LINEAR PREDICTIVE SPECTRAL ANALYSIS VIA THE L_p NORM

By

JAMES EDWIN SCHROEDER

Bachelor of Science in Electrical Engineering
University of Iowa
Iowa City, Iowa
1976

Master of Science in Electrical Engineering
University of Iowa
Iowa City, Iowa
1978

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December 1985

Thesis
1985
S381l
cop.2

OKLAHOMA STATE UNIVERSITY LIBRARY

LINEAR PREDICTIVE SPECTRAL

ANALYSIS VIA THE

L_p NORM

Thesis Approved:

Thesis Advisor

Dean of the Graduate College

1248647

PREFACE

This study involves linear predictive spectral analysis under the general L_p norm; both one dimensional and two dimensional spectral estimation algorithms are developed. The objective in this study is determination of frequency resolution capability for various L_p normed solutions to linear predictive spectral estimation equations. A modified residual steepest descent algorithm is utilized to generate the required solution. The research presented in this thesis could not have been accomplished without the support of the Oklahoma State University Research Consortium For Well Log Data Enhancement Via Signal Processing. The member companies of this consortium include Amococ Production Company, Arco Oil and Gas Company, Cities Service Oil and Gas Corporation, Conoco, Exxon, IBM, Mobil Research and Development, Phillips Petroleum Corporation, Sohio Petroleum Company, and Texaco.

The author wishes to express his appreciation to his major adviser, Dr. Rao Yarlagadda, for his guidance and assistance during the course of this research. Appreciation is also expressed to other committee members, Dr. Zuhair Al-Shaieb, Dr. Randy Reininger, and Dr. David Soldan for their constructive criticism during this study.

Special thanks are given to Ms. Barbara Caldwell and Mrs. Kelly Whitfield for their tireless assistance in the preparation of this manuscript. Finally, and most importantly, this study could not been completed without the encouragement and patience of Sarah Lowe-Schroeder, the very real motivation behind this entire project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

Motivation

Spectral estimation is an important research area for geophysical signal processing, especially as applied to seismic data; other possible geophysics applications include the full wave acoustic log. A frequency - wavenumber (F-K) spectral display may aid in identifying and interpreting the various propagating waves contained within sonic log data; this technique may be especially effective if an array of data is available from the logging tool. For seismic applications the F-K plane offers a convenient means of separating signal and noise components; two dimensional (2-D) filtering can be applied to remove undesireable noise and/or interfering signals. Both seismic data and full wave acoustic data generally possess a limited spatial aperture due to physical constraints or design criteria; the sonic logging tool, for example, is designed with a small spatial aperture to achieve sufficient vertical resolution.

It is known that Fourier spectral estimation techniques exhibit poor resolution on limited record length data, since resolution is proportional to the observation interval (a form of Heisenberg's uncertainty principle). The past 15 years have seen the emergence of powerful new spectral estimation algorithms designed to improve resolution over that available from the Fourier transform, especially

when using short data records. These alternative methods (termed parametric or model based) include maximum entropy spectral estimation (MESE), maximum likelihood spectral estimation (MLSE), and auto-regressive or linear predictive techniques (AR or LP). In one dimension LP and MESE are equivalent, however this is not true for 2-D spectral estimation; indeed MESE becomes very complicated and involves solving a set of non-linear equations (Lang and McClellan, 1982). LP extends naturally to the 2-D case however and provides a convenient framework for developing spectral estimation algorithms that take advantage of the two dimensional structure of array data.

2-D LP algorithms have emerged the past several years based upon least squares (LS) solutions. This is natural since least squares solutions are easy to generate and analytically tractable. Many problems, however, are not amenable to a least squares solution; for example, if the data are contaminated by impulsive noise a LS approach will weight the impulses equally with the good data and produce poor results. What is required in this case is a L_1 solution (in general L_p, $1 \leq p \leq 2$) which will on the average result in a better estimator since the "outliers" will be ignored rather than be given equal weight. Although general L_p solution methods have computational and analytical disadvantages, several rapidly converging L_p algorithms have been developed recently and research directed at 2-D spectral estimation utilizing the L_p norm is justified.

Two dimensional linear predictive spectral estimation via the L_p norm is currently being researched. For the impulsive noise case it has been shown that an L_p ($p = 1$) solution to a set of linear prediction equations may offer increased frequency resolution over that obtainable

via L_2 techniques. An improved formulation of the residual steepest descent algorithm (RSD) was used to generate the required L_1 solution.

Other research included within this thesis includes a comparison of the L_p (p = 1) 1-D spectral estimator against a variety of noise sources. The noise types under investigation are: Gaussian, Rayleigh, uniform, and impulsive, Laplacian, and Cauchy. Another section of this thesis will consider the application of the proposed 2-D spectral estimation algorithm to synthetic acoustic well log data. As this data type possesses a sufficiently long time record, discrete Fourier transform (DFT) techniques will be blended with the LP (L_p, p = 1, 2 norm) method in order to simplify computations; namely, a DFT is applied in the time dimension and LP analysis is applied in the spatial dimension.

Overview

Chapter II of this report will review the field of spectral estimation and highlight a number of applications where spectral estimation has found widespread use. Emphasis will be placed on applications of AR or LP techniques as that is the focus of this report. Applications of spectral estimation to geophysics will also be covered in a section of Chapter II. A review of linear prediction, especially applied to spectral estimation, will be given in Chapter III. Chapter IV presents the results of applying an L_1 normed solution to the 1-D LP spectral estimation problem. Results of an L_1 normed solution for a 2-D LP spectral estimator are given in Chapter V. Additionally, in Chapter V, DFT techniques are blended with LP Methods, in order to reduce the computation burden inherent in 2-D spectral estimation. Finally, in

Chapter VI, the conclusions reached in this research are summarized, and future areas of research are indicated.

Four appendices are included in this thesis. In Appendix A the sinusoidal frequency locations are calculated that result in a singular linear prediction matrix. The separability of a discrete Fourier transform and linear prediction techniques as applied to two dimensional data is shown in Appendix B. Appendix C presents a simple example in which an L_p normed solution to a matrix equation is calculated; a few interesting characteristics of non L_2 normed solutions are demonstrated. Finally, in Appendix D, an alternate least squares spectral representation of time series data is developed; namely, the Fourier-Bessel series expansion.

CHAPTER II

SPECTRAL ESTIMATION SURVEY

Spectral Estimation Overview

The history of Fourier analysis dates back to the eighteenth century when Gauss made use of the trigonometric Fourier series in his orbital mechanics work. Although the Fast Fourier Transform (FFT) is widely attributed to Cooley and Tukey (1965), apparently similar formulas were in use as long ago as 1754. A detailed account of the historical development of the FFT can be found in Heideman, et al (1984). Since 1965, with widespread dissemination of the Cooley-Tukey FFT algorithm, Fourier analysis has found application in numerous engineering and scientific fields. In Robinson (1982), a historical perspective of spectrum estimation is available covering the time frame 600 B.C. to the present!

Possibly, Schuster (1898) originated the field of spectral analysis with his attempt to fit sunspot data to a Fourier series in order to detect periodicities within the data. Wiener (1930) utilized the Fourier transform to study the harmonic properties of stochastic processes, allowing a spectral interpretation of random data. Khinchine (1934), as well, related the autocorrelation function of a random process to the power spectral density, apparently independently of Wiener. Although Wiener and Khinchine developed the theory necessary to analyze

random processes (or deterministic data) via the Fourier transform, workable techniques were not developed until 1958. Blackman and Tukey (1958) developed a spectral estimation method for discrete data based upon the autocorrelation function. In their method, a data window is first applied to the estimated auto correlation function (i.e. computed from the raw data) and a Fourier transform applied to the windowed autocorrelation sequence in order to estimate the spectral density. With rediscovery of the FFT algorithm by Cooley and Tukey (1965), spectral estimates could more efficiently be computed via the periodogram method originally proposed by Schuster (1898). The FFT can be directly applied to the data set (or windowed data if desired) and the spectral estimate is given by the resultant magnitude squared. Much has been written about the FFT based periodogram method of spectral estimation and its various modifications; see, for example, Bingham et al (1967), Brigham and Morrow (1967), Brigham (1974), Jenkins and Watts (1968), Welch (1967), and Welch (1977) to name just a few.

Although the FFT based periodogram approach is computationally efficient and produces sufficiently accurate spectral estimates in many cases, difficulties arise that may preclude use of this technique when insufficient data samples are available. A fundamental limitation of FFT based techniques is frequency resolution. As is well known, frequency resolution is proportional to data length (a form of Heisenberg's uncertainty principle); therefore, poor frequency resolution is an unavoidable byproduct of limited record length data. Since the data is necessarily of finite duration, another limitation of the FFT method, termed "leakage", arises. Spectral leakage refers to the phenomena of spectral energy in one spectral band spilling over (leaking) into other

spectral bands; as a result, spectral energy may be indicated where none exists. The leakage problem is unavoidable with finite length data records (Brigham, 1974), although the problem may be alleviated to some extent via windowing techniques. Windowing (tapering) the data, although reducing leakage, reduces the available frequency resolution of the "main lobe." Unfortunately, short data records result in greater leakage problems, due to increased sidelobe amplitudes, further complicating the problem of decreased frequency resolution previously mentioned.

In order to overcome the intrinsic limitations of the FFT based methods (periodogram or Blackman-Tukey), other methods have been sought to estimate the spectral density of discrete data, especially limited record length data. These newer spectral estimation methods, often termed "Modern Spectral Analysis," are model based or parametric (i.e. a function of the data) in contrast to the FFT based methods that are data independent or non-parametric. Since information (known or assumed) about the discrete data is incorporated into the spectral estimation algorithm, it is reasonable to expect some improvement in performance; usually increased frequency resolution results, although at the price of added algorithm complexity. The increase in complexity of parametric spectral estimation algorithms over that required by non-parametric (FFT based) is quite significant and it pays to ascertain whether the increased frequency resolution is necessary. In many cases of interest, however, especially when faced with a short data record (4 to 8 data samples is not unheard of), a modern spectral analysis algorithm may be necessary to effect the desired frequency resolution.

Beginning in the late 1960's, parametric spectral estimation research began in at least two distinct quarters: geophysical data processing and statistical estimation theory. Burg (1967, 1968, 1970) developed the theory of maximum entropy (ME) spectral analysis. Since truncation of the autocorrelation function results in "smearing" of the spectral estimate, a fundamental limitation of the Blackman-Tukey method, Burg (1967) proposed extending the autocorrelation function beyond the known lags in some statistical manner. The appropriate criteria, Burg argued, is the principle of "maximum entropy," or most randomness. In other words, the unknown autocorrelation function lags are estimated with the least possible constraints imposed upon them. Clearly, setting the unknown lags to zero, as is done in the Blackman-Tukey method, is not optimum. The maximum entropy method (MEM) of spectral estimation offers increased frequency resolution over that obtainable via FFT based techniques. Additionally, the Levinson-Durbin recursion, (Levinson, 1947; Wiggins and Robinson, 1965; and Durbin, 1960), may be applied to the MEM in order to efficiently calculate the unknown predictor coefficients required to form the spectral estimate. MEM, until recently, has resisted extension to higher dimensions. Lang (1981) and McClellan and Lang (1982) have proposed a solution to this difficult nonlinear problem. A comprehensive survey of the multidimensional spectral estimation problem can be found in McClellan (1982), however, this subject will be covered in greater detail in a later section of this report. As a final remark, it should be mentioned that the so called, "Burg Algorithm," Burg (1967), is distinctly different from MEM spectral estimation. The Burg algorithm will be covered

elsewhere in this thesis. Additionally, a complete treatment of MEM can be found in Jaynes (1982).

Independently of Burg, Parzen (1968) proposed the idea of utilizing autoregressive (AR) modelling as a spectral estimation technique. AR modeling has its origins in economic time series forecasting: see Walker (1931), Yule (1927), or the more accessible Box and Jenkins (1970). The Yule-Walker equations (Box and Jenkins, 1970) form a linear relationship between the AR parameters and the autocorrelation function of the assumed data model, thus the entire body of statistical time series theory can be applied to this spectral estimation method. The Yule-Walker equations may be efficiently solved by the Levinson-Durbin algorithm; as a result, calculation of the AR coefficients is relatively simple. Additionally, the AR parameters may be calculated directly from the data (solution of the Yule-Walker equations first requires autocorrelation estimates) using techniques from the linear prediction theory that have found extensive application in the fields of speech processing and geophysical predictive deconvolution. Therefore, as a complement to the statistical time series approach, any techniques available from linear prediction theory can be applied to AR spectral estimation. AR spectral estimation from the viewpoint of linear prediction will be covered in more detail in the separate section of this thesis; extensive surveys of linear prediction can be found in, for example, Makhoul (1975) and Schroeder (1984). Obviously, any of the linear predictive coding techniques prevalent in speech processing may be applied to AR spectral estimation (Rabiner and Schafer, 1978; or Markel and Gray, 1976). Within the science of geophysical data processing, specifically predictive deconvolution via least squares fil-

tering techniques, a large body of theoretical and experimental work is available (Robinson and Treitel, 1980; Robinson, 1967; Robinson, 1967; or Claerbout, 1976). Spectral estimation within the geophysical community will be covered in some detail in another section of this thesis. Underlying the application of AR spectral estimation to a specific problem is the appropiate choice of model order and model selection (Nitzberg, 1979; or Gutowski, Robinson, and Treitel, 1978) and will be covered in a later section. Before closing this brief discussion of AR spectral estimation, it should be said that Van den Bos (1971) has shown one-dimensional (1-D) MEM equivalent to the AR spectral estimator. Burg (1972) developed a theoretical relationship between the MEM spectral estimator and the maximum likelihood method (MLM) which proved that MEM offers increased frequency resolution. In higher dimensions, however, AR and MEM are not equivalent.

AR and MEM have found extensive use in many diverse fields such as direction finding (Gabriel, 1980; Thorvaldsen, Waterman, and Lee, 1980), oceanography (Holm and Hovem, 1979), environmental modeling (Hacker, 1978), biomedicine (Gersch and Yonemoto, 1977), radio astronomy (Wernecke and D'Addario, 1977), image reconstruction (Hsu, 1975), sonar (Haykin, 1985), radar (Haykin, 1985; Haykin, 1979; Gibson, Haykin, and Kesler, 1979; Kesler and Haykin, 1978), and geophysics (Barrodale and Erickson, 1980; Burg, 1967; Claerbout, 1976; Griffiths and Prieto-Diaz, 1977; La Coss, 1976; Landers and La Coss, 1977; McDonough, 1974; Ulrych and Clayton, 1976). A recent compilation (Haykin, 1985) focuses on the fields of exploration seismology, sonar, radar, radio astronomy, and tomographic imaging and as such contains, in one volume, a diverse application of AR and MEM spectral estimation (as well as many other

methods). A comprehensive survey by Kay and Marple (1981) contains a detailed treatment of many spectral estimation methods: Periodogram and Blackman-Tukey methods, modeling approach, transfer function modeling techniques, AR/LP methods, MEM, moving average (MA) methods, autoregressive moving average (ARMA) techniques, Pisarenko Harmonic Decomposition (PHD) estimator, Prony methods, and the maximum Likelihood method (MLM). In addition to the detailed technical presentation offered, many diverse applications of spectral estimation are developed and/or indicated; a summary of algorithm complexity for the various spectral estimators is particularly useful.

Remaining sections of this thesis will survey the applications and developments of spectral estimation with the field of geophysical data processing, present a more detailed treatment of AR spectral estimation from the linear prediction approach, develop an L_1 normed linear predictive spectral estimator (1-D and 2-D), and blend the FFT method with the 1-D AR/LP technique. This hybrid method (DFT/LP) provides a computationally efficient 2-D spectral estimator with increased frequency resolution over that obtainable with 2-D FFT techniques.

Spectral Estimation in Geophysics

This section will cover several of the more common applications of spectral analysis to geophysical data processing. Where appropiate, past research efforts from the geophysical literature concerning spectral estimation advancements, will be indicated. Areas of geophysical spectral estimation applications covered in this section include seismic data analysis, acoustic well log analysis, and earthquake detection. Inevitably, some overlap will occur between this

section and the previous section; although this will lead to some redundancy in presentation, it is unavoidable due to the pervasive nature of the field of spectral estimation.

Use of the F-K plane for the analysis of seismic data has been routine practice for many years. For example, in Burg (1964), the F-K representation is used to indicate the distribution of signal and noise components of their assumed theoretical seismic model. Another example of F-K plane utilization can be found in Green, Frosch, and Romeny (1965); here, an F-K representation aids in theoretical analysis of the experimental Large Aperture Seismic Array (LASA). Although Burg (1967) developed a high resolution parametric spectral estimator, namely the previously discussed ME method, other researchers were equally active. Capon, Greenfield, and Kolker (1967), in an F-K analysis of the LASA, formulated the multidimensional Maximum Likelihood Method (MLM) of two dimensional spectral estimation. Somewhat later, Capon (1969), published details of the ML spectral estimation method including analytical expressions for the means and variances of both conventional and high resolution (here, MLM) F-K estimates. The effect of the 2-D ML spectral estimator is such that an optimal (in the mean square sense) bandpass filter is formed for each desired wavenumber; significantly, uniform array spacing is not required. Although the ME spectral estimator possesses higher resolution than ML (see Burg, 1972), applying the MEM to a nonuniformly spaced array requires an interpolation step; MLM spectral estimation algorithms may be applied directly to nonuniformly spaced arrays with little added complexity. Velocity and frequency properties of seismic noise structure were analyzed via F-K methods in La Coss, Kelly, and Toksoz (1969); in this work a frequency domain beam

forming (FDBF) method of spectral estimation was utilized in order to generate F-K data. Additionally, the mean and variance of the FDBF technique were derived. La Coss (1971) reviews Fourier based F-K estimators and presents results of an experimental investigation of two high resolution methods: Capon's MLM and Burg's MEM. Also, in this work, use of F-K analysis in the study of long period seismic noise is indicated. No general conclusions were drawn concerning the relative performance of ME vs ML. Woods and Lintz (1973) applied the MLM of spectral estimation to the case of two correlated plane waves impinging on a seismic array; a claim of arbitrarily high frequency resolution is made in the limit of zero background noise and unlimited computational complexity. As a final comment concerning F-K spectral estimation of seismic array data, the reader is referred to the MEM tutorial by McDonough (1974). The separability of the ME spectral estimator is noted, thus if sufficient data is available in one dimension (usually the "time" dimension) a DFT may be applied in time and ME estimation applied in "space" with considerable reduction in algorithm complexity. Additionally, McDonough notes that an interpolation step may be inserted in order to handle the nonuniformly spaced array problem.

In addition to the numerous applications of F-K analysis to seismic data interpretation and/or filtering, another potential use of the F-K plane arises in sonic well logging. The recent geophysical literature contains significant work aimed towards the interpretation of the full waveform acoustic log; generally, the desired goal in these efforts is separation of the compression wave components from the overlapping shear wave. Although most research seems to be directed at non F-K plane

methods, recent work (Parks, McClellan, and Morris, 1983) indicates that the F-K plane may be of some value to the geologist as an aid in the interpretation of sonic logs in cases where an array of data is available. More will be said about the results of Parks et al, later; first, however, the extent of some current research concerning sonic well log analysis will be indicated. Utilizing "semblance" processing, a normalized correlation function, coupled with controlled window moveouts, Willis and Toksoz (1983), Cheng, Toksoz, and Willis (1981), and Cheng and Toksoz (1983) have successfully extracted formation shear velocity information from sonic log data. Correlation techniques implemented in the frequency domain were used by Ingram, Morris, Macknight, and Parks (1981) in order to determine shear velocities from acoustic waveform logs. Other recent efforts targeted specifically towards shear velocity determination include Minear and Fletcher (1983), Dennis and Wang (1984), Chen and Willen (1984), Tanner and Koehler (1969), and Kimball and Marzetta (1984). Typically, one or two traces were utilized in these studies, rather than an array (perhaps of size 6 or 8) of data traces. If an array of acoustic well log data is available, perhaps currently an unrealistic assumption, F-K plane analysis may offer some advantages in separating the various wave components of interest. Preliminary work by Parks, Morris, and Ingram (1982) indicated potential use of F-K plane spectral analysis to separate compressional and shear wave components, at least if an array of data were available. Synthetic acoustic well log data was utilized (real-axis integration model of Tsang and Rader, 1979), therefore array data could readily be generated. This early research continued and results presented in Parks, McClellan, and Morris (1983) demonstrated

the separation of multiple components, including the compressional head wave, casing arrival, nondispersive cement wave, stoneley wave (mode 0), mode 1, and a shear wave. The F-K data, in both reports just mentioned, were generated by performing a DFT along the time dimension (sufficient data is available in time for required frequency resolution) and applying a Prony spectral estimator in the spatial dimension. The spatial dimension is necessarily of limited aperture to ensure acceptable vertical resolution and as such requires a high resolution spectral estimation technique. Although their F-K plane method has not been tested on anything other than synthetic data, the results appear encouraging.

Before leaving the sonic log application of spectral analysis, it is worth mentioning one other promising research area that, however, is not pursued in this thesis. Modern spectral estimators are known to perform rather poorly on data that violate the assumptions of the underlying model (for example, AR) or that contain transient phenomena such as a first order exponential decay; acoustic well log data exhibit these nonstationary characteristics and can be expected to cause some modeling problems. The Burg algorithm, for example, is applied to a signal with an exponential decay in the work by Swingler (1979) and is shown to exhibit relatively poor performance. A modified Burg algorithm, (Nikias and Scott, 1981 or Scott and Nikias, 1982), incorporating an energy weighting criteria to the linear prediction equations, offers improved performance when tested against an envelope modulated sinusoidal test signal. In Nikias and Scott (1983), a "covariance least squares" spectral estimation technique is developed which resulted in improved performance (frequency resolution and robustness); here, the class of

test signals consisted of a first order exponential decay, cosine function plus an additive transient, and an envelope modulated signal. For analyzing sonic log array data, the combination of DFT in time and the covariance least squares (CLS) algorithm of Nikias and Scott (1983) in space, may provide improved spectral estimation performance.

In addition to the numerous applications of F-K plane analysis of seismic data cited, and the potential application of F-K analysis to sonic log data, other areas of geophysical data processing make use of modern spectral estimation methods as well. No attempt will be made to exhaustively list all potential applications of spectral analysis to geophysical data, however, a few references will be provided for the interested reader. Landers and La Coss (1977) apply Burg's algorithm to three different data sets of geophysical interest. First, the log spectrum of a short period seismogram (i.e. cepstrum) is estimated in order to determine the depth of a seismic event. The second application involves analysis of earthquake rate occurrence in order to check for possible periodicities related to known astronomical and/or terrestrial rotational periods. A final data set, analyzed for spectral content, was extracted from core samples of ocean bottom sediment. In this application, core displacement corresponds to "geologic" time and periodicities of various chemical parameters are correlated with ice age occurrence and past solar heating conditions. Finally, Griffiths and Prieto-Diaz (1977) apply the Burg MEM of spectral estimation to earthquake time series data; additionally, the earthquake data is analyzed via an adaptive AR method developed by Griffiths (1975).

Multichannel/Multidimensional Spectral Estimation

Many of the references cited in the previous section concerned either multichannel or multidimensional spectral estimation. This is especially the case for geophysical data applications of spectral estimation, naturally, with the emphasis placed on array processing. In this section, an attempt will be made to highlight research concerned with multichannel and multidimensional spectral estimation that has not been mentioned previously; in fact, emphasis will be placed on work reported the past 5 to 10 years, reflecting the current level of interest in higher dimensional spectral estimation.

Fortunately, a recently compiled and comprehensive tutorial is available on the topic of multidimensional spectral estimation; namely, McClellan (1982). Seven types of estimators are discussed in some detail: Fourier, separable, data extension, MLM, MEM, AR, and Pisarenko methods. In Haykin (1985), for obvious reasons, the emphasis is on multidimensional methods and provides a convenient summary of recent work in a single volume. Here, the focus will be primarily on linear predictive (AR) research and MEM.

If sufficient data are available, the obvious choice for a spectral estimator is an FFT based method, such as a windowed autocorrelation or periodogram technique. In 2-D fewer windows are available, (see Huang, 1972), and the periodogram method suffers the same maladies in 2-D as in 1-D: excessive sidelobe leakage and poor resolution when analyzing short data records.

For data collected from a line array, it often occurs that sufficient data is available in time, but since a relatively small number of array elements are generally used, the spatial dimension is charac-

terized by a rather small aperture. Therefore, a computationally efficient estimator results if a DFT is applied in the time dimension, followed by a high resolution method in the spatial dimension, (see Joyce, 1979 or Parks, McClellan, and Morris, 1983 for examples). Joyce blended the DFT with Burg's MEM and Parks et al applied the DFT in time and performed a Prony analysis in space. See also McDonough (1974) for a DFT/MEM estimator. Since the output of a DFT is complex, the high resolution estimator applied in the spatial dimension is required to handle complex valued inputs.

Another method of increased frequency resolution consists of first extrapolating the available data in both dimensions, prior to performing conventional Fourier analysis via FFT techiques. Frost and Sullivan (1979) extend their 2-D data via the Burg algorithm and apply a 2-D DFT to the extrapolated data to form the spectral estimate. Joyce (1979) extrapolates the data in one dimension, then applies a DFT in the extended dimension; the remaining dimension, however, is accomodated via MEM as mentioned previously. Another extrapolation method is available in Roucos and Childers (1979).

MLM was originally posed as a multidimensional spectral estimator, (see Burg, 1969), however, the resulting frequency resolution is known to be inferior to that of the MEM (Burg, 1972). Despite lower available frequency resolution, the MLM is simpler to implement, especially for arbitrary array spacing, and remains a viable spectral estimation method. Lim and Dowla (1983) developed an improved formulation of the MLM, based upon the known relation between the MEM and the MLM, (Burg, 1972), that offers increased spectral resolution. They claim, also, that the algorithm complexity is about that of a conventional ML esti-

mator. Marzetta (1983) re-solves the ML estimation problem by maximizing the power of a planewave at each wavenumber, rather than the planewave complex amplitude.

Although in 1-D ME and AR are equivalent, this is not true in the m-D problem, possibly due to the lack of time ordering in the 2-D plane, (Marzetta, 1978). Significant mathematical detail is available concerning the m-D MEM spectral estimation problem, (McClellan, 1982); here, however, the purpose is to highlight a few key efforts. Extensions of the MEM to multichannel spectral estimation were developed initially, (see Strand, 1977 or Morf, Vieira, Lee, and Kailath, 1978), however, formulation of an m-D MEM solution has been difficult. Apparently, the form of the 2-D ME spectral estimator was discovered by Burg, (unpublished, see Woods, 1976). A solution to the 2-D MEM problem is given by Woods (1976), based upon a proof of the existence of a 2-D discrete Markov field. Dickinson (1980), however, claims that such 2-D Markov spectral estimates need not exist, thus the viability of this spectral estimator may be in doubt. In Ulrych and Walker (1981), several methods of 2-D spectral estimation are discussed that approximate the true 2-D MEM solution; additionally, a least squares estimate of the 2-D autocorrelation matrix is developed that results in increased frequency resolution. Lim and Malik (1981) proposed an iterative algorithm for ME power spectrum estimation that is computationally efficient (FFT based), however, convergence is not guaranteed. Their algorithm, significantly, does not require uniformly spaced array sampling. Recently, a convergent 2-D MEM algorithm has been developed, (Lang and McClellan, 1982) that does not require uniform array spacing. Additionally, a necessary and sufficient condition is

derived for the existence and uniqueness of the the MEM spectral estimate. A potential disadvantage of the algorithm by Lang and McClellan (1982) is the required estimate of the autocorrelation function; Sharma and Chellappa (1984) have developed a model based 2-D MEM estimator that circumvents this step. A different application of 2-D ME techniques can be found in Wernecke and D'Addario (1977), in which a ME solution is applied to the image reconstruction problem.

Although the AR spectral estimate and the ME spectral estimate are different in higher dimensions, the AR method based on LP equations is an attractive alternative. The extension of 1-D linear prediction to 2-D is not difficult and as a result there is currently considerable research directed towards 2-D LP spectral estimation. Difficulties do arise, however, such as frequency bias and non-circular symmetric estimates; disadvantages will be pointed out as specific references are mentioned.

Jackson and Chien (1979) formulated a 2-D LP spectral estimation algorithm in order to simultaneously estimate frequency and bearing. With a single quadrant filter, they found that spectral peaks exhibited a definite skew. By calculating two sets of prediction error filter coefficients (say from quadrants one and two) and combining the results in a circular symmetric manner to form a single estimate, the frequency skew was diminished. Marzetta (1980) considers the selection of a prediction filter mask and develops a class of 2-D minimum mean square-linear prediction error filters that may be solved by a 2-D Levinson's algorithm. Therrien (1981) applies the multichannel Levinson's recursion to solve a set of 2-D normal equations, although the resulting 2-D prediction error filter may be unstable. In addition to an AR based

2-D spectral estimator, Cadzow and Ogino (1981) develop an autoregressive moving average (ARMA) model based 2-D estimator. The resulting spectral estimator is empirically shown to exhibit superior frequency resolution relative to Fourier based methods; additionally, an ARMA model generally requires fewer parameters than an AR model to achieve comparable statistical data representation. Kumaresan and Tufts (1981) combine the work of Jackson and Chien (1979) and Ulrych and Clayton (1976) into a maximum length prediction error filter formulation that is subsequently solved via the minimum norm solution (this formulation is underdetermined). Simulation results are presented that indicate their method will resolve two closely spaced plane waves in Gaussian noise. The applicability of representing array date via single quadrant AR models has been studied by Tjøstheim (1981) from a theoretical statistical time series viewpoint, also, a practical spectral estimation algorithm is developed and demonstrated on simulated array data. Chellapa and Sharma (1983) fit noncausal spatial AR models to array data in order to estimate the 2-D spectrum; empirical evidence indicates that fewer parameters are required compared to causal AR modeling to achieve the same frequency resolution. The computation burden, though not excessive, of 2-D AR spectral estimation is reduced via decimation in Zou and Lin (1984) with little loss of frequency resolution ability. Application of an AR lattice parameter model to 2-D spectral estimation is considered in Kayran, Parker, and Klich (1984) that forms the estimate from the autocorrelation matrix. The resulting algorithm was quite efficient (approximately 5 to 10 seconds on an IBM 3033) and no frequency bias was observed. Improved algorithm robustness (resistance to outliers) was achieved by Sharma and Chellappa (1984) and

involves fitting a 2-D noncausal spatial autoregressive model to the given data.

Finally, other 2-D spectral estimation methods will now be indicated, generally without comment, in order to indicate the broad scope of current research efforts. Extensions of Pisarenko's decomposition method can be found in McClellan (1982) and Lang and McClellan (1982). 2-D spectral estimation via unconstrained minimization of estimated covariance recursion error is investigated in Nikias et al (1982) and Nikias and Raghuveer (1983). In Durrani and Chapman (1983) eigenfilter methods for 2-D spectral estimation are presented, which can be viewed as an extension of Pisarenko's method. Another eigenvector technique can be found in Kumaresan and Tufts (1983). A multidimensional digital filtering approach, (Halpeny and Childers, 1975), is applied to the wavefront decomposition problem. Nawab, Dowla, and La Coss (1984) have developed a time averaged covariance method to estimate 2-D spectrums. A principle component algorithm was utilized in Rao and Kung (1984) for a state space approach to 2-D spectral estimation. Another application by Miao and Chen (1984) uses singular value decomposition techniques for spectral estimation. Obviously, many dozens of references concerning 2-D spectral estimation could be listed; the intent here, however, is to demonstrate the variety of active research within this field.

Linear Predictive Spectral Estimation Review

Autoregressive (AR) or linear predictive (LP) methods as applied to spectral estimation have been previously mentioned. Although some overlap will result, the purpose of this section is to highlight some of the more recent work in this area. Additionally, several topics, such

as line splitting and model order selection, not previously discussed, will be mentioned. Chapter III of this report presents a detailed mathematical treatment of linear predictive spectral estimation; the focus in this section will be held to a broad brush literature survey. Although this section is limited to AR and/or LP techniques, it should be remembered that in 1-D AR/LP and MEM are equivalent. Linear prediction (and AR) are very active fields of research, especially with applications to spectral estimation, speech processing, and predictive deconvolution, and as a result comprehensive tutorials are available. The reader is referred to Kay and Marple (1981) for a spectral estimation viewpoint, Makhoul (1975) or Markel and Gray (1976) for speech applications, Schroeder (1984) with a slight philosophical orientation, or Robinson and Treitel (1980) for geophysical applications.

All pole models (AR) have their origins in statistical time series analysis, (see Box and Jenkins, 1970) based upon the Yule-Walker (Y-W) autocorrelation technique (Makhoul, 1975 or Kay and Marple, 1982). Basically, the Y-W method uses biased estimates of the autocorrelation lags calculated from the data set in the Y-W normal equations. As the autocorrelation matrix of the Y-W normal equations is Toeplitz the Levinson recursion may to used to efficiently compute the AR coefficients. A positive-definite autocorrelation matrix is guaranteed if biased autocorrelation estimates are used, however, frequency resolution is decreased, a phenomena termed "line splitting" may result, and the frequency estimate may exhibit bias. In Marple (1975) a discussion of frequency bias and poor resolution can be found, and Kay and Marple (1979) consider the line splitting problem inherent in AR spectral

estimation with an assumed Toeplitz structure to the autocorrelation matrix.

The Y-W method has the disadvantage of first requiring estimates of the autocorrelation lags, an added computational step that in addition to introducing possible numerical instabilities (Barrodale and Erickson, 1980), may be avoided by calculating the AR coefficients directly from the available data. This improvement was recognized early on by Burg (1967). The so called "Burg Algorithm", distinct from Burg's MEM (although the two estimators are mathematically equivalent), estimates the AR parameters directly from data samples (no autocorrelation lag estimates required) by minimizing in a least squares sense the forward and backward linear prediction errors (one step predictor). Burg further constrained the AR parameters to satisfy the Levinson recursion, which guarantees a stable all pole prediction error filter. Unfortunately, this constraint, though guaranteeing a stable prediction filter (necessary if the AR model will be used for data prediction), causes frequency bias and line splitting in the resulting spectral estimate (Chen and Stegen, 1974; Fougere, 1976). A theoretical treatment of frequency bias in AR spectral estimation (noiseless case) can be found in Swingler (1980) and experimental results for noisy data are presented in Chen and Stegen (1974). Another disadvantage of the Burg algorithm is apparent when applied to non-stationary signals (Swingler, 1979; Nikias and Scott, 1981); generally poor spectral estimates are obtained in such cases.

In an attempt to overcome the frequency bias and line splitting problems of the Burg algorithm (a result of the Levinson constraint or enforced Toeplitz structure), Ulrych and Clayton (1976) and Nuttal

(1976) suggested the forward-backward least squares method with the Levinson's constraint removed. Marple (1980) demonstrates that less frequency bias and no line splitting results with application of such an algorithm; additionally, a computationally efficient recursive algorithm (including detailed flowchart) is presented.

In another development, with the introduction of energy weights (EW) into the Burg algorithm, Nikias and Scott (1981) proposed an algorithm that improved the frequency bias and line splitting difficulties noted previously. Further details of the E-W method are available in Scott and Nikias (1982). Nikias and Scott (1983), furthermore, improved the performance of the E-W algorithm, especially for non-stationary and/or transient signals, by removal of the Levinson constraint; this algorithm has been designated (by Nikias and Scott) as the covariance least squares (CLS) method.

Other approaches to eliminating the frequency bias and line splitting problem, inherent in the Burg algorithm, have been developed. Already mentioned is the technique developed by Marple (1980) that is nearly as efficient (computationally) as the Burg algorithm, however, this algorithm demonstrates less frequency bias; no line splitting has been observed. Additionally, Fortran source code is included, which should result in further testing of this algorithm. Kumaresan and Tufts (1980) set the number of linear prediction equations equal to the number of sinusoids, increased the number of filter coefficients to the maximum available, and utilized the minimum norm solution to the resulting underdetermined set of equations. This method is shown to offer improved performance over that available from the FBLP method of Nuttal (1976), Ulrych and Clayton (1976), and Marple (1980) at low signal to

noise ratios. As the so called Kumaresan-Prony (K-P) method requires a matrix inversion of order equal to the number of sinusoids that are expected (or desired), difficulties may arise in a complex signal scenario. Marple (1983) developed a fast algorithm for solving the pseudo-inverse operation required in the K-P method and further tests the algorithm against complex test signals; specifically, poor spectral estimates were obtained when applied to a broad band signal. Additional theoretical work concerning the K-P method may be found in Tufts and Kumaresan (1982).

The previous AR spectral estimation techniques are widely used and computationally efficient. Direct least squares solution methods have not generally been applied to the prediction equations in order to generate the AR parameters. The reasons usually cited are the increased computational complexity and a potentially non minimum phase prediction filter. If a minimum phase filter is required (not a mathematical necessity), poles outside the unit circle may be "reflected" back inside the unit circle(see Claerbout, 1976; Atal and Hanauer, 1971). In Barrodale and Erickson (1980) direct algorithms for least squares linear prediction are presented, with considerable numerical analysis considerations, based upon Cholesky's method (see Lawson and Hanson, 1974) for solving overdetermined normal equations. Their algorithm also incorporates a dynamic choice of model order; the recursion is terminated via the Akaike criteria (Akaike, 1970). Determination of model order and model selection is always a difficulty in applying AR techniques to real data. See Gutowski, Robinson, and Treitel (1978) for detailed consideration of this practical difficulty. (A new model order determination strategy is given in Fougere, 1985). The least squares

approach of Barrodale and Erickson (1980) is demonstrated to result in less frequency bias and line splitting when compared to the Burg algorithm. Additionally, the Fortran source code is listed in a separate part (Part II) of the same reference for the convenience of other researchers.

Such direct approaches to AR parameter estimation as just mentioned lead to consideration of solutions other than the L_2 norm (least squares). In general, an L_p ($1 \leq p \leq 2$) normed solution may be generated and potential advantageous in the spectral estimation problem may be possible in certain cases. For example, an L_1 solution is known to be more robust in the presence of impulsive noise than the L_2 solution (L_2 weights all data and noise equally) and may result in improved spectral estimation performance in such cases, (see Claerbout and Muir, 1973). An L_1 solution to an overdetermined system of equations, based upon a modification of the simplex method, is given in Barrodale and Roberts (1974). More recently (Yarlagadda, Bednar, and Watt, 1985), efficient iterative solutions to a set of overdetermined equations in the L_p ($1 \leq p \leq 2$) norm have been developed. This algorithm forms the basis of the L_1 spectral estimation work presented in later chapters of this report. It is remarked here that little work has been accomplished concerning the application of L_1 normed solutions to spectral estimation; an example may be found in Levy et al. (1982) in which the simplex algorithm was invoked. Other L_1 norm applications via the simplex algorithm are given in Mammone, Wang, and Gay (1985) and Garcia-Gomez and Alcazar-Fernandez (1985). In Figueiras-Vidal et al. (1985) the simplex algorithm is used to generate an L_1 solution to a set of over determined ARMA equations. Spectral extrapolation via an L_1 norm

(simplex algorithm) is postulated in Mammone (1983). As previously mentioned, an efficient version of the basic simplex algorithm may be found in Barrodale and Roberts (1974). The simplex algorithm, however, is restricted to L_1 solutions (the more general L_p problem can not be considered), is computationally inefficient, and requires significant memory storage availability. Although direct solutions of LP equations currently are unpopular in spectrum estimation research, this is not the case in other disciplines. As an example, linear regression analysis, which also involves solution of an overdetermined system of equations, is a research field in which the L_p ($1 \leq p \leq 2$) norm is considered extensively. A good starting point from which to survey linear regression research, especially concerning non L_2 norm criteria, is Arthanari and Dodge (1981); see also Huber (1981). As a general indication of the level of research activity directed towards L_p solutions (including the normed L_p space, $1 \leq p \leq 2$; $0 \leq p \leq 1$, a non normed space; and the Chebyshev criteria or L_∞), the reader's attention is directed to the following reference sampling from linear regression oriented literature: Appa and Smith (1973) on L_1 and Chebyshev estimation, Barrodale and Young (1966) for L_1 and L_∞ (Chebyshev) approximations, Barrodale and Roberts (1970) on L_p solution approximations, Ekblom and Henriksson (1969) over L_p criteria in parameter estimation, and McCormick and Sposito (1976) covering L_1 estimation. It is expected that many results concerning the application of the L_p norm to linear regression will find application to spectrum estimation, now that the computation burden has been significantly reduced (eg. Yarlagadda et al., 1985 and Barrodale and Erickson, 1980).

