet (the length of the wavelet is assumed known), and, simultaneously, the reflectivity sequence $r_0(t)$. Since convolution in the time domain is equivalent to multiplication in the frequency domain, the frequency domain version of (4.1) is

$$X(f) = W(f)R(f), \quad (4.2)$$

where upper-case letters imply the Fourier transform of the lower-case letters. An estimate of the wavelet can be obtained by

$$\hat{W}(f) = \frac{X_0(f)}{R_{log}(f)},$$

$R_{log}(f)$ being the Fourier transform of the well log reflectivity and $X_0(f)$ the Fourier transform of the seismogram. The wavelet $\hat{w}(t)$ obtained by this procedure will fit the data when convolved with $r_{log}(t)$, but this will shed no light on the true wavelet or the true reflectivity. Additional constraints placed on the wavelet and choosing the wavelet that optimizes some norm will perhaps yield an estimate $\hat{w}(t)$ that will not satisfy the data exactly when convolved with $r_{log}(t)$, but may give some measure of the difference between $r_{log}(t)$ and $r_0(t)$. *Oldenburg, et al.* (1981) suggested calculating the ‘smoothest’ wavelet constrained to have zero area that fits the data, where ‘smoothest’ is defined as having least energy in its second derivative. One characteristic feature of the seismic wavelet is that it is approximately both time and bandlimited. This feature has thus far not been used as a constraint but may be incorporated by means of modelling the

wavelet as a linear combination of $K = 2T(b - a)$ modified prolate spheroidal functions, as described in the last chapter, where T is the length of the wavelet and a and b are its lower and upper frequency limits.

2. The Prolate Spheroidal Wavelet

Estimates of wavelets are very simple to compute using prolate spheroidal functions. It is necessary to find the expansion coefficients of the wavelet estimate $w_p(t)$ such that

$$w_p(t) = \sum_{k=0}^{K-1} \alpha_k \phi_k(c_a, c_b; t), \quad (4.3)$$

where $c_a = \pi Ta$, $c_b = \pi Tb$. Calculation of the α_k 's is relatively easy: a synthetic seismogram produced by convolving $r_{log}(t)$ with $w_p(t)$ is

$$\begin{aligned} x_p &= \int_{\tau_1}^{\tau_2} w_p(\tau) r_{log}(t - \tau) d\tau \\ &= \int_{\tau_1}^{\tau_2} \sum_{k=0}^{K-1} \alpha_k \phi_k(c_a, c_b; \tau) r_{log}(t - \tau) d\tau \\ &= \sum_{k=0}^{K-1} \alpha_k \int_{\tau_1}^{\tau_2} \phi_k(c_a, c_b; \tau) r_{log}(t - \tau) d\tau \\ &= \sum_{k=0}^{K-1} \alpha_k \zeta_k(t). \end{aligned}$$

$\zeta_k(t)$ is the convolution of the k^{th} -order prolate spheroidal function with $r_{log}(t)$. The α_k 's are found by minimizing

$$\epsilon^2 = \int_t [x_0(t) - x_p(t)]^2 dt = \int_t [x_0(t) - \sum_{k=0}^{K-1} \alpha_k \zeta_k(t)]^2 dt$$

with respect to α_k . It is easy to show that this minimization yields

$$\sum_{k=0}^{K-1} \alpha_k \int_t \zeta_k(t) \zeta_j(t) dt = \int_t x_0(t) \zeta_j(t) dt$$

or, more succinctly,

$$\sum_{k=0}^{K-1} \alpha_k Z_{kj} = g_j.$$

The expansion coefficients for the wavelet, therefore, may be found by solving a $K \times K$ symmetric matrix equation with K usually less than 15. If nothing else, this is a computationally efficient way of finding an estimate of the wavelet, and the result has some desirable properties: while $w_p(t)$ may not be the 'smoothest' wavelet estimate, it is composed of a relatively small number of smooth basis functions, and so is expected to be smooth. Also, by choosing the lower frequency limit $a > 0$, the zero-area constraint will automatically be met, and the final wavelet is guaranteed to exist in the frequency band of the true wavelet. Finally, $x_p(t)$ may not fit the data exactly, and the difference between the recorded and synthetic seismograms may contain information about the difference between $r_{log}(t)$ and $r_0(t)$. This is illustrated by the following examples.

3. Example 1 - A Land Wavelet

In this example, a wavelet with three extrema was convolved with a reflectivity sequence consisting of 10 delta functions, and the result of this convolution is plotted in Figure 4.1(a). Noise with standard deviation $\sigma_s = 0.1$ was added to this seismogram, and the noisy seismogram is plotted in Figure 4.1(b). The true reflectivity used is plotted in Fig. 4.1(c), and for an estimate of the true reflectivity, some of the spikes were shifted from their true positions. The negative peak at $t = 13$ was moved to $t = 10$, and the peaks at $t = 90$ and $t = 120$ were moved to $t = 94$ and $t = 125$, respectively. White noise with standard deviation $\sigma_r = 0.1$ was then added to the modified reflectivity (Fig. 4.1(d)), and an attempt to recover the wavelet from this modified reflectivity and the noisy seismogram was made by implementing the procedure described in the previous section. The length of the wavelet ($t = 25$) was assumed known (in practice, this can be estimated by finding the lag where the autocorrelation function of the trace decays to noise level), and a periodogram of the noisy seismogram was taken to find the frequency band of the wavelet. The values for the lower and upper limits in the band were $a = .01$, $b = .11$ (Note that the Nyquist is .5)

The difference between the 'observed' seismogram $x_0(t)$ and the synthetic seismogram $x_p(t)$ obtained by convolving the estimated wavelet with the modified reflectivity is shown in Figure 4.2(a). Note that $x_p(t)$ does not match $x_0(t)$, and that the largest differences occur immediately after the position of the spikes which had been shifted. This misfit may be used as a criterion to determine whether the approximate reflectivity $r_{log}(t)$ is a good estimate of $r_0(t)$ or not. The true and estimated wavelets are shown in Fig. 4.2(b).

Using the information in 4.2(a), it is possible that the approximate reflectivity could be perturbed until a best fit to the data was found. By shifting the peaks back to their true position, and leaving the wrong amplitudes and the noise in the reflectivity, the misfit between synthetic and 'observed' seismograms is reduced considerably (Fig. 4.3(a)). The wavelet estimate from this second, better approximation to the true reflectivity is shown with the true wavelet in Figure 4.3(b).

4. Example 2 - A Marine Wavelet

The same reflectivity as in Example 1 was used, but this time a marine-like wavelet with 8 extrema was convolved with it. Three spikes of the true reflectivity were moved from their true positions, and noise was added to the seismogram and the modified reflectivity, as in Example 1. Figures 4.4, 4.5, and 4.6 are analogous to 4.1, 4.2, and 4.3. Note that the misfit in 4.5(a) is of roughly the same proportions as the seismogram itself, which implies that the choice of basis functions for the expansion constrains the space of possible solutions in such a manner that a wavelet that would fit the data when convolved with the modified reflectivity cannot be found. If the spikes are shifted back to their true positions, the $x_p(t)$ calculated will fit the data much better (Fig. 4.6(a)), and quite a good wavelet estimate is obtained (Fig. 4.6(b)).

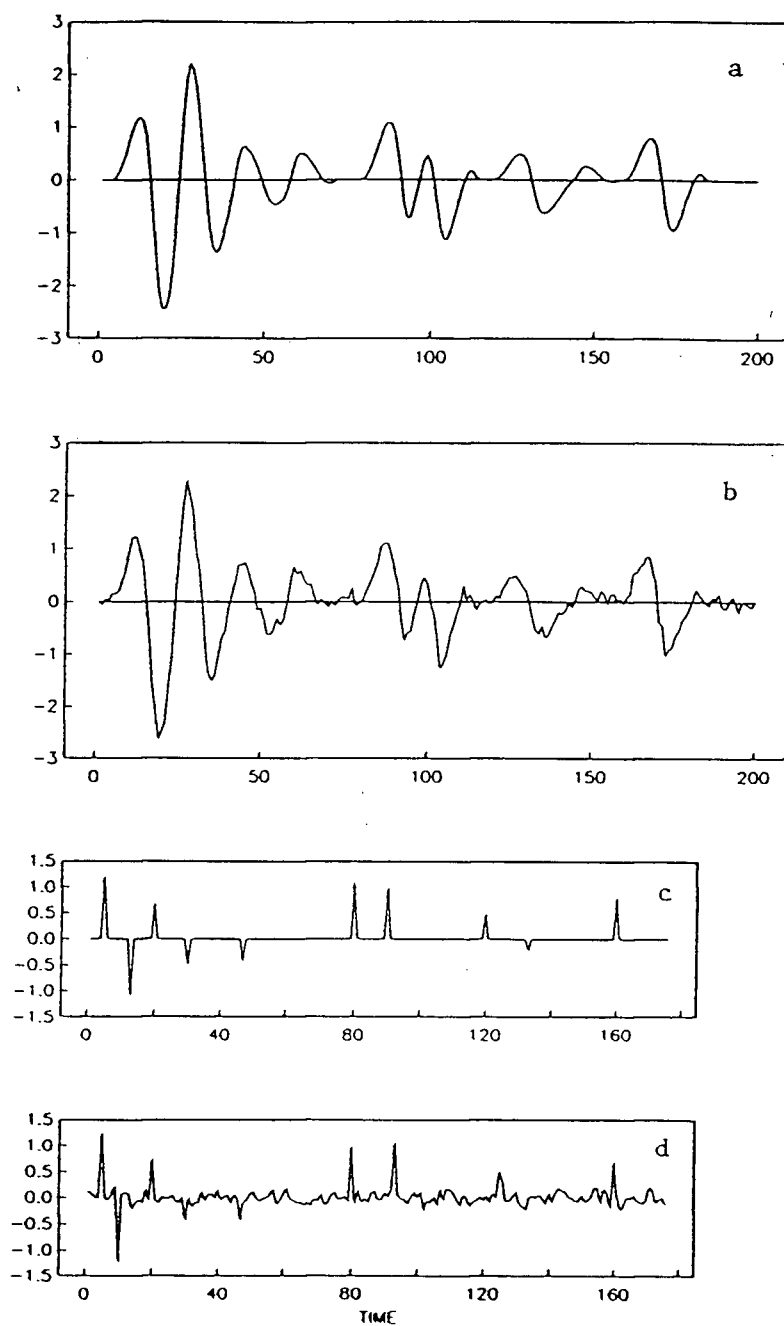


Figure 4.1- (a) The true seismogram, Example 1. (b) The noisy seismogram. (c) The true reflectivity. (d) The modified reflectivity used for the wavelet estimation scheme.