Another area of increasing interest with respect to linear predictive spectral estimation is that of singular value decomposition (SVD) methods; just a few references will be briefly mentioned here for the interested reader. An SVD tutorial, in the context of linear systems theory, can be found in Klema and Laub (1980). Successful applications of SVD to spectral estimation are given in Tufts, Kumaresan, and Kirsteins (1982) and Tufts and Kumaresan (1982).

CHAPTER III

LINEAR PREDICTION/AUTOREGRESSIVE MODELING

Introduction

The spectrum estimation algorithms that will be developed in Chapter IV, based upon an L_p ($p = 1$) solution to a system of overdetermined equations, start with linear prediction equations formed from the available data set. This chapter will develop the mathematical foundation of the linear prediction equations that will be utilized in the following chapter. For historical reasons, and to demonstrate the link between AR modeling and LP, the Yule-Walker (Y-W) equations will be first presented. Next, an efficient computational method of solving the Y-W equations, the so called Levinson-Durbin recursion, which avoids matrix inversion, will be discussed. Following presentation of the Levinson-Durbin recursion, linear prediction equations will be developed, and two popular solution techniques will be indicated: Burg's algorithm and the forward backward linear prediction (FBLP) method. Finally, direct L_p normed solution of LP equations will be discussed; at this point the material in Chapter IV will follow logically and directly.

Yule-Walker Equations

An Autoregressive model of order p can be expressed as

$$x_n = -\sum_{k=1}^{p} a_k \, x_{n-k} + w_n \quad , \tag{3.1}$$

where, w_n, represents the error, and a_k are termed the AR parameters. Note that the present value of this process is given by a weighted combination of p past values plus an error term, thus the process is "regressed" (from its statistical roots) on to itself; hence, the term "autoregressive process" has been applied to this representation. Of course, without the w_n term Equation (3.1) bears a striking resemblance to a linear prediction equation, however, that subject will be covered separately. From Box and Jenkins (1970), the AR power spectral density may be expressed as

$$P(f) = \frac{\sigma^2}{\left| 1 + \sum_{k=1}^{p} a_k \exp(-j2\pi kf) \right|^2} \tag{3.2}$$

with σ^2 as the variance of the error term, and a_k are the AR parameters from Equation (3.1). Equation (3.2) may also be written in z-transform notation, with $z = \exp(j2\pi f)$, to highlight the engineering terminology of "all pole model" that is applied to an AR process:

$$P(z) = \left. \frac{\sigma^2}{\left| 1 + \sum_{k=1}^{p} a_k \, z^{-k} \right|^2} \right|_{z = e^{j2\pi f}} \tag{3.3}$$

From the fundamental theorem of algebra, it is known that the denominator of Equation (3.3) has exactly p roots (i.e. poles); with the roots designated z_k, alternatively, Equation (3.3) may be expressed in factored form as

$$P(z) = \frac{\sigma^2}{\left| \sum_{k=1}^{p} (1 - \frac{z_k}{z}) \right|^2} \quad , \tag{3.4}$$

which explicity shows the all pole structure of an AR process. Note that a unit sampling interval has been assumed for notational simplicity, therefore, $0 \leq f \leq 1/2$. The important idea here, from the spectrum estimation viewpoint, is that one only needs to estimate the parameters $\{a_1, a_2, \ldots, a_p, \sigma^2\}$ and apply Equation (3.2) in order to form P(f). The denominator of Equation (3.2) may be evaluated directly or, for p sufficiently large, an FFT algorithm may be desired, since (3.2) has the same form as the DFT. Since the Y-W equations form a linear relationship between the autocorrelation function and the AR parameters, they may be used in order to estimate the unknown a_k. Derivation of the Y-W equations will generally follow Box and Jenkins (1970), although using notation generally found in engineering literature, and the AR parameter sign convention implicit in equation (3.1). The autocorrelation is defined as

$$R_k = E[x_{n-k}x_n]. \tag{3.5}$$

Since, from Equation (3.1),

$$x_n = -a_1x_{n-1} - a_2x_{n-2} - \cdots - a_px_{n-p} + w_n \text{ , then}$$

$$x_{n-k}x_n = -a_1x_{n-k}x_{n-1} \; a_2x_{n-k}x_{n-2} - \cdots - a_px_{n-k}x^*_{n-p} + x_{n-k}w_n. \tag{3.6}$$

Taking the expected value of both sides of Equation (3.6) and applying the autocorrelation definition of Equation (3.5) results in

$$R_k = -a_1R_{k-1} - a_2R_{k-2} - \cdots - a_pR_{k-p} \text{ , for } k > 0 \tag{3.7}$$

Note that for $k > 0$, x_{n-k} and W_n are uncorrelated, thus $E[x_{n-k}W_n) = 0$,

and Equation (3.7) holds. The case of k = 0 will be handled later. If we substitute k = 1, 2, ..., p into Equation (3.7) we obtain a set of linear equations for $a_1, a_2, \ldots, a_p$ in terms of $R_1, R_2, \ldots, R_p$ that are termed the Yule-Walker equations. Expanding Equation (3.7) as just indicated leads to the equation set

$$\begin{aligned} R_1 &= -a_1 R_0 - a_2 R_1 - \cdots - a_p R_{p-1} \\ R_2 &= -a_1 R_1 - a_2 R_0 - \cdots - a_p R_{p-2} \\ &\vdots \qquad \vdots \qquad \vdots \qquad\qquad \vdots \\ R_p &= -a_1 R_{p-1} - a_2 R_{p-2} - \cdots - a_p R_0 \end{aligned} \tag{3.8}$$

In practice, R_k are generally unknown, and must be replaced by autocorrelation estimates. With $\underline{a} = (a_1, a_2, \ldots a_p)^T$, $\underline{R} = (R_1, R_2, \ldots, R_p)^T$, and matrix R defined as

$$R = \begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{p-1} \\ R_1 & R_0 & R_1 & \cdots & R_{p-2} \\ \vdots & & & & \\ R_{p-1} & R_{p-2} & & \cdots & R_0 \end{bmatrix} ,$$

Equation (3.8) can be expressed more compactly in matrix notation as

$$R\ \underline{a} = -\underline{R} \ . \tag{3.9}$$

Equation (3.9) represents p equations in p unknowns and therefore may be solved directly for the AR parameters, $\underline{a}$, which results in

$$\underline{a} = -R^{-1}\ \underline{R} \tag{3.10}$$

The matrix inversion in Equation (3.10) may be avoided by use of the Levinson-Durbin recursion, which will be illustrated in a later section

of this chapter. The awkward notation used in Equations (3.9, 3.10) serves to emphasize that vector, $\underline{R}$, is comprised of certain elements of matrix, R plus R_p; this fact, along with the Hermitian and Toeplitz structure of matrix R allows application of the Levinson-Durbin recursion. When, k = 0, in Equation (3.6), another contribution from the expectation operation is $E(w_n w_n) = \sigma^2$; thus, Equation (3.7) becomes

$$R_0 = -a_1 R_{-1} - a_2 R_{-2} - \ldots - a_p R_{-p} + \sigma^2, \; k = 0. \tag{3.11}$$

Therefore, once the AR parameters have been computed from Equation (3.10), Equation (3.11) may be used to calculated σ^2, if an estimate of the "true" power spectral density via Equation (3.2) is desired, rather than just the spectral peak locations.

Levinson - Durbin Recursion

Since use of Equation (3.10), developed in the previous section, assumes a priori knowledge of AR model order and, furthermore, requires a matrix inversion, other means of solution are desired. One widely used technique involves application of the Levinson-Durbin recursion (Levinson, 1947; Durbin, 1960) to Equation (3.10). The recursion will be developed and illustrated by way of a simple example before the general recursive formulae are given.

In the previous section theoretical autocorrelations of lag k were denoted by R_k; this section will use the notation, r_k, for an estimated autocorrelation at lag k. Also, an additional subscript will be added to the AR coefficients, a_k, in order to indicate model order; for example, a_{p2} is the second AR parameter from a p^{th} order model. The

complete AR parameter set is $\{a_{p1}, a_{p2}, \ldots, a_{pp}\}$, for an AR(p) model specification. With the notation just specified, the Y-W equations may be written

$$r_k = -a_{p1}r_{k-1} - a_{p2}r_{k-2} - \cdots$$
$$-a_{pp}r_{k-p}\ , \quad k = 1, 2, \ldots, p. \tag{3.12}$$

For an AR(2) process, Equation (3.12) becomes ($r_o = 1$):

$$r_2 = -(a_{21}r_1 + a_{22})$$
$$r_1 = -(a_{21} + a_{22}r_1). \tag{3.13}$$

An AR(3) process is expressed as

$$r_3 = -(a_{31}r_2 + a_{32}r_1 + a_{33})$$
$$r_2 = -(a_{31}r_1 + a_{32} + a_{33}r_1)$$
$$r_1 = -(a_{31} + a_{32}r_1 + a_{33}r_2). \tag{3.14}$$

The last two equations of Equation set (3.14) may be solved for coefficients a_{31} and a_{32} in terms of a_{33}. In matrix form the solution is

$$\begin{bmatrix} a_{31} \\ a_{32} \end{bmatrix} = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}^{-1} \begin{bmatrix} -r_2 - a_{33}\, r_1 \\ -r_1 - a_{33}\, r_2 \end{bmatrix}. \tag{3.15}$$

Rewritting Equation (3.15) results in

$$\begin{bmatrix} a_{31} \\ a_{32} \end{bmatrix} = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}^{-1} \begin{bmatrix} -r_2 \\ -r_1 \end{bmatrix} - a_{33} \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{3.16}$$

From Equation (3.13) we find

$$\begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} r_1 & 1 \\ 1 & r_1 \end{bmatrix}^{-1} \begin{bmatrix} -r_2 \\ -r_1 \end{bmatrix} , \tag{3.17}$$

which upon substitution into Equation (3.16) leads to

$$\begin{bmatrix} a_{31} \\ a_{32} \end{bmatrix} = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} + a_{33} \begin{bmatrix} a_{22} \\ a_{21} \end{bmatrix} . \tag{3.18}$$

Expanding Equation (3.18), we have

$$\begin{aligned} a_{31} &= a_{21} + a_{33}\, a_{22} \\ a_{32} &= a_{22} + a_{33}\, a_{21} . \end{aligned} \tag{3.19}$$

Next, solve the first equation from Equation set (3.14) for a_{33}, thus

$$a_{33} = -r_3 - a_{31} r_2 - a_{32} r_1 ,$$

which, upon substitution of a_{31} and a_{32} from Equation (3.19), and solving for a_{33}:

$$a_{33} = \frac{-(r_3 + a_{21}\, r_2 + a_{33}\, r_1)}{1 + a_{21}\, r_1 + a_{22}\, r_2} . \tag{3.20}$$

In general, the recursion may be expressed

$$a_{pk} = a_{p-1,k} + a_{pp}\, a_{p-1,p-k} , \quad k=1, 2, \ldots, p-1 \tag{3.21}$$

$$a_{pp} = \frac{-(r_p - \sum_{k=1}^{p-1} a_{p-1,k}\, r_{p-k})}{1 + \sum_{k=1}^{p-1} a_{p-1,k}\, r_k} \tag{3.22}$$

A final important point will be made concerning the coefficient, a_{pp}. In the statistical literature, the last coefficient, a_{pp}, is termed the

partial autocorrelation coefficient (PARCOR). Thus, for example, if the true process is AR(q), p = 1, 2, ..., q, defines a set of partial autocorrelations coefficients $\{a_{11}, a_{22}, \ldots, a_{qq}\}$. It is known that (see Box and Jenkins, 1970) the partial autocorrelation coefficients, a_{pp}, are non zero for p less than or equal to q and zero for p greater than q. Also, for N data points, the estimated partial autocoefficients of order q + 1 and higher are approximately independent with variance

$$V_{ar}\{a_{pp}\} \simeq 1/N \ , \ p \geq q + 1 \ .$$

The partial autocorrelation coefficients, therefore, may be used as a guide for choosing the "correct" AR model order. A slightly different formulation, though equivalent, of the Levinson-Durbin recursion is given in Kay and Marple (1981); specifically the denominator of Equation (3.22) is formulated as a function of s^2 and partial autocorrelation coefficient, a_{pp}. In the engineering literature, the partial autocorrelation coefficients, $\{a_{11}, a_{22}, \ldots, a_{qq}\}$ are often termed reflection coefficients. Additionally, it can be shown (Lang and McClellan, 1979) that a necessary and sufficient condition for a positive semidefinite autocorrelation matrix is $|a_{pp}| \leq 1$, p = 1, 2, ..., q. $|a_{pp}| \leq 1$, p = 1, 2, ..., q is also a necessary and sufficient condition that the poles of the prediction error filter lie on or within the unit circle (Lang and McClellan, 1979).

Linear Prediction

The autoregressive process, AR(p), expressed in Equation (3.1) can also be cast into a linear prediction equation. Consider the case where x_n is predicted from a linear combination of the previous p data samples

(hence, the term linear prediction). Denoting the predicted value by x_n, we have

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} \quad . \tag{3.23}$$

The error between the actual value, x_n, and the predicted value, $\hat{x}_n$, is given by

$$e_n = x_n - \hat{x}_n$$

$$e_n = x_n + \sum_{k=1}^{p} a_k x_{n-k} \quad , \tag{3.24}$$

where e_n is known as the residual. The parameter set $\{a_1, a_2, \ldots, a_p\}$ may be chosen to minimize the prediction error power, E_p, defined as

$$E_p = E\,(e_n^2)$$

$$E_p = E\{(x_n - \hat{x}_n)^2\} \quad . \tag{3.25}$$

Substituting Equation (3.24) into Equation (3.25) results in

$$E_p = E\{x_n + \sum_{k=1}^{p} a_k x_{n-k}\}^2 \quad , \tag{3.26}$$

which may be minimized by setting

$$\frac{\partial E_p}{\partial A_i} = 0, \quad 1 \leq i \leq p. \tag{3.27}$$

Carrying out the differentiation indicated in Equation (3.27) leads to the normal equations

$$E(x_n x_{n-i}) = -\sum_{k=1}^{p} a_k E(x_{n-k} x_{n-i}), \quad 1 \leq i \leq p. \tag{3.28}$$

For a stationary process x_n, we have

$$E(x_{n-k}\, x_{n-i}) = R_{|i-k|} \quad , \tag{3.29}$$

with R_i denoting the process autocorrelation. Under a stationary assumption, Equation (3.28) becomes

$$R_i = - \sum_{k=1}^{p} a_k R_{|i-k|} , \quad i = 1, 2, \ldots, p. \tag{3.30}$$

The minimum average error is

$$E_{p\ min} = E(x_n^2) + \sum_{k=1}^{p} a_k E(x_n x_{n-k})$$

$$E_{p\ min} = R_0 + \sum_{k=1}^{p} a_k R_k \ . \tag{3.31}$$

The error sequence e_n is uncorrelated with prediction values x_n and, further, e_n is a white process if x_n was generated by an AR(p) process. A comparison of Equation (3.30) with Equation (3.10) and a comparison of Equation (3.31) with Equation (3.11) demonstrates the equivalence of AR parameter identification via the Y-W equations and linear prediction of an AR process.

As seen in the previous sections, the Y-W equations that resulted from either the AR parameter identification approach or from linear prediction theory may be solved for the unknown AR coefficients using estimates of the autocorrelation lags. Several methods exist, however, that do not require autocorrelation estimates and operate directly on the raw data. One method uses forward and backward linear prediction with a Levinson's constraint (Burg, 1967), while the other technique combines forward and backward linear prediction (Ulrych and Clayton, 1976; Nuttall, 1976) without the Levinson's constraint. Both methods will be briefly highlighted, as they are perhaps the most popular in use. Detailed derivation may be found in the indicated references; here, however, the key equations will be summarized, (see Kay and Marple, 1981).

The Burg algorithm (Burg, 1967) minimizes the sum of the forward and backward prediction error energies under a constraint that the AR

parameters satisfy the Levinson recursion. Let the predicted sample be given by a linear combination of p previous data samples (N total samples), then the forward prediction error is (from Equation 3.24):

$$e_{pn} = x_n - \hat{x}_n$$

$$e_{pn} = x_n + \sum_{k=1}^{p} a_{pk} X_{n-k}$$

$$e_{pn} = \sum_{k=0}^{p} a_{pk} X_{n-k} \quad , \tag{3.32}$$

where $a_{po} = 1$ and n is defined for $p \leq n \leq N - 1$. If the process is wide sense stationary the backward prediction parameters are the complex conjugates of the forward prediction coefficients, thus the backward prediction error may be written (* denotes complex conjugate)

$$b_{pn} = \sum_{k=0}^{p} a_{pk}^{*} x_{n-p+k} \quad , \tag{3.33}$$

with $p \leq n \leq N - 1$ and $a_{po} = 1$. In order to estimate the AR parameters, Burg minimized the sum of the forward and backward prediction error energies

$$E_p = \sum_{n=p}^{N-1} |e_{pn}|^2 + \sum_{n=p}^{N-1} |b_{pn}|^2 \tag{3.34}$$

subject to the Levinson constraint

$$a_{pk} = a_{p-1,k} + a_{pp} a_{p-1,p-k}^{*} \tag{3.35}$$

over the AR model order 1, 2, ..., p. As details of the Burg algorithm are quite complex they will be omitted. A complete listing of the necessary recursions (the equations are readily programmable) and a flowchart is available in Kay and Marple (1981). The Burg algorithm is

known to exhibit line splitting and frequency bias (see Kay and Marple, 1979; Chen and Stegan, 1974; Fougere, 1976). These problems have been reduced or eliminated by a method introduced by Ulrych and Clayton (1976) and Nuttall (1976) in which the forward and backward prediction error energies are minimized, as in the Burg algorithm, but with the Levinson constraint removed. This forward-backward linear prediction (FBLP) algorithm follows by setting the derivatives of E_p with respect to AR parameters a_{p1} to a_{pp} to zero. Thus.

$$\frac{\partial E_p}{\partial a_{pi}} = 0, \; i = 1, 2, \ldots, p \tag{3.36}$$

which results in

$$2 \sum_{j=0}^{p} a_{pj} \, r_p(i, j) = 0 \quad , \tag{3.37}$$

where

$$r_p(i, j) = \sum_{k=0}^{N-p-1} (x_{k+p-j} x^*_{k+p-i} + x_{k+i} \, x^*_{k+j}), \; 0 \leq i, j \leq p \tag{3.38}$$

The minimum prediction error energy is

$$E_p = \sum_{j=0}^{p} a_{pj} \, r_p(0,j) \quad . \tag{3.39}$$

The p normal equations from (3.37) and (3.39) may be expressed in matrix form as

$$R_p \, \underline{a}_p = \underline{E}_p \quad , \tag{3.40}$$

where $\underline{a}_p = (1, a_{p1}, \ldots, a_{pp})^T$,

$\underline{E}_p = (E_p, 0, \ldots, 0)^T$, and

$$R_p = \begin{bmatrix} r_p(0,0) & \cdots & r_p(0,p) \\ \vdots & & \\ r_p(p,0) & \cdots & r_p(p,p) \end{bmatrix} .$$

Equation (3.40) may be solved by Gaussian elimination or via a fast recursive algorithm by Marple (1980). An iterative least squares solution has been developed by Barrodale and Erickson (1980) for Equation (3.40). Iterative least squares (or other L_p normed solutions for that matter) have the additional advantage of less round off error than Levinson's recursion for p large (Barrodale and Erickson, 1980).

Iterative Linear Prediction Parameter Estimation

The so called direct methods, such as the Burg algorithm and the FBLP method briefly covered previously, are used extensively. For one thing, the solution techniques are based on the Levinson-Durbin recursion, which in addition to possessing computational efficiency, allows the often difficult choice of model order to proceed in a logical manner. Various termination rules, for example Akaike's criteria, provide convenient means for stopping computations when an "optimal" AR model order is reached. Curiously, statisticians may consider the Levinson-Durbin recursion numerically unstable (see Box and Jenkins, 1970) and suggest caution if such a recursion is invoked. The Levinson-Durbin recursion, however, is used extensively in engineering applications and such numerical instabilities have apparently not been reported. Possibly, due to the widespread success of such recursive solutions to AR/LP equations, iterative methods have not found much application within engineering spectrum estimation research. The

potential use of iterative methods, however, is indicated in Makhoul (1975).

Many iterative methods from the field of numerical analysis are available, including gradient methods, steepest descent techniques, Newton's method, the conjugate gradient method, and the stochastic approximation method (Hildebrand, 1974). Wang and Treitel (1973) used gradient methods in the design of digital Wiener filters; such a problem, naturally, is quite similar to AR modeling and parameter estimation. Barrodale and Erickson (1980) reported excessive rounding errors from application of the Levinson-Durbin recursion based Burg algorithm and sought to develop an iterative least squares solution to the normal equations; an efficient implementation of the numerically stable Cholesky method was developed and applied to the spectrum estimation problem. The new technique reportedly resulted in less frequency bias and line splitting compared to the Burg algorithm. L_p ($1 \leq p \leq 2$) solutions of normal equations for predictive deconvolution of seismic wavelets were considered by Yarlagadda, Bednar and Watt (1985). Efficient algorithms were developed based upon the iterative reweighted least squares (IRLS) algorithm and the residual steepest descent (RSD) method. There may be good reasons for consideration of the L_p ($1 \leq p \leq 2$) norm. L_1 solutions, for example, are known to be more robust in the presence of data contaminated by outliers (Huber, 1981; Claerbout and Muir, 1973); in such cases iterative techniques are required to generate the appropiately normed solution, whatever disadvantages may exist relative to the recursive direct method of solution (Burg algorithm). Estimation of AR parameters by iterative techniques for spectral estimation applications has not been rigorously

pursued, probably due to the additional computational burden. Additionally, the AR parameters calculated via iterative methods need not result in a minimum phase prediction error filter, which may be a problem in some applications. If the location of the frequencies to be estimated is the end goal, the prediction error filter need not be stable. If the estimated AR parameters will eventually be utilized in an AR model, however, perhaps for data extrapolation, then filter stability is necessary. Atal and Hanauer (1971) present a technique for reflecting any poles generated by a LS method inside the unit circle, with the amplitude response left unchanged, if filter stability is required.

That the AR parameter estimation problem can be recast into a set of linear algebraic equations will be demonstrated. Equation (3.23), with prediction filter length m and n total data points, may be written in matrix form as

$$X \underline{a} = \underline{y} \tag{3.41}$$

where

$$X = \begin{bmatrix} x_m & x_{m-1} & \cdots & x_1 \\ x_{m+1} & x_m & \cdots & x_2 \\ \vdots & & & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-m} \end{bmatrix} ,$$

$$\underline{a} = (a_1, a_2, \ldots, a_m)^T , \quad \text{and}$$

$$\underline{y} = -(x_{m+1}, x_{m+2}, \ldots x_n)^T .$$

Note that if $n > 2m$ this system is overdetermined and in general no $\underline{a}$ will exist that satisfies equation (3.41). With a residual vector $\underline{e} = \underline{y} - X \underline{a}$ defined, with $n > 2m$, and assuming matrix X has rank m, a solution

is known that minimizes the residual sum of squares $\underline{e}^T\underline{e}$. The LS solution to equation (3.41) in this case is

$$\underline{a} = (X^TX)^{-1} X^T\underline{y} \tag{3.42}$$

Equations (3.39) and (3.40) could be utilized to estimate the AR parameters, or alternatively LS iterative algorithms applied (e.g. Barrodale and Erickson, 1980), however, use of forward prediction equations alone may result in a non positive definite autocorrelation function (see Ulrych and Clayton, 1976). The FBLP method discussed in an earlier section of this report guarantees a positive definite autocorrelation matrix, furthermore, the F-B equations are easily incorporated into equation (3.41) by redefining X, $\underline{a}$, and $\underline{y}$. Again, for a prediction filter of order m, and n data points, we have $X\,\underline{a} = \underline{y}$ (equation 3.41 repeated) with

$$X = \begin{bmatrix} x_m & x_{m-1} & \cdots & x_1 \\ x_{m+1} & x_m & \cdots & x_2 \\ \vdots & & & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-m} \\ x_2^* & x_3^* & \cdots & x_{m+1}^* \\ x_3^* & x_4^* & \cdots & x_{m+2}^* \\ \vdots & & & \vdots \\ x_{n-m+1}^* & x_{n-m+2}^* & \cdots & x_n \end{bmatrix} \tag{3.43}$$

$\underline{a} = (a_1, a_2, \ldots, a_m)^T$, and

$\underline{y} = -(x_{m+1}, x_{m+2}, \ldots, x_n, x_1^*, x_2^*, \ldots, x_{n-m}^*)^T$.

Thus, the FBLP equations of Ulrych and Clayton (1976) and Nuttall (1976) can be expressed in the linear algebra form of $X\underline{a} = \underline{y}$. Note that the system is overdetermined for $n > 3m/2$. Equations of the form of (3.41)

are readily solvable in the L_p ($1 \leq p \leq 2$) norm via algorithms developed by Yarlagadda, Bednar, and Watt (1985) and the effect of non L_2 solutions upon the spectral estimation problem can be ascertained. Specific algorithms used to generate L_p ($p = 1$) normed solutions will be presented in the next chapter when applied to specific spectral estimation problems.

CHAPTER IV

1-D LP (L_p Normed) SPECTRAL ESTIMATION

Introduction

Spectral estimation has applications to many fields such as radar, sonar, radio astronomy, and seismic prospecting to name a few. If enough data are available and the frequency spacing adequate, discrete Fourier transform (DFT) techniques are preferred for resolving the sinusoids. However, if the spacing of the sinusoids is within the resolution limits of the DFT, as is often the case with a limited data record length, other methods must be used to extract frequency information.

A number of modern spectral estimation methods (i.e. parametric or model based) are available for use; these include the maximum likelihood method (MLM), the maximum entropy method (MEM), and the auto-regressive method (AR) or linear predictive (LP) techniques. A comprehensive summary of modern spectral estimation methods can be found in Kay and Marple (1981). The AR based forward-backward linear prediction (FBLP) method introduced in Nuttal (1976) and Ulrych and Clayton (1976) is particularly amenable to an L_1 solution and forms the basis of this work.

L_2 solutions are widely used and offer computational and theoretical advantages; it is well known, however, that for certain noise types, for example impulsive noise, the L_2 solution is not optimal. For

the impulsive noise case an absolute value error criteria applied to the linear prediction equation (i.e. an L_1 solution) offers better results than an L_2 solution. A mean square error criteria equally weights the data and as a result a single bad data point (impulsive noise) may have a disproportionately large effect on the solution. By contrast an L_1 solution will tend to reject a few bad data points (outliers) and result in a solution that is more representative of the original data. For the spectral estimation problem the end result may be increased frequency resolution. Levy et al (1982) have presented an application of the L_1 norm to spectral estimation based upon the simplex algorithm; however, the simplex algorithm is restricted to generating an L_1 solution and furthermore, possesses computational disadvantages. The residual steepest descent (RSD) algorithm can be used to generate L_p solutions (generally $1 \leq p \leq 2$), however, convergence has been a problem. A modification of the RSD algorithm (Yarlagadda et al., 1985) has eliminated the convergence problem and a viable method is now available for generating L_p normed solutions.

Unpublished work by Schroeder and Yarlagadda has shown that increased frequency resolution is possible for linear predictive spectral estimation utilizing an L_1 norm solution to the prediction equations. Specifically, an L_1 solution demonstrated increased frequency resolution for the case of two 2-D sinusoids in impulsive noise; other types of noise were not considered. This chapter presents an L_1 solution to a set of overdetermined 1-D linear prediction equations, the FBLP method of (Nuttal, 1976; Ulrych and Clayton, 1976), and tests the resulting spectral estimator against the case of two sinusoids in noise; the noise considered in this work is Gaussian, Rayleigh, uniform, impulsive, Laplacian, and Cauchy.

L_1 Solution To Forward-Backward Linear Prediction Equations

The FBLP method provides a convenient framework for applying the RSD algorithm to the linear prediction spectral estimation problem. As the RSD algorithm requires an overdetermined system of equations, one can choose the prediction filter length such that this condition is satisfied. Additionally, for the case of two sinusoids in noise, the necessary computations are minimal since the filter length can be kept quite short. It has been shown, (Kumaresan, 1982), that for M sinusoids and N data values the prediction error filter length, L, should be between M and (N - M/2); L = N/2 or L = N/3 is often used as a rough rule of thumb in selecting L (see Ulrych and Clayton, 1976). Since the focus of this section is on a relative comparison of an L_1 solution under various noise types rather than to produce the highest possible resolution at low signal to noise ratios (SNR) the prediction filter length has been set to the minimum allowable, L = M (here M = 2) for computational simplicity.

The test data (N = 8 data points) is generated from:

$$x(n) = a_1 \exp(\omega_1 n) + a_2 \exp(\omega_2 n) + w(n), \; n = 1,2,\ldots N \qquad (4.1)$$

where a_1 and a_2 are in general unknown complex numbers. In this work, both have been set to one for simplicity. w(n) are independent complex noise samples. Applying the prediction filter to the N data samples without going off the data segment, in the forward and backwards direction the prediction error equations can be written (see equation 3.41 with vector $\underline{a}$ now denoted $\underline{g}$):

$$\begin{bmatrix} x(L) & x(L-1) & \cdots & x(1) \\ x(L+1) & x(L) & \cdots & x(2) \\ \vdots & \vdots & & \vdots \\ x(N-1) & x(N-2) & \cdots & x(N-L) \\ x^*(2) & x^*(3) & \cdots & x^*(L+1) \\ x^*(3) & x^*(4) & \cdots & x^*(L+2) \\ \vdots & \vdots & & \vdots \\ x^*(N-L+1) & x^*(N-L+2) & \cdots & x^*(N) \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_L \end{bmatrix} = - \begin{bmatrix} x(L+1) \\ x(L+2) \\ \vdots \\ x(N) \\ x^*(1) \\ x^*(2) \\ \vdots \\ x^*(N-L) \end{bmatrix}$$

or more compactly, (4.2)

$$\underline{X}\underline{g} = \underline{b}, \tag{4.3}$$

where (*) corresponds to complex conjugation.

The prediction coefficients are given by:

$$g = (g_1, g_2, \ldots, g_L)^T. \tag{4.4}$$

The prediction error filter is given by:

$$H(z) = 1 + \sum_{k=1}^{L} g_k z^{-k} . \tag{4.5}$$

The frequency estimate is calculated from:

$$S(Z) = \frac{1}{|H(z)|^2} , \quad z = e^{j2\pi f} \tag{4.6}$$

For this work, L = 2, M = 2, and N = 8, thus equation (4.2) represents an overdetermined system with matrix X having dimensions 12 x 2, column vector, $\underline{g}$, length 2, and column vector, $\underline{b}$, length 12. L_p analysis can now be applied directly to equation (2) via the complex RSD algorithm (Yarlagadda et al,. 1985) summarized below.

Complex RSD Algorithm

1. $\underline{g}(0) = (X^*X)^{-1}X^*\underline{b}$

2. $\gamma_i(k) = \text{sgn}(|X\underline{g}(k)| - |\underline{b}|_i \cdot |(X\underline{g}(k) - \underline{b})_i|^{(p-1)}$

3. $\underline{r}(k) = X\underline{g}(k) - \underline{b}$
4. Minimize with respect to Δ_k

   $E(k) = \underline{r}(k) - \Delta_k X(X^*X)^{-1}X^* \underline{y}(k)$
5. $\underline{g}(k+1) = \underline{g}(k) - \Delta_k(X^*X)^{-1}X^* \underline{y}(k)$
6. Go to stop 2 or stop if convergence is achieved.

Here, X^* corresponds to complex conjugate transpose. All matrices and vectors in the RSD algorithm just presented are defined over the complex field as mentioned previously; this allows use of the analytic signal in the computer simulations. In particular, step 2 of the RSD algorithm has been modified such that the magnitude of the residual vector is tested by the sgn(.) function. The matrix inverse operation is defined over the complex field and no modification is required. Note that in this case with L = 2, calculating $(X^*X)^{-1}$, required in step 1, of the RSD algorithms is trivial; in a more complex scenario the inverse computation could be avoided employing Cholesky's algorithm. The minimization in step 4 involves only vectors and as such is not difficult; in this work the iterative reweighted least squares (IRLS) algorithm (Yarlagadda et al., 1985) was used to calculate Δ_k. The minimization of the Δ_k factor using IRLS techniques has typically required six or seven iterations, however, since only vectors are involved the computations involved are minimal. The RSD algorithm generally converges within two or three iterations. This method of generating an L_p solution, namely RSD coupled with IRLS as previously explained, is computationally quite efficient.

In the following section simulation results will be presented to verify that the L_1 solution to the linear prediction equations allows the detection of two sinusoids in noise. Additionally, in some cases an

L_1 solution of two sinusoids will be shown to provide greater resolution than that obtainable via L_2 techniques.

Simulation Results

An 8 point test data segment was chosen to realistically represent a rather short data segment; thus N = 8. The unknown complex constants, a_1 and a_2, have been set to one ($a_1 = a_2 = 1$). Since the goal of this section is to compare an L_1 based spectral estimator with an L_2 based spectral estimator in a variety of noise backgrounds and not to arrive at a lower bound on performance related to SNR, a very short filter length has been chosen. Here, the prediction filter length is set to two, L = 2. Six different noise types have been simulated for this work: Gaussian, Rayleigh, uniform, impulsive, Laplacian, and Cauchy. For the case of Gaussian, Rayleigh, and uniform noise, SNR = 30 dB; the impulsive noise consists of a single complex spike with both real and imaginary parts having a value of 2.0. A unit sampling period was chosen for convenience which results in the maximum frequency range of 0 to .5Hz. All spectral data in Figures 9-40. are plotted on the interval (0. - .25) Hz with a resolution of .001 Hz. The relative frequency spacing between the two sinusoids was chosen such that the L_2 based spectral estimator just failed to resolve the two complex sinusoids for each noise type; the result of applying the L_1 based estimator is then given for comparison. It must be remembered that this data represents just one realization of what is a random process; i.e. the sinusoidal locations are random variables with unknown mean, variance, and underlying probability density functions (pdf). Note that in all cases the frequency spacing of the two sinusoids exceeds the resolution

available from the DFT (1/8 = .125 Hz). As the prediction error filter has just two coefficients Equation (4.5) is evaluated directly (a DFT is more efficient for longer filters); Equation (4.6) is used to form an estimate of the sinusoid locations.

In this work uniform noise is generated from the FORTRAN RAN(.) function call. Gaussian noise is derived from uniform noise, based upon the Central Limit Theorem, by summing twelve (12) independent samples of uniform noise. Rayleigh noise can be generated from Gaussian noise using the probability density function transformation, $Y = \sqrt{X_1^2 + X_2^2}$, where X_1 and X_2 are N(0,1). Laplacian noise is generated from unifrom noise by the transformation $Y = \sqrt{2}\ \text{sign}(x)\ \log\ (1 - 2|x|)$ and Cauchy noise by the transformation $Y = .3183099\ \tan\ (\pi x)$. The power is normalized in all cases by scaling the random samples by the factor $(1/\sigma^2)$.

Prior to looking at data which compares L_2 normed spectral estimation with L_1 normed spectral estimation, it is instructive to first compare DFT techniques with LP methods. An L_2 normed solution will be used for the model based estimator and the DFT based periodogram method is used for the nonparametric estimator. The eight sample data record is padded with zeros to 512 points prior to application of the DFT. From Figure 1 it can be seen that use of a DFT on short record length data has difficulties, even with widely spaced sinusoids in moderate noise (f_1 = .1 Hz, f_2 = .4 Hz with 10 dB Gaussian noise). Although the two spectral peaks are easily discernable, the excessive sidelobe amplitude, due to a short data record, gives rise to the possibility of a third spectral peak at .25 Hz. By comparison, as seen in Figure 2 , the LP spectral estimator unambiguously depicts the

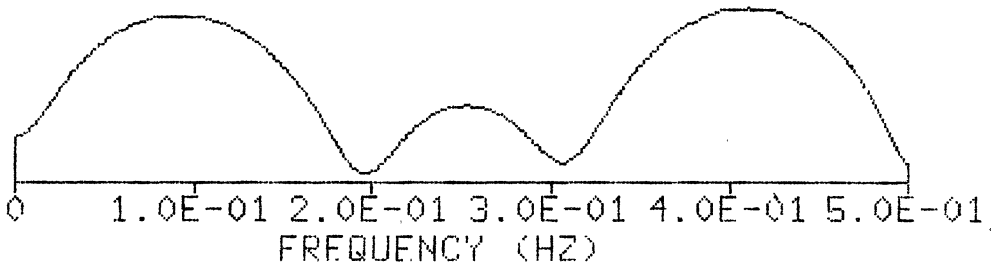

Figure 1. 1-D DFT Spectral Estimate (N = 8; f_1, f_2 = .1, .4 Hz; 10 dB Gaussian noise)

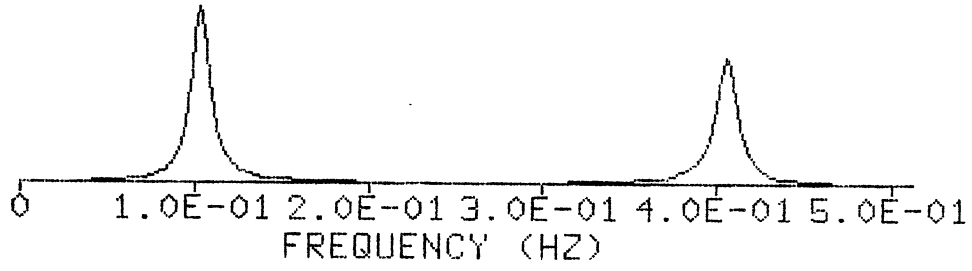

Figure 2. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .4 Hz; 10 dB Gaussian noise)

correct spectrum. In the data just presented, although 10 dB of Gaussian noise was added, the difficulty of the DFT method resulted from limited record length data rather than signal to noise ratio. This case is repeated with no noise to demonstrate this; compare Figure 3 (DFT method) with Figure 4 (LP method). Next, the relative poor resolution capabilities of a DFT spectral estimator will be shown. For this example, f_1 = .2 Hz, f_2 = .27 Hz, and 20 dB of Gaussian noise has been added. From Figure 5 it can be seen that the DFT method is unable to resolve the spectral peaks, however, the LP estimator is able to resolve the spectral peaks (Figure 6). Again, this example is repeated with no noise added. The DFT method is still unable to resolve the peaks (Figure 7), but the LP method easily resolves the spectral peaks (Figure 8). Indeed, for this no noise case, within the limits of computer word precision the peaks tend towards a pair of impulses. The remaining data sets offer a comparison of an L_2 normed LP method and an L_1 normed estimator in Gaussian and non Gaussian noise.