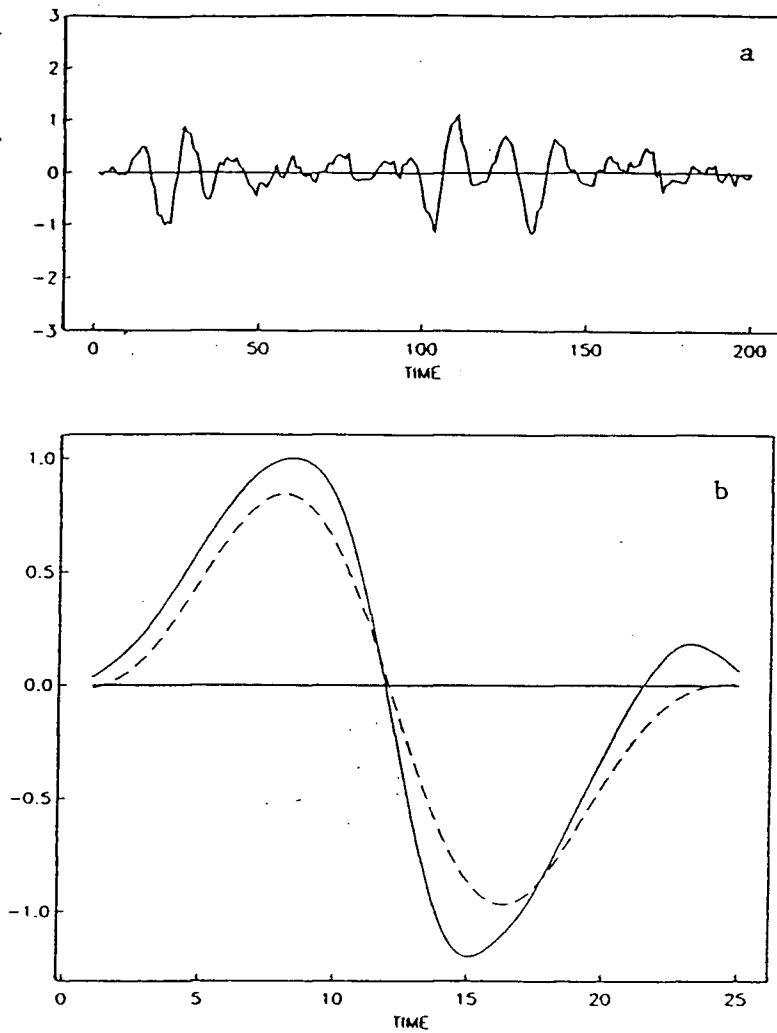


Figure 4.2- (a) The difference between the 'observed' and the synthetic seismogram from the calculated wavelet and the modified reflectivity. (b) The true (solid) and estimated (dashed) wavelets.

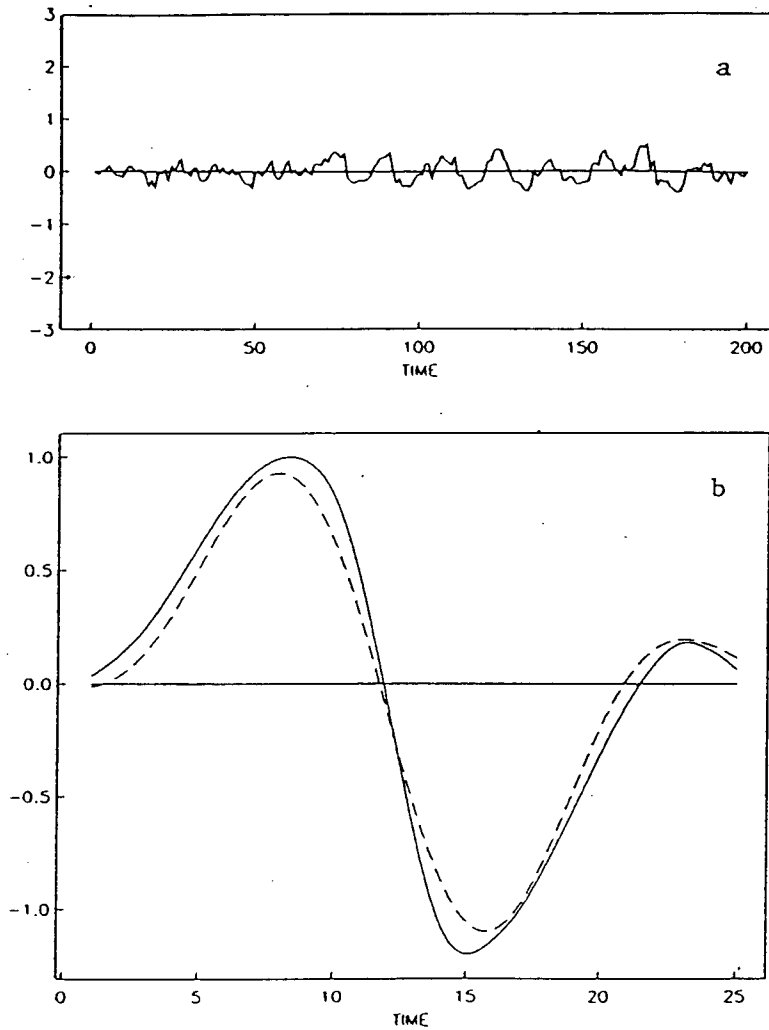


Figure 4.3- (a) The difference between the 'observed' and the synthetic seismogram obtained from the calculated wavelet and the improved estimate of the reflectivity. (b) The true (solid) and estimated (dashed) wavelets.

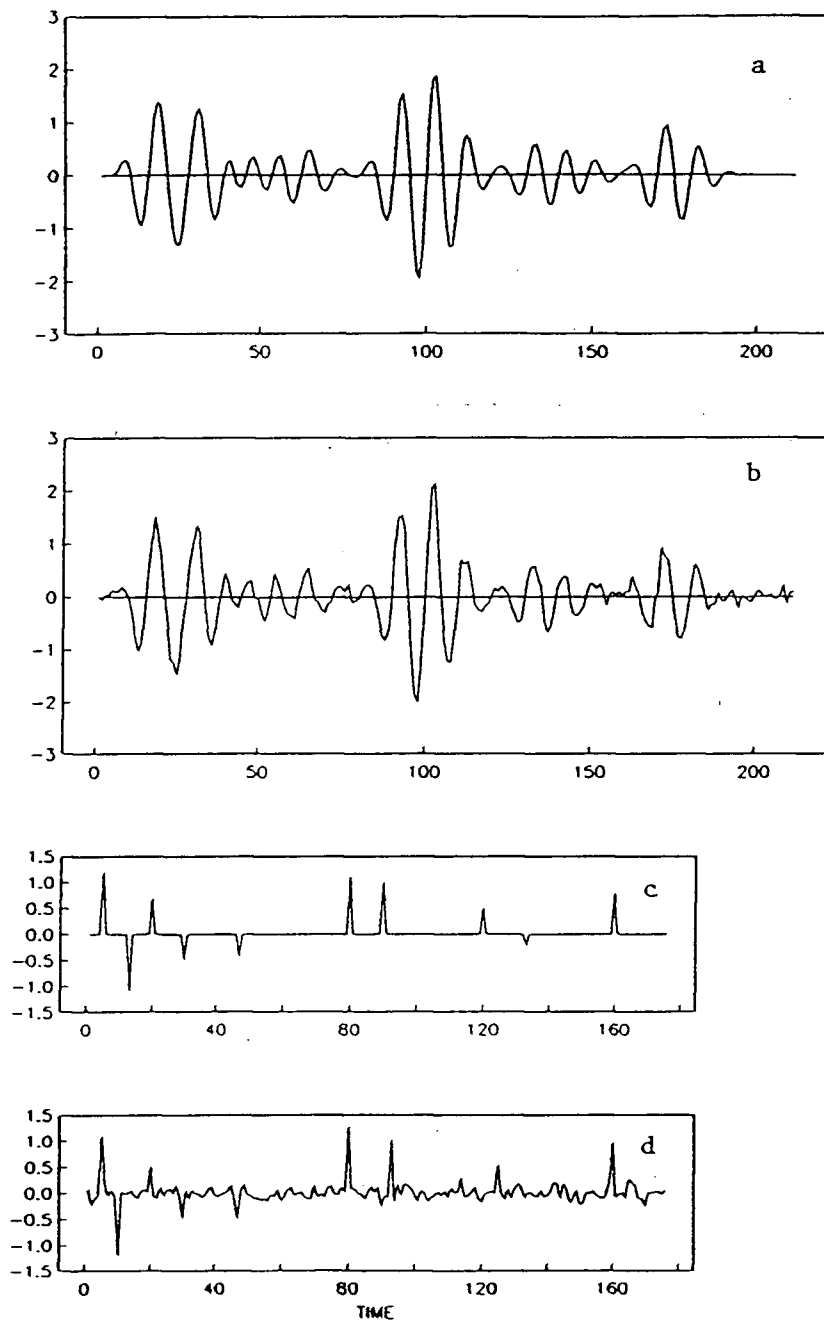


Figure 4.4- (a) The true seismogram from the 'marine' wavelet. (b) The noisy seismogram. (c) The true reflectivity. (d) The modified reflectivity used for wavelet estimation.