For the case depicted in Figures 9 and 10, f_1, f_2 = .15, .18 Hz and the noise is impulsive (a single complex impulse). The results of an L_2 based solution is shown in Figure 9 and it can be seen that the sinusoids can not be resolved. In contrast, the L_1 based estimator easily separates the two sinusoids as seen in Figure 10. This is not surprising, of course, since the L_1 norm is known to reject outliers, while the L_2 norm gives equal weight to all data values. With Gaussian noise (SNR = 30 dB) added to the test data (Frequency spacing f_1, f_2 = .15/.185 Hz) the results from applying the L_2 based estimator is shown in Figure 11; the results of applying the L_1 estimator can be seen in Figure 12. Again, the sinusoids are unresolved when using the L_2

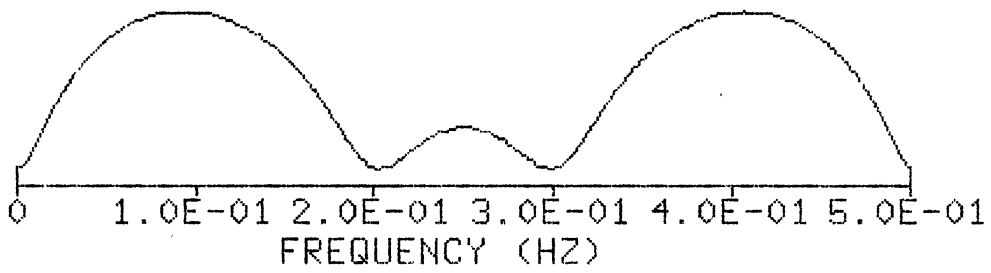

Figure 3. 1-D DFT Spectral Estimate (N = 8; f_1, f_2 = .1, .4 Hz; No noise)

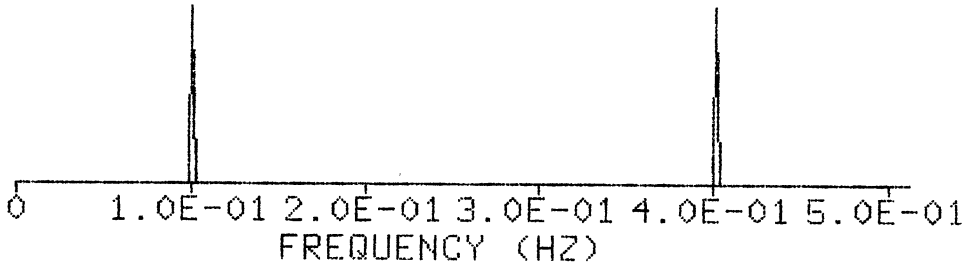

Figure 4. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .4 Hz; No noise)

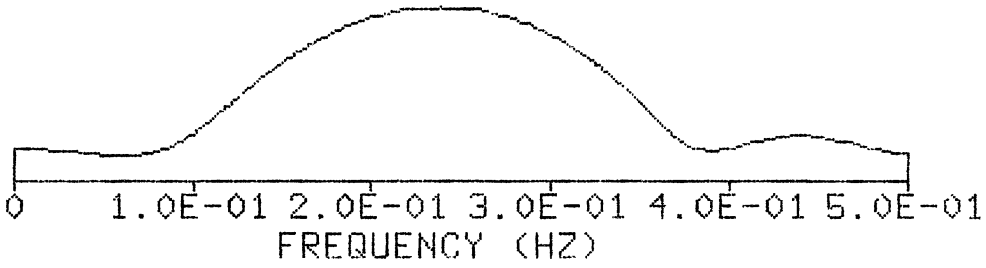

Figure 5. 1-D DFT Spectral Estimate (N = 8; f_1, f_2 = .2, .27 Hz; 20 dB Gaussian noise)

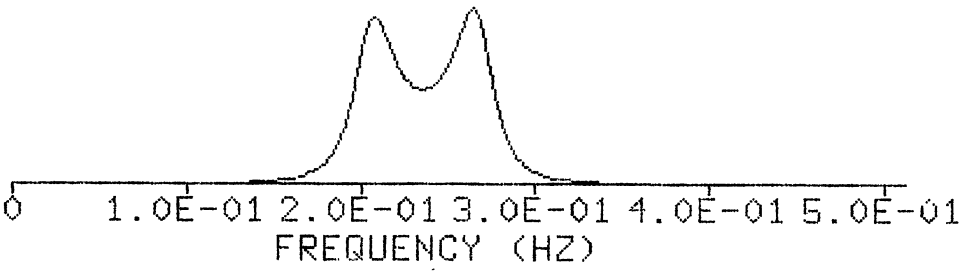

Figure 6. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .2, .27 Hz; 20 dB Gaussian noise)

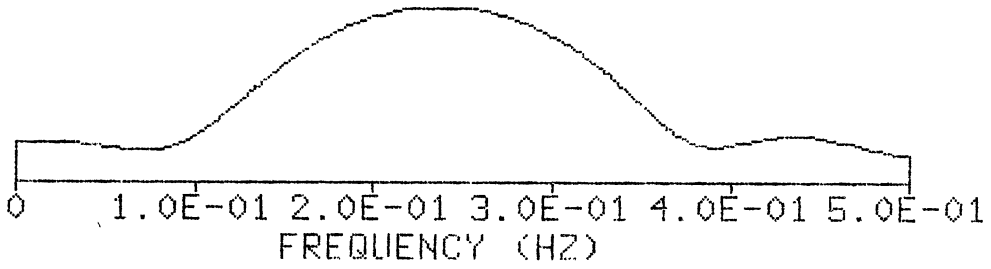

Figure 7. 1-D DFT Spectral Estimate (N = 8; f_1, f_2 = .2, .27 Hz; No noise)

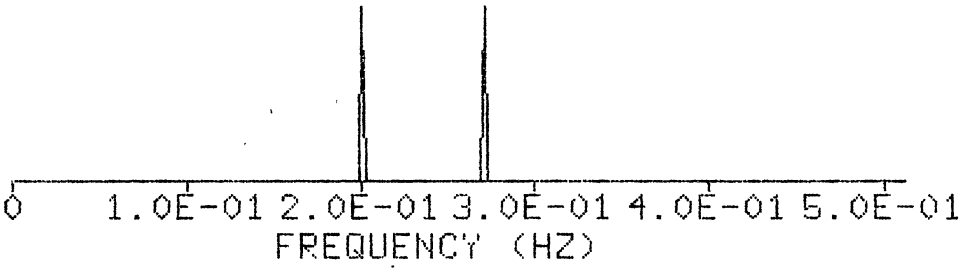

Figure 8. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .2, .27 Hz; No noise)

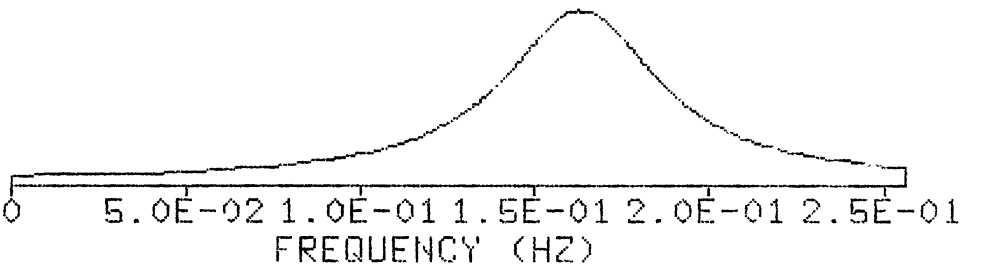

Figure 9. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .18 Hz; Impulsive noise)

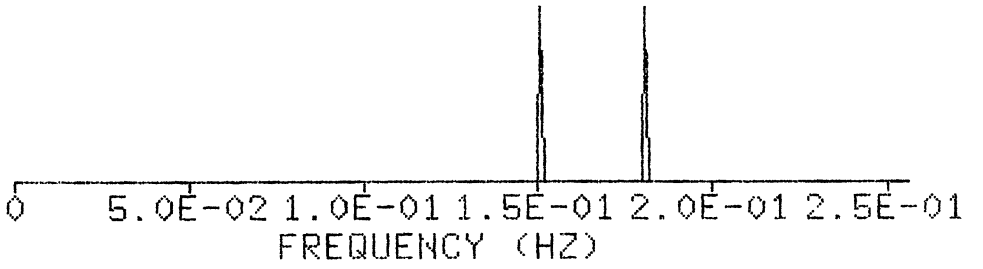

Figure 10. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .18 Hz, Impulsive noise)

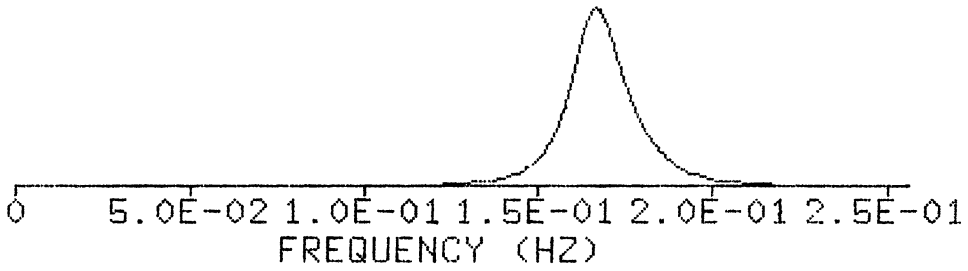

Figure 11. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .185 Hz; 30 dB Gaussian noise)

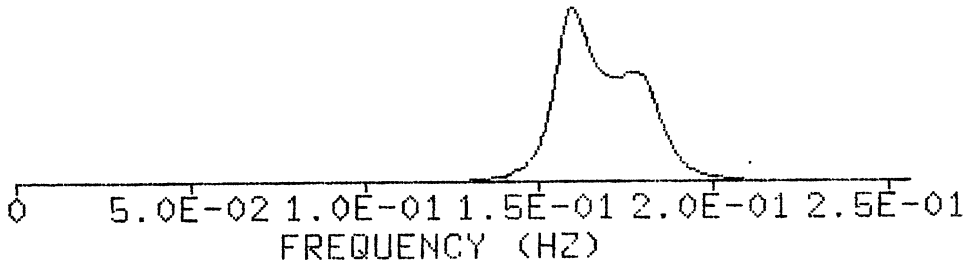

Figure 12. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .185 Hz; 30 dB Gaussian noise)

estimator, while the L_1 estimator is just able to resolve the sinusoids. Figures 13 and 14 depict the results of applying an L_2/L_1 estimator respectively to the test data with uniform noise (SNR = 30 dB) added. Here f_1, f_2 = .15, .187 Hz. Again, the L_2 solution can not resolve the sinusoids while the L_1 estimator shows the possible presence (weakly) of a second sinusoid. The case of two complex sinusoids plus Rayleigh noise proved to be a somewhat more difficult case to handle for both estimators (Figures 15 and 16). The frequency spacing in the example is f_1, f_2 = .15, .20 Hz; at this spacing the L_2 estimator is unable to clearly resolve the two sinusoids, however the L_1 estimator does not indicate conclusively the presence of a second sinusoid. In fact both estimators result in approximately the same performance for this test case. The case of two complex sinusoids plus additive Laplacian noise will be used to illustrate the effect of choosing various p values for the L_p normed solution. For this example f_1, f_2 = .1, .17 Hz and SNR = 30 dB; p = 2.5, 2.0, 1.5, 1.1, 1.0, and .9. Upon viewing, in succession Figures 17-22, it can be seen that as p tends towards p = 1.0 the peaks become increasingly distinct, indicating perhaps that an L_1 norm may be preferred over an L_2 normed solution for data exhibiting Laplacian noise.

It is quite instructive at this point to consider that no noise case; for this test, f_1, f_2 = .1, .11 Hz, and p = 2.0, 1.5, and 1.0. From Figure 23 and Figure 24 it is seen that the L_p normed (p = 1.0, 1.5) solution results in the detection of both sinusoids. The L_2 normed solution, however, is unable to separate the sinusoids, as seen in Figure 25.

This has been stated previously, but it bears repeating; all data

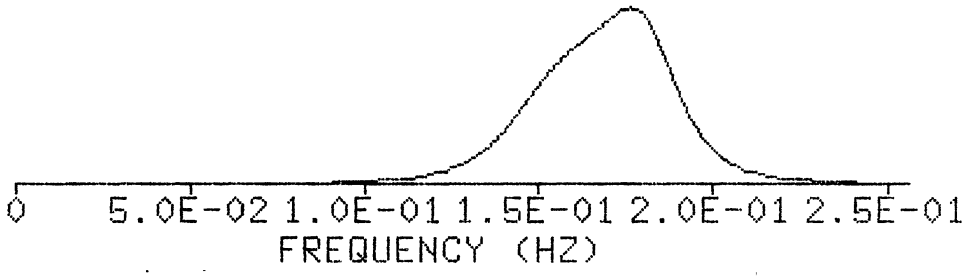

Figure 13. 1-D LP (L_2 Norm) Spectral Estimate (N = 8, f_1, f_2 = .15, .187 Hz; 30 dB Uniform noise)

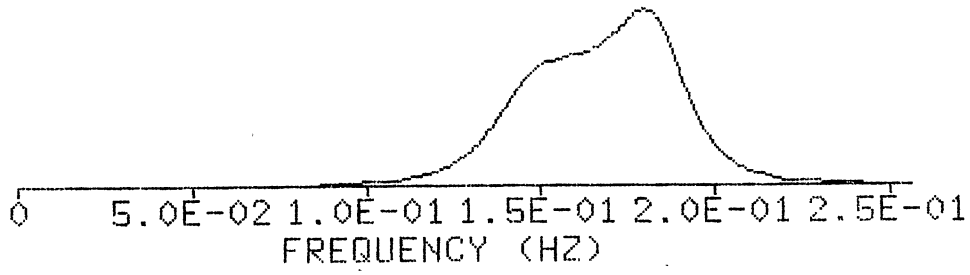

Figure 14. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .187 Hz; 30 dB Uniform noise)

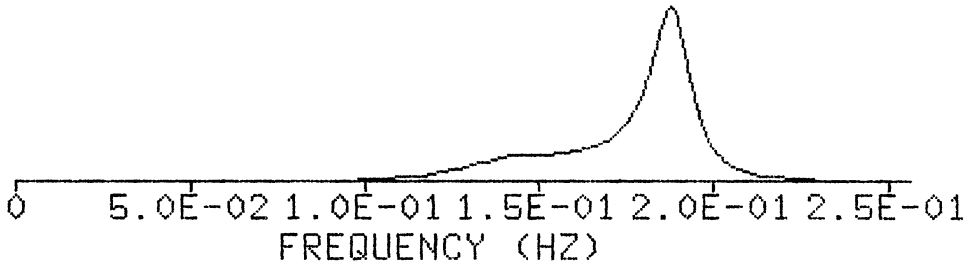

Figure 15. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .2 Hz; 30 dB Rayleigh noise)

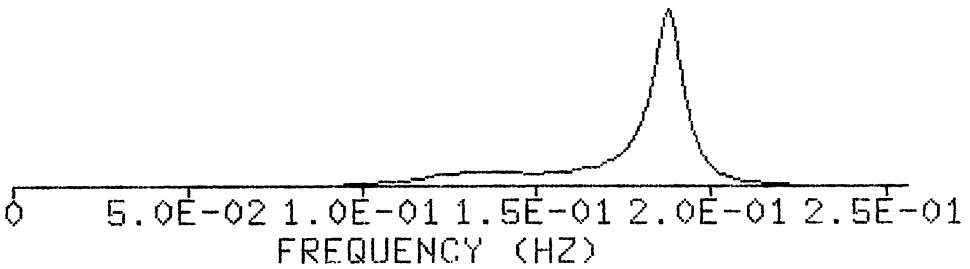

Figure 16. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .15, .2 Hz; 30 dB Rayleigh noise)

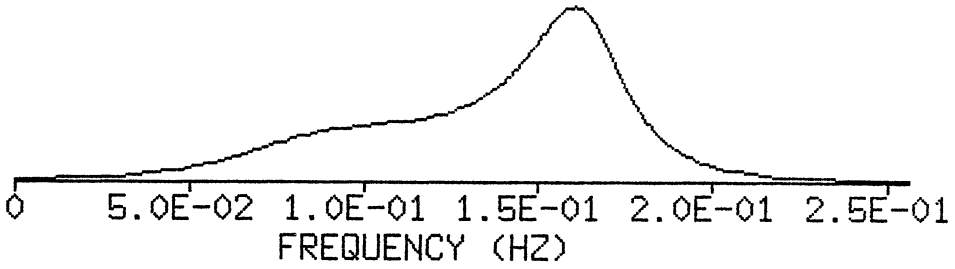

Figure 17. 1-D LP ($L_{2.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

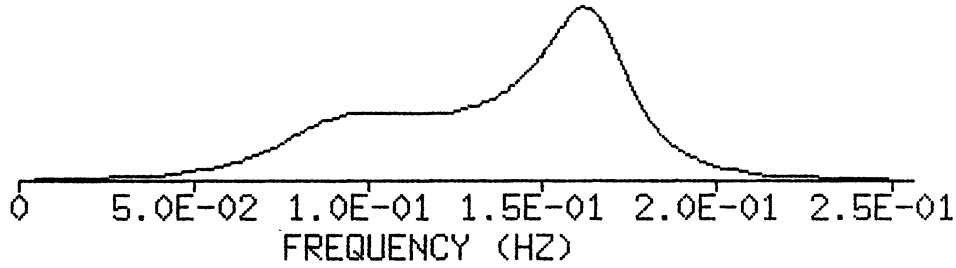

Figure 18. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

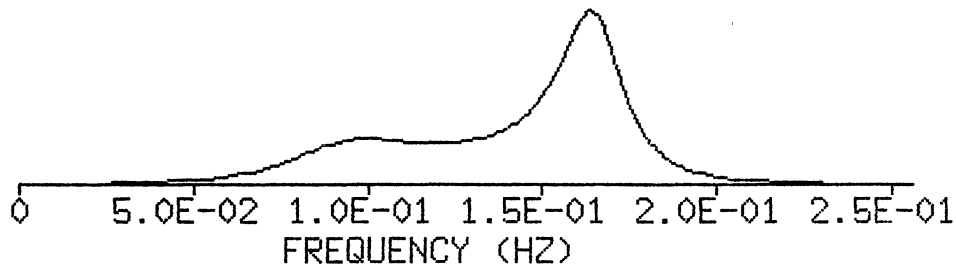

Figure 19. 1-D LP ($L_{1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

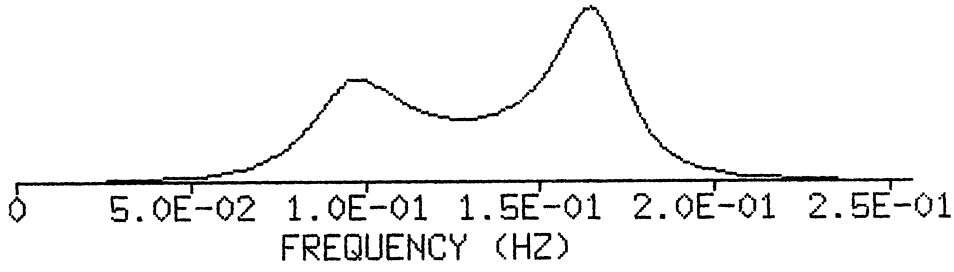

Figure 20. 1-D LP ($L_{1.1}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

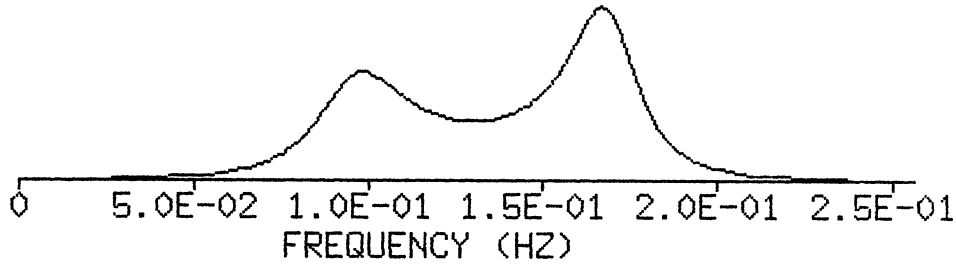

Figure 21. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

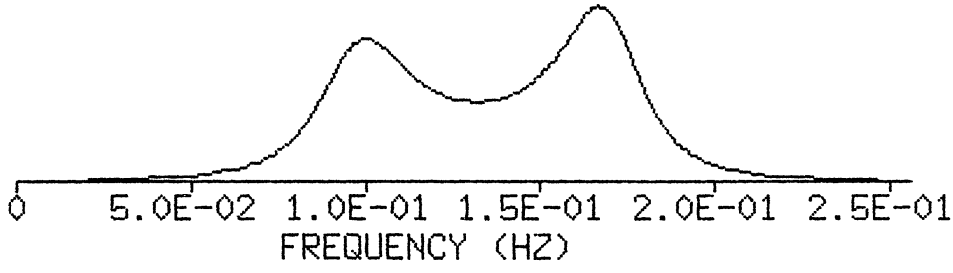

Figure 22. 1-D LP (L_9 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .17 Hz; 30 dB Laplacian noise)

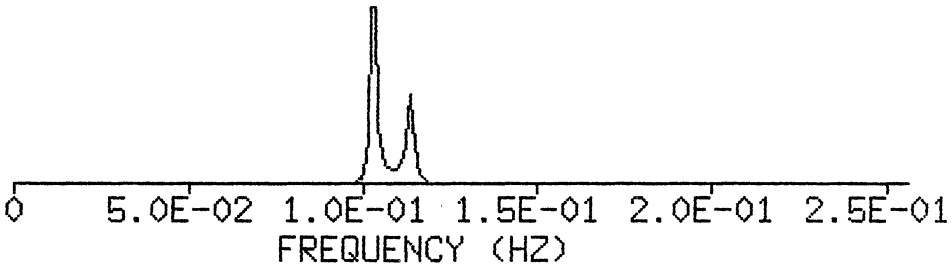

Figure 23. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .11 Hz; No noise)

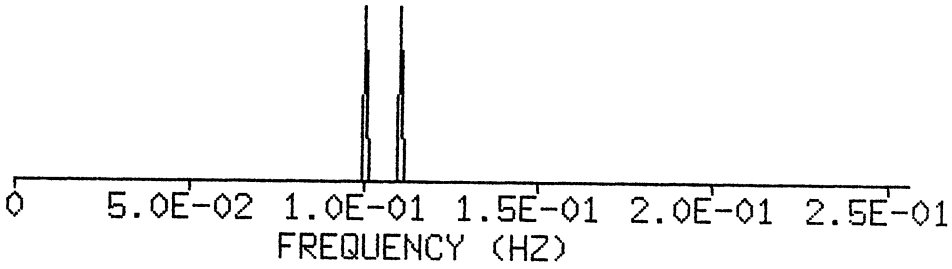

Figure 24. 1-D LP ($L_{1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .11 Hz; No noise)

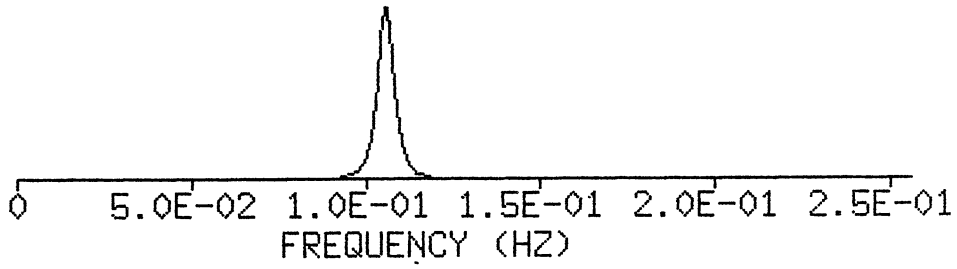

Figure 25. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .11 Hz; No noise)

so far (Figures 1-25) are the result of a single experiment. Obviously, however, the location of f_1 and f_2 is a random variable with unknown mean and variance. Linear predictive spectral estimation is non-linear and analytical expressions for $E\{f_i\}$ and $E\{f_i^2\}$ have not been found. One hopes, naturally, that $E\{f_i\} = f_i$ and the $E\{f_i^2\}$ is not too large. As even asymptatic expressions are not currently available, confidence in the resulting frequency requires repeated experiments to be performed with independent noise sequences. Such repeated experiments, sometimes termed "Monte-Carlo" methods, are not entirely rigorous, however, this exercise will serve to develop some feel for expected outcomes. In any event, there is very little recourse, until further theoretical results are produced.

For this data, f_1, f_2 = .1, .2 Hz, SNR = 30 dB, and the number of independent trials is 100. L_2 and L_1 normed solutions to the prediction equations are generated. The noise types include: Gaussian, Uniform, Rayleigh, Laplacian, and Cauchy. It should be noted that for Cauchy noise, stating SNR = 30 dB has no particular meaning since the second order (and greater) moments for this distribution are undefined. The Cauchy pdf is valid, however, and L_2 normed results may be compared to L_1 normed results for the same relative noise "power". The Gaussian noise case is depicted in Figure 26 and Figure 27; the frequency estimates for both L_1 and L_2 normed solutions are apparently unbiased with a concentrated variance. The L_1 normed result (Figure 27) shows a single "wild point" estimate, but this is probably of little statistical significance. The results of applying the L_2 and L_1 normed solution in the present of uniform noise is shown in Figure 28 and Figure 29. Although both estimators seem unbiased, the L_1 normed solution has a

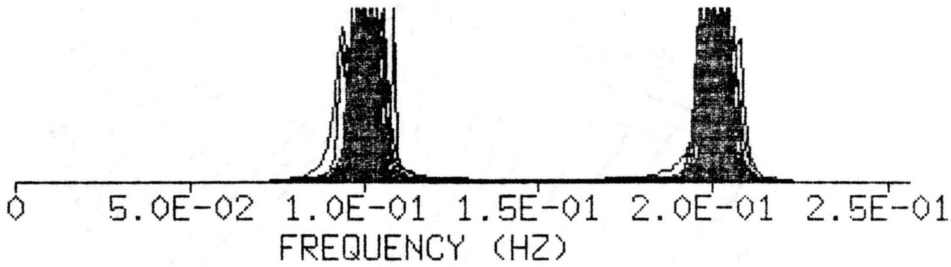

Figure 26. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Gaussian noise; 100 Monte Carlo trials)

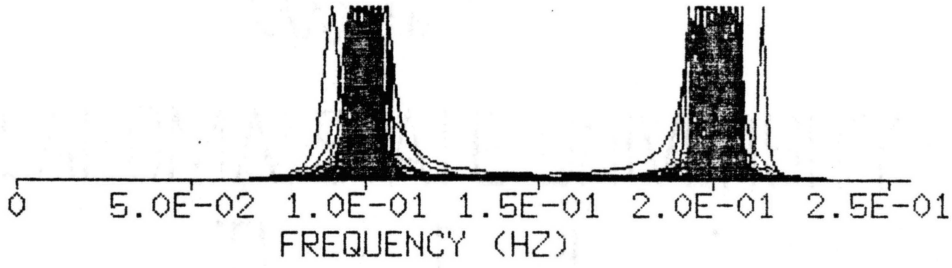

Figure 27. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Gaussian noise; 100 Monte Carlo trials)

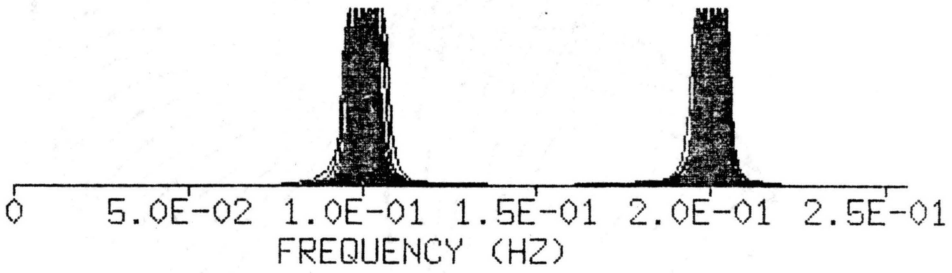

Figure 28. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Uniform noise; 100 Monte Carlo trials)

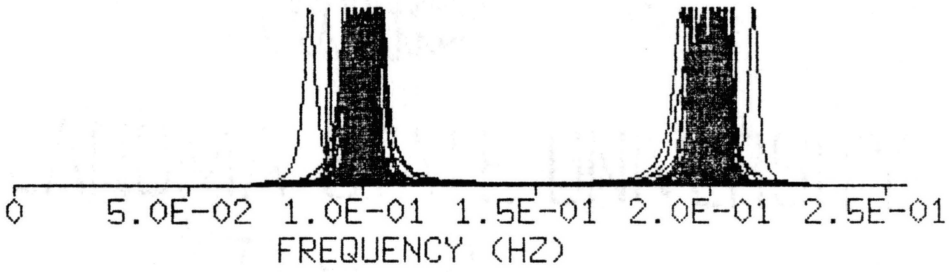

Figure 29. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Uniform noise; 100 Monte Carlo trials)

somewhat greater variance; since the L_∞ norm is preferred for uniformly distributed errors, perhaps this is not too surprising. From Figure 30 and Figure 31 it appears that the addition of Rayleigh noise results in biased frequency estimation for both L_2 and L_1 normed solutions. Also, from Figure 31, it can be seen that invoking an L_1 normed solution results in a slightly greater variance, especially around f_2. It is not known for sure why the estimated sinusoid locations appear biased, however, it may be due to the fact that the Rayleigh distribution is highly skewed with no negative values. It was hoped that in the presence of Laplacian noise an L_1 normed solution would prove superior since choosing an L_1 norm effects a maximum likelihood (ML) estimate in the parameter estimation problem. Indeed, individual examples that demonstrate the increased frequency resolution possibility from an L_1 normed solution are easy to generate as mentioned previously. Comparing Figure 32 (L_2 normed solution) with Figure 33 (L_1 normed solution) it can be seen that both estimators appear to be unbiased. Also, if the graphs are overlaid, the estimated variances are the same. Unfortunately, however, the L_1 normed estimator exhibits several "wild points". At present, this phenomena has no explanation; fortunately, this problem is not severe and represents only about two percent of the trials. As seen from Figure 34 and Figure 35 both L_1 and L_2 normed estimators performed about the same. The estimates of sinusoid location are apparently unbiased. Both solutions result in a higher estimator variance than the previous noise sources; however, that is probably not a fair comparison for reasons previously mentioned.

The next five data plots (Figure 36-40) have been added for reader interest, although the results have no theoretical foundation; more will

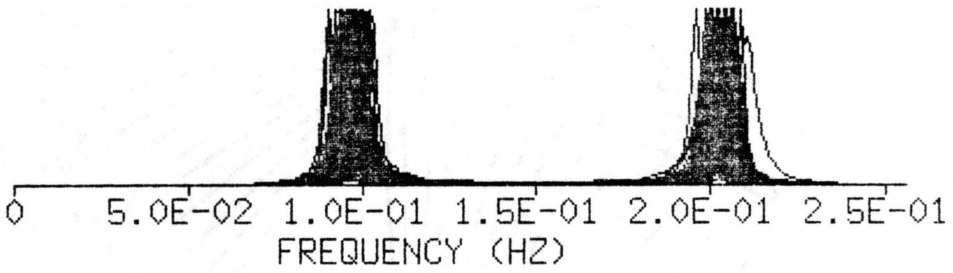

Figure 30. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Rayleigh noise; 100 Monte Carlo trials)

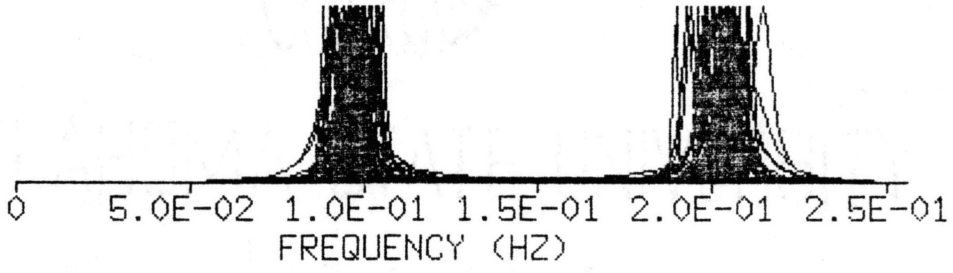

Figure 31. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Rayleigh noise; 100 Monte Carlo trials)

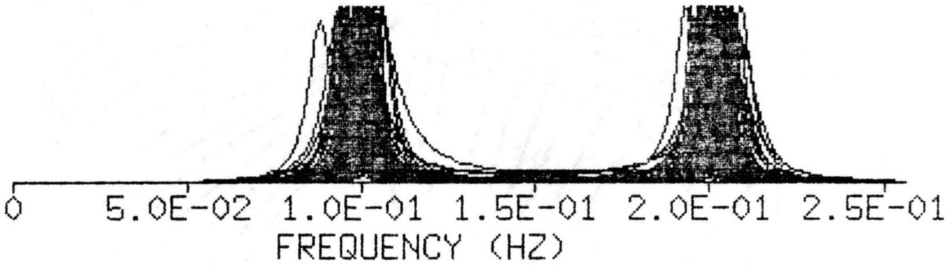

Figure 32. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Laplacian noise; 100 Monte Carlo trials)

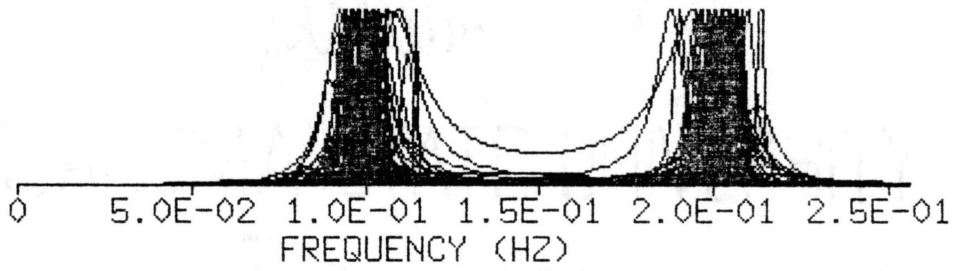

Figure 33. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Laplacian noise; 100 Monte Carlo trials)

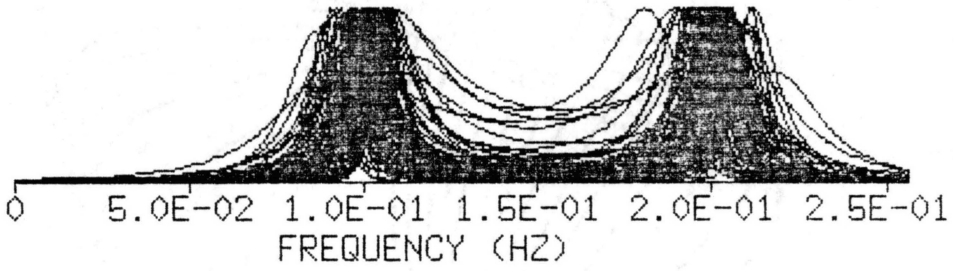

Figure 34. 1-D LP (L_2 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Cauchy noise; 100 Monte Carlo trials)

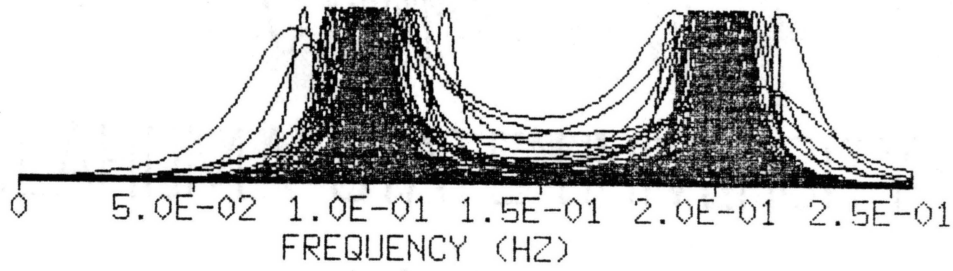

Figure 35. 1-D LP (L_1 Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Cauchy noise; 100 Monte Carlo trials)

be said about valid norm space and convergence properties later. In Figures 36-40 an L_p "norm" (p = -1.5) has been used to generate estimates of the prediction coefficients. The term "norm" has been enclosed in parenthesis since p < 1 is not strictly a normed space because the Holder inequality is violated (Royden, 1968). The RSD algorithm still converges, however, and even though we no longer have a firm theoretical foundation, we may still proceed to generate a spectral estimate with p = -1.5. This has been demonstrated for the five noise sources used previously via Monte Carlo techniques (100 runs). For the cases of Gaussian, Uniform, Rayleigh, and Laplacian noise, (SNR = 30 dB), the spectral estimates of f_1, f_2 = .1, .2 are quite stable (Figure 36-40). There is no apparent frequency bias and the sample variance appears quite small. The Cauchy noise case, however, exhibits significantly greater variance, although the estimate appears unbiased. The reason for the higher sample variance in the presence of Cauchy noise is unclear; as mentioned previously though, second order and greater monments of a Cauchy ditribution are undefined which results in an ambiguous SNR. This data has bee included primarily to demonstrate the robustness of the modified RSD algorithm and to hopefully cause the reader to consider such improper values of p such as p = -1.5. Perhaps further research in this direction may someday lead to fruitful results and even new theoretical developments within the field of measure theory.

An important result of the L_p "normed" (p = -1.5) solution to a set of linear prediction equations is that the RSD algorithm seems to be very robust with such a value of p. This result is not entirely expected since the L_p space (p < 1) is not a normed space (Royden, 1968)

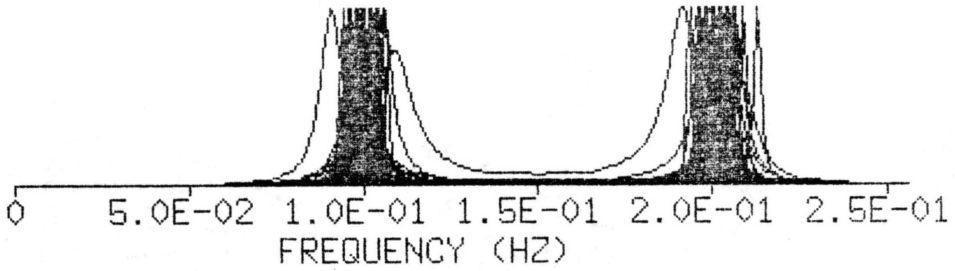

Figure 36. 1-D LP ($L_{-1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Gaussian noise; 100 Monte Carlo trials)

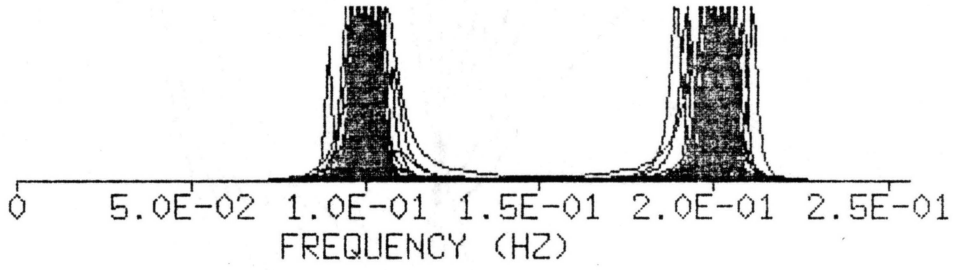

Figure 37. 1-D LP ($L_{-1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Uniform noise; 100 Monte Carlo trials)

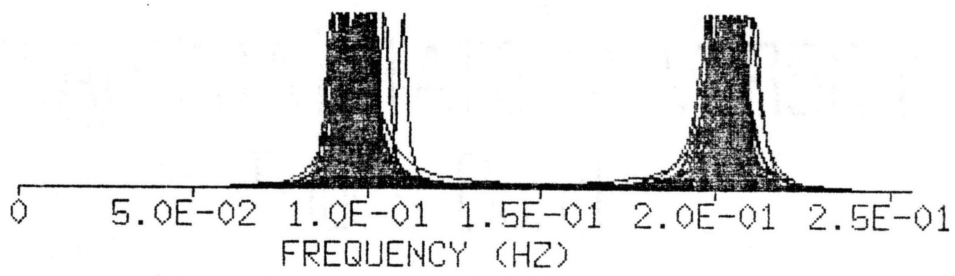

Figure 38. 1-D LP ($L_{-1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Rayleigh noise; 100 Monte Carlo trials)

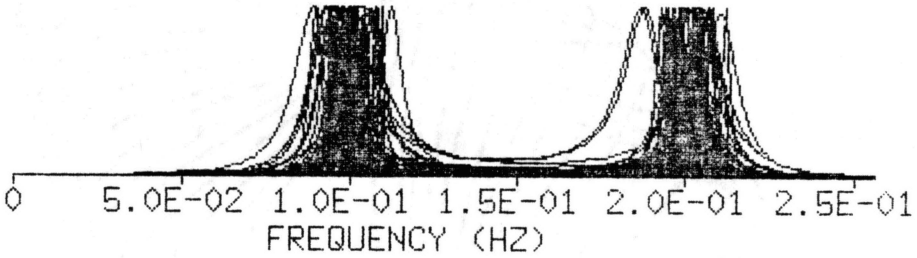

Figure 39. 1-D LP ($L_{-1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Laplacian noise; 100 Monte Carlo trials)

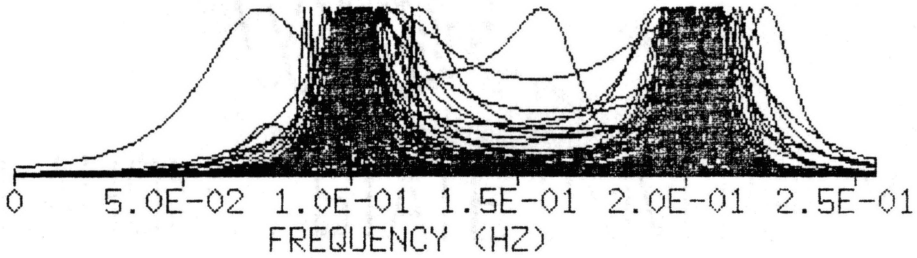

Figure 40. 1-D LP ($L_{-1.5}$ Norm) Spectral Estimate (N = 8; f_1, f_2 = .1, .2 Hz; 30 dB Cauchy noise; 100 Monte Carlo trials)

and any convergence proof (for example Fletcher, Grant, and Hebden, 1971) is no longer theoretically valid. As far as formal convergence proofs are concerned, little theorectical work has been accomplished; most researchers apparently depend upon empirical evidence in order to demonstrate that convergence is guaranteed. One exception is the IRLS algorithm briefly mentioned in an earlier section. This algorithm has been analyzed rather extensively by Byrd and Payne (1979), however the report is not generally available. An L_p norm convergence analysis (continuous functions, although valid for the discrete case) has been published by Fletcher, Grant, and Hebden (1979) that indicates possible divergence for $p \geq 3$. The analysis is quite tedious and as such will not be reproduced here; the material is easily obtainable in the open literature for the interested reader. Their analysis is fairly general and should be applicable for most least squares type iterative schemes, particularly IRLS. Setting $p = 3$ in this work caused divergence as predicted by the convergence analysis just mentioned.

Additionally, Tables 1, 2, and 3 summarize the statistical results of the Monte Carlo simulations (100 runs, SNR = 30 dB) for the five noise cases. The sample mean, sample variance, and 95% confidence limits are computed and tabulated. The confidence interval is computed under the large sample limiting distribution

$$\frac{\bar{x} - \mu}{S/\sqrt{n - 1}} \longrightarrow N(0, 1) \tag{4.7}$$

with

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{4.8}$$

and

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\bar{x})^2 \ . \tag{4.9}$$

Thus Equation (4.7) may be solved for μ and the 95% confidence interval is computed from

$$P_r\{a < \frac{\bar{x} - \mu}{S/\sqrt{n-1}} < b\} = .95 \tag{4.10}$$

From a standard normal table, $a \simeq -2.0$ and $b \simeq +2.0$, which results in

$$\bar{x} - \frac{2S}{n-1} < \mu < \bar{x} + \frac{2S}{n-1} \ . \tag{4.11}$$

From Table 1 (L_2 normed solution) it can be seen that the estimates appear nearly unbiased, except for the Rayleigh noise case. Rayleigh noise is especially difficult to handle since all noise samples are positive, hence a non-zero mean noise vector. In fact, for Rayleigh noise, the 95% confidence interval does not include the known frequency locations of f_1, f_2 = .1, .2 Hz. Some bias is also observed for the case of Laplacian noise, though not as severe as the Rayleigh, however, an explanation is not readily apparent. The Laplacian distribution is symmetric with a zero mean, therefore one expects that the frequency estimates should be reltively unbiased. Perhaps further simulations would be required to clarify this case. An L_1 normed solution to the linear prediction equations also performed quite well. As noted previously, the variance is slightly higher than when an L_2 normed solution is invoked. For Laplacian noise, the L_1 normed solution exhibited slightly less frequency bias than the L_2 normed solution, although the variance is greater, however, this is probably not statistically significant. The 95% confidence interval does include the known sinusoidal frequency locations, however, in the L_1 normed case.

TABLE I

L_2 NORMED SPECTRAL ESTIMATE DATA

Noise Type (30 dB)	f_1	S_{f_1}	95% C.I.	f_2	S_{f_2}	95% C.I.
Gaussian	.0998	.0026	.0993 .1003	.2003	.0024	.1998 .2008
Uniform	.1000	.0021	.0996 .1004	.2000	.0020	.1996 .2004
Rayleigh	.0967	.0024	.0962 .0972	.2023	.0025	.2018 .2028
Laplacian	.0989	.0033	.0982 .0996	.2011	.0031	.2005 .2017
Cauchy	.0994	.0051	.0984 .1004	.2010	.0055	.1999 .2021

TABLE II

L_1 NORMED SPECTRAL ESTIMATE DATA

Noise Type (30 dB)	f_1	S_{f_1}	95% C.I.	f_2	S_{f_2}	95% C.I.
Gaussian	.0997	.0032	.0991 .1003	.2003	.0036	.1996 .2010
Uniform	.0997	.0036	.0990 .1004	.1998	.0034	.1991 .2005
Rayleigh	.0970	.0035	.0963 .0977	.2024	.0041	.2016 .2032
Laplacian	.0994	.0047	.0985 .1003	.2008	.0045	.1999 .2017
Cauchy	.0986	.0075	.0971 .1001	.2007	.0083	.1990 .2024

TABLE III

$L_{-1.5}$ NORMED SPECTRAL ESTIMATE DATA

Noise Type (30 dB)	f_1	S_{f_1}	95% C.I.	f_2	S_{f_2}	95% C.I.
Gaussian	.0998	.0032	.0992 .1004	.2004	.0037	.1997 .2011
Uniform	.0999	.0033	.0992 .1006	.2001	.0035	.1994 .2008
Rayleigh	.0971	.0035	.0964 .0978	.2023	.0032	.2017 .2029
Laplacian	.0992	.0047	.0983 .1001	.2008	.0047	.1999 .2017
Cauchy	.0999	.0089	.0981 .1017	.2002	.0090	.1984 .2020

Again, as was mentioned earlier, the results for the L_p ($p = -1.5$) "norm" have no theoretically foundation. The fact that the RSD and IRLS algorithms converged, however, may eventually prove to be quite significant. In any event, the statistical data is summarized in Table 3 for the interested reader. In fact, for whatever reason, the $L_{-1.5}$ "normed" solution performed better than the L_1 or L_2 normed solution for the case of Cauchy noise. At this time no explanation is available for this phenomena.

From the above we conjecture that there exists a space called a "pseudo normed space," which we define as a space where-in L_p algorithms converge. Further research in this would be very fruitful.

Conclusions and Future Research

A computational simple algorithm (RSD) that allows solutions with a L_p norm ($1 \leq p \leq 1$) to be generated has been applied to a set of overdetermined linear prediction equations. For the case of impulsive noise the L_1 solution has been shown to be clearly superior (with frequency resolution as a comparison criteria) to the L_2 solution. In the presence of Gaussian noise, the L_1 solution in this case performed slightly better than the L_2 solution, however the Monte-Carlo simulations did not prove this out. The uniform and Laplacian noise cases show an L_1 solution to possibly offer a minor increase in resolution capabilities; the Rayleigh noise case is inconclusive. Again, it must be stated that this data is a result of a single experiment. As pointed out in the data discussions, although, single experiments may easily be found that show the L_1 norm to be superior (in two sinusoid resolution ability), we simply can not generalize from the

specific. The Monte-Carlo simulations bear this point out (albeit non-rigorously) as was demonstrated. A problem with Monte-Carlo type experiments is that one never really knows when enough trials have been run. For example, the two sinusoids in Laplacian noise case under an L_1 normed spectral estimator exhibited a single "wild point", over the course of 100 runs. Upon running the exact same experiment for 500 trials, however, results in still the same single "wild point". And that is the dilemma: is it fair to ignore that single case or not.

Since in general the RSD algorithm can be utilized to obtain a solution with the L_p norm criteria, this method is not restricted to studying the L_1 norm. Additional work in applying the RSD algorithm to the linear predictive spectral estimation method has considered the more general L_p problem, $1 \leq p \leq 2$. Possibly, intermediate values of p, say $p = 1.2$, will blend the characteristics of the L_1 and L_2 norm in some advantageous way. Additionally, $p = 1$ may produce an unstable prediction error filter in some cases. A somewhat higher value of p may result in a stable filter and still retain the desired characteristics of an L_1 normed solution. So far, with the limited experimentation performed to date, it appears that an L_p (p "near" 1.0) normed solution possesses characteristics similar to an L_1 normed solution.

One final comment is in order in this conclusion section concerning the performance of the L_1 normed spectral estimator against two sinusoids plus Laplacian noise. As was pointed out in the literature survey section, the L_1 normed solution can be shown to be the maximum likelihood parameter estimator in the presence of Laplacian noise. This being the case it is natural to expect that the calculated linear prediction coefficients (under an L_1 norm) will be a more accurate

parametric representation of the underlying process. A better model characterization may possibly lead to improved spectral resolution. This was not observed, however, and a quick comment is probably in order. In this simulation the data length consisted of just eight points. Such a short length is obviously not going to result in accurate noise statistics, however, increasing the data length negates the reason for justifying the added complexity of parametric spectral estimators in the first place!

CHAPTER V

Introduction

As pointed out previously, two-dimensional spectral estimation has important applications in many fields such as geophysics, radar, sonar, and radio astronomy. If the frequencies are spaced sufficiently far apart and if enough data are available, 2-D discrete Fourier transform (DFT) techniques are adequate for resolving the sinusoids. However, if the spacing of the sinusoids is within the resolution limits of a DFT, as is often the case with a limited data record length, other methods must be used to extract frequency and wavenumber locations.

A number of modern spectral estimation techniques that have been applied successfully to the one-dimensional problem have been extended to multiple dimensions; these include the maximum-likelihood method (MLM), the maximum-entropy method (MEM), and the auto-regressive method (AR). In contrast to the one-dimensional problem, multiple-dimensional MEM and AR methods are different. A summary of current research efforts in multi-dimensional spectral estimation can be found in the comprehensive survey by McClellan (1982). A simultaneous frequency and wavenumber estimation technique introduced by Kumaresan and Tufts (1981) based upon 2-D linear prediction is especially attractive from a computational point of view; their method forms the prediction error filter via a minimum norm solution.

As pointed out earlier, L_2 solutions are widely used and offer computational and theoretical advantages, however it is well known that for certain noise types, for example impulsive noise, the L_2 solution is not optimal. In cases where impulsive noise is present an L_1 (i.e. absolute value) error criteria is better suited as a minimization criteria. Due to the increased computational complexity of computing L_1 solutions this error criteria has not been extensively researched, especially with respect to multi-dimensional spectral estimation. One application of the L_1 norm, however, was presented by Levy et al (1982) that utilized the simplex algorithm to generate the required solution. The residual steepest descent (RSD) algorithm (Huber, 1981) can be used to generate L_p solutions (generally $1 \leq p \leq 2$), however, convergence has been a problem. As noted in Chapter IV, a modification of the RSD algorithm (Yarlagadda et al., 1985) has eliminated the convergence problem and a viable method is now available for generating L_p norm solutions.

The remainder of this chapter presents an L_1 solution of the 2-D linear prediction method for frequency-wavenumber estimation (formulated in Kumaresan and Tufts, 1981) based upon the RSD algorithm (Yarlagadda et al., 1985). The next section of this chapter will briefly summarize the 2-D linear prediction method given for the case of two 2-D complex sinusoids in noise and the following section will develop the problem reformulation necessary to utilize the RSD algorithm. Simulation results will be presented that show the improved resolution possible with an L_1 solution. As an alternative to the 2-D LP formulation, a hybrid technique that blends a DFT along one dimension and LP methods (L_1 and L_2 norm) along the remaining dimension. This procedure offers

the computational advantages of the DFT (assuming sufficient data are available in the dimension for the required resolution) plus the high resolution capabilities in the dimension with limited record length data.

2-D Linear Prediction Equations

In Kumaresan and Tufts (1980) a computationally efficient method of extracting complex sinusoids in noise is developed based upon a minimum norm solution of an underdetermined set of linear prediction equations. A key feature of this technique is that the maximum size matrix inverse required is equal to the number of complex sinusoids assumed present within the data set; thus a considerable computational savings can be achieved since the number of sinusoids present is typically rather small. Therefore, with two sinusoids in noise as a test example a 2x2 inverse is required. A straight-forward extension of this method from one-dimension to two-dimensions is given in Kumaresan and Tufts (1981).

The relevant equations comprising the 2-D Linear Prediction formulation will now be given. Note that two prediction error filters are formed, each based upon a forward prediction equation and a backward prediction equation across opposite corners of the data array (valid for detecting two complex sinusoids only). Thus, each corner of the data array is utilized as a predicted sample with the remaining data used as support for the prediction filter (see Figure 41). The two prediction filters are then combined in a circular symmetric manner in order to form the frequency-wavenumber estimate.

The test data array is generated from:

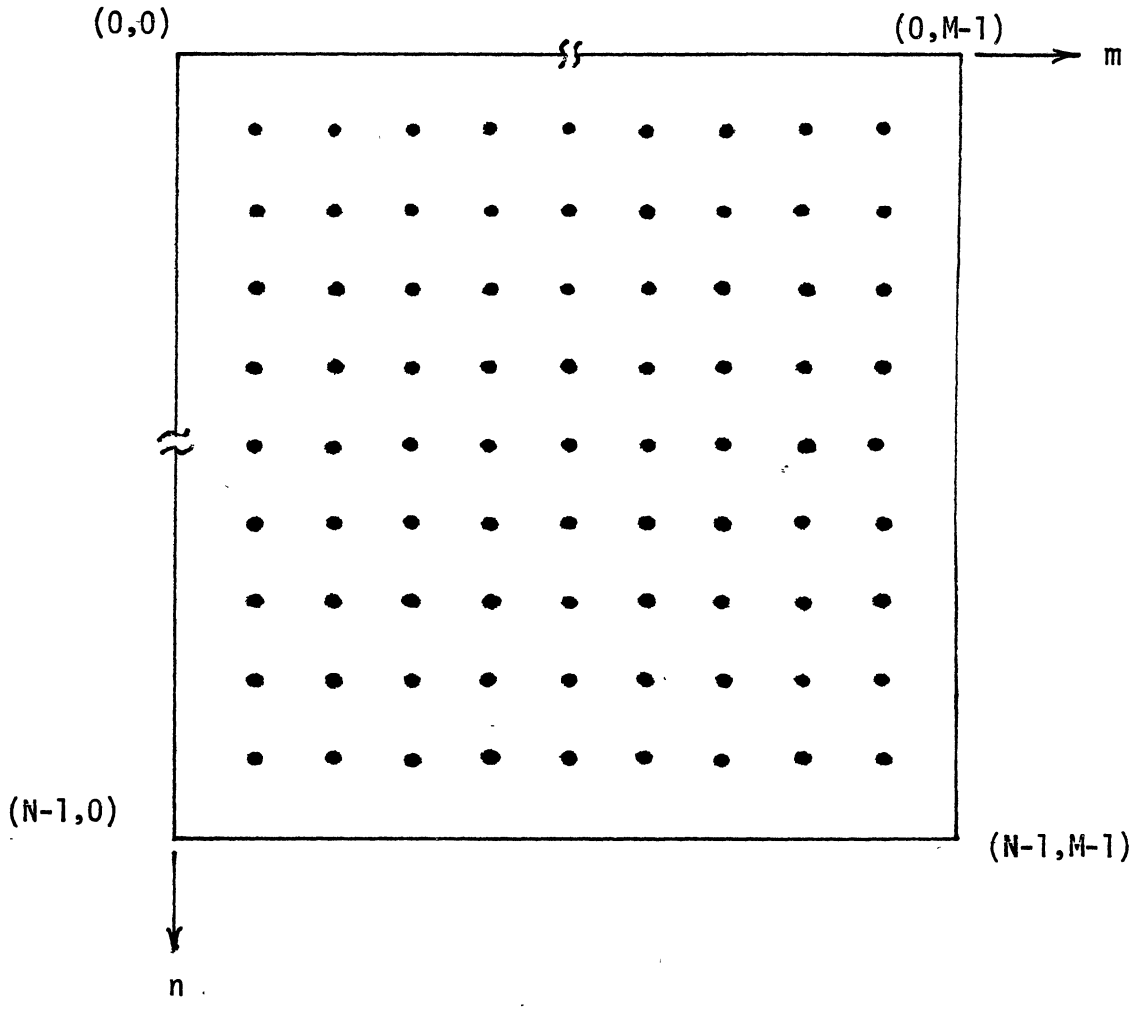

Figure 41. Data support for prediction filters H_1 and H_2

$$y(n,m) = A_1 \exp(j(\omega_1 n + \omega_2 m)) + A_2 \exp(j(\omega_3 n + \omega_4 m)) + W(n,m) \qquad (5.1)$$

where A_1 and A_2 are unknown complex numbers. In this work, both have been set to one for simplicity. W(n,m) are independent complex noise samples. With n the space index and m the time index, ω_1 and ω_3 are unknown wavenumbers and ω_2 and ω_4 are the unknown frequencies. N and M are the total number of spatial and time samples respectively.

The prediction equations are given by:

H_1 filter: $Y_1 \underline{a} = \underline{y_1}$

$$\begin{bmatrix} y(N-1, M-2) & y(N-1, M-3) & \cdots & y(0,0) \\ y^*(0,1) & y^*(0,2) & \cdots & y^*(N-1,M-1) \end{bmatrix} \underline{a} = \begin{bmatrix} y(N-1, M-1) \\ y^*(0,0) \end{bmatrix} \qquad (5.2)$$

H_2 filter: $Y_2 \underline{b} = \underline{y_2}$

$$\begin{bmatrix} y(0, M-2) & y(0,M-3) & \cdots & y(N-1,0) \\ y^*(N-1,1) & y^*(N-1,2) & \cdots & y^*(0,M-1) \end{bmatrix} \underline{b} = \begin{bmatrix} y(0, M-1) \\ y^*(N-1,0) \end{bmatrix} \qquad (5.3)$$

The prediction coefficients for the H_1 filter are given by:

$$\underline{a} = [a_{01}, a_{02}, \ldots, a_{N-1,M-1}]^T \qquad (5.4)$$

and the prediction coefficients for the H_2 filter are given by:

$$\underline{b} = [b_{01}, b_{02}, \ldots, b_{N-1,M-1}]^T . \qquad (5.5)$$

The asterisk (*) denotes complex conjugation.

The prediction error filters are given by:

$$H_1 = 1 - \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} a_{ik} z_1^{-i} z_2^{-k} \qquad (5.6)$$

$$H_2 = 1 - \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} b_{ik}\, z_1^{+i}\, z_2^{-k} \qquad (5.7)$$

with $i, k \neq 0$. It has been shown (Kumaresan, 1982) that for noiseless data $H_1(e^{j\omega_1}, e^{j\omega_2})$ and $H_2(e^{j\omega_3}, e^{j\omega_4})$ are equal to zero with $\underline{a}$ and $\underline{b}$ calculated as the minimum norm solution as follows (here * is complex conjugate transpose):

$$\underline{a} = Y_1^*[Y_1\ Y_1^*]^{-1}\underline{y_1} \qquad (5.8)$$

$$\underline{b} = Y_2^*[Y_2\ Y_2^*]^{-1}\underline{y_2}. \qquad (5.9)$$

The simultaneous frequency-wavenumber estimate is then calculated from:

$$S(z_1,z_2) = \frac{1}{|H_1(z_1,z_2)|^2 + |H_2(z_1,z_2)|^2} \qquad (5.10)$$

L_1 Solution of 2-D LP Equations

It is desired to reformulate the 2-D linear prediction equations in order that the RSD algorithm may be used to generate an L_1 solution. This is necessary since the RSD method requires an over-determined system of equations, while the 2-D linear prediction equations are under-determined. Alternatively the simplex algorithm could be used to solve the set of under-determined prediction equations directly, however the goal here is to make use of the rapidly converging RSD algorithm. Additionally, use of the RSD algorithm will allow investigation of the general L_p norm with $1 \leq p \leq 2$. For example, choosing $L_{1.2}$ may result in a blending of the characteristics of an L_1 norm and the L_2 norm.

Consider the set of under-determined equations:

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_Q \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} , \tag{5.11}$$

or more compactly $Y\ \underline{a} = \underline{d}$. For obvious reasons we assume that the matrix Y has a full rank; i.e., the rank here is 2. In linear prediction terminology the vector $\underline{a}$ is the array of prediction coefficients, d_1 and d_2 are the predicted samples, and matrix Y contains the data samples used to estimate d_1 and d_2. Equation (5.11) may be written:

$$\begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} y_{13} & y_{14} & \cdots & y_{1Q} \\ y_{23} & y_{24} & \cdots & y_{2Q} \end{bmatrix} \begin{bmatrix} a_3 \\ a_4 \\ \vdots \\ a_Q \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} . \tag{5.12a}$$

which can be written in symbolic form

$$Y_1\ \underline{\alpha_1} + Y_2\ \underline{\alpha_2} = \underline{d} \tag{5.12b}$$

where

$$\underline{\alpha_1}^T = [a_1, a_2],\ \underline{\alpha_2}^T = [a_3, a_4, \ldots, a_Q],\ \underline{d}^T = [d_1, d_2] \tag{5.12c}$$

and the matrices Y_1 and Y_2 are implicitly defined by comparing (5.12a) with (5.12b). For the following discussion we will assume without losing any generality that Y_1 is nonsingular. If Y_1 is a singular matrix, then we need to find two columns in Y that are independent and rewrite (5.12a), where now Y_1 will have the new columns. For two sinusoids plus no noise the occurrence of singularity is predictable (see Appendix A).

Solving (5.12b), we have

$$\underline{\alpha_1} = -Y_1^{-1} Y_2 \underline{\alpha_2} + Y_1^{-1} \underline{d} \tag{5.13}$$

Equation (13) can be augmented and is

$$\begin{bmatrix} \underline{\alpha_1} \\ \underline{\alpha_2} \end{bmatrix} = \begin{bmatrix} -Y_1^{-1}Y_2 \\ I \end{bmatrix} \underline{\alpha_2} + \begin{bmatrix} Y_1^{-1}\underline{d} \\ 0 \end{bmatrix} \tag{5.14}$$

where I is an identity matrix of order (Q-2). It is clear that minimization of the vector $\underline{\alpha}^T = [\underline{\alpha_1}^T \ \underline{\alpha_2}^T]$ simply corresponds to solving the overdetermined system of Q equations in (Q-2) unkowns:

$$\begin{bmatrix} -Y_1^{-1} Y_2 \\ I_{Q-2} \end{bmatrix} \underline{\alpha_2} = \begin{bmatrix} -Y_1^{-1} \underline{d} \\ 0 \end{bmatrix} . \tag{5.15a}$$

For simplicity, we will write this in the symbolic form

$$X \underline{b} = \underline{c} \tag{5.15b}$$

where $\underline{b} = \underline{\alpha_2}$.

Equation (5.15b) has the general form of an overdetermined system of equations and L_p analysis methods can be applied to solve for $\underline{\alpha_2}$.

Before considering the solutions, we want to point out that the coefficient matrix X in (5.15b) has a special form, i.e., the last (Q-2) rows corresponds to an identity matrix. This allows an efficient implementation for L_p solutions. An additional problem is that the entries in X and $\underline{c}$ in (5.15b) are over the complex field. L_p solutions are computed using the RSD algorithm presented in Yarlagadda et al. (1985). This algorithm is modified to fit the complex case at hand and the steps are summarized below. Although this algorithm is given in Chapter IV, it is repeated here for the reader's convenience.

Complex RSD Algorithm

1. $\underline{b}(0) = (X^*X)^{-1}X^* \underline{c}$
2. $q_i(k) = \text{sgn}\ (|X\ \underline{b}(k)| - |\underline{c}|)_i \cdot |(X\ \underline{b}(k) - \underline{c})_i|^{(p-1)}$
3. $\underline{r}(k) = X\ \underline{b}(k) - \underline{c}$
4. Minimize with respect to Δ_k

   $E(k) = \underline{r}(k) - \Delta_k\ X(X^*X)^{-1}X^*\ \underline{\gamma}(k)$
5. $\underline{b}(k+1) = \underline{b}(k) - \Delta_k(X^*X)^{-1}X^*\ \underline{\gamma}(k)$
6. Go to step 2 or stop if convergence is achieved.

It is clear from these steps that the inverse of (X^*X) is required. Since the matrix has a special structure, a simple procedure can be used to invert this matrix. The matrix identity used in the simulation, termed the matrix inversion lemma, can be found in Brogan (1974) and is given here for easy reference:

$$[P^{-1} + H^TQH]^{-1} = P - PH^T\ [HPH^T + Q]^{-1}HP. \tag{5.16}$$

With the matrix X in (5.15b) partitioned as

$$X = \begin{bmatrix} X_1 \\ I_{Q-2} \end{bmatrix} \tag{5.17}$$

$(X^*X)^{-1}$ can be expressed as

$$(X^*X)^{-1} = [X_1^* \vdots I_{Q-2}] \begin{bmatrix} X_1 \\ I_{Q-2} \end{bmatrix}^{-1} \tag{5.18a}$$

or

$$(X^*X)^{-1} = [X_1^*X_1 + I_{Q-2}]^{-1}. \tag{5.18b}$$

The identity given by (16) may now be applied to the right side of (5.18b) by choosing $P = I_{m-2}$, $Q = I_2$, and $H = X_1$. Thus, by application

of (5.16), we now have

$$[X_1^* X_1 + I_{Q-2}]^{-1} = I_{Q-2} - X_1^*(X_1 X_1^* + I_2)^{-1} X_1. \qquad (5.19)$$

Note that the large inverse on the left side of (5.19) is replaced by a 2x2 inverse on the right side of (5.19).

The minimzation in step 4 involves only vectors and as such is not difficult. In this work the iterative reweighted least squares (IRLS) algorithm was used to solve for Δ_k; IRLS is discussed in considerable detail in Yarlagadda et al. (1985).

It has been found that the RSD algorithm converges within two or three iterations making the computational aspect quite efficient for a solution of this complexity. The minimization of the Δ_k factor using IRLS techniques has typically required six or seven iterations, however, since only vectors are involved the computations required are minimal. Again, note that just a 2x2 invese is required by the above implemention.

In the following section simulation results will be presented to verify that this L_1 formulation allows the detection of two complex sinusoids in noise. Additionally, in the presence of impulsive noise, the L_1 solution will be shown to offer increased resolution over that obtainable via the minimum norm solution.

Simulation Results

A 16x16 data array was chosen as a test example to realistically represent a rather short data record length; thus N = M = 16. For simplicity the constants A_1 and A_2 in (5.1) are set to one (A_1=A_2=1). In order to highlight the advantages of an L_1 solution as compared to an

L_2 solution, the noise is chosen to be impulsive; specifically, two complex impulses are added to the data at n = m = 8 and n = m = 12 with the amplitude of both real and imaginary components equal to 2.8 approximately. The impulse amplitude has a negligible effect on the L_1 results, however the L_2 solution is very sensitive to the impulse amplitude and a value of 2.8 was chosen to allow reasonable results from L_2. The L_2 solution can easily be forced to fail completely by arbitrarily increasing the amplitude of the noise spikes (see Figures 42 and 43), however the intent here is to compare the relative spectral resolution properties of L_1 vs. L_2. A unit sampling period was chosen for convenience which results in a frequency range of 0 to 0.5 Hz (fs = 1 Hz) and a wavenumber range of 0 to 0.5 cycles/unit sampling distance. Four sets of complex sinusoidal frequencies that have been used to demonstrate the relative resolution capabilities of L_1 vs. L_2:

Case 1. $\omega_1 = (2\pi)(.125)$, $\omega_2 = (2\pi)(.2)$,
$\omega_3 = (2\pi)(.125)$, $\omega_4 = (2\pi)(.23)$

Case 2. $\omega_1 = (2\ )(.125)$, $\omega_2 = (2\pi)(.2)$,
$\omega_3 = (2\ )(.125)$, $\omega_4 = (2\pi)(.22)$

Case 3. $\omega_1 = (2\pi)(.125)$, $\omega_2 = (2\pi)(.2)$,
$\omega_3 = (2\pi)(.125)$, $\omega_4 = (2\pi)(.21)$

Case 4. $\omega_1 = (2\pi)(.125)$, $\omega_2 = (2\pi)(.2)$,
$\omega_3 = (2\pi)(.125)$, $\omega_4 = (2\pi)(.205)$

Note that the frequency/wavenumber spacings in all cases exceed the resolution limit available via discrete Fourier transform methods (Δf = 1/16=.0625). With N=M=16 the prediction errors filters, H_1 and H_2, have

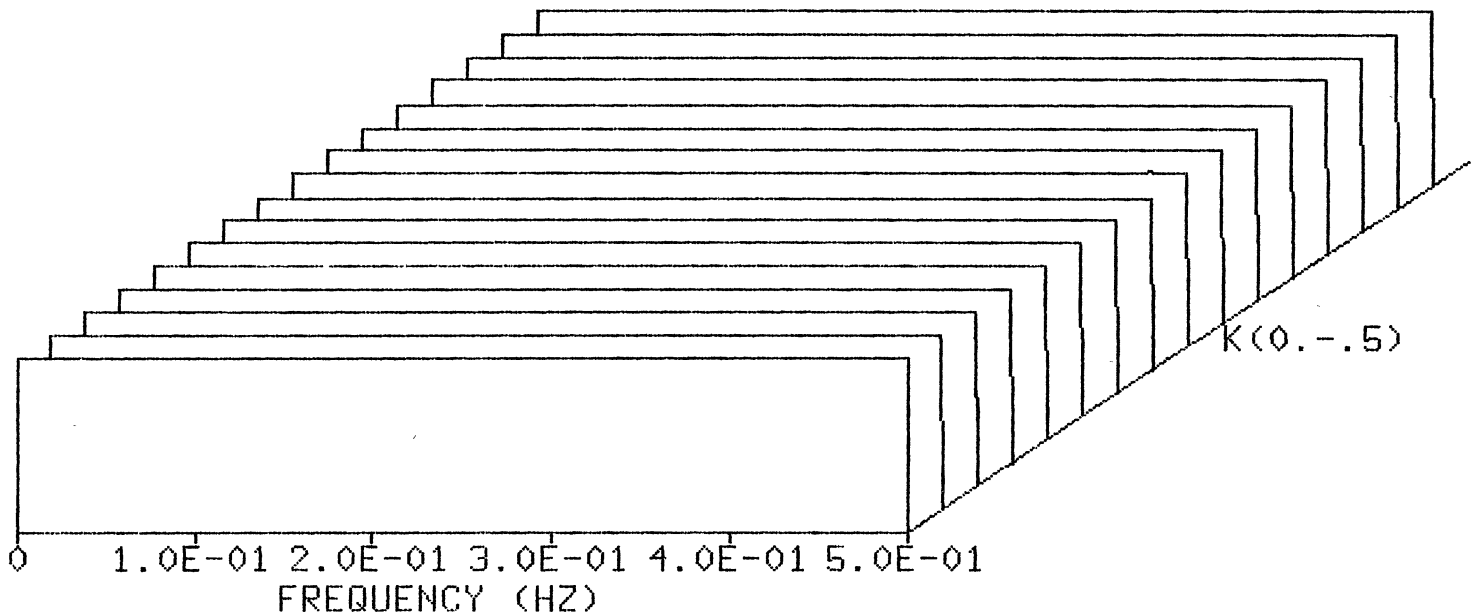

Figure 42. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; Impulsive noise)

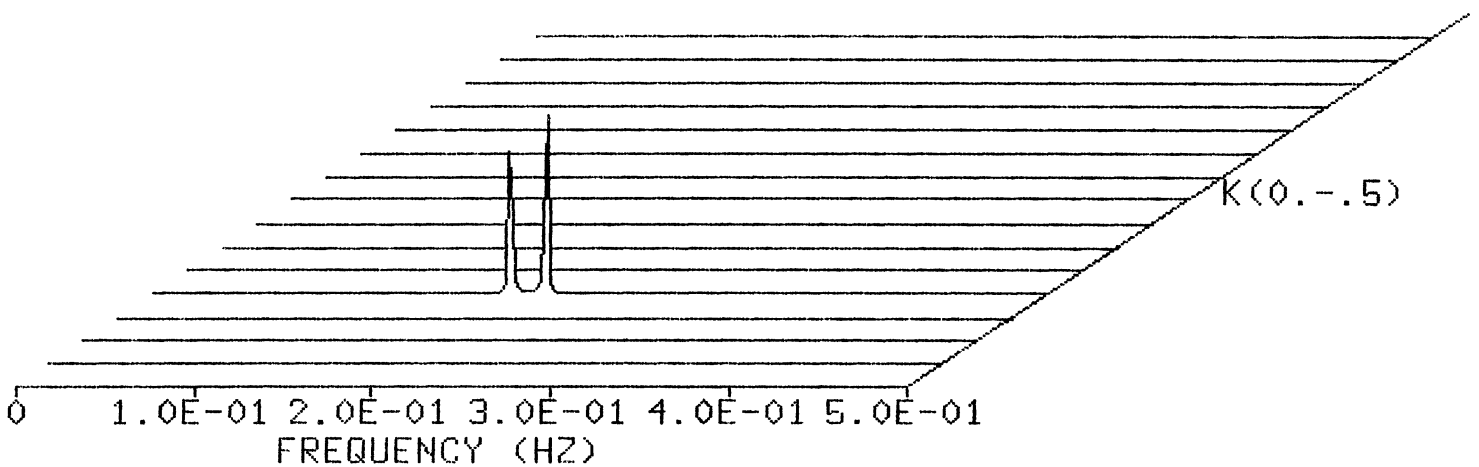

Figure 43. 2-D LP (L_1 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; Impulsive noise)

255 coefficients (NM-1); H_1 and H_2 are evaluated from (5.6) and (5.7) via a 2-D DFT. Equation (5.10) is used to form the final simultaneous frequency-wavenumber estimate.