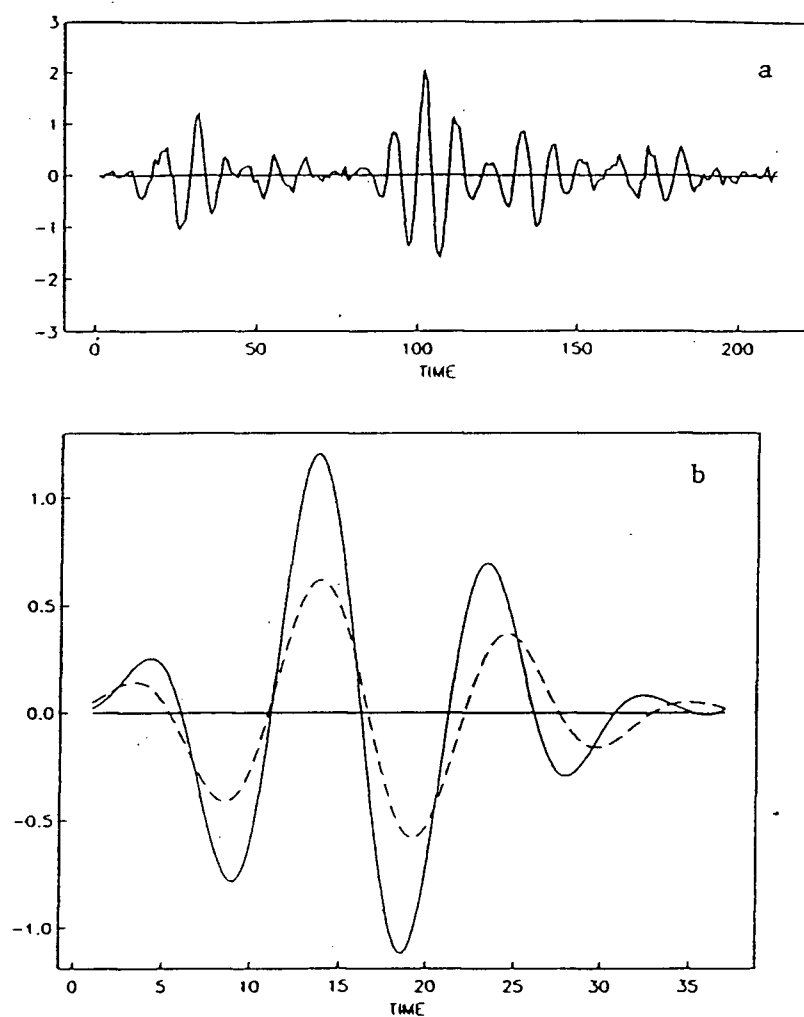


Figure 4.5- (a) The difference between the 'observed' and sythetic seismo-gram from initial reflectivity and wavelet estimates. (b) The true (solid) and estimated (dashed) wavelets.

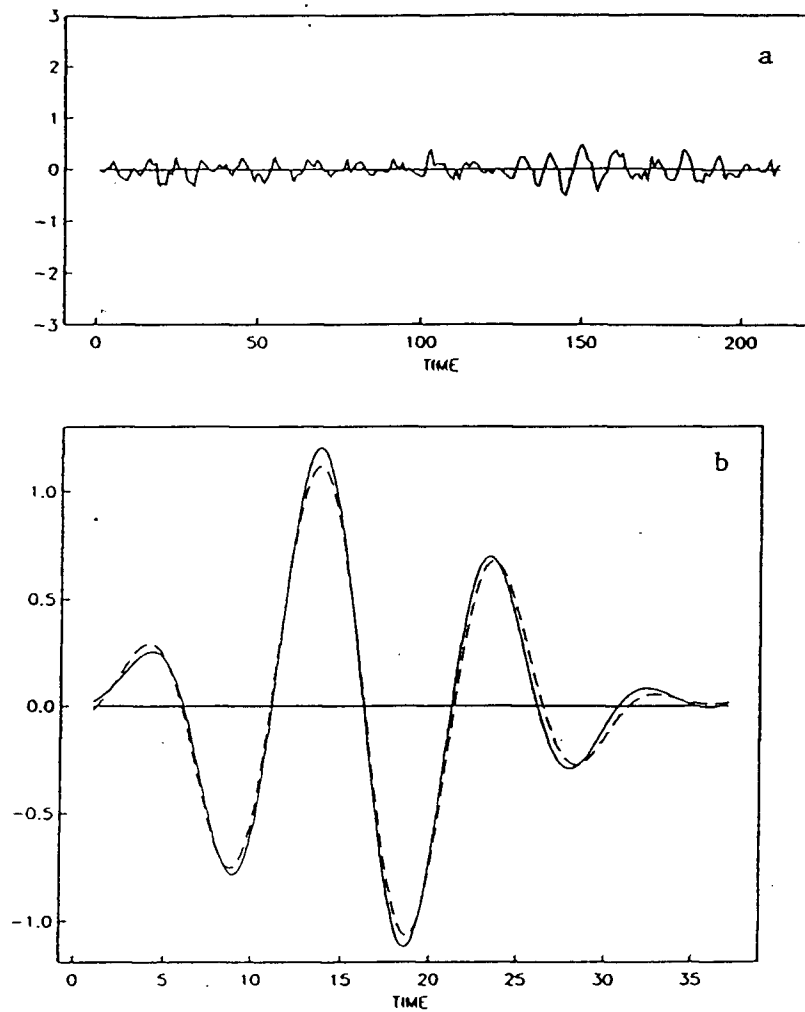


Figure 4.6- As in Figure 4.5, but from a noisy reflectivity estimate with the peaks in the correct positions.

5. Further Suggestions and Conclusions

Of course, the problem remains of *how* to perturb $r_{log}(t)$ to reduce the misfit between $x_p(t)$ and $x_0(t)$. This problem is too broad to be discussed in depth here. One possibility would be to divide the seismogram into different segments, and to try to minimize the local misfit by shifting the spike positions, then the amplitudes. When this misfit is minimized, one could put the segments back together and extract a wavelet and reflectivity from the entire trace. Another possibility might be to use the estimated wavelet to obtain a new reflectivity estimate by spectral division, compute a new estimate of the wavelet, and thereby iterate toward a solution where the misfit is minimized. The reason this can be done is because the choice of basis functions does not guarantee that the estimated wavelet convolved with the approximate reflectivity will fit the data. However, any procedure that depends on an initial guess may converge to the wrong solution or not converge at all.

In any case, the method for wavelet estimation presented here is computationally very efficient, most of the computational expense being in the calculation of the basis functions. As seen in Appendix B, even this computation is simple for wavelets of time-bandwidth product less than 7.5. If there is some previous information about the underlying geology, this presents a viable alternative to other methods of deconvolution.

*...no hay nada mejor
que un monte, un rancho, y un lucero,
cuando se tiene un "te quiero"
y huele a sendas en flor...*

—Alfredo Espino, ‘ Un Rancho y Un Lucero’

CONCLUSION

Prolate spheroidal functions have special properties that make them very useful for processing geophysical signals. Perhaps the most important of these properties, and the one exploited in this study, is that the PSF's are timelimited and highly concentrated in the frequency domain. Of the three problems investigated here, the PSF's present attractive solutions in two.