As was done in the previous chapter, prior to presenting data demonstrating the potential advantages of L_1 normed solutions versus L_2 normed solutions, a comparison of the DFT method versus an LP technique will be presented. With this limited size data array (16 x 16) the 2-D DFT spectral estimator will be seen to perform rather poorly (excessive sidelobes and low frequency resolution) relative to an LP 2-D spectral estimator. With widely spaced sinusoids (f_1 = .125 cycles/foot, f_2 = .1875 Hz; f_3 = .25 cycles/foot, f_4 = .3125 Hz) plus 0 dB Gaussian noise a 2-D DFT technique exhibits high sidelobe amplitudes, although the two peaks may be seen without too much difficulty (Figure 44; of course the locations are known a priori). In Figure 45, for the same noise and frequency spacings, the results of applying a 2-D LP estimator is given; the improvement is obvious and needs no further comment. This experiment is repeated with the same frequency spacing, however, now the noise (Gaussian) has been reduced so that the signal to noise ratio is increased to 10 dB. From Figures 46 (DFT method) and 47 (LP method) it can be seen that high sidelobes are present from the DFT; the LP spectral estimator, however, results in sharply defined peaks with such a relatively high signal to noise ratio. Next, the relative frequency resolution cpabilities of the 2-D DFT method of the LP estimator will be compared, first with 10 dB signal to noise ratio (Gaussian noise), then with the signal to noise ratio increased to 30 dB; the frequencies are: f_1 = .125 cycles/foot, f_2 = .2 Hz, and f_3 = .125 cycles/foot, f_4 = .22 Hz. For the 10dB SNR case, the 2-D DFT is unable to resolve the

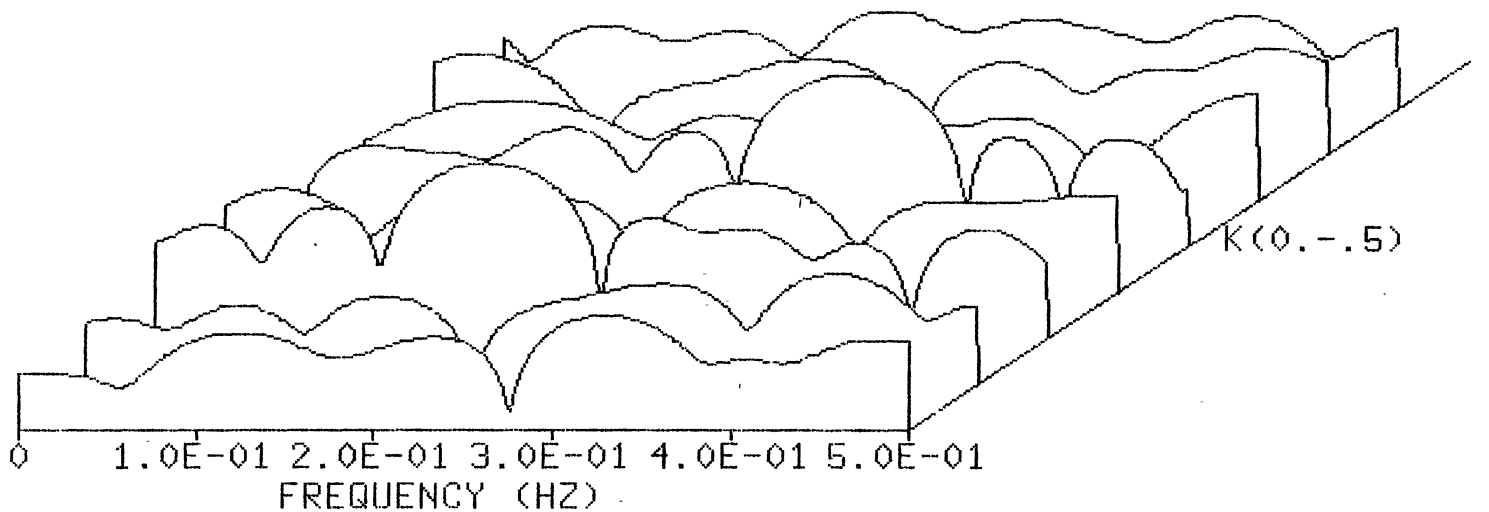

Figure 44. 2-D DFT Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 0. dB Gaussian noise)

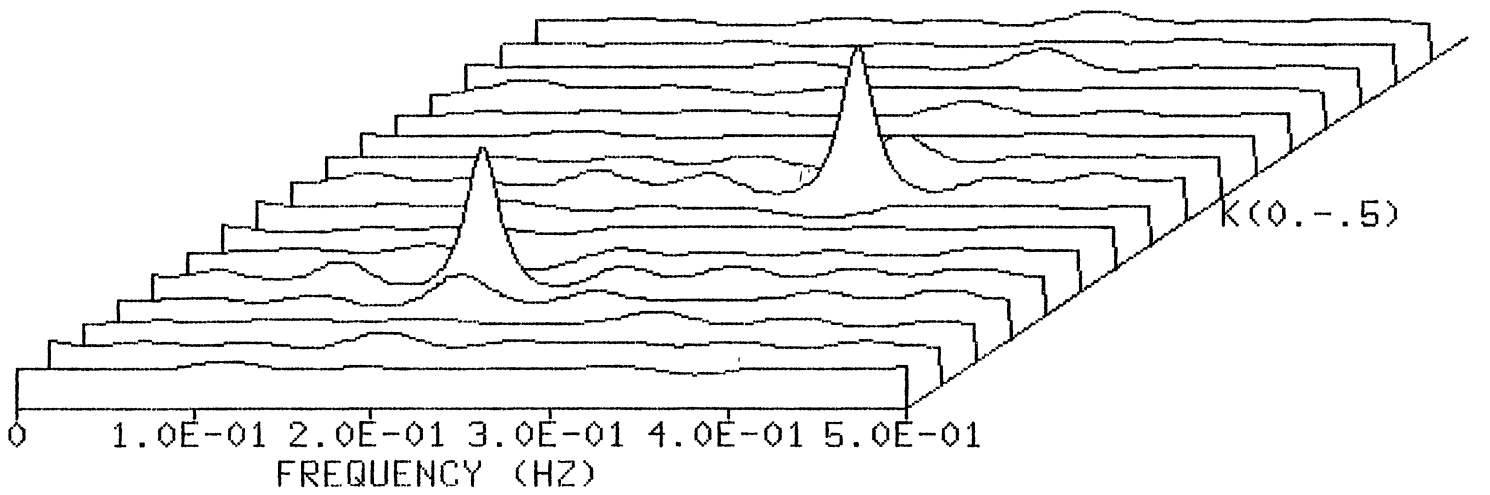

Figure 45. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 0 dB Gaussian noise)

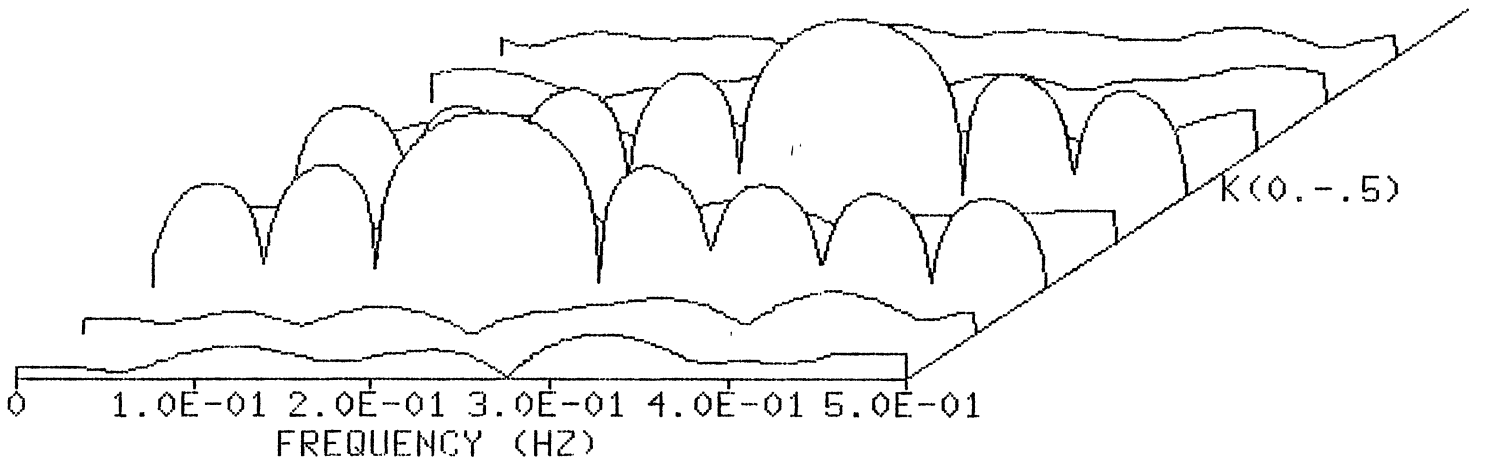

Figure 46. 2-D DFT Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 10 dB Gaussian noise)

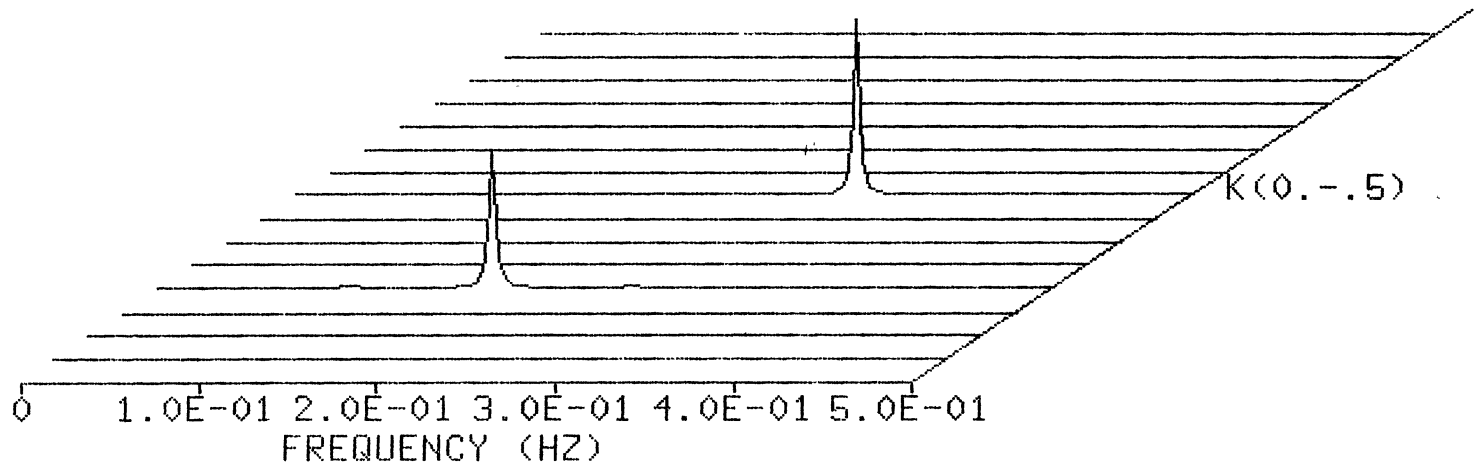

Figure 47. 2-D Lp (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 10 dB Gaussian noise)

closely spaced temporal frequencies (Figure 48), however, the 2-D LP estimator separates the sinusoids (Figure 49). With a 30 dB SNR, the results are the same (compare Figures 50 and 51).

For case 1, the L_2 formulation results in the frequency-wavenumber estimate shown in Figure 52 as compared to Figure 53. in which an L_1 solution was used. Both L_1 and L_2 methods resolve the spectral peaks, although the L_1 solution is charaterized by a much sharper peak. Figure 54. shows that for the spectral spacing of case 2 the L_2 formulation is unable to resolve the peaks. L_1, however, still resolves the spectral peaks of case 2 as can be seen in Figure 55. Case 3 is presented in Figures 56 and 57. Naturally the peaks are merged for the L_2 solution, (Figure 56) but an L_1 solution resolves the spectral peaks (Figure 57). Finally from Figure 58 (case 4) it is seen that the peaks have merged with the L_1 solution. Thus, in this example, an L_1 solution of the 2-D LP equations suggested in Kumaresan and Tufts (1981), is able to resolve spectral peaks at a closer spacing than that achievable via the minimum norm solution.

Although the results of applying an L_p (p=1 and p=2) normed spectral estimator to the case of sinusoid detection in non Gaussian noise has been considered in some detail previously, a small set of data will be presented here to demonstrate the effect of non-Gaussian noise on a 2-D LP (L_2 normed) spectral estimator. The test frequencies are that of case 2 and the SNR is 10 dB. In Figures 59 and 60 it is seen that the two sinusoids are resolvable for Gaussian noise and uniform noise. With Rayleigh noise, however, the 2-D LP spectral estimator is unable to resolve the two sinusoids (Figure 61); thus the underlying noise statistics play an important role in LP based spectral estimation.

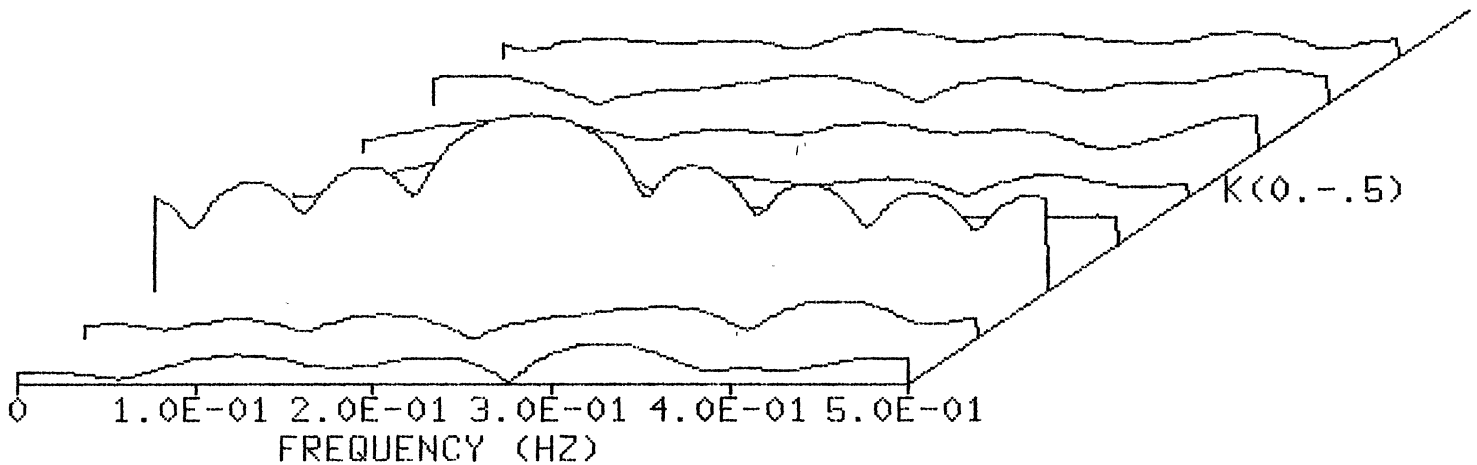

Figure 48. 2-D DFT Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 10 dB Gaussian noise)

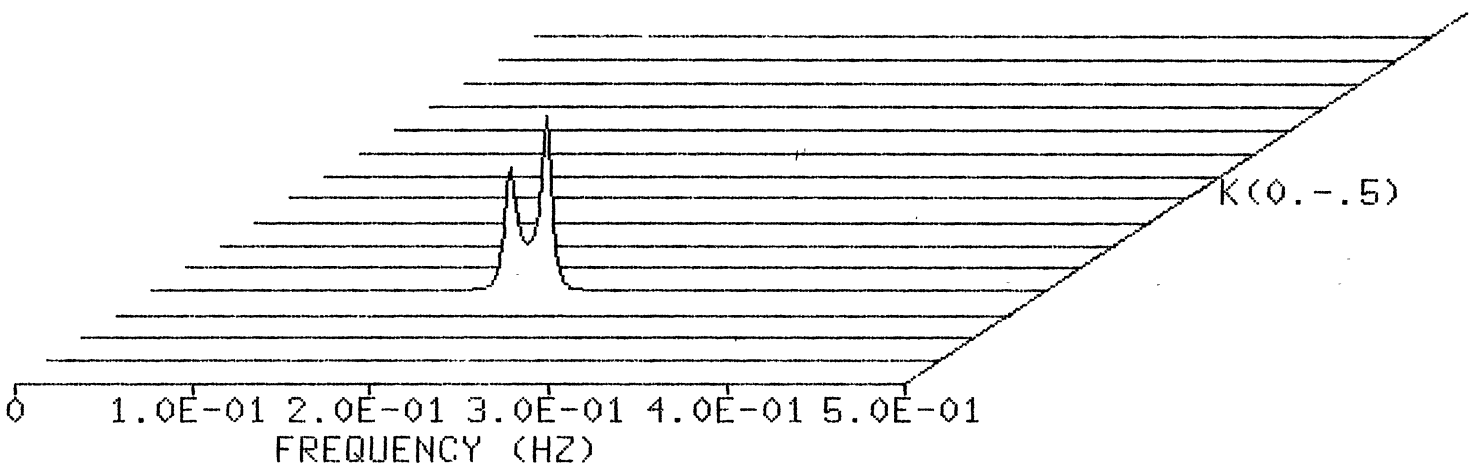

Figure 49. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 10 dB Gaussian noise)

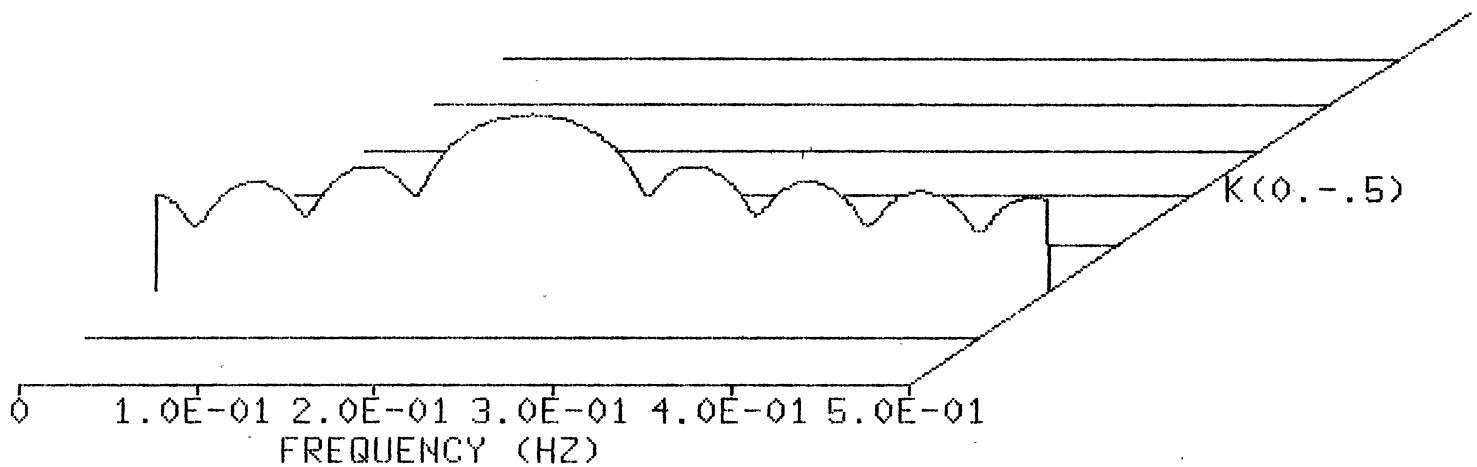

Figure 50. 2-D DFT Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 30 dB Gaussian noise)

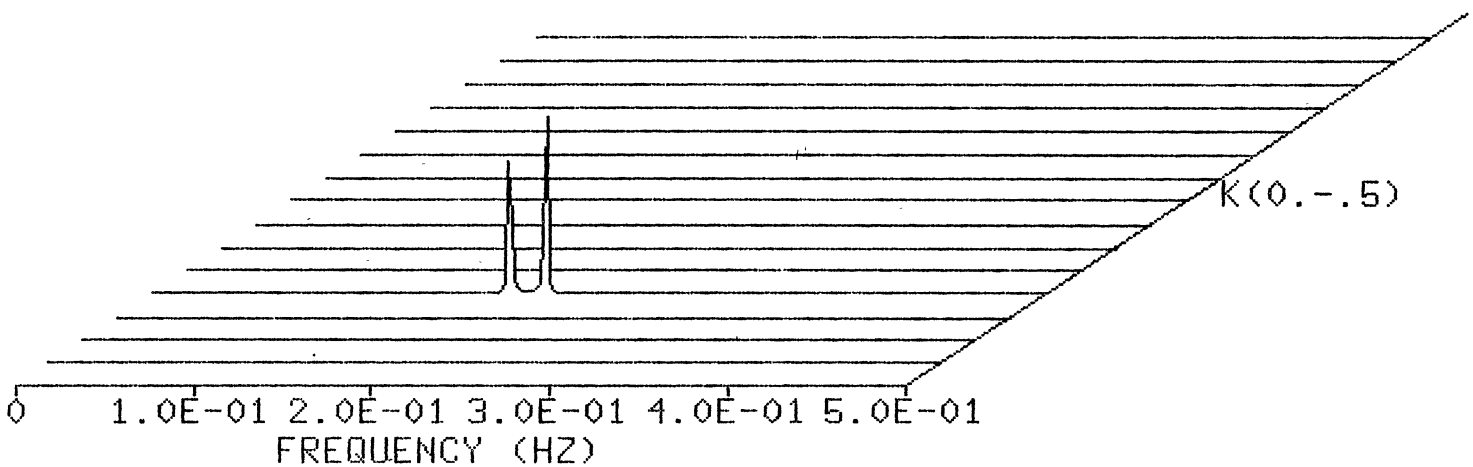

Figure 51. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 30 dB Gaussian noise)

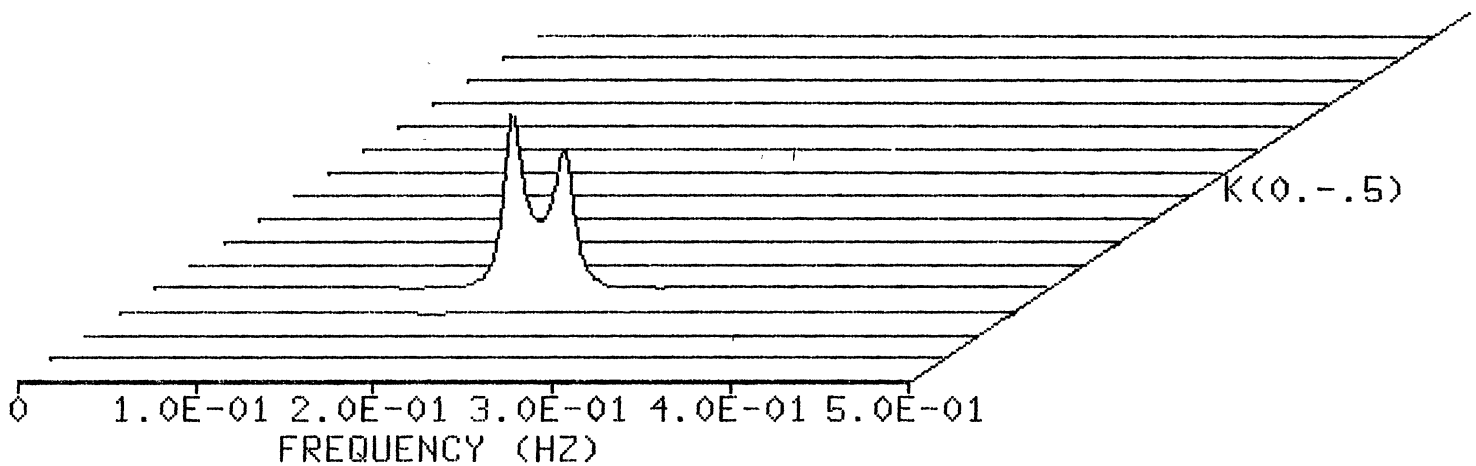

Figure 52. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .23 Hz; Impulsive noise)

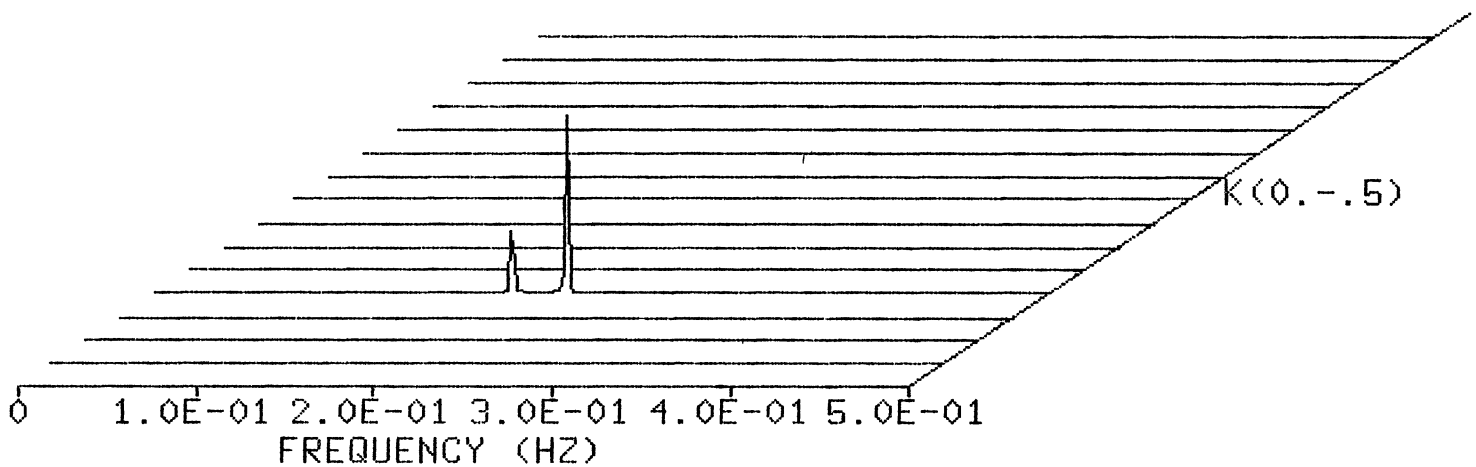

Figure 53. 2-D LP (L_1 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .23 Hz; Impulsive noise)

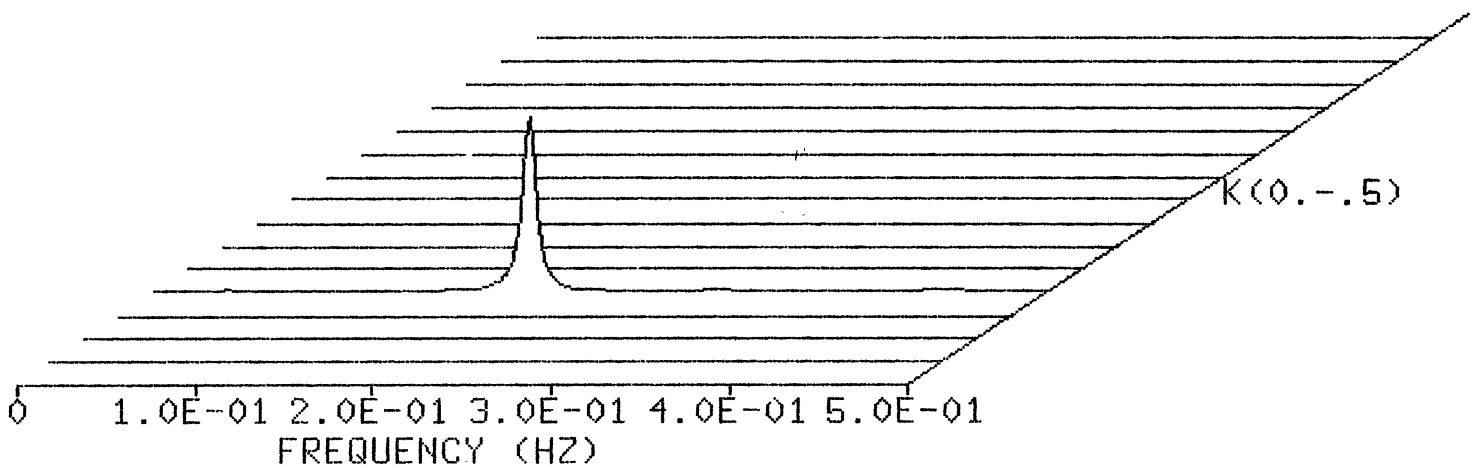

Figure 54. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; Impulsive noise)

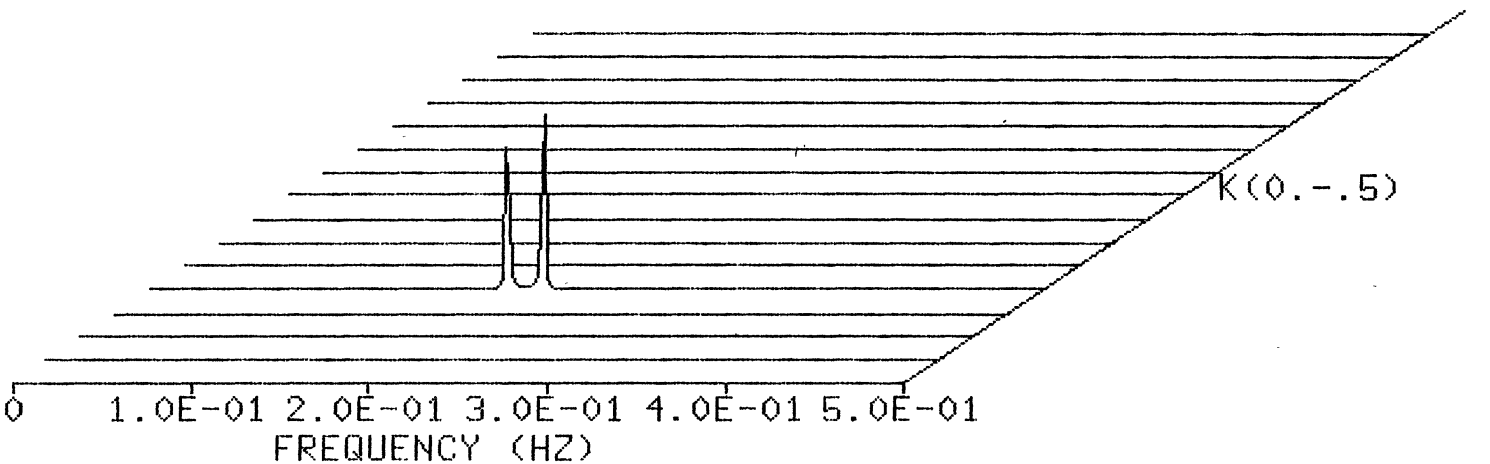

Figure 55. 2-D LP (L_1 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 =.125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz, Impulsive noise)

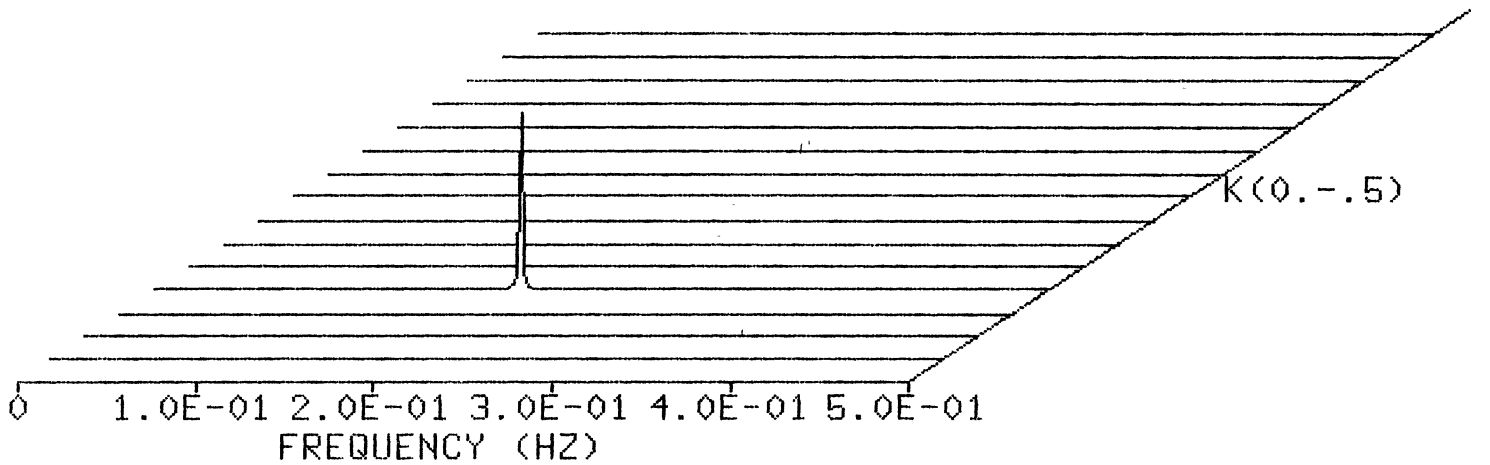

Figure 56. 2-D LP (L_2 Norm) Spectral Estimate (N, M =16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .21 Hz; Impulsive noise)

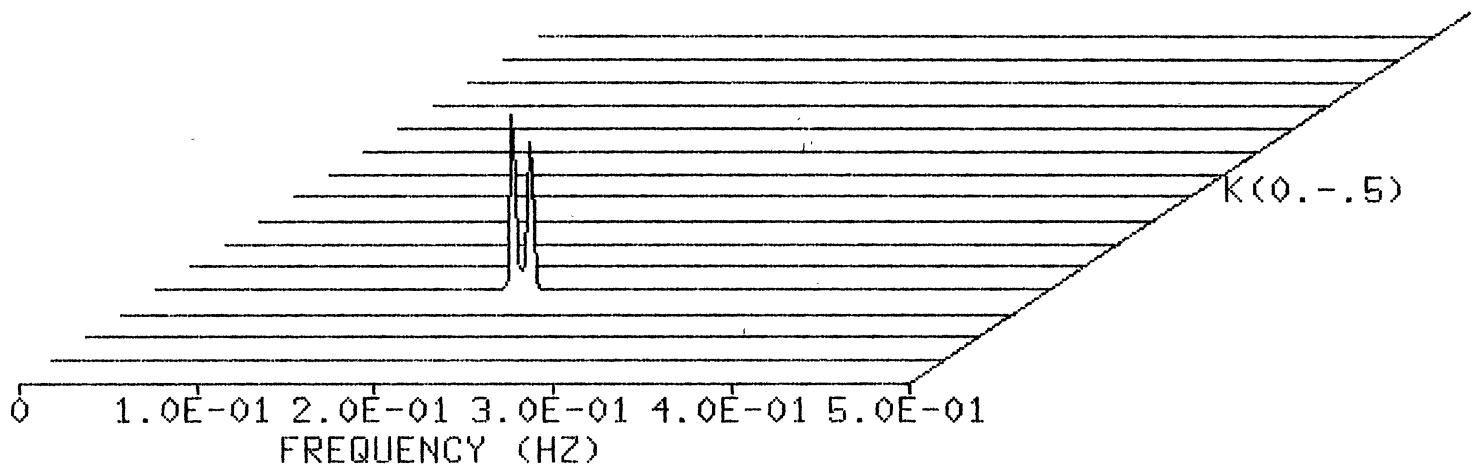

Figure 57. 2-D LP (L_1 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .21 Hz; Impulsive noise)

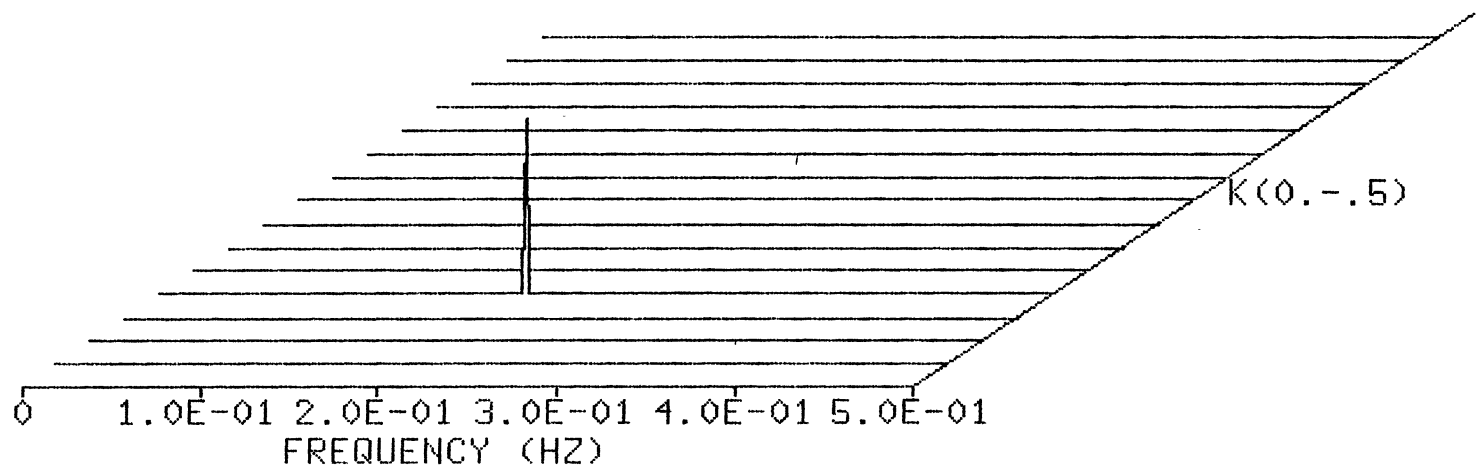

Figure 58. 2-D LP (L_1 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2Hz; f_3, f_4 = .125 cycles/foot, .205 Hz; Impulsive noise)

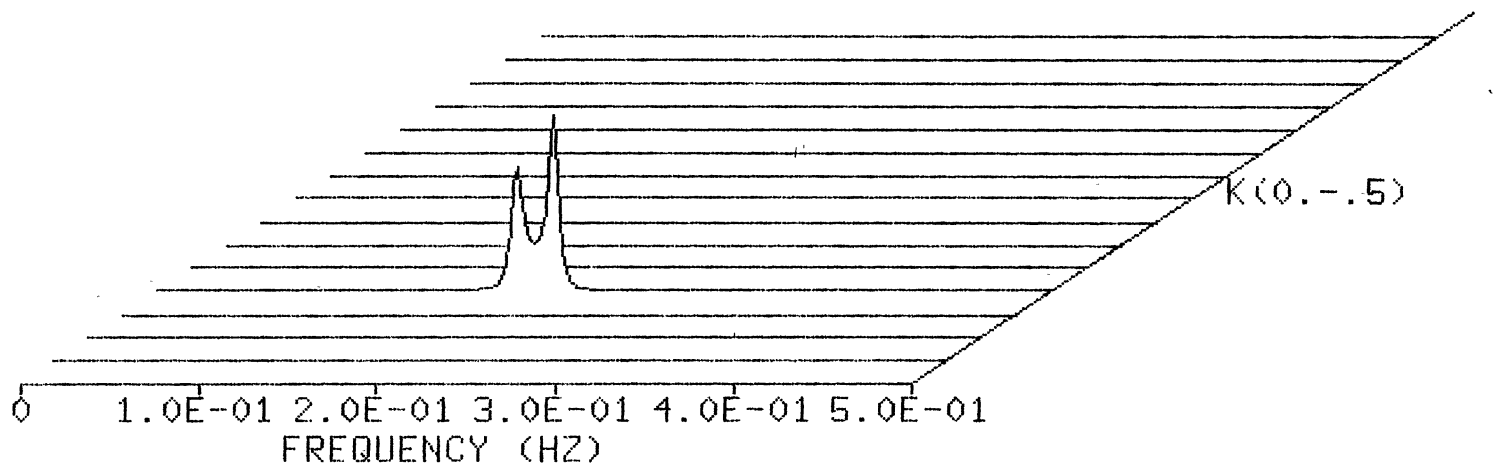

Figure 59. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 10 dB Gaussian noise)

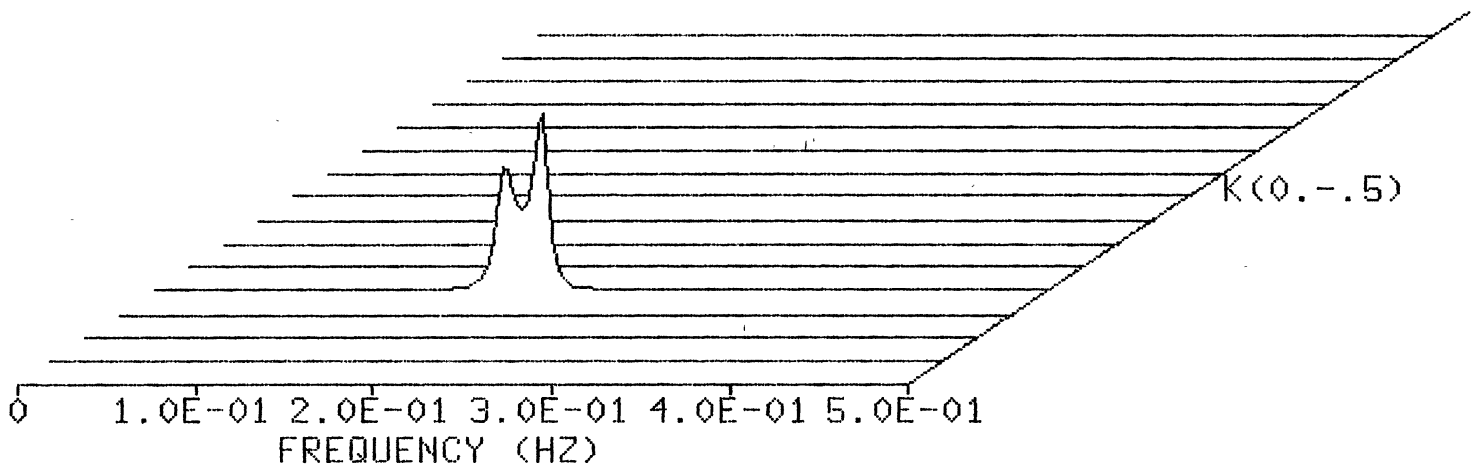

Figure 60. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz; 10 dB Uniform noise)

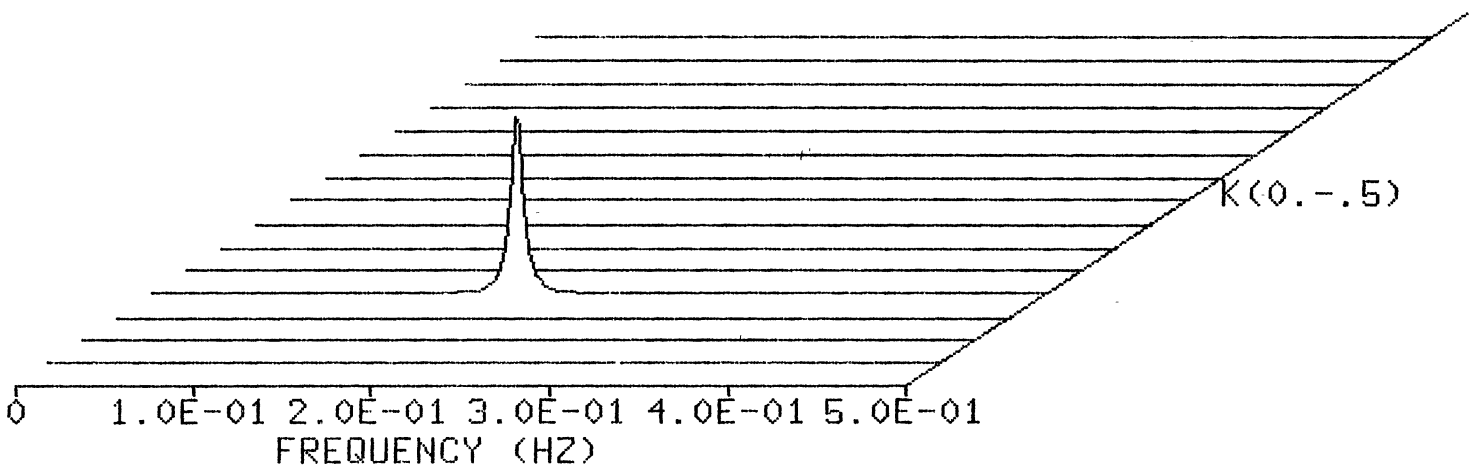

Figure 61. 2-D LP (L_2 Norm) Spectral Estimate (N, M = 16, 16; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .22 Hz, 10 dB Rayleigh noise)

For illustrative purposes, the efficiency of the proposed algorithm is demonstrated using $|\Delta_k|$ as a measure of convergence. Consider case 1; $|\Delta_k|$ for k = 1,2 are respectively given by .98 and .00 with the i^{th} component $(\underline{b}(0))_i = 1 + j1$ selected as the initial vector. For $(\underline{b}(0))_i = -5000 - j3000$ (obviously an unrealistic initial value) $|\Delta_k|$, k = 1,2,3, are given as 4040., .003, and .0. A plot of the convergence data is given in Figure 62.