Since data tapers must be highly concentrated in frequency, a weighted average of spectra obtained by tapering the data with different low-order prolates can yield a useful nonparametric method of spectrum estimation for data whose spectra are known to be slowly varying. If the spectra vary rapidly, the large time-bandwidth products used for the data tapers will smooth over details in the spectrum, so that the APS spectrum is not as useful in these cases. The multiple taper technique can also be used to test for the presence of sinusoids, and this may be most important in attempting to find sinusoidal components in noisy data. Furthermore, this harmonic analysis technique can be applied to nonstationary signals by subdividing the time series into a collection of short, overlapping segments. The most encouraging results obtained for nonstationary data were observed when looking for exponentially decaying sinusoids, where both the

frequencies and the damping constants were unknown. However, compared to other techniques of spectrum estimation, the prolate spheroidal method suffers from a great computational burden, and so should only be used when very accurate frequency domain information is desired.

When working with signals whose time-bandwidth product is less than 7.5, the material presented in Chapter III for SNR enhancement may prove very useful. By separating the data into components which are frequency-limited in the signal band and components which are not it is possible to distinguish between signal and noise even inside the signal band, something that cannot be done with conventional Fourier expansions. This procedure may still be applied to signals with large time-bandwidth products, but because of computational difficulties one would have to deal with some rather complicated asymptotic formulae [*Slepian* (1978), *Slepian* (1965)], so the computational simplicity of the procedure would disappear.

The last application examined in this thesis was wavelet estimation. This particular method involves expanding the wavelet by PSF's, where the expansion coefficients are found using an estimate of the true reflectivity. Computation of the wavelet estimate is efficient, and the main computational limitation (inversion of a matrix) remains the same regardless of the amount of data used. The most interesting feature, though, is that the estimated wavelet convolved with the estimated reflectivity will not fit the data if the reflectivity estimate is different from the true one. This may be used as a criterion to judge how close the estimated reflectivity is to the true one, and it could form part of a general iterative deconvolution process. The details of this iteration must still be worked out, and since this would involve a major project in itself, only the very basics were studied here.

All things considered, the PSF's are an important set of functions for processing geophysical data, and they should receive more attention in the future. This thesis demonstrates the versatility of these special functions, with hope that some of the ideas presented here will become useful. *Vale.*

REFERENCES

- Abramowitz, M., and Stegun, I.A., *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.
- Akaike, H., 'Statistical Predictor Identification', *Ann. Inst. Statist. Math.* **22**, 1970, 203-217.
- Bartlett, M.S., 'Periodogram Analyses and Continuous Spectra', *Biometrika* **37**, 1950, 1-16.
- Beyer, W.H. (editor), *CRC Standard Mathematical Tables*, CRC Press Inc., Boca Raton, Florida, 1984.
- Blackman, R.B., and Tukey, J.W., *The Measurement of Power Spectra*, Dover Publications, New York, 1958.
- Cervantes Saavedra, Miguel de, *El Ingenioso Hidalgo Don Quijote de la Mancha*, Editorial Castalia, Madrid, 1978.
- Darío, Rubén, 'Canción de Otoño en Primavera', *Obras Completas (Poesías)*, Editorial Anaconda, Buenos Aires, 1958.
- Espino, Alfredo, 'Un Rancho y Un Lucero', *Jícaras Tristes*, UCA/Editores, San Salvador, El Salvador, 1981.
- Flammer, C., *Spheroidal Wave Functions*, Stanford University Press, Stanford, Calif., 1957.
- García Márquez, Gabriel, *Cien Años de Soledad*, Espasa-Calpe, Madrid, 1982.
- Garland, G.D., *Introduction to Geophysics*, W.B. Saunders Co., Toronto, 1982.
- Griffiths, L.J., and Prieto-Diaz, R., 'Spectral Analysis of Natural Seismic Events Using Autoregressive Techiques', *IEEE Trans. G.E.* **15**, 1977, 13-25.
- Hermann, R.B., 'Computer Programs in Earthquake Seismology', Dept. of Earth and Atmospheric Sciences, Saint Louis University, 1978.
- Hermann, R.B., 'Some Aspects of Bandpass Filtering of Surface Waves', *Bull. Seism. Soc. Am.* **63**, 663-671.
- Hoel, P.G., *Introduction to Mathematical Statistics*, J. Wiley and Sons, Toronto, 1984.
- Jenkins, G.M., and Watts, D.G., *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 1968.
- Kanasewich, E.R., *Time Sequence Analysis in Geophysics*, University of Alberta Press, Edmonton, Alberta, 1973.