A potentially serious disadvantage of this direct 2-D LP spectral estimation method is that of algorithm complexity. Even though the data size is quite small (16 x 16), reltively large matrices are involved (255 x 255) due to the algorithm structure. Large matrices require significant storage requirements that may severaly tax the resources of mini-computer systems. Additionally, numerical calculations such as multiplication and addition, require excessive execution times. For example, on a VAX 11/750 computer system, this 2-D spectral estimation method typically required approximately three and one-half minutes. With other users logged on this time would increase proportionately. In the next section, a hybrid method that blends a DFT with LP is formulated that significantly reduces computational complexity.

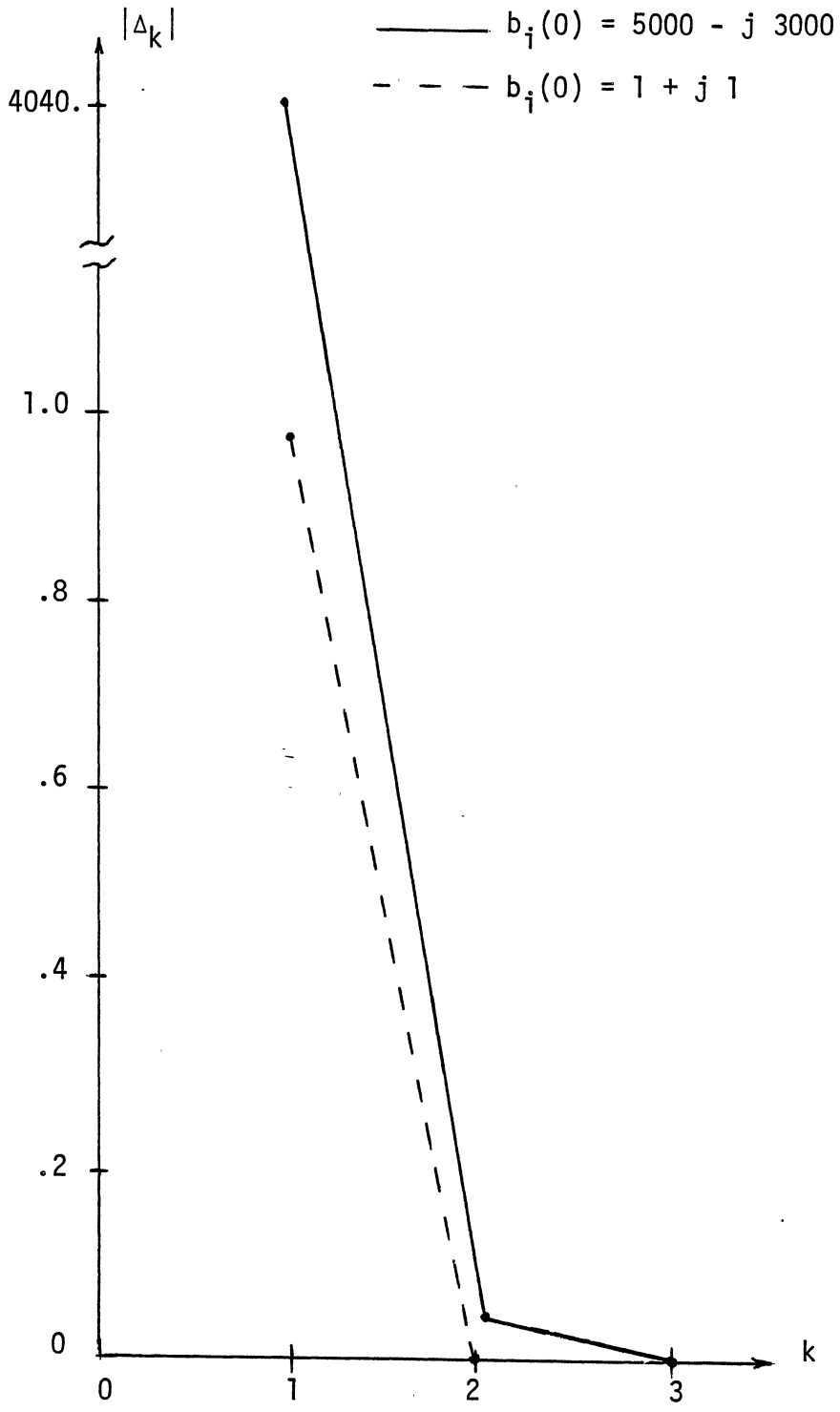

Figure 62. RSD Algorithm Convergence Example

Separable DFT/LP Spectral Estimation

The previous section presented a 2-D LP spectral estimator that was computationally inefficient and numerically ill conditioned; this section develops an improved method of 2-D spectrum estimation. In two dimensional spectral estimation it is often the case that one dimension possesses a significant number of data samples (e.g. 256 samples), while very few (e.g. 8 samples) data samples are available in the other dimension. This situation arises in applications where a relatively short spatial array is utilized in order to collect data that is a function of time. Such a data collection scenario is prevalent in geophysical data processing, however, it is equally applicable to other disciplines: Sonar arrays, for example, are typically of small aperture, while collecting numerous time samples. Whatever the application it is apparent that a sufficient number of time samples are available for application of the DFT. For long data records the DFT (periodogram method) provides more than adequate frequency resolution and minimal sidelobe leakage; additionally, the computational efficiency of FFT algorithms is well known. A high resolution spectral estimation technique, however, is required in the spatial dimension due to the limited number of available data samples. As the DFT output is complex valued, regardless of the high resolution technique selected, the estimator must be able to handle complex signals; this restriction should not generally be a problem and no difficulties have been noted in this work. The DFT/AR method was apparently first used by Jackson, Joyce, and Feldkamp (1979), where the Burg algorithm was utilized to calculate the AR parameters required for a high resolution spatial frequency estimate. Joyce (1979) considered the problem of limited data

samples in both dimensions, extrapolated the data in one dimension via AR methods prior to applying a DFT to the extrapolated dimension, and finally, applied the Burg algorithm in the second dimension (complex output of DFT of extrapolated data). In this work a DFT will be applied in the time dimension and an AR/LP estimator will be applied to the DFT output to form the spatial frequency estimate. LP parameter estimates will be generated by the iterative L_p normed solution method discussed in Chapter III that was applied to specific problems in Chapters IV and previous sections of this chapter.

Before the results of this hybrid DFT/LP method are presented, the required steps of this algorithm will be highlighted. The following steps are necessary:

step 1: Calculate a 1-D DFT of each line of the data array in the long dimension (time).

$$H(z_1,m) = Z_n\{x(n,m)\}, \quad n = 1, 2, \ldots, N$$
$$m = 1, 2, \ldots, M$$

where Z_n { } denotes a z-transform in the subscripted variable, evaluated on the unit circle, $z_1 = \exp(j2\pi f)$.

step 2: Using $H(z_1,m)$ as the complex input, calculate estimates of the AR parameters, a_k.

step 3: For each complex data set from DFT along other dimension, form the spectral estimate in the conventional manner (e.g. equation 3.2).

After step 3, the result is a 2-D spectral plane, with resolution in the long (time) dimension determined by the DFT and resolution in the short

(spatial) dimension determined by the AR/LP estimator. Again, the AR parameters are calculated iteratively via an L_p ($1 \leq p \leq 2$) normed solution (i.e. the modified RSD and IRLS algorithms). Consult Appendix B for additional detail on the separable LP/DFT estimator.

Results of applying this hybrid method to the test case of two 2-D sinsoids in Gaussian noise will now be given. The data size used for testing against a 2-D DFT based spectral estimator is 16 x 512, however, the DFT/LP spectral estimator was applied to an 8 x 512 data array in order to ensure that the spatial dimension was quite short. Here, it was not desired that the DFT estimator fail completely so the slightly larger data array was utilized. The two sinusoids are defined by the frequency pairs f_1 = .125 cycles/foot, f_2 = .1875 Hz, and f_3 = .25 cycles/foot and f_4 = .3125 Hz; SNR is 30 dB. In this high SNR case, with widely spaced sinusoids and large data array, the DFT technique resolves the sinusoids (Figure 63). The DFT/LP method as applied to the same data performs equally well, as seen in Figure 64. Lowering the SNR to 0 dB, however, causes the DFT estimator to exhibit a high variance spectrum with obscured spectral peaks (Figure 65), while from Figure 66 it is seen that the DFT/LP method resolves the two sinusoids. Additional sinusoids may be detected with higher order filters. From Figure 67 it is seen that a third order filter will detect three sinusoids and from Figure 68 the four sinusoid case is shown (fourth order filter). Also, a fifth component (obviously spurious) may be seen in Figure 68; since the data record length here equals 8, a filter length of 4 is a bit two long. A general rule of thumb is to restrict the filter length to values less than one-third the data length.

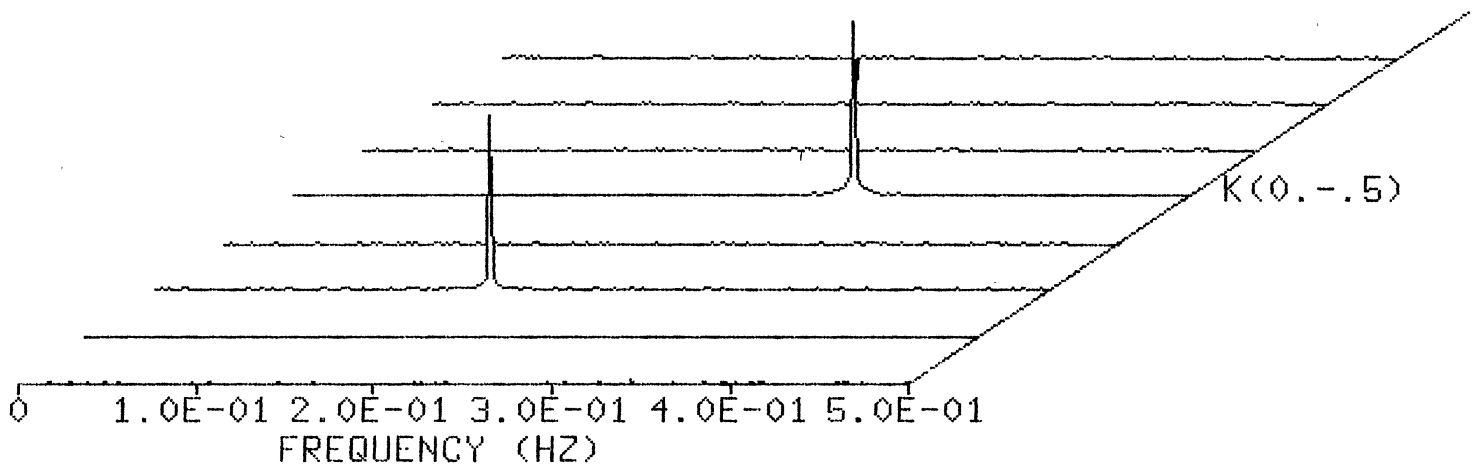

Figure 63. 2-D DFT Spectral Estimate (N, M = 16, 512; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 30 dB Gaussian noise)

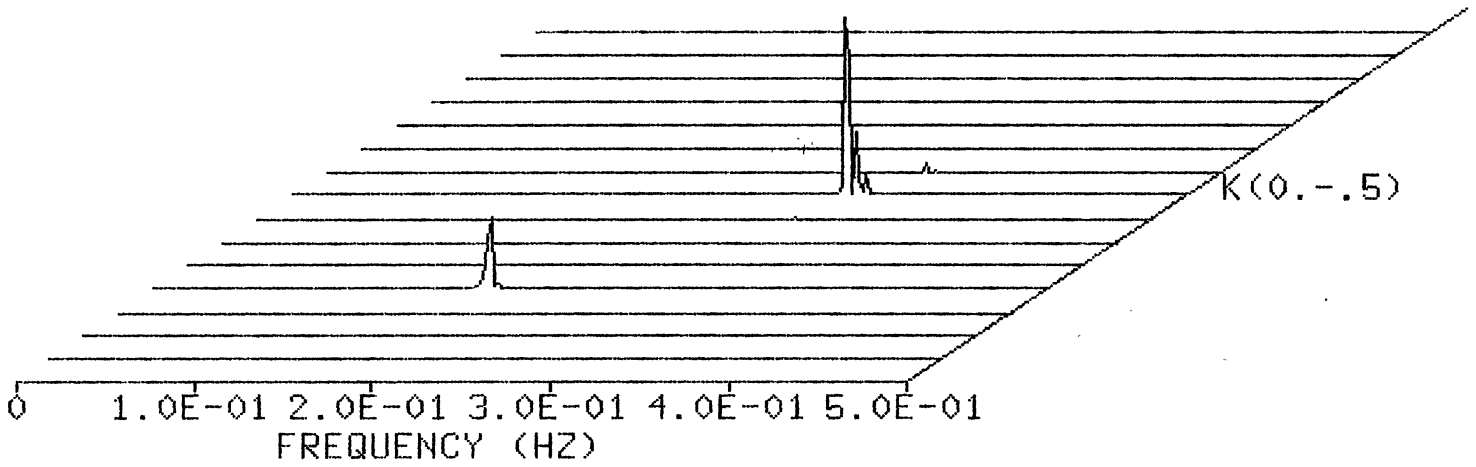

Figure 64. DFT/LP (L_2 Norm) Spectral Estimate (N, M = 8,512; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 30 dB Gaussian noise)

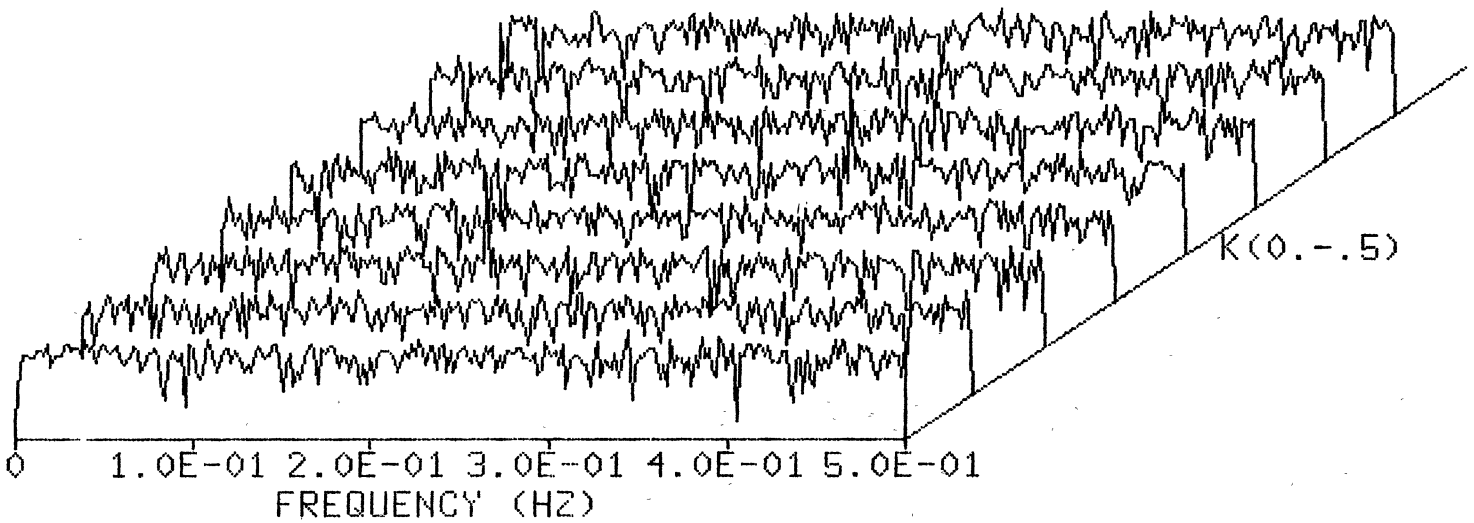

Figure 65. 2-D DFT Spectral Estimate (N, M = 16, 512; f_1, f_2 = .125 cycles/foot, .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 0 dB Gaussian noise)

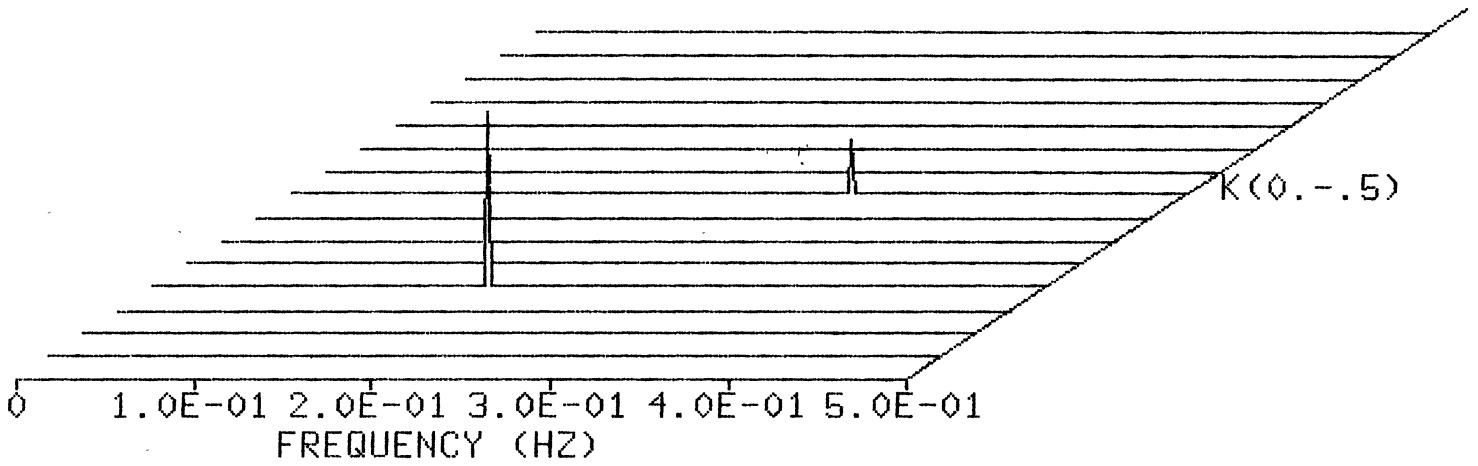

Figure 66. DFT/LP (L_2 Norm) Spectral Estimate (N, M = 8, 512; f_1, f_2 = .125 cycles/foot; .1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; 0 dB Gaussian noise)

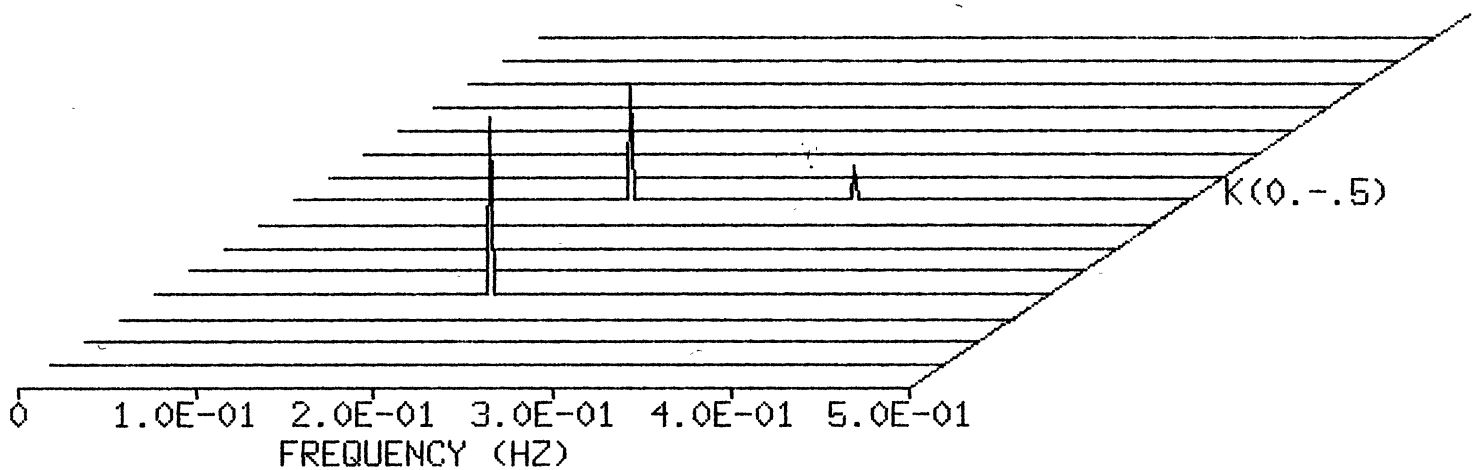

Figure 67. DFT/LP (L_2 Norm) Spectral Estimate (N,M = 8, 512; f_1, f_2 = .125 cycles/foot,.1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; f_5, f_6 = .25 cycles/foot, .1875 Hz; 0 dB Gaussian noise)

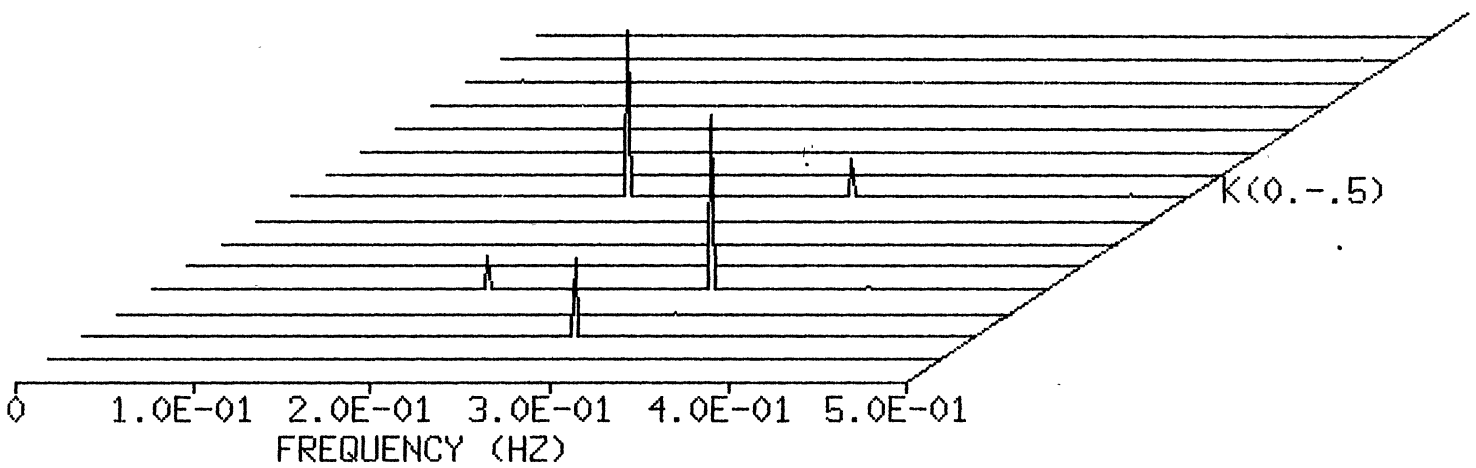

Figure 68. DFT/LP (L_2 Norm) Spectral Estimate (N,M = 8, 512; f_1, f_2 = .125 cycles/foot,.1875 Hz; f_3, f_4 = .25 cycles/foot, .3125 Hz; f_5, f_6 = .25 cycles/foot, .1875 Hz; f_7, f_8 = .125 cycles/foot, .3125 Hz; 0 dB Gaussian noise)

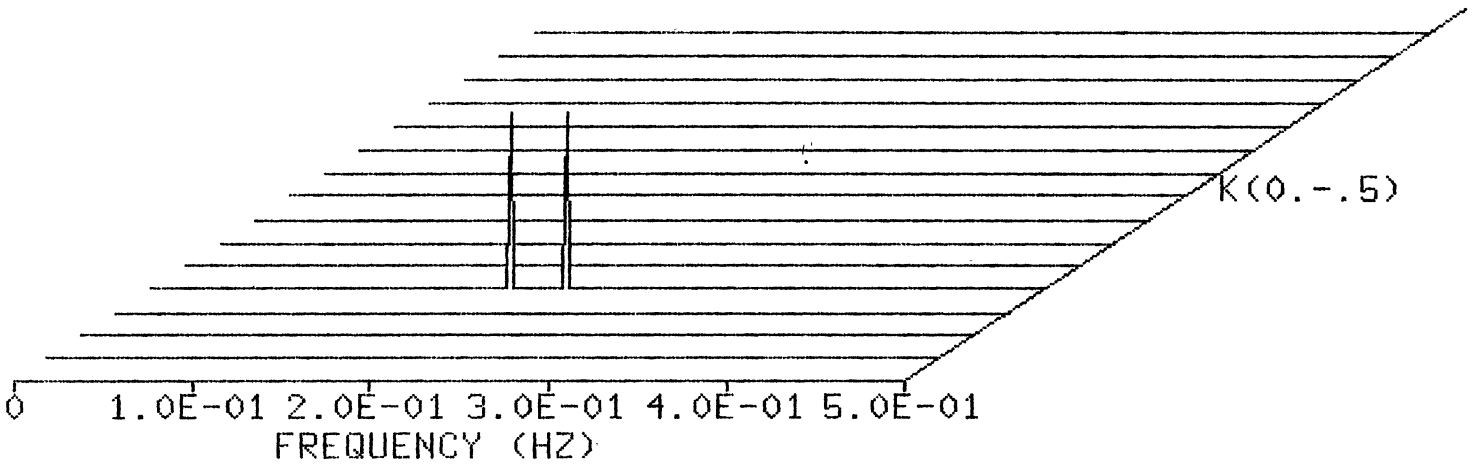

Figure 69. DFT/LP (L_1 Norm) Spectral Estimate (N, M = 8, 512; f_1, f_2 = .125 cycles/foot, .2 Hz; f_3, f_4 = .125 cycles/foot, .23 Hz; 20 dB Gaussian noise)

In the previous section it was noted that a direct 2-D LP formulation resulted in extensive execution time. By contrast, the hybrid method developed in this section executes significantly faster. Although, the execution is a function of RSD algorithm converge, typically about 30 second is required to process an 8x512 data array on a VAX 11/750 computer system.

Application of F-K Analysis to the Acoustic Well Log

The instrumentation used in sonic well logging consists of a long wireline type tool that contains a transmitter and a receiver or array of receivers. A single transmitter and receiver is depicted in Figure 70, however we are interested in the situation where a small number of receivers are mounted on a single tool. As this type of logging tool is moved through the borehole, the transmitter emits pulses of acoustical energy at discrete instants of time that is limited only by the time required for the acoustical reverberation to die out within the borehole fluid. Since these acoustical pressure waves travel within the borehole surrounded by a geological formation, information about the surrounding formation such as porosity is encoded into the propogating waves. The array of receivers serves to sense the acoustical pressure variations occurring from the propagating waves. The signals collected from the array of receivers (both a function of space and time) are digitized and trasmitted to the surface via cable and stored for later analysis by the geologist. Thus, the data collected from a sonic well logging tool forms a two dimensional space-time data array and frequency-wavenumber

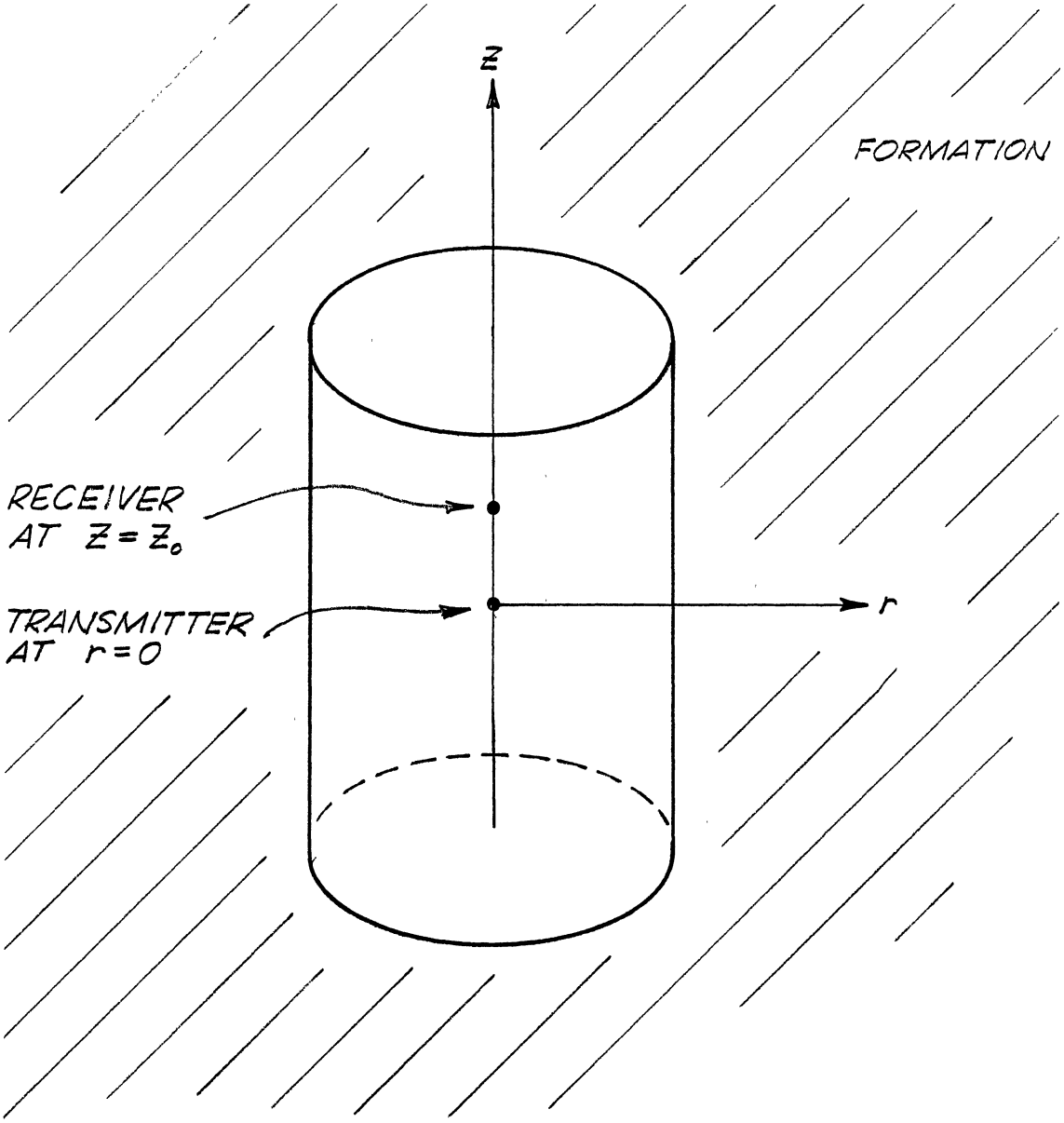

Figure 70. Acoustic Well Log Borehole Geometry

analysis may be applied to aid in interpretation of formational characteristics.

In any array signal processing problem such as just described compromises must be made. In order to increase available spatial frequency resolution it is desired to maintain as long an array as possible; it is well known that frequency resolution increases with aperture. Unfortunately, the surrounding geologic formation is not likely to be homogeneous over a very long array length. In fact, adequate vertical resolution enabling the geologist to make an accurate determination of formation characteristics versus depth requires a short receiver array. The vertical resolution that is important to a geologist decreases with increased array lengths. Another problem, of course, is the limited number of receivers that is practical to mount on such a logging tool; thus we have the problem of a short array span (aperture) plus a limited number of data samples. Fortunately all of the temporal data samples are available and we do not have a limited time aperture problem to deal with. One solution to improving a data array characterized by limited data availability in one dimension has been mentioned in the previous section; namely, a hybrid approach whereby a DFT is applied in the time dimension and a higher resolution parametric spectral estimator is applied in the spatial dimension. Thus, the more complex techniques are only used in the dimension with limited data availability greatly reducing computational complexity.

Basically two types of waves propagate within a borehole: guided waves that propagate in the fluid and refracted waves that travel through the formation. A fast formation has the property that the formation shear velocity is greater than the fluid velocity; in such a

case a shear wave is produced that is somewhat slower than the compressional waves. Unfortunately, the shear wave does not always exist. The velocities of these various waves may be related by the geologist to various properties of the surrounding formation. It may be remarked in passing that the guided fluid waves are dispersive and include the so called stoneley wave as well as other propagating modes. If the borehole is cased, it is also possible for waves to propagate within the casing; this situation becomes quite complicated and tends to further increase the difficulty of obtaining an accurate frequency-wavenumber representation. As mentioned in the literature survey section, much research is currently directed at accurately modelling the highly nonlinear phenomena of acoustical waves propagation in various borehole types. In the frequency-wavenumber plane dispersive waves (velocity is a function of frequency) manifest themselves as curved lines while nondispersive waves appear as straight lines. Recall that velocity is the inverse of the slope of a line in the frequency-wavenumber plane (1/sec / 1/ft = ft/sec).

Figure 71 and Figure 72 show typical array data characteristic of the acoustic well log; this data was generated by a synthetic modelling program, however, and did not arise from a real borehole. The nonstationary and transient characteristics of this type of data are obvious. A convential 2-D DFT generated spectral estimate (16 traces used) is shown in Figure 73. It is difficult to interpret a 2-D spectrum from a 3-D perspective plot, however, a contour plotting package is not currently available. Obviously, a trained geologist well versed in the subtle idiosyncrasies of acoustic well log data, would be invaluable here. In Figure 74 the results of applying DFT/LP (L_2

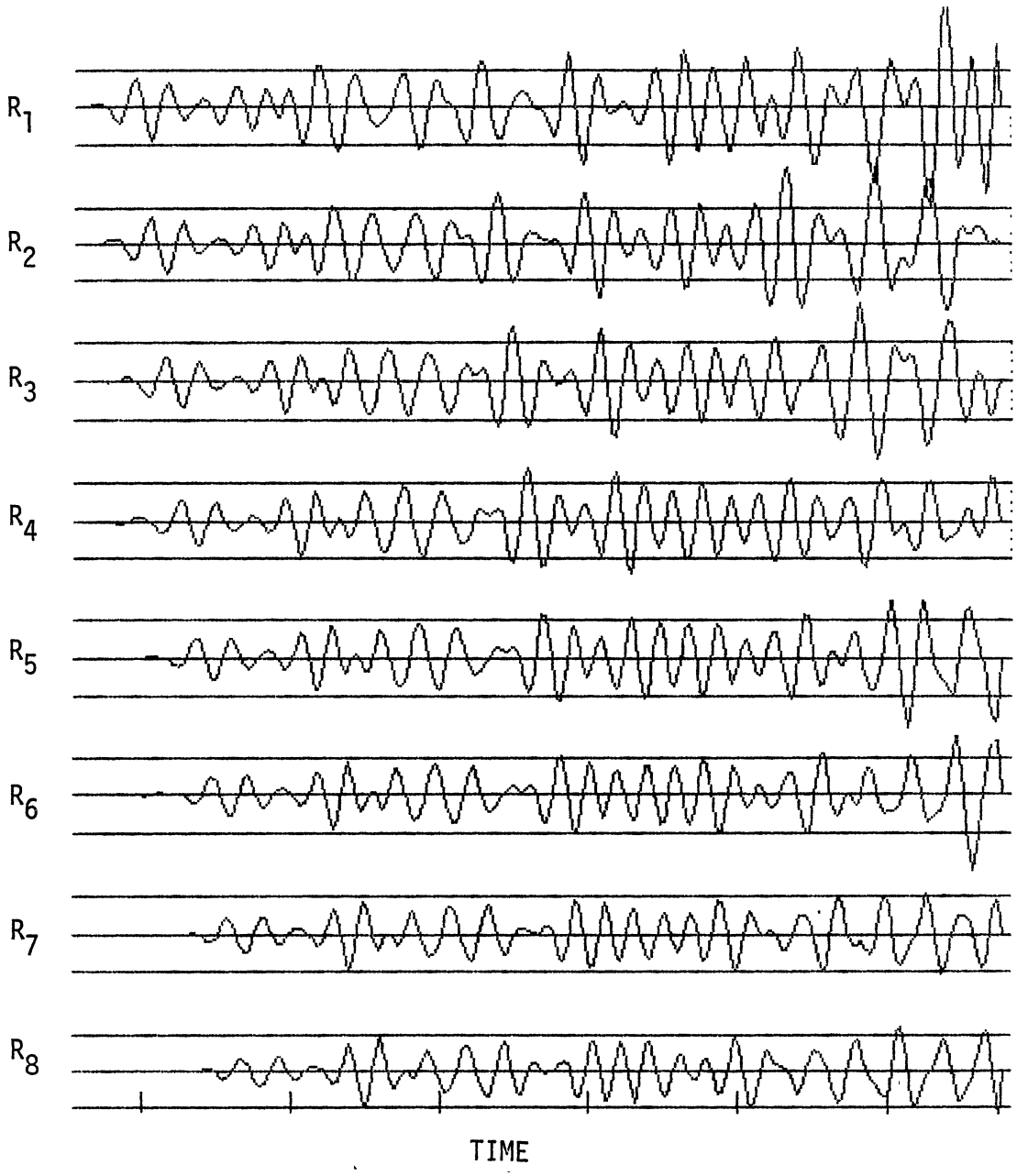

Figure 71. Acoustic Well Log Synthetic Data (Traces 1 - 8)

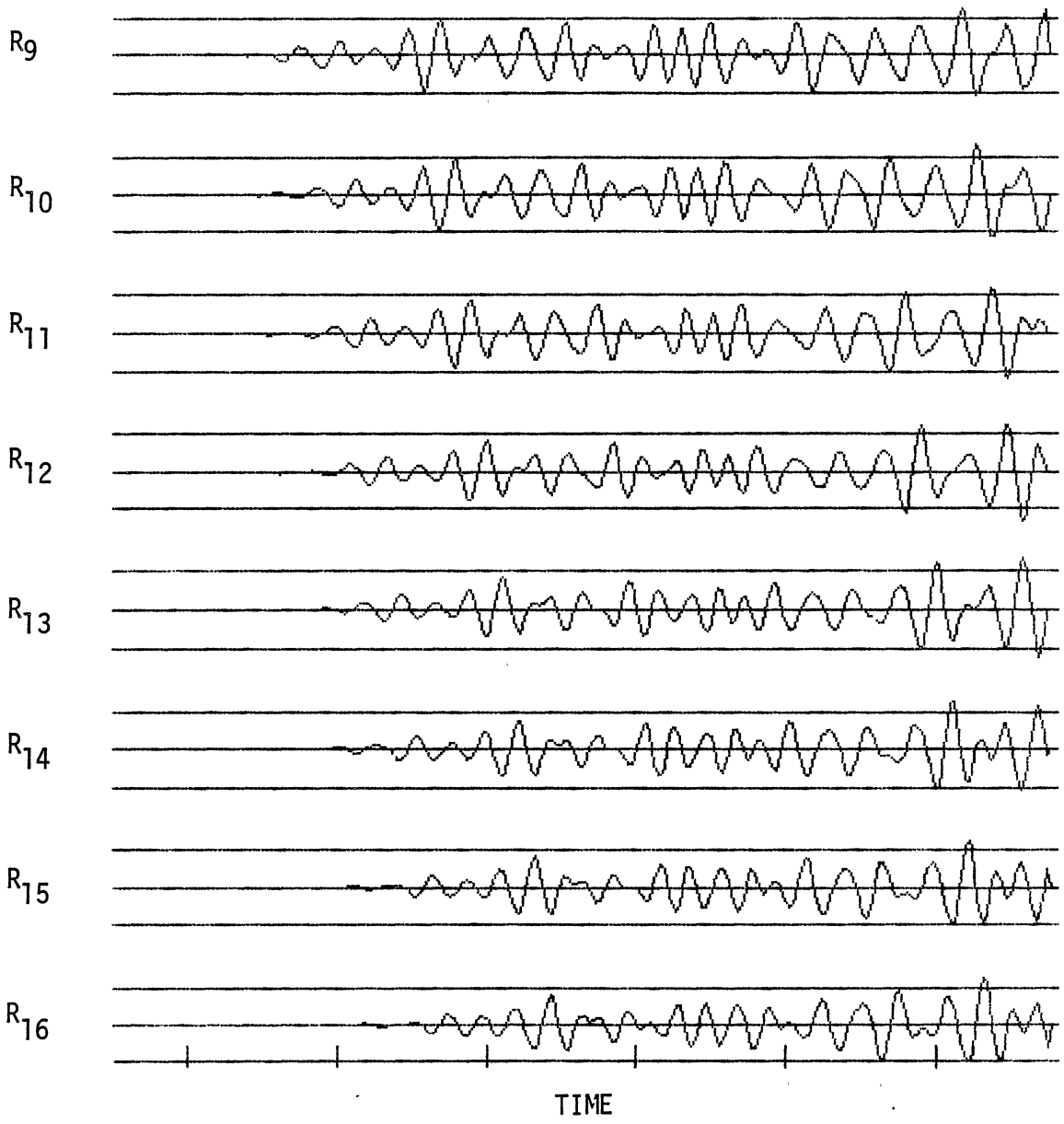

Figure 72. Acoustic Well Log Synthetic Data (Traces 9 - 16)

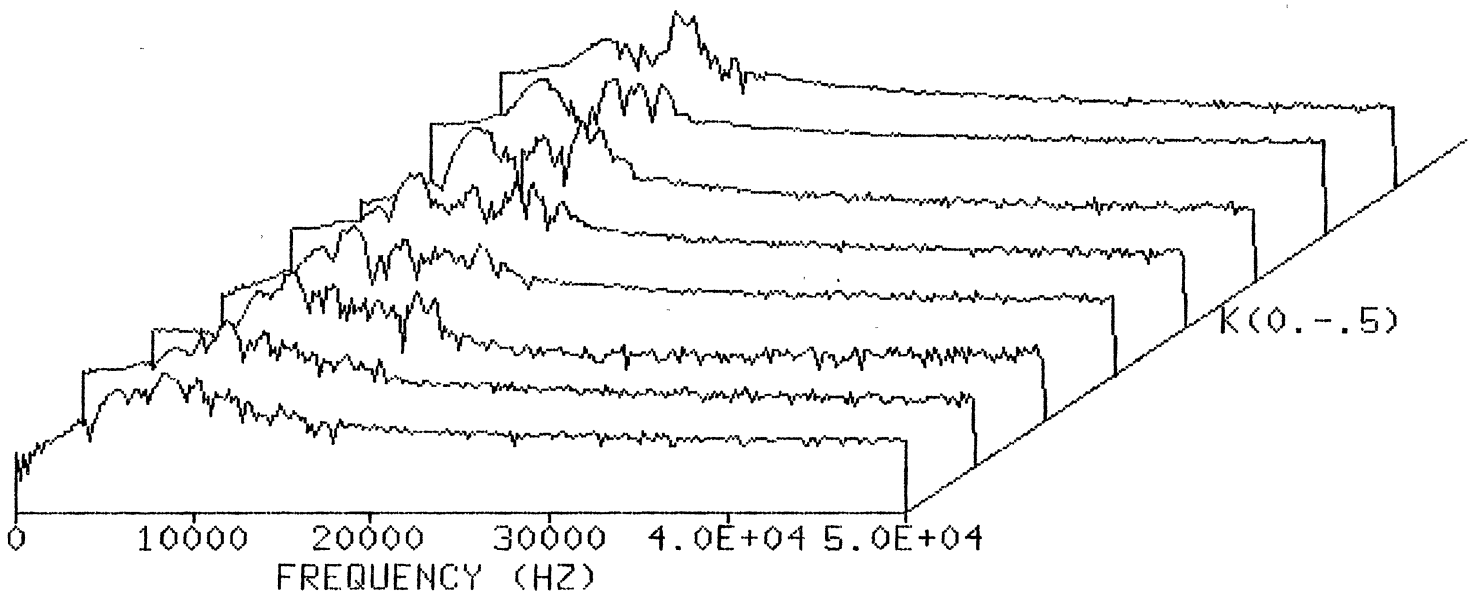

Figure 73. 2-D DFT Spectral Estimate (N, M = 16, 512; synthetic acoustic well log data)

normed) spectral estimator can be seen. As the filter length (2) is quite short the spectrum can not be compared directly to Figure 74, however, the general trend of the main spectra (visible in the DFT plot) is apparent. The DFT/LP (L_1 normed) estimator was also tried with results shown in Figure 75. A slightly higher order filter may be required, however, the viability of this technique has been demonstrated. Additionally, it is expected that a data window moving at the same rate as an expected velocity component may be used to isolate one or two components for more detailed frequency-wavenumber analysis.

Conclusions and Future Research

A computationally simple 2-D technique of simultaneous frequency-wavenumber estimation introduced in Kumaresan and Tufts (1981) has been modified in order to incorporate the L_1 norm minimization criteria. For the case of impulsive noise the L_1 norm solution has been shown to result in increased frequency resolution. Use of the RSD algorithm from Yarlagadda et al. (1985) allows an L_1 solution to be calculated without undue computational burden. Additionally, only a 2x2 (in general LxL for L sinusoids) matrix inverse is required for the detection of two (L) complex sinusoids in noise. Since this spectral estimation algorithm has been formulated assuming a complex space, it may easily be blended with DFT techniques if a sufficiently long data record is available in one dimension (typically the case for the "time" dimension), thus achieving a further simplification in algorithm complexity. Since in general the RSD algorithm can be utilized to obtain a solution with the L_p norm criteria, this method is not restricted to studying the L_1 norm. Current research efforts are focused on a better understanding of

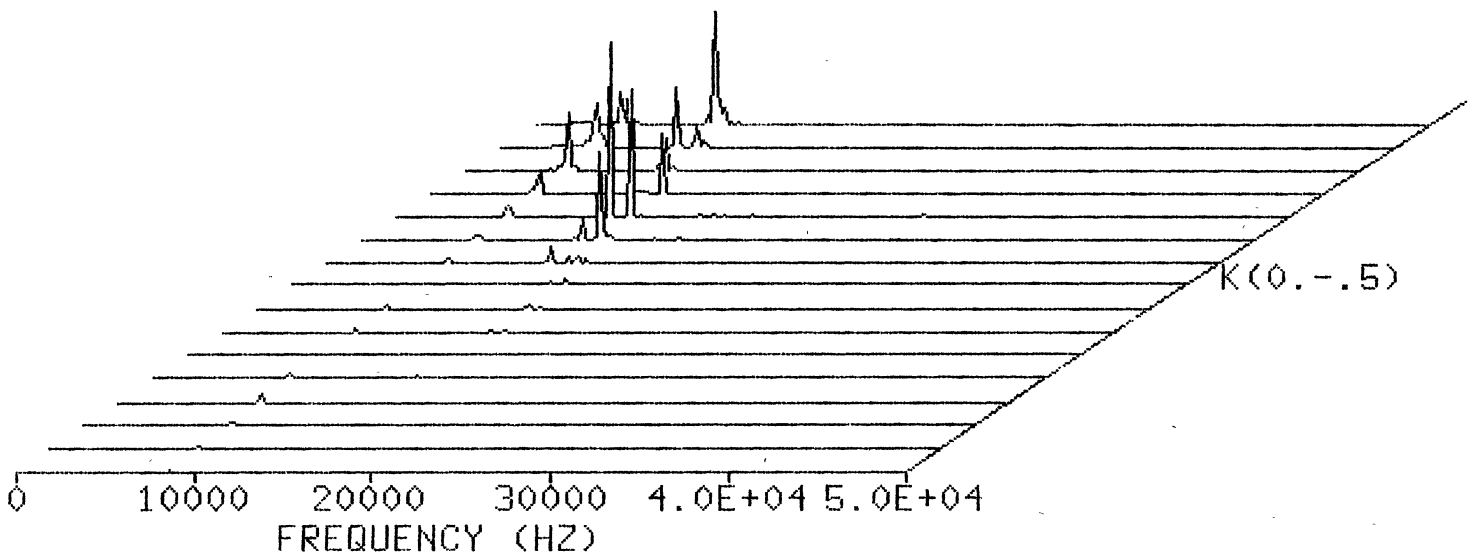

Figure 74. DFT/LP (L_2 Norm) Spectral Estimate (N, M = 8, 512; synthetic acoustic well log data)

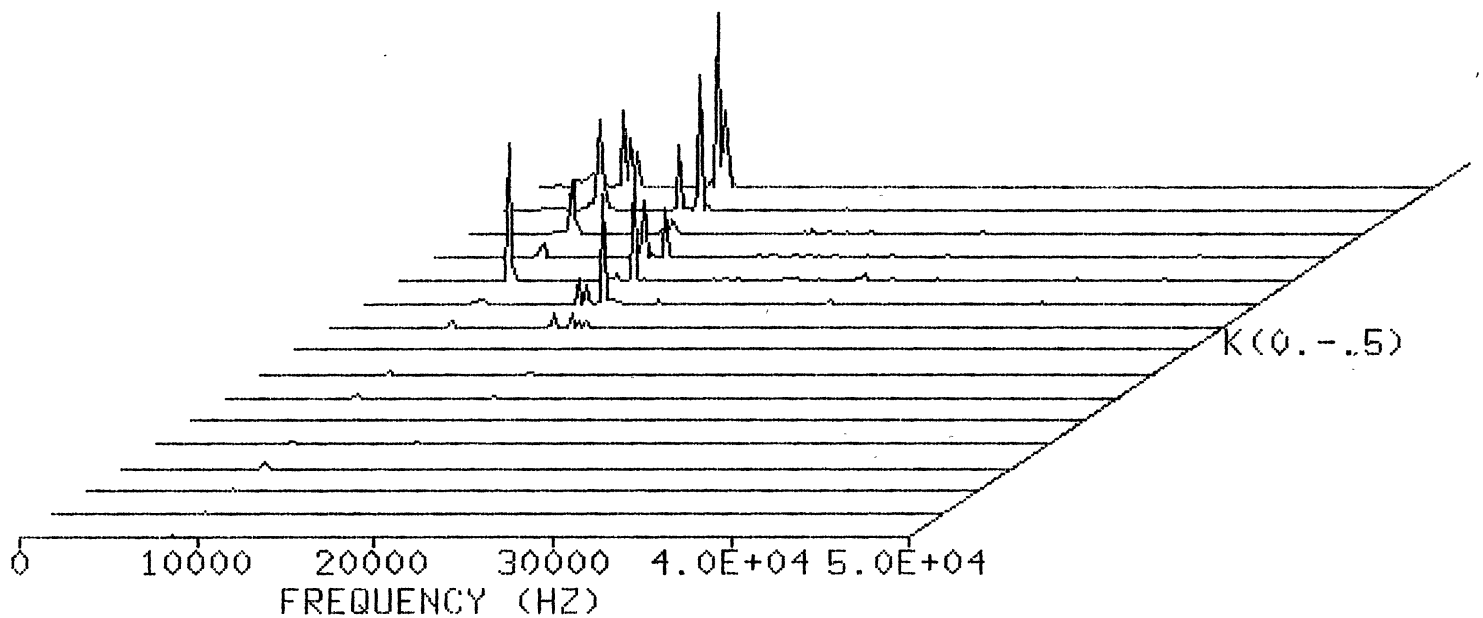

Figure 75. DFT/LP (L_1 Norm) Spectral Estimate (N, M = 8, 512; synthetic acoustic well log data)

the L_1 norm as applied to spectral estimation and consideration of the L_p norm problem, $1 \leq p \leq 2$. It is hoped that intermediate values of p, $1 \leq p \leq 2$, will offer the best of L_1 and L_2. Also, it has been shown that the computational burden may be reduced by combining a DFT with LP methods; in this technique, the computationally expensive high resolution estimator (LP estimator) is only applied where it is necessary (i.e. the dimension possessing a limited amount of data).

CHAPTER VI

CONCLUSION

The purposes of this investigation were threefold: consideration of least squares iterative solutions for estimating AR/LP parameters, comparison of various L_p normed ($1 \leq p \leq 2$) solutions of AR/LP equations, and development of a viable method for application of L_p normed AR/LP spectrum estimation algorithms to 2-D array data. Although the ultimate test of any new spectral estimation scheme is success against realistic signal environments, for example a relatively broad band spectrum plus multiple sinusoids, at this stage of development the case of two sinusoids in noise (obviously simplistic) was utilized exclusively. There are sufficient difficulties during the research and development of new spectral estimation techniques as applied to the simple two sinusoids in noise problem that inclusion of a complex signal scenario would be unnecessarily chaotic.

It has been shown that iterative least squares type numerical methods are a viable means of estimating the AR/LP parameters. Iterative algorithms used in this study include the iterative reweighted least squares (IRLS) algorithm and a modified residual steepest descent (RSD) method. The convergence properties of the modified RSD algorithm have been excellent, generally just two or three iterations resulted in a desired solution. With such a fast convergence the resulting algorithms are computationally efficient removing one of

the major obstacles for consideration of iterative methods as a means of estimating AR/LP parameters, prior to forming the spectral estimate. Stability of the resulting prediction error filter has not been addressed. For one thing, the prediction error filter does not require stability for application to spectral estimation; also, it is somewhat complex to check stability, although, efficient methods are available should a stable filter be required (for example parametric signal representation).

Despite the minor drawbacks of iterative methods, such as an unstable prediction error filter and an increased computational burden, an important advantage is the ability to consider the general L_p ($1 \leq p \leq 2$) norm. Of particular interest may be the L_1 norm (absolute value error criteria) since such an error criteria is known to be very robust in the presence of impulsive noise or data that are contaminated by outliers. By contrast, the L_2 (least squares) norm exhibits rather poor performance on data containing impulsive noise or outliers since all data points (good or bad) receive equal weighting. In this work, it has been shown that choosing the L_1 norm for estimating AR/LP parameters results in a spectral estimator that demonstrates markedly improved performance (compared to an L_2 based estimator) when applied to data with impulsive noise (outliers). Additionaly, spectra generated via an L_1 normed solution exhibited a "peakier" shape than that generated via L_2 methods.

The L_p normed solution methods were also successfully applied to the 2-D spectral estimation problem. Two 2-D implementations were considered: direct application to a small sized data array and a hybrid method that combined a discrete Fourier transform (DFT) with the

proposed L_p normed AR/LP spectral estimator. For a small data array contaminated by impulsive noise, an L_1 based estimator was able to resolve two closely spaced sinusoids to a greater degree than the corresponding L_2 solution. Although a very small data array was used, the computation load was quite heavy; additionally, the problem was ill conditioned, which may limit practicality of the technique. Since it often occurrs that a relatively long data record is available in one dimension of 2-D data (often the time dimension), DFT techniqes may be blended with with high resolution methods to greatly simplify computations. It was found that the L_p iterative method blended easily with the DFT method which results in a computationally efficient 2-D spectral estimator that may be used in cases where sufficient data are available along one dimension in order for a DFT to provide adequate resolution.

Future Research

As always there is significant work left undone. An important aspect of any spectral estimation algorithm is performance against complex signal environments; the transition from the two sinusoids in noise case to multiple sinusoids and/or broadband signals plus unknown noise statistics can be very difficult and is usually very ad hoc. Selection of the appropiate model order can be difficult if the true spectrum is unknown, although a few guidelines are available. Most likely, when applying an L_p normed iterative solution to the spectral estimation problem, one will need to resort to hypothesis testing against the partial autocorrelation coefficients in order to select a model order, as the Levinson-Durbin recursion is not available.

Clearly, a Levinson-Durbin type recursion would be helpful in selecting an appropiate model order when estimating AR/LP parameters via L_p normed iterative methods. A possible solution is to first apply a conventional L_2 normed estimator, perhaps the FBLP method, and terminate according to some criteria (for example Akaike's criteria). The resulting L_2 solution if then used as the initial condition for the L_p normed RSD algorithm.

Another area of importance is the variance and bias of the frequency estimates available from this L_p method. Excessive frequency bias or line splitting has not been noted in any cases encountered to date, however, it probably would be advisable to exhaustively exercise this new spectral estimator. Undoubtedly there is much that is not understood concerning the reliability of the spectral estimates via the L_p method and the only means available to pin down the variance of these estimates is an experimental investigation with uncorrelated noise sequences.

This research has only been concerned with the L_1 norm. In general, the modified RSD algorithm may be used to generate L_p ($1 \leq p \leq 2$) normed solutions; it may be that an $L_{1.2}$ normed solution offers some advantages. If a stable prediction filter is required, for example, perhaps $L_{1.1}$ or $L_{1.2}$ may be sufficiently resistant to outliers yet retain stability. Statisticians working within the field of linear regression use the so called Chebyshev criteria (also called MINMAX or the L norm). It may be possible, via the RSD algorithm, to approximate this norm with say $p = 10$. L_p criteria ($0 \leq p < 1$) have also been studied within the context of linear regression; some work is being done with $p < 0$. Unfortunately, L_p space $0 \leq p < 1$ is not a linear normed

space. The general field of error criteria selection from the statistical literature may be fertile ground for new spectral estimation research.

REFERENCES

[1] A. Akaike, "Statistical Predictor Identification," Ann. Inst. Stat. Math. Vol. 22, pp. 203-217, 1970.

[2] N. Andersen, "On The Calculation of Filter Coefficients for Maximum Entropy Spectral Analysis," Geophysics, Vol. 39, No. 1, pp 69-72, February 1974.

[3] G. Appa and C. Smith, "On L_1 and Chebyshev Estimation," Math. Program., Vol. 5, pp. 73- , 1973.

[4] T.S. Arthanari and Y. Dodge, Mathematical Programming In Statistics, John Wiley and Sons, New York, 1981.

[5] B.S. Atal and S.L. Hanauer, "Speech Analysis And Synthesis By Linear Prediction Of The Speech Wave," J. Acous. Soc. Am., Vol. 50, pp. 637-655, 1971.

[6] I. Barrodale and R.E. Erickson, "Algorithms for least-squares linear prediction and maximum entropy spectral analysis - Part I: Theory and Part II: FORTRAN Program," Geophys., vol. 45, pp. 405-446, Mar. 1980.

[7] I. Barrodale, F.D.K. Roberts, and D.R. Hunt, "Computing Best L_p Approximations by Functions Nonlinear In the Parameter," Computer J., Vol. 13, No. 4, p. 382-386, November 1970.

[8] I. Barrodale and F.D.K. Roberts, "Applications of Mathematical Programming to L_p Approximation," in Nonlinear Programming, Rosen, Mangasarian, and Ritter, Eds, Academic Press, New York, 1970.

[9] I. Barrodale and F.D.K. Roberts, "Solution of An Overdetermined System of Equations in the L_1 Norm," Comm. ACM, Vol. 17, No. 6, pp. 319, June 1974.

[10] I. Barrodale and F.D.K. Roberts, "An Efficient Algorithm For Discrete L_1 Linear Approximation With Linear Constraints," SIAM J. Numer. Anal., Vol. 15, No. 3, pp. 611, June 1978.

[11] I. Barrodale and A. Young, "Algorithms For Best L_1 and L_∞ Linear Approximations On A Discrete Data Set," Numer. Math., Vol. 8, pp. 295-306, 1966.

[12] R.H. Bartels, A.R. Conn, and J.W. Sinclair, "Minimization Techniques For Piecewise Differentiable Functions: The L_1 Solution to an Overdetermined Linear System," SIAM J. Numer. Anal., Vol. 15, No. 2, pp. 224-241, April 1978.

[13] C. Bingham, M.D. Godfrey, and J.W. Tukey, "Modern techniques of power spectrum estimation," IEEE Trans. Audio Electroacoust., vol. AU-15, pp. 56-66, June 1967.

[14] R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra From the Point of View of Communications Engineering. New York: Dover, 1959.

[15] G.E.P. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day, 1970.

[16] Bracewell, R., The Fourier Transform and its Applications, NY: McGraw-Hill, 1964.

[17] E.O. Brigham and R.E. Morrow, "The fast Fourier transform,"IEEE Spectrum, vol. 4, pp. 63-70, Dec. 1967.

[18] E. Oran Brigham, The Fast Fourier Transform. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[19] W.L. Brogan, Modern Control Theory, Quantum Publishers, Inc., New York, NY, 1974.

[20] Brunol, J. and P. Chavel, "Fourier Transformation of Rotationally Invariant two-Variable Functions: Computer Inplementation of the Hankel Transform", Proc. IEEE, Vol. 65, pp. 1089-1090, 1977.

[21] J.P. Burg, "Three-Dimensional Filtering With An Array Of Seismometers," Geophysics, Vol. 29, No. 5, pp. 693-713, October 1964.

[22] J.P. Burg, "Maximum entropy spectral analysis," In Proc. 37th Meeting Society of Exporation Geophysicists (Oklahoma City, OK), Oct. 31, 1967.

[23] ----, "A new analysis technique for time series data," NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, Aug. 12-23, 1968.

[24] ----, "New concepts in power spectral estimation," in Proc. 40th Annu. Int. Society of Exploration Geophysicists (SEG) Meeting (New Orleans, LA) Nov. 11, 1970.

[25] ----, "The relationship between maximum entropy and maximum likelihood spectra," Geophys., vol. 37, pp. 375-376.

[26] R.H. Byrd and D.A. Payne, "Convergence of the Iteratively Reweighted Least Squares Algorithm for Robust Regression," Tech. Report 313, Johns Hopkins University, June, 1979.

[27] J.A. Cadzow and K. Ogino, "Two-dimensional spectral estimation," IEEE Trans. Acoust., Speech Signal Proc., vol. ASSP-29, pp. 96-401, June 1981.

[28] Candel, S.M., "Dual Algorithms for Fast Calculation of the Fourier Bessel Transform," IEEE Trans ASSP, Vol. -29, pp. 963-972, 1981.

[29] Candel, S.M., "Fast Computation of Fourier-Bessel Transform, "Proc. IEEE ICASSP, Vol. 3, pp. 2076-2079, 1982.

[30] Candel, S.M., "An Algorithm for the Fourier-Bessel Transform," Comp. Phys. Comm., Vol. 23, pp. 343-353, 1981.

[31] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," Proc. IEEE, Vol. 57, No. 8, pp. 1408-1418. August 1969.

[32] J. Capon, R.J. Greenfield, and R.J. Kolker, "Multidimensional Maximum Likelihood Processing Of A Large Aperture Seismic Array," Proc. IEEE, Vol. 55, pp. 192-211, 1967.

[33] Cavanagh, E. and B. Cook, "Numberical Evaluation of Hankel Transforms Via Gaussian-Laguerre Polynomial Expansions, "IEEE Trans. ASSP, Vol. ASSP-28, pp. 361-366, 1979.

[34] R. Chellapa and G. Sharma, "Two-Dimensional Spectral Estimation Using Spatial Autoregressive Models," Proc. ICASSP '83, Boston, MA, pp. 855-857.

[35] Chen, C.S., and K. Gopalan, "An Experiment of Fourier-Bessel Decomposition of Speech Signal, ASSP Spectrum Estimation Workshop, Nov. 1983, p. 93.

[36] C.H. Cheng, and M.N. Toksoz, "Determination of Shear Wave Velocities In Slow Formations," SPWLA Tenety-Fourth Annual Logging Symposium, June 23-26, 1981.

[37] S.T. Chen and D.E. Willen, "Shear Wave Logging In Slow Formations," SPWLA Twenty-Fifth Annual Logging Symposium, June 10-13, 1984.

[38] W.Y. Chen and G.R. Stegen, "Experiments with maximum entropy power spectra of sinusoids," J. Geophys. Re., vol. 79. no. 20, pp. 3019-3022, July 1974.

[39] J.F. Claerbout, Fundamentals of Geophysical Data Processing. New York: McGraw-Hill: 1976.

[40] J.F. Claerbout and R. Muir, "Robust Modeling with Erratic Data," Geophysics, Vol. 38, No. 5, pp. 826-844, October 1973.

[41] J.W. Cooley and J.W Tukey, "An Algorithm For the Machine Computation of Complex Fourier Series," Math Comp., Vol 19, No. 2, pp. 297-301, April 1965. Reprinted in Digital Signal Processing,

L.R. Rabiner and C.M. Rader, Eds., pp. 223-227, New York: IEEE Press, 1972.

[42] J.R. Dennis and S.Y. Wang, "Real Time Shear Wave Logging," SPWLA Twenty-Fifth Annual Logging Symposium, June 10-13, 1984.

[43] B.W. Dickinson, "Two-Dimensional Markov Spectrum Estimates Need Not Exist," IEEE Trans. On Information Theory, IT-26, No. 1, pp. 120-129, January, 1980.

[44] F. Durbin, "The fitting of time series models," Rev. Inst. Int. de Stot., vol. 28, pp. 233-244, 1960.

[45] T.S. Durrani and R. Chapman, "Eigenfilter Methods For 2-D Spectral Estimation," Proc. ICASSP '83, pp. 863-866.

[46] H. Ekblom and S. Henriksson, "L_p-Criteria for the Estimation of Local Parameters," SIAM J. Appl. Math., Vol. 17, No. 6, pp. 1130-1141, November 1969.

[47] A.R. Figueiras-Vidal, J.R. Casar-Corredera, R. Garcia-Gomez, and J.M. Paez-Borrallo, "L_1 Norm Versus L_2 Norm minimization in parametric spectral analysis: a general discussion," Proc. ICASSP '85, Tampa, FL, pp. 304-307, 1985.

[48] R. Fletcher, J.A. Grant, and M.D. Hebden, "The Calculation of Linear Best L_p Approximations," Computer J., Vol. 14, No. 3, 1971.

[49] P.F. Fougere, "Spectrum Model-Order Determination Via Significant Reflection Coefficients," Proc. ICASSP ' 85, Tampa, FL, pp. 1345-1347.

[50] P.F. Fougere, E.J. Zawalick, H.R. Radoski, "Spontaneous line splitting in maximum entropy power spectrum analysis,: Phys. of the Earth and Planetary Inter., vol. 12, pp. 201-207, 1976.

[51] O.L. Frost and T.M. Sullivan, "High resolution two-dimensional spectral analysis,: in Proc. ICASSP 79 (Washington, DC, Apr. 1979) pp. 673-676.

[52] W.F. Gabriel, "Spectral analysis and adaptive array super-resolution techniques," Proc. IEEE, vol. 68, pp. 654-666, June 1980.

[53] R. Garcia-Gomez, and J.M. Alcazar-Fernandez, "A linear programming approach to multipulse speech coding," Proc. ICASSP '85, Tampa, FL, pp. 953-956, 1985.

[54] Gerardi, F.R., "Application of Mellin and Hankel Transform to Networks with Time-Varying Parameters," IRE Trans. On Circuit Theory, pp. 197-207, June 1959.

[55] W. Gersch and J. Yonemoto, "Automatic classification of EEG's: A parametric model new features for classification approach," in Proc. 1977 Joint Automatic Control Conf. (San Francisco, CA) June 22-24, 1977, pp. 762-769.

[56] J. Gibson, S. Haykin, and S.B. Kesler, "Maximum entropy (adaptive) filtering applied to radar clutter,: in Rec. 1979 IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 166-169.

[57] Gopalan, K. and C.S. Chen, "Numerical Evaluation of Fourier-Bessel Expansion," ICASSP, April 14-16, 1983, Boston, MA.

[58] P.E. Green, R.A. Grosch, and C.F. Romeny "Principles Of An Experimental Large Aperture Seismic Array (LASA)," Proc. IEEE, vol. 53, No. 12, pp. 1821-1833, December 1965.

[59] L.J. Griffiths, "Rapid Measurement of Digital Instantaneous Frequency," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol ASSP-23, pp. 207-22, April 1975.

[60] L.J. Griffiths and R. Prieto-Diaz, "Spectral analysis of natural seismic events using autoregressive techniques," IEEE Trans. Geosci. Electron., vol. GE-15, pp. 13-25, Jan. 1977.

[61] P.R. Gutowski, E.A. Robinson, and S. Treitel, "Spectral Estimation: Fact or Fiction," IEEE Trans. On Geoscience Electronics, Vol. GE-16, No. 2, pp. 80-84, April 1978.

[62] C.S. Hacker, "Autoregressive and transfer function models of mosquito populations," in Time Series and Ecological Processes, H.M. Shugat, Ed. Boulder, CO: SIAM, 1978, pp. 294-303.

[63] O.S. Halpeny and D.G. Childers, "Composite Wavefrom Decomposition Via Multidimensional Digital Filtering Of Array Date," IEEE Trans. On Circuits And Systems, Vol. CAS-22, No. 6, pp. 552-562, June 1975.

[64] S.S. Haykin, Ed., Nonlinear Methods of Spectral Analysis. New York: Springer-Veriag, 1979.

[65] M.T. Heideman, D.H. Johnson, and C.S. Burrus, "Gauss And The History of The Fast Fourier Transform," IEEE ASSP Magazine, pp. 14-21, October, 1984.

[66] F.B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill, New York, 1974.

[67] S. Holm and J.M. Hovem, "Estimation of scalar ocean wave spectra by the maximum entropy method," IEEE J. Ocean. Eng., Vol. OC-4, pp. 76-83, July 1979.

[68] M. Hsu, "Maximum entropy principle and its application to spectral analysis and image reconstruction," Ph.D. dissertation, Ohio State Univ., 1975.

[69] T.S. Huang, "Two-dimensional windows," IEEE Trans. Audio Electroacoust., vol AU-20, pp. 88-90, Mar. 1972.

[70] P.J. Huber, Robust Statistics, John Wiley and Sons, New York, 1981.

[71] J.D. Ingram, C.F. Morris, E.E. Macknight, and T.W. Parks, "Direct Phase Determination of Shear Velocities From Acoustic Waveform Logs," Fifty-First SEG Meeting, Los Angeles, California, Octover 11-15, 1981.

[72] L.B. Jackson and H.C. Chien, "Frequency and bearing estimation by two-dimentional linear prediction," in Proc. ICASSP 79 (Washington, DC, Apr. 1979), pp. 665-668.

[73] P.L. Jackson, L.S. Joyce, and G.B. Feldkamp, "Application of Maximum Entropy Frequency Analysis to Synthetic Aperture Radar," Proc. RADC Spectrum Estimation Workshop, Rome, NY, 1978, pp. 217-225.

[74] E.T. Jaynes, "On The Rationale of Maximum Entropy Methods," Proc. IEEE, Vol. 70, No. 9, pp. 939-952, September 1982.

[75] G.M. Jenkins and D.G. Watts, Spectral Analysis and Its Applications. San Francisco, CA: Holden-Day, 1968.

[76] Johansen, H.K. and L. Sorensen, "Fast Hankel Transforms," Geophys. Prospecting, Vol. 27, pp. 876-901, 1979.

[77] R.H. Jones, "Autoregressive Order Selection," Geophysics, Vol. 41, No. 4, pp. 771-773, August 1976.

[78] L.S. Joyce, "A separable 2-D autoregressive spectral estimation algorithm," in Proc. ICASSP 79 (Washington, DC, Apr. 1979), pp. 677-680.

[79] S.M. Kay and S.L. Marple Jr., "Spectrum Analysis - A Modern Perspective," Proc. IEEE, Vol. 69, No. 11, November 1981.

[80] S.M. Kay and S.L. Marple, "Sources of and remedies of spectrum line-splitting in AR spectrum analysis," in Proc. IEEE Int. Conf. ASSP, pp. 151-154, 1979.

[81] A.H. Kayran, S.R. Parker, and D.J. Klich, "Two-Dimensional Spectral Estimation With Autoregressive Lattice Parameters,: Proc. ICASSP '84, pp. 4.1.1-4.1.2.

[82] S.B. Kesler and S.S Haykin, "The maximum entropy method applied to the spectral analysis of radar clutter," IEEE Trans. Inform. Theory, vol. IT-24, pp. 269-272, Mar. 1978.

[83] A. Ya. Khinchin, "Korrelationstheorie der Stationaren Stochastischen Prozess," Math. Annalen, vol. 109, pp. 604-615, 1934.

[84] C. Kimball and T. Marzetta, "Semblance Processing of Borehole Acoustic Array Data," Geophysics, Vol. 49, no. 3, March 1984, pp. 274-281.

[85] V.C. Klema and A.J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," IEEE Trans. Automatic Control, Vol. AC-25, No. 2, pp.164-176, April 1980.

[86] R. Kumaresan, "Estimating The Parameters of Exponentially Damped/Undamped Sinusoids In Noise," Ph.D. dissertation, University of Rhode Island, Kingston, RI, August 1982.

[87] R. Kumaresan and D.W. Tufts, "Improved Spectral Resolution III: Efficient Realization," Proc. IEEE, Vol. 68, No. 10, pp. 1354-1355, Octover 1980.

[88] R. Kumaresan and D. Tufts, "A Two-Dimensional Technique for Frequency-Wavenumber Estimation," Proc. IEEE, Vol. 69, No. 11, pp. 1515-1517, November 1981.

[89] R. Kumaresan and D.W. Tufts, "Estimating The Angles of Arrival of Multiple Plane Waves," IEEE Trans. On Aerospace and Electronic Systems, Vol. AES 19, No. 1, pp. 134-139, January 1983.

[90] R.T. Lacoss, "Data adaptive spectral analysis methods," Geophysics, Vol. 36, pp. 661-675, Aug. 1971.

[91] R.T. Lacoss, E.J. Kelly, and M.N. Toksoz, "Estimation of Seismic Noise Structure Using Arrays," Geophysics, Vol. 34, No. 1, pp. 21-38, February 1969.

[92] T.E. Landers and R.T. Lacoss, "Some geophysical applications of autoregressive spectral estimates," IEEE Trans. Geosci. Electron., vol. GE-15, pp. 26-32, Jan. 1977.

[93] S.W. Lang, "Spectral Estimation For Sensor Arrays," Ph.D. Dissertation, M.I.T., Combridge, MA, August, 1981.

[94] S.W. Lang and J.H. McClellan, "A simple proof of stability for all pole linear prediction models," Proc. IEEE, Vol. 67, No. 5, pp. 860-861, May 1979.

[95] S.W. Lang and J.H. McClellan, "Spectral estimation for sensor arrys,: in Proc. 1st ASSP Workshop on Spectral Estimation (Hamilton, Ont., Canada, Aug. 1981), pp. 3.2.1-3.2.7.

[96] ----, "MEM spectral estimation for sensor arrays," in Proc. Int. Conf. on Digital Signal Processing (Florence, Italy, Sept. 1981), pp. 383-390.

[97] ----, "The extension of Pisarenko's method to multiple dimensions," in Proc. ICASSP 82 (Paris, France, May 1982), pp. 125-128.

[98] S.W. Lang and J.H. McClellan, "Multidimensional MEM Spectral Estimation," IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP-30, No. 6, pp. 880-887, December 1982.

[99] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems, Englewood Cliffs, NJ, Prentice-Hall, 1974.

[100] N. Levinson, "The Wiener (root mean square) error criterion in filter design and prediction," J. Math Phy., vol. 25, pp. 261-178, 1947.

[101] S. Levy, C. Walker, T. Ulrych, and P.K. Fullagor, "A Linear Programming Approach to the Estimation of the Power Spectra of Harmonic Processes," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-30, No. 4, pp. 675-679, August 1982.

[102] J.S. Lim and N.A. Malik, "A new algorithm for two-dimensional maximum entropy power spectrum estimation," IEEE Trans. Acoust., Speech, Signal Proc., vol. ASSP-29, pp. 401-413, June 1981.

[103] Luke, Y.L., Integrals of Bessel Functions, NY: McGraw-Hill, 1962.

[104] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, vol. 63, No. 4, pp. 561-580, April, 1975.

[105] R.J. Mammone, "Spectral extrapolation of contrained signals," J. Opt. Soc. Am., Vol 73, No. 11, November, 1983.

[106] R.J. Mammone, K. Wang, and S. Gay, "LPC speech analysis using the L_1 norm," Proc. ICASSP '85, Tampa, FL, pp. 485-488, 1985.

[107] J.D. Markel and A.H. Gray Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.

[108] L. Marple, Jr., "Frequency resolution of high resolution spectrum analysis techniques," in Proc. 1st RADC Spectrum Estimator Workshop, pp. 19-35, 1975.

[109] L. Marple, "A new autoregressive spectrum analysis algorithm," IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-26, no. 4, Aug. 1980.

[110] S.L. Marple, Jr., "A Fast Computational Algorithm For And Performance Of THe Kumaresan-Prony Method of SPectrum Analysis," Proc. ICASSP '83, Boston, MA, pp. 1419-1421, 1983.

[111] T. L. Marzetta, "A linear prediction approach to two-dimensional spectral factorization and spectral estimation," Ph.D. dissertation, M.I.T., Combridge, MA, Feb. 1978.

[112] J.H. McClellan, "Multidimensional Spectral Estimation," Proc. IEEE, Vol. 70, No. 9, pp. 1029-1039, September 1982.

[113] J.H. McClellan and S.W. Lang, "Multidimensional MEM Spectral Estimation," presented at the Int. Conf. on Spectral Analysis and its Use In Underwater Acoustics, London, England, April, 1982.

[114] G.F. McCormick and V.A. Sposito, "Using the L_2 Estimator in L_1 Estimation," SIAM J. Numer. Analy., Vol. 13, No. 3, pp. 337-343, June 1976.

[115] R.N. McDonough, "Maximum entropy spatial processing of array data," Geophysics, vol. 39, pp. 843-851, Dec. 1974.

[116] N. Miao and Z.Z. Chen, "Application of SVD to 2-D Spectral Estimation," Proc. ICASSP '84, pp. 4.2.1-4.2.4.

[117] J.W. Minear and C.R. Fletcher, "Full-Wave Acoustic Logging," SPWLA Twenty-Fourth Annual Logging Symposium, June 27-30, 1983.

[118] M. Morf, A. Vieira, D.T.L. Lee, and T. Kailath, "Recursive Multichannel Maximum Entropy Spectral Estimation," IEEE Trans. On Geoscience Electronics, Vol. GE-16, No. 2, April 1978.

[119] S.H. Nawab, F.U. Dowla, and R.T. LaCoss, "A New Method For Wideband Sensor Array Processing," Proc. ICASSP '84, pp. 237-240.

[120] C.L. Nikias and P.D. Scott, "Improved spectral resolution by energy-weighted prediction method," Proc. IEEE Int. Conf. ASSP, vol. 2, pp. 469-499, Mar. 1981.

[121] C.L. Nikias and M.R. Raghuveer, "A New Class of High Resolution and Robust Multidimensional Spectral Estimation Algorithms," Proc. ICASSP '83, pp. 859-862.

[122] C.L. Nikias, and P.D. Scott, "The Covariance Least - Square Algorithm For Spectral Estimation of Processes of Short Data Length," IEEE Trans. on Geoscience and Remote Sensing, Vol. Ge-21, No. 2, pp. 180-190, April 1983.

[123] C.L. Nikias, P.D. Scott, and J.H. Siegel, "A New Robust 2-D Spectral Estimation Method and its Application in Cardiac Data Analysis," Proc. ICASSP '82, pp. 729-731.

[124] R. Nitzberg, "Spectral estimation: An impossibility?" Proc. IEEE, vol. 67, pp. 437-439, Mar. 1979.

[125] Nunez, P.L., "Representation of Evoked Potentials by Fourier-Bessel Expansion," IEEE Trans. Biomedical Eng., Vol. BME-20, pp. 372-374, 1973.

[126] A.H. Nuttal, "Spectral estimation by means of overlapped fast Fourier transform processing of windowed data," NUSC Tech. Rep. 4169, New London, CT, Oct. 13, 1971.

[127] ----, "Spectral analysis of a univariate process with bad data points, via maximum entropy, and linear predictive techniques," Naval Underwater Systems Center, Tech. Rep. 5303, New London, CT, Mar. 26, 1976.

[128] Oppenheim, A.V., G.V. Frisk, and D.R. Martinez, "An Algorithm for the Numberical Evaluation of the Hankel Transform,: Proc. IEEE, Vol. 66, pp. 264-265, 1978.

[129] Oppenheim, A.V., G.V. Frisk, and D.R. Martinez, "Computation of the Hankel Transform Using Projections," J. Acoustical Soc. Amer., Vol. 68, pp. 523-529, 1980.

[130] Papoulis, A., Signal Analysis, NY: McGraw-Hill, 1977.

[131] T.W. Parks, C.F. Morris, and J.D. Ingram, "Velocity Estimation From Short - Time Temporal And Spatial Frequency Estimates," Proc. ICASSP '82, pp. 399-402, 1982.

[132] T.W. Parks, J.H. McClellan, and C.F. Morris, "Algorithms For Full-Waveform Sonic Logging," ASSP Spectrum Estimation Workship II, November 1983, Tampa, Florida, pp. 186-191.

[133] E. Parzen, "Mathematical considerations in the estimation of spectra," Technometrics, vol. 3, pp. 167-190, May 1961.

[134] ----, "Statistical spectral analysis (single channel case) in 1968," De. Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 11, June 10, 1968.

[135] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, New Yor, 1978.

[136] D.V.B. Rao and S.Y. Kung, "A State Space Approach For The 2-D Harmonic Retrieval Problem," Proc. ICASSP '84, pp. 4.10.1-4.10.3.

[137] E.A. Robinson, "Predictive decomposition of time series with application to seismic exploration," Geophysics, col. 32, pp. 418-484, June 1967.

[138] E.A. Robinson, "A Historical Perspective of Spectrum Estimation," Proc. IEEE, Vol. 70, No. 9, pp. 885-907, September 1982.

[139] E.A. Robinson and S. Treitel, Geophysical Signal Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[140] J.B. Rosen, O.L. Mangasarian, and K. Ritter, Eds., Nonlinear Programming, Academic Press, New York, 1970.

[141] S. Roucos and D.G. Childers, "A Two-Dimensional Maximum Entropy

Spectral Estimator," Proc. ECASSP, 1979, pp. 669-672.

[142] M.R. Schroeder, "Linear Prediction, Entropy, and Signal Analysis," IEEE ASSP Magazine, pp. 3-26, July 1984.

[143] J. Schroeder and R. Yarlagadda, "Two Dimensional Linear Predictive Spectral Estimation Via the L_1 Norm,' Submitted to IEEE Trans. Acoustics, Speech, and Signal Processing.

[144] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," Terrestrial Magnetism, vol. 3 pp. 13-41, Mar. 1898.

[145] P.D. Scott and C.L. Nikias, "Energy-weighted linear predictive spectral estimation: A new method combining robustness and high resolution," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, no. 2, Apr. 1982.

[146] Siegman, A.E., "Quasi Fast Hankel Transform," Opt. Letters, Vol. 1, pp. 13-15.

[147] G. Sharma and R. Chellappa, "An Iterative Algorithm for Robust 2-D Spectrum Estimation," Proc. ICASSP '84, pp. 4.4.1-4.4.4.

[148] G. Sharma and R. Chellappa, "A Model Based Approach For 2-D MEPS Analysis," Proc. ICASSP, 1984, pp. 4.5.1-4.5.4.

[149] Slepian, D., H.J. Landau, and H.O. Pollack, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty Principle I and II," Bell Systems Tech., J., Vol. 40, No. 1, pp. 43-84.

[150] Sneddon, I.N., Special Function of Mathematical Physics and Chemistry, NY: Interscience Published, Inc., 1956.

[151] O.N. Strand, "Multichannel Complex Maximum Entropy (Autoregressive) Spectral Analysis," IEEE Trans. On Automatic Control, Vol. AC-22, No. 4, August, 1977, pp. 634-640.

[152] D.N. Swingler, "Frequency errors in MEM processing," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, no. 2, pp. 257-259, Apr. 1980.

[153] D.N. Swingler, "A comparison between Burg's maximum entropy method and nonrecursive technique for the spectral analysis of deterministic signals," J. Geophys. Res., vol. 84, no. Ba, pp. 679-685, Feb. 1979.

[154] M.T. Taner and F. Koehler, "Velocity Spectra-Digital Computer Derivation and Applications of Velocity Functions," Geophysics, Vol. 34, pp. 859-881.

[155] C.W. Therrien, "Relations between 2-D and multichannel linear prediction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-29, pp. 454-456, June 1981.

[156] T. Thorvaldsen, A.T. Waterman Jr., and R.W. Lee, "Maximum entropy angular response patterns of microwave transhorizon signals," IEEE Trans. Autennas Propagat., Vol. AP-28, pp. 722-724, Sept. 1980.

[157] D. Tjøstheim, "Autoregressive Modelling and Spectral Analysis of Array Data in the Plane, IEEE Trans. On Geoscience And Remote Sensing, Vol. GE-19, No. 1, January 1981.

[158] Tranter, C.J., Bessel Function with some Physical Applications, London: The English University Press Ltd., 1968.

[159] L. Tsang and D. Rader, "Numerical Evaluation of the Transient Acoustic Waveform Due to a Point Source In a Fluid - Filled Borehole," Geophysics, Vol. 44, pp. 1706-1720, 1979.

[160] D.W. Tufts and R. Kumaresan, "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," Proc. IEEE, Vol. 70, No. 9, pp. 975-989, September 1982,

[161] D.W. Tufts and R. Kumaresan, "Singular Value Decomposition and Improved Frequency Estimation Using Linear Prediction," IEEE Trans. On Acoustics, Speech, and Signal Processing, Vol. ASSP=30, No. 4, pp. 671-657, August 1982.

[162] D.W. Tufts, and R. Kumaresan, and I. Kirsteins, "Data Adaptive Signal Estimation By Singular Value Decompostion of a Data Matrix," Proc. IEEE, Vol. 70, No. 6, pp. 684-685, June 1982.

[163] T.J. Ulrych and T.N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," Rev. Geophys., Vol. 13, No. 1, pp. 183-200, February 1975.

[164] T.J. Ulrych and R.W. Clayton, "Time Series Modelling and maximum Entropy," Phys. Earth Planet. Interiors, Vol. 12, pp. 188-200, August 1976.

[165] T.J. Ulrych and C.J. Walker, "High Resolution 2-Dimensional Power Spectral Estimation," in Applied Time Series Analysis II, Findley, Ed., New York: Academic, 1981, pp. 71-99.

[166] A. Van DenBos, "Alternative interpretation of maximum entropy spectral analysis," IEEE Trans. Inform. Theory, vol. IT-17, pp. 493-494, July 1971.

[167] G. Walker, "On periodicity in series of related terms," Proc. Roy. Soc. London, Series A, vol. 131, pp. 518-532, 1931.

[168] R.J. Wang and S. Treitel, "The determination of digital Wiener filters by means of gradient methods," Geophysics, Vol. 38, No. 2, pp. 310-326, April 1973.

[169] P.D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodogrms," IEEE Trans. Audio Electroacoust., vol. AU-15, pp. 70-73, June 1967.

[170] P.D. Welch, "On the variance of time and frequency averages over modified periodograms," in Rec. 1977 IEEE Int. Conf. Acoustics, Speech and Signal Processing (Hartford, CT), May 9-11, 1977, pp. 58-62.

[171] S.J. Wernecke and L.R. D'Addario, "Maximum entropy image reconstruction," IEEE Trans. Comput., vol. C-26, pp. 351-364, Apr. 1977.

[172] N. Wiener, "Generalized harmonic analysis," Acts Mathematics, vol. 55, pp. 117-258, 1930.

[173] R.A. Wiggins and E.A. Robinson, "Recursive solution to the multichannel filtering problem" J. Geophysical Res., vol. 70. pp. 1885-1891, Apr. 1965.

[174] M.E. Willis and M.N. toksoz, "Automatic P and S Velocity Determination From Full Waveform Digital Acoustic Logs," Geophysics, vol. 48, no. 12, pp. 1631-1644, December 1983.

[175] J.W. Woods, "Two dimensional discrete Markovian fields," IEEE Trans. Inform. Theory, vol. IT-18, pp. 232-240, Mar. 1972.

[176] ----, "Two-dimensional Markov spectral estimation," IEEE Trans. Inform. Theory, vol. IT-22, pp. 552-559, Sept. 1976.

[177] J.W. Woods and P.R. Lintz, "Plane Waves at Small Arrays," Geophysics, vol. 38, no. 6, pp. 1023-1041, December 1973.

[178] R. Yarlagadda, J.B. Bednar, and T.L. Watt, "Fast Algorithms for L_p Deconvolution," IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No. 1, pp. 174-182, February 1985.

[179] G.U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," Phillosophical Trans. Royal Soc. London, Series A, vol. 226, pp. 267-298, July 1927.

[180] L. Zou and B. Lin, "Improvement of Resolution and Reduction of Computation In 2-D Spectral estimation Using Decimation," Proc. ICASSP '84, pp. 4.7.1-4.7.3.

APPENDIXES

APPENDIX A

In this appendix an equation relating the singularity condition of the linear prediction matrix to sinusoidal frequency location will be derived. With a deterministic test signal (noiseless case) given by

$$y(n,m) = \exp[j(\omega_1 n + \omega_2 m)] + \exp[j(\alpha\omega_1 + \beta\omega_2)], \tag{A.1}$$

where α and β are constants relating the sinusoids locations, choose $n=n_1$, $m=m_1$ for one sinusoid and $n=n_2$, $m=m_2$ for the other sinusoid location. From equation (5.3), it is seen that the linear prediction matrix is

$$Y = \begin{bmatrix} \cdots\ y(n_1,\ m_1)\ \cdots\ y(n_2,\ m_2)\ \cdots \\ \cdots\ y^*(N\text{-}1\text{-}n_1,\ M\text{-}1\text{-}m_1)\ \cdots\ y^*(N\text{-}1\text{-}n_2,\ M\text{-}1\text{-}m_2)\ \cdots \end{bmatrix}, \tag{A.2}$$

and the 2x2 partition of Y, denoted here as matrix P is

$$P = \begin{bmatrix} y(n_1,\ m_1) & y(n_2,\ m_2) \\ y^*(N\text{-}1\text{-}n_1,\ M\text{-}1\text{-}m_1) & y^*(N\text{-}1\text{-}n_2,\ M\text{-}1\text{-}m_2) \end{bmatrix}. \tag{A.3}$$

In order to arrive at the desired result, we set det (P) = 0, that is

$$y(n_1,\ m_1)y^*(N\text{-}1\text{-}n_1,\ M\text{-}1\text{-}m_1) - y(n_2,m_2)y^*(N\text{-}1\text{-}n_1,\ M\text{-}1\text{-}m_1) = 0 \tag{A.4}$$

with y(n,m) evaluated from equation (A.1). Although the algebra is somewhat messy, straightforward manipulations lead to

$$\exp(j\theta_1) - \exp(j\theta_2) + \exp(j\theta_3) - \exp(j\theta_4) = 0$$

with

$$\theta_1 = \omega_1[n_1 + \alpha n_2 + \alpha(1 - N)] + \omega_2[m_1 + \beta m_2 + \beta(1 - M)]$$

$$\theta_2 = \omega_1[n_1 + \alpha n_2 + (1 - N)] + \omega_2[m_1 + \beta m_2 + (1 - M)]$$

$$\theta_3 = \omega_1[\alpha n_1 + n_2 + (1 - N)] + \omega_2[\beta m_1 + m_2 + (1 - M)]$$

$$\theta_4 = \omega_1[\alpha n_1 + n_2 + \alpha(1 - N)] + \omega_2[\beta m_1 + m_2 + \beta(1 - M)]. \qquad (A.5)$$

Solving Equation (A.5) and equating real and imaginary parts results in the set of equations

$$\cos(\theta_1) + \cos(\theta_3) = \cos(\theta_2) + \cos(\theta_4)$$

$$\sin(\theta_1) + \sin(\theta_3) = \sin(\theta_2) + \sin(\theta_4) , \qquad (A.6)$$

with θ_i defined as in Equation (A.5), that must be satisfied in order that det(P) = 0. By application of the trigonometric identity

$$\cos(A) + \cos(B) = 2\cos[1/2(A + B)] \cos [1/2(A - B)] \qquad (A.7)$$

to the left and right side of the first Equation in (A.6) the following condition for det (P) = 0 results in

$$\omega_2(\beta - 1)(M - 1) = \omega_1(1 - \alpha)(N - 1) \pm l\, 2\pi, \qquad (A.8)$$

$$l = 0, 1, 2, \ldots$$

Upon substituting $\omega = 2\pi f$ into Equation (A.8) an equivalent expression is obtained:

$$(\beta - 1)(M - 1)f_2 + (\alpha - 1)(N - 1)f_1 = \text{an integer} . \qquad (A.9)$$

Since $f_3 = \alpha f_1$ and $f_4 = \beta f_2$, Equation (A.9) may also be written

$$(M-1)(f_4 - f_2) + (N-1)(f_3 - f_1) = \text{an integer.} \qquad \text{(A.10)}$$

Solving the second equation in Equation set (A.6) involving the sin(•) function leads to the same results summarized in Equations (A.8, A.9, or A.10) and will not be repeated here. Thus, Equations (A.8, A.9, or A.10) may be used to predict the combinations of f_1, f_2, f_3, f_4 that will to a singular linear prediction equation matrix. In practice, however, with noise added, the singularity problem has not been noted.

APPENDIX B

In this appendix, the separability of a discrete Fourier transform (DFT) and the linear prediction spectral estimation method as applied to a sum of two-dimensional sinusoids. Although this derivation will consider a test signal composed of two sinusoids for algebraic and notational simplicity, the extension to multiple sinusoids is striaghtforward. Essentially three basic steps are required to complete the derivation: first, with the two-dimensional signal considered as an N x M matrix, N distinct M point DFT's will be applied along the data matrix rows; second, closed form expressions are developed for the M point summations that result from application of the DFT; and third, after application of a DFT to the rows of the data matrix plus appropiate simplification, the resulting two-dimensional function will be shown to have significant sinusoidal components located only at the frequency pairs defining that set of two-dimensional sinusoids of the test signal. Thus, a linear prediction spectral estimation algorithm (or any other method for the matter) may be applied to the columns of the complex DFT result and the pair of sinusoids will be correctly located in the frequency-wavenumber (F-K) plane. The derivation follows.

The pair of two-dimensional sinusoids are modelled as:

$$y(n,m) = \exp(j(\omega_3 n + \omega_2 m)) + \exp(j(\omega_3 n + \omega_4 m)), \qquad (B.1)$$

which, in the F-K plane, gives rise to a pair of impulses (or relative maxima with finite data records) at coordinates (ω_1, ω_2) and (ω_3, ω_4). A one-dimensional M-point DFT is given by

$$X(z) = \sum_{m=1}^{M} x(m)\, z^{-m}, \tag{B.2}$$

with $z = \exp(j\omega)$. Applying Equation (B.2) to the rows of the data matrix defined by Equations (B.1), $n = 1,2,\ldots, N$, $m = 1,2,\ldots, M$ results in

$$Y(n,\omega) = \sum_{m=1}^{M} (\exp(j(\omega_1 n + \omega_2 m)) + \exp(j(\omega_3 n + \omega_4 m))) \cdot \exp(-j\omega m). \tag{B.3}$$

Factoring out terms not involved in the summation defined by (B.3) and combining common exponentials leads to

$$Y(n,\omega) = \exp(j\omega,n) \cdot \sum_{m=1}^{M} \exp(jm(\omega_2 - \omega) + \exp(j\omega_3 n) \cdot \sum_{m=1}^{M} \exp(jm(\omega_4 - \omega)). \tag{B.4}$$

This completes the first step of the derivation and Equation (B.4) defines the M-point DFT along the rows of a data matrix described by $y(n,m)$. The complex exponential summations in (B.4) have a closed form that simplifies this development considerably. The following identity will be used:

$$\sum_{k=1}^{K} r^k = \frac{r - r^{k+1}}{1 - r}, \quad r \neq 1, \tag{B.5}$$

with $r = \exp(j(\omega_i - \omega))$.

Thus,

$$\sum_{m=1}^{M} \exp(jm(\omega_i - \omega)) = \frac{\exp(j(\omega_i - \omega)) - \exp(j(\omega_i - \omega))^{M+1}}{1 - \exp(j(\omega_i - \omega))} \qquad (B.6)$$

Equation (B.6) simplifies by re-expressing (B.6) in magnitude-phase format. The details are somewhat tedious, though straightforward, and therefore will be omitted. In polar form, then, Equation (B.6) becomes

$$\sum_{m=1}^{M} \exp(jm(\omega_i - \omega)) = \frac{\sin(M(\frac{\omega_i - \omega}{2}))}{\sin(\frac{\omega i - \omega}{2})} \exp(j(\frac{M+1}{2})(\omega_i - \omega)) \qquad (B.7)$$

Now Equation (B.4), which is the result of applying a DFT to the rows of $y(n,m)$, can be written

$$Y(n,\omega) = K_1 \exp(j\omega_1 n) + K_2 \exp(j\omega_3 n)$$

with

$$K_1 = \frac{\sin(M(\frac{\omega 2 - \omega}{2}))}{\sin(\frac{\omega 2 - \omega}{2})} e^{j\frac{M+1}{2}(\omega_2 - \omega)}$$

$$K_2 = \frac{\sin(M(\frac{\omega_4 - \omega}{2}))}{\sin(\frac{\omega_4 - \omega}{2})} e^{j\frac{M+1}{2}(\omega_4 - \omega)} . \qquad (B.8)$$

The magnitude of K_i has the functional form of

$$K = \frac{\sin(Mx)}{\sin(x)} , \qquad (B.9)$$

which peaks sharply at x = 0 with value K = M; for $x \neq 0$, the magnitude falls off very rapidly. This will be illustrated with more clarity a bit later. First, however, we will expres $Y(n,\omega)$ from equation (B.8) with $\omega = \omega_2, \omega_4$, which are the columns of $Y(n,\omega)$ that define the two-dimensional sinusoids located in the F-K plane at (ω_1, ω_2) and (ω_3, ω_4). Picking these values of results in

$$Y(n,\omega_2) = M\, e^{j\omega_1 n} + K\, e^{j\omega_3 n}$$

and

$$Y(n,\omega_4) = K^*\, e^{j\omega_1 n} + M\, e^{j\omega_3 n} ,$$

with

$$K = \frac{\sin[M(\frac{\omega_4 - \omega_2}{2})]}{\sin(\frac{\omega_4 - \omega_2}{2})}\, e^{j\frac{M+1}{2}(\omega_4 - \omega_2)} , \qquad (B.10)$$

Note that K^* is the complex conjugate of K and $\omega_2 \neq \omega_4$. Next, a simple sketch will be given that will show $|k| << M$ for practical values of $(\omega_4 - \omega_2)$; it is useful to recall at this point that the frequency resolution imposed by application of a DFT is $\Delta\omega = 2\pi/M$. In Figure 76, $|k|/M$ is plotted as a function of $(\frac{\omega_4 - \omega_2}{2})$. As a practical matter, $(\frac{\omega_4 - \omega_2}{2})$ must be greater than the unit frequency resolution of 2 /M; that is

$$(\frac{\omega_4 - \omega_2}{2}) > \pi/M \qquad (B.11)$$

and it can be seen that $|k| << M$. Therefore, Equation (B.10) becomes

$$Y(n,\omega_2) \simeq M\, e^{j\omega_1 n}$$

$$Y(n,\omega_4) \simeq M\, e^{j\omega_3 n}\ , \qquad \text{(B.12)}$$

and any desired spectral estimation method may be applied to $Y(n,\omega_i)$ to locate the peaks at ω_1 and ω_3, which in the F-K plane results in a peak at (ω_1, ω_2) and (ω_3, ω_4). In this work, a linear prediction spectral estimator is applied to $Y(n, \omega_i)$, $n = 1,2,\ldots, N$; ω_i is defined by $\omega_i = k\, 2\pi/M$, $k = 0,1,\ldots, M/2 - 1$.

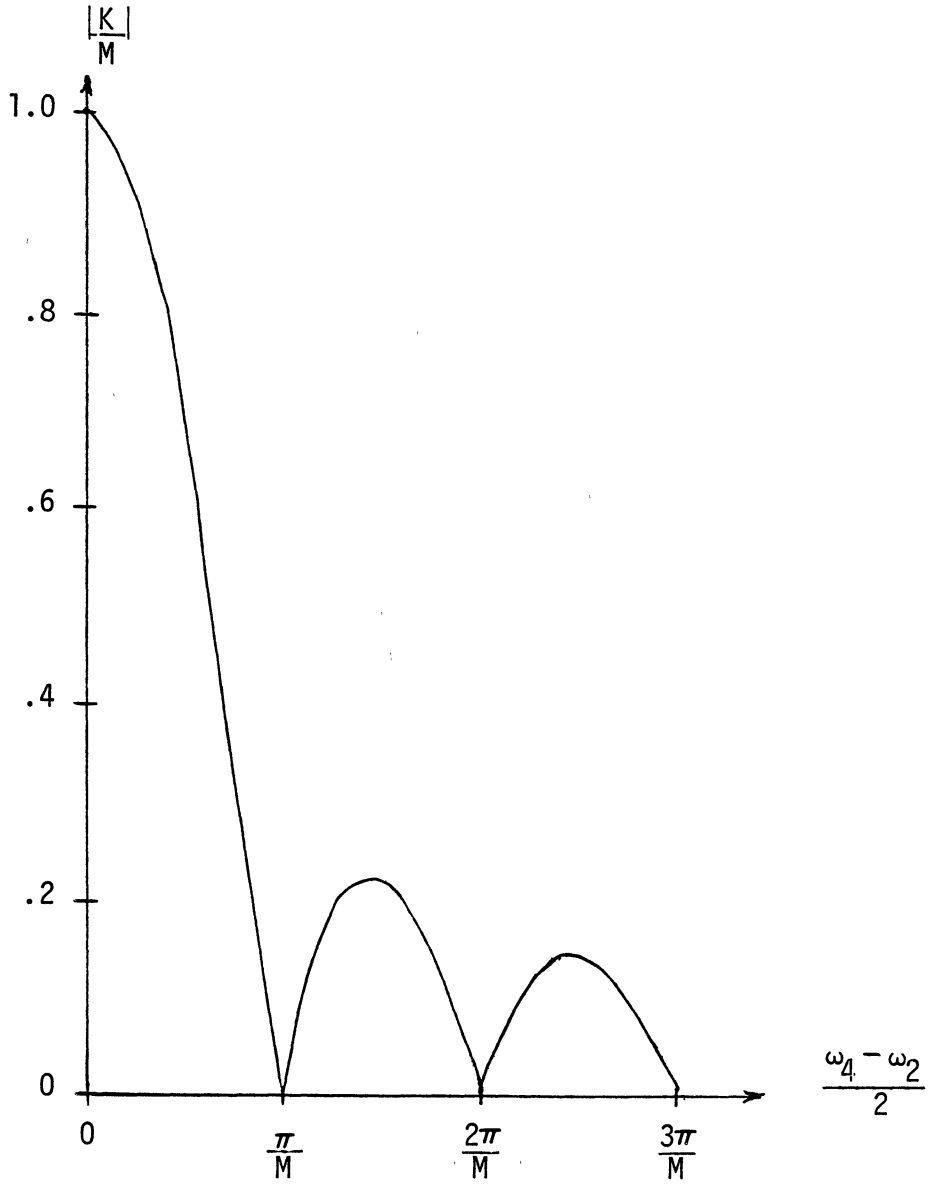

Figure 76. Sampling Function sin (mx)/sin (x)

APPENDIX C

Many physical problems of interest to the engineer or applied mathematician may be described by a system of equations such as described by Equation (C.1).

$$X \underline{a} = \underline{y} \tag{C.1}$$

In (C.1) X is a matrix and $\underline{a}$ and $\underline{y}$ are vectors; depending on the specific problem formulation, the resulting dimensions of X, $\underline{a}$, and $\underline{y}$ may lead to a system of equations that is underdetermined, overdetermined, or exactly determined. Thus with X a matrix of dimension mxn, $\underline{a}$ a vector of length n, and $\underline{y}$ a vector of length M, for $m > n$, the system of equations is overdetermined, for $m < n$ the system is underdetermined, and with $m = n$, the system is exactly determined. Obviously, in the exactly determined case, matrix X has full rank, i.e. rank X is $m = n$. For the underdetermined system, matrix X has rank m and for the overdetermined case matrix X is rank n. In this thesis in which Equation (C.1) is used to represent a linear predictive spectral estimation formulation vector $\underline{y}$ contains the "predicted" data values and vector $\underline{a}$ is the unknown prediction coefficients. Once the linear prediction coefficients have been estimated it is an easy matter to form a spectral estimate from the so called prediction error filter. In the deconvolution problem the prediction coefficients are utilized to form a filter that attempts to remove undesirable effects of the data collection system from the data samples. For the curve fitting problem, vector $\underline{a}$ may represent relative weights applied to a certain choice of

basis functions. Numerous examples may be listed, however, in all cases vector $\underline{a}$ needs to be calculated in some manner. Probably, the most common (and easiest) method of calculating $\underline{a}$ is by means of a least squares solution since analytic solutions are known and easily calculated. For an overdetermined system of equations, the solution to Equation (C.1) is given by

$$\underline{a} = (X^T X)^{-1} X^T \underline{y}. \tag{C.2}$$

For an underdetermined system of equations, $\underline{a}$ is known to be

$$\underline{a} = X(XX^T)^{-1} X^T \underline{y} \tag{C.3}$$

If the system of equations represented by (C.1) is exactly determined, the solution is simply

$$\underline{a} = X^{-1} \underline{y} \ . \tag{C.4}$$

Obviously, X^{-1}, $(X^T X)^{-1}$, and $(XX^T)^{-1}$ must exist or one needs to improve the numerical conditioning of the data before a solution is feasible. Although least squares solutions are quite popular and effective, the thrust of this thesis has been consideration of general L_p $(1 \leq p \leq 2)$ normed solution to Equation (C.1). Unfortunately, analytic solutions to Equation (C.1) for $p \neq 2$ are not possible, however, in simple cases the error minimization may be carried out explicitly for a general L_p $(1 \leq p \leq 2)$ norm. This will be demonstrated next for the simplest possible example.

In this example matrix X is 2x1, vector $\underline{a}$ is 1x1 (i.e. a scalar), and vector $\underline{y}$ is 2x1. A general L_p $(1 \leq p \leq \infty)$ normed solution to (C.1) may easily be calculated by taking the partial derivatives of an appropiate error measure, setting the derivates equal to zero, and solving the resultant equation(s) for p. With the dimensions of this example Equation (C.1) may be written

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad a \quad = \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{C.5}$$

which shows the components of X, a, and $\underline{y}$ explicitly facilitating the differentiation process. An L_p norm is calculated by minimizing an error measure, E, as given by

$$E = \sum_i ||\epsilon_i||^p \tag{C.6}$$

where the i^{th} component of the error residual is

$$\epsilon_i = (X\,\underline{a} - \underline{y})_i \,. \tag{C.7}$$

In this simple example

$$E = (X_1\, a - y_1)^p + (X_2\, a - y_2)^p \tag{C.8}$$

Taking the partial derivative of E with respect to a in (C.8) and setting the result equal to zero leads to

$$\frac{\partial E}{\partial a} = 0 = p(X_1\, a - y)^{p-1} X_1 + p(X_2\, a - y_2)^{p-1} X_2 \,. \tag{C.9}$$

Simplifying Equation (C.9) and solving for a results in

$$a = [X_1 X_i^{\frac{1}{p-1}} - X_2(-X_2)^{\frac{1}{p-1}}]^{-1} \cdot [X_1^{\frac{1}{p-1}} y_1 - (-X_2)^{\frac{1}{p-1}}\, y_2] \tag{C.10}$$

which is an L_p ($1 \leq p \leq \infty$) normed solution to Equation (C.5). Five special cases of p will now be considered: p = 2., p = , p = 3, p = 1.5, and p = 0. The case of p = 0 is not strictly valid since $p \neq 0$ is assumed in the calculations, however the result is interesting. Of course, p < 1 does not constitue a normed space as pointed out earlier in this thesis.

For p = 2, Equation (C.10) becomes

$$a = (X_1^2 + X_2^2)^{-1}(X_1\, Y_1 + X_2\, Y_2^2)$$

or

$$a = (X^T X)^{-1} X^T y \tag{C.11}$$

as expected and here mainly serves a partial verification check on the minimization calculations.

An $L_p(p = \infty)$ norm is also known as the Chebyshev norm and is commonly used in filter design. Statistician seem to favor the L_∞ norm in problems for which uniform noise is expected. A characteristic of the Chebyshev (L_∞) norm is an error vector exhibiting "equal ripple" and this will be clearly seen in this simple example. With $p = \infty$ Equation (C.10) is now

$$a = (X_1 - X_2)^{-1}(y_1 - y_2) \tag{C.12}$$

which, since the quantity $(X_1 - X_2)$ is a scalar, may be expressed

$$a = \frac{y_1 - y_2}{x_1 - x_2} \quad . \tag{C.13}$$

Note that

$$\epsilon_1 = X_1\, a - y_1 \tag{C.14}$$

$$\epsilon_2 = X_2\, a - y_2 \tag{C.15}$$

and Equation (C.13) results from setting $\epsilon_1 = \epsilon_2$ and solving for a; thus the "equal error (ripple)" characteristic of an L_∞ norm.

In case 3, p = 3, a difficulty is noted. The solution for a with p = 3 is

$$a = \frac{X_1^{1/2} y_1 - (-X_2)^{1/2} y_2}{X_1 (X_1)^{1/2} - X_2 (-X_2)^{1/2}} \tag{C.16}$$

and the square root of $(-X_2)$ may present a difficulty if a is required to be a real valued coefficient.

With p = 1.5, another peculiarity is noted. Now Equation (C.10) has solution

$$a = \frac{X_1^2 y_1 - X_2^2 y_2}{X_1^3 - X_2^3} . \qquad (C.17)$$

Obviously, from Equation (C.17) it can be seen that $X_1 \neq X_2$ to prevent a from going to infinity, a decidedly pathelogical solution.

Finally, in case 5, consider $p = 0$ (though $p \neq 0$ is assumed!), which leads to

$$a = 1/2(y_1/X_1 + y_2/X_2). \qquad (C.18)$$

Thus a is the average of the two possible solutions for a considering the rows of matrix X individually. What this means is an open question and one needs to be careful that no generalizations are attempted!

APPENDIX D

SIGNAL DECOMPOSITION

Introduction

Although this thesis was concerned with primarily linear predictive spectral estimation, this appendix is included to provide an alternative "spectral" characterization technique. Fourier series coefficients are known to provide a least squares fit to a function expanded into a series of sinusoidal basis functions and as such may be concerned an L_2 spectral estimation method. Other series, such as the Fourier-Bessel series considered in this appendix, thus may in this generalized spectral characteristic viewpoint be considered as a non sinusoidal "spectral" estimate. "Spectral" is enclosed in quotes to emphasize the fact that the physical interpretation of Fourier-Bessel coefficients is uncertain.

In many cases it may not be desireable or even practical to represent a signal by its sample values directly or by an analytical function if a suitable function is available. For example, a signal may be determined by time domain sample values when the parameters of interest are better describable within the frequency domain. Many practical signals are highly redundant, both image and speech signals fall into this category, and it may be desirable and possibly necessary to represent the signal with a fewer number of samples for economy of storage and/or transmission bandwidth limitations. Generally signals are

processed or filtered to "improve" low signal-to-noise ratios or to emphasize or de-emphasize certain characteristics. Voice signals, for example, possess information content within a relatively narrow range of frequencies and therefore can be successfully filtered to remove noise energy outside of the frequency range of interest. Whatever the desired goal the processing of signals can many times be carried out more efficiently in another domain than that of the original signal. An obvious example here with the advent of hardware Fast Fourier Transform (FFT) devices is the widespread frequency domain processing of naturally occurring time domain signals. The field of speech processing has benefited tremendously from the ability to represent voice signals in domains other than the time domain. Pattern recognition techniques rely on the ability to generate a set of coefficients from the raw data (time domain samples) that are more compact (i.e. fewer samples) and hopefully are more closely related to the signal characteristics of interest. Clearly, if one is interested in frequency content, a Fourier series representation packs the frequency information in to fewer samples (Fourier series coefficients) than a time domain representation. For these reasons, and many others not mentioned, the theory of signal decomposition by means of series representation is important to such applications as seismic, speech and image processing.

Series Representations

Possibly the first example of using a series representation dates back to 1753 when D. Bernoulli achieved success in expressing the functional form of a vibrating string as the series (Bracewell, 1964):

$$f(x,t) = A_1 \sin(x) \cos(at) + A_2 \sin(2x) \cos(2at) + \ldots$$

The idea of representing an arbitrary function as a sum of sinusoids was quite new and controversial at the time and opposition to the correctness of such a series or possibly any series was voiced by Euler and Lagrange. Lagrange, in fact, publicly disagreed with Fourier's claim in 1807 that series representations were possible and it was not until 1829 that rigorous treatment of this idea was initiated by Dirichlet. Well, as they say, the rest is history. Since those beginnings many important mathematical theories have resulted such as the invention of the Riemann integral and research continues unabated today. For example, the well known Dirichlet conditions are sufficient only and research continues for necessary and sufficient conditions.

The theory of series representation of an arbitrary signal is more general then expressing a signal as a sum of sinusoids. In fact, any orthonormal set of basis functions can be used to represent some arbitrary function. For example, if we define an orthogonal set of functions as follows:

$$\int_{-\infty}^{\infty} \phi_m(t)\ \phi_n(t)dt = \begin{matrix} 1, & m = n \\ 0, & m \neq n \end{matrix} \tag{D.1}$$

the function f(t) can be written

$$f(t) = \sum_{n=0}^{\infty} C_n\ \phi_n(t) \tag{D.2}$$

where

$$C_n = \int_{-\infty}^{\infty} f(t)\ \phi_n(t)dt \tag{D.3}$$

This result can easily be demonstrated (disregarding convergence considerations) by multiplying both sides of (2) by $\phi_n(t)$ and integrating each side over all values of time, t. If we restrict f(t) to signals pos-

sessing finite energy and bandlimited frequency spectra a useful property can be stated: The energy, E, of f(t) is given by

$$E = \int_{-\infty}^{\infty} f^2(t)dt = \sum_{n=0}^{\infty} C_n^2 < \infty \qquad (D.4)$$

This is merely a re-statement of Parseval's well known formula concerning Fourier series coefficients.

Although the generalized form of series representation is useful for the construction of mathematical proofs, we are more interested in specific choices of the basis function, $\phi_n(t)$. Obviously, choosing

$$\phi_n(t) = e^{jn\omega t} \qquad (D.5)$$

results in a Fourier series, but many other functions have found use also. If f(t) is only available over a finite segment of time $(-\tau, \tau)$, a realistic assumption, it may be desirable to concentrate the energy within this interval. Denoting the concentrated energy by the fractional energy ratio

$$E = \frac{\int_{-\infty}^{\infty} |\phi_n(t)|^2 \, dt}{\int_{-\infty}^{\infty} |\phi_n(t)|^2 \, dt}$$

it can be shown (Papoulis, 1977) that E is maximized for $\phi_n(t)$ corresponding to prolate spheroidal functions. This choice of basis functions has been investigated thoroughly by Slepien, et al. (1961) and has found use within many areas of digital signal processing and filter design.

Another possible choice for $\phi_n(t)$ is a family of Bessel Functions, which results in an expansion termed the Fourier-Bessel Series. In this case, choosing a zero order Bessel function for illustration,

$$\phi_n(t) = J_0(\lambda_n t) \tag{D.6}$$

and f(t) can then be expressed as

$$f(t) = \sum_{n=1}^{\infty} C_n J_0(\lambda_n t)$$

The Fourier-Bessel (F-B) series has found applications in optics and acoustics; additionally, certain unique properties of this series may be useful to the speaker identification problem and acoustical well log signal analysis. Properties of the F-B series will be covered in more detail within another section of this appendix.

A few possible choices of basis functions have been briefly mentioned, however generally speaking the choice is very problem specific. The exponential kernel is an obvious and attractive choice. For one thing, most people are quite familiar and comfortable with Fourier series theory; also, the Fourier series possesses some very nice analytical properties such as shift invariance that make the various mathematical manipulations inevitably required much simpler. Fourier series theory includes the happy result that the series coefficients are given by a discrete Fourier transform; thus coefficient generation is an easy process with the numerous FFT algorithms that abound. By contrast, prolate spheroidal functions are more complicated to generate and their use may be difficult to justify in many cases, even though possessing possible theoretical advantages in a particular application. As we will see later, calculation of the Fourier-Bessel series coefficients requires computation of a Hankel transform, which until recently greatly diminished consideration of this series for potential applications. Fast Hankel transforms (FHT) have now been developed which allow

computation of F-B coefficients at a speed somewhat greater than Fourier coefficients; this should result in increased use of the F-B expansion. One possible application of the F-B series is the speaker identification problem, however, discussion of this topic will be deferred until the F-B series has been presented in more detail. Another signal type of interest is obtained from an acoustic well log and the F-B series expansion may prove useful in recovering geologic information concerning the surrounding rock formation.

The remaining sections of the appendix will discuss the F-B series in some detail, computation of the F-B series via a FHT, and some simulation results obtained to test the feasibility and accuracy of a F-B series signal representation.

Fourier-Bessel Series

Bessel functions arise as solutions of the differential equation

$$x^2y'' + xy' + (x^2 - n^2)y = 0, \quad n \geq 0 \tag{D.7}$$

which is called Bessel's differential equation. The general solution of (D.7) is given by

$$y = C_1 J_n(x) + C_2 Y_n(x) \tag{D.8}$$

where $J_n(x)$ is called a Bessel function of the first kind of order n and $Y_n(x)$ is called a Bessel function of the second kind of order n. Bessel functions are expressable in series form; for example $J_n(x)$ can be written

$$J_n(x) = \sum_{r=0}^{\infty} \frac{(-1)^r (x/2)^{n+2r}}{r!\,\Gamma(n + r + 1)} \tag{D.9}$$

and in particular

$$J_o(x) = 1 - \frac{x^2}{2^2} + \frac{x^4}{2^2\,4^2} - \frac{x^6}{2^2\,4^2\,6^2} + \ldots \tag{D.10}$$

It can be readily shown that Bessel functions are orthogonal with respect to the density function x. This can be seen by computing

$$\int_0^1 xJ_n(\alpha x)J_n(\beta x)dx = \frac{\beta J_n(\alpha)J_n{}^1(\alpha) - \alpha J_n(\beta)J_n{}^1(\alpha)}{\alpha^2 - \beta^2} \tag{D.11}$$

and

$$\int_0^1 xJ_n{}^2(\alpha x)dx