- Kaplan, W., *Advanced Mathematics for Engineers*, Addison-Wesley, Reading, Mass., 1981.
- Landau, H.J., and Pollak, H.O., 'Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-II', *Bell Syst. Tech. J.* **40**, 1961, 65-84.
- Landau, H.J., and Pollak, H.O., 'Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-III', *Bell Syst. Tech. J.* **41**, 1962, 1295-1336.
- Lindberg, C.R., *Multiple Taper Spectral Analysis of Terrestrial Free Oscillations*, Ph.D. Thesis, University of California, San Diego, 1986.
- McMechan, George A., and Ottolini, Richard, 'Direct Observation of a $p - \tau$ Curve in a Slant Stacked Wave Field', *Bull. Seis. Soc. Am.* **70**, No. 3, 1980, 775-789.
- Meixner, J., and Schafke, F.W., *Mathieusche Funktionen und Sphäroidfunktionen*, Julius Springer, Berlin, 1954.
- Miller, K.S., 'Complex Linear Least Squares', *SIAM Review* **15**, No. 4, 1973.
- Morse, P.M., and Feshbach, H., *Methods in Theoretical Physics*, McGraw-Hill, New York, 1953.
- Neruda, Pablo, 'Unidad', *Obras Completas*, Losada, Buenos Aires, 1962.
- Oldenburg, D.W., Levy, S., and Whittall, K.P., 'Wavelet Estimation and Deconvolution', *Geophysics* **46**, No. 11, 1981, 1528-1542.
- Papoulis, A., *Signal Analysis*, McGraw-Hill, New York, 1977.
- Parzen, E., 'Mathematical Considerations in the Estimation of Spectra', *Techometrics* **3**, 1961, 157-189.
- Priestley, M.B., *Spectral Analysis and Time Series*, Academic Press, London, 1981.
- Robison, E.A., and Treitel, S., *Geophysical Signal Analysis*, Prentice-Hall, New Jersey, 1980.
- Salarrué, 'La Botija', *Cuentos de Barro*, Ministerio de Educación (Publicaciones), San Salvador, El Salvador, 1981.
- Schultz, Philip S., and Claerbout, Jon F., 'Velocity Estimation and Downward Continuation by Wavefront Synthesis', *Geophysics* **43**, No. 4, 1978, 691-714.
- Sheriff, R.F., and Geldart, L.P., *Exploration Seismology*, Cambridge University Press, Cambridge, 1981.
- Slepian, D., 'Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V', *Bell Syst. Tech. J.* **55**, 1978, 1371-1429.
- Slepian, D., 'Some Asymptotic Expansions for Prolate Spheroidal Wave Functions', *J. Math. and Phys.* **44**, No. 2, 1965, 99-140.
- Slepian, D., 'Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-IV', *Bell Syst. Tech. J.* **43**, 1964, 3009-3057.
- Slepian, D., and Pollak, H.O., 'Prolate Spheroidal Functions, Fourier Analysis, and Uncertainty-I', *Bell Syst. Tech. J.* **40**, 1961, 43-64.

- Stratton, J.A., Morse, P.M., Chu, L.J., Little, J.D.C., and Corbató, F.J., *Spheroidal Wave Functions*, The Technology Press of M.I.T., Cambridge, Mass., and John Wiley and Sons, New York, 1956.
- Thomson, D.J., 'Spectrum Estimation and Harmonic Analysis', *IEEE Proceedings* **70**, No. 9, 1982, 1055-1096.
- Ulrych, T.J., 'Maximum Entropy Power Spectrum of Truncated Sinusoids', *J. Geophys. Res.* **77**, 1972, 1396-1400.
- Ulrych, T.J., and Bishop, T.N., 'Maximum Entropy Spectral Analysis and Autoregressive Decomposition', *Rev. Geophys. and Space Physics* **13**, 1975, 183-200.

APPENDIX A

The Fourier Transform of a Discretized ‘Boxcar’

A discretized ‘boxcar’ is

$$x_n = \begin{cases} 1, & \text{if } |n| \leq N; \\ 0, & \text{otherwise} \end{cases}$$

and its Fourier transform is given by

$$X(f) = \sum_{n=-N}^N e^{-i2\pi f n}$$

when $\Delta t = 1.0$. If we let $a = e^{i2\pi f N}$, then $X(f)$ becomes

$$X(f) = a \sum_{m=0}^M e^{-i2\pi f m}, \quad M = 2N + 1.$$

Recognizing that the summation is a geometric series with common ratio $e^{-i2\pi f}$, and that the sum of such a series with ratio z ($|z| < 1$) is

$$S_M = \sum_{m=0}^M z^m = \left(\frac{1 - z^M}{1 - z} \right),$$

we have

$$\begin{aligned} X(f) &= a \left(\frac{1 - e^{-i2\pi f M}}{1 - e^{-i2\pi f}} \right) \\ &= a \frac{(e^{i2\pi f M/2} - e^{-i2\pi f M/2})e^{-i2\pi f M/2}}{(e^{i\pi f} - e^{-i\pi f})e^{-i\pi f}} \end{aligned}$$

and since

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i},$$

$X(f)$ becomes

$$\begin{aligned} X(f) &= a \frac{\sin 2\pi f M/2}{\sin \pi f} (e^{-i2\pi f (M-1)/2}) \\ &= a \frac{\sin 2\pi f (N + \frac{1}{2})}{\sin \pi f} (e^{-i2\pi f N}) \\ &= \frac{\sin 2\pi f (N + \frac{1}{2})}{\sin \pi f}. \end{aligned}$$

This last expression is known as a Dirichlet kernel. Note that it has characteristics of a ‘sinc’ function at low frequencies, but also accounts for the aliasing that is present when digitized data are considered.

APPENDIX B

Computation of the DPSS's

When L is small, it is easy to compute the DPSS's by finding the eigenfunctions of the Toeplitz matrix, however, when L becomes large, this is inconvenient. To find a better procedure, we find the eigenfunctions in the discrete case and interpolate to the number of points desired. The integral equation that defines the continuous-time prolate spheroidal functions is

$$\lambda_k(c)\phi_k(c;t) = \int_{-1}^1 \phi_k(c;t') \frac{\sin c(t' - t)}{\pi(t' - t)} dt', \quad (B1)$$

and can be reduced to a system of linear equations by means of Gauss-Legendre quadrature [see *Abramowitz and Stegun* (1965), *Beyer* (1984)], so by using J^{th} -order quadrature, (B1) is reduced to a system of linear equations of the form

$$\lambda_k(c)\phi_k(c;t) \sim \sum_{j=1}^J w_j \phi_k(c;t) \frac{\sin c(t_j - t)}{\pi(t_j - t)}, \quad (B2)$$

where t_j and w_j are the j^{th} abscissa and weight, respectively, of the Gauss-Legendre quadrature formula. Now, defining

$$\Psi_{kj} = \sqrt{w_j} \phi_k(c, t_j)$$

and

$$A_{mj} = \sqrt{w_m w_j} \frac{\sin c(t_m - t_j)}{\pi(t_m t_j)},$$

the modified eigenfunctions Ψ_{kj} can be found by finding the eigenvalues and eigenfunctions of a $J \times J$ Toeplitz matrix:

$$\lambda_k(c) \Psi_{km} = \sum_{j=1}^J A_{mj} \Psi_{kj}.$$

Now the DPSS's may be interpolated to the desired number of points by using the 'sinc' function that composes the kernel.

This interpolation does not require that the DPSS's have an odd number of points, as equation (1.19) does, but in fact the DPSS's can be computed directly from a Toeplitz matrix even for even numbers of points. An equivalent equation to (1.19) is

$$\lambda_k(LW) \phi_n^k(LW) = \sum_{m=1}^L \frac{\sin 2\pi W(m-n)}{\pi(m-n)} \phi_m^k(LW), \quad (B3)$$

because shifting the indices only introduces a phase shift, but the amplitude spectrum of the functions remains the same. In equation (B3), L may be taken to be an even number, and for this reason (B3) was used to compute the DPSS's throughout the entire study.

The computational limitations of finding the DPSS's through any form of matrix equation can be disastrous. For any time-bandwidth product, the eigenvalues corresponding to the higher-order eigenvectors are very small, and when several fall below machine precision (10^{-16} in the UBC computer), the associated eigenvectors cannot be computed using any of the methods outlined here. However, usually only the lower-order functions are needed, and therefore this does not represent a serious difficulty. The most serious problem comes when the large eigenvalues begin to get very close to unity. For large time-bandwidth products, and small values of k , $\epsilon_k(c) = 1 - \lambda_k(c)$ falls below machine precision, and that also causes computational problems, since the computer cannot resolve the difference between the few largest eigenvalues. For large time-bandwidth products ($TW > 7.5$), therefore, the low-order functions cannot be computed through an eigenvalue-eigenvector problem, and it is these that are of crucial importance in the consideration of bandlimited functions. Other ways of computing these lower-order functions, perhaps through asymptotic formulae [*Slepian* (1978), *Slepian* (1965)], must be studied.

APPENDIX C

The Cramér Spectral Representation

Consider a stationary, zero-mean, stochastic process $x(t)$ for $|t| \leq T/2$, and its periodic extension $x_p(t)$,

$$x_p(t) = x(t) \quad \text{for } |t| \leq T/2$$

$$x_p(t + nT) = x(t), \quad n = 0, \pm 1, \pm 2, \dots$$

The Fourier integral

$$X_p(f) = \int_{-T/2}^{T/2} x_p(t) e^{-i2\pi ft} dt \tag{C1}$$

does not converge to a finite number as $T \rightarrow \infty$, because

$$\int_{-\infty}^{\infty} |x(t)| dt$$

does not converge for a stationary process [Priestley (1981), sec. 4.5].

To look for an alternative spectral representation, we examine the function

$$Z_p(f) = \int_{-\infty}^f X_p(f') df'. \quad (C2)$$

Notice that

$$\Delta Z_p(f_n) = \Delta Z_p(f_{n+1}) - \Delta Z_p(f_n) \simeq X_p(f_n) \Delta f_n \quad (C3)$$

remains well-behaved as $T \rightarrow \infty$, because as $|X_p(f_n)| \rightarrow \infty$, $\Delta f_n \rightarrow 0$, and so $\Delta Z_p(f_n)$ converges to a finite number for $T \rightarrow \infty$.

This suggests representing $x_p(t)$ by a Fourier-like series

$$x_p(t) = \sum_{n=-\infty}^{\infty} e^{i2\pi f_n t} \Delta Z_p(f_n),$$

and as $T \rightarrow \infty$ this converges to a Stieltjes integral:

$$x(t) = \int_{-\infty}^{\infty} e^{i2\pi f t} dZ(f). \quad (C4)$$

Equation (C4) is called the Cramér spectral representation of $x(t)$. If

$$E[x(t)] = 0,$$

then for (C4) to hold, it must be that

$$E[dZ(f)] = 0. \quad (C5)$$

Also, from (C3) and the definition of spectral density function $S(f)$,

$$S(f) = \lim_{T \rightarrow \infty} E \left[\frac{|X_p(f)|^2}{T} \right],$$

then

$$E \left[|\Delta Z_p(f_n)|^2 \right] \simeq E \left[|X_p(f_n) \Delta f_n|^2 \right] = E \left[\frac{|X_p(f_n)|^2}{T} \right] \Delta f_n.$$

Letting $T \rightarrow \infty$,

$$E \left[|dZ(f)|^2 \right] = S(f) df. \tag{C6}$$

For further discussion of generalized integration, see *Priestley* (1981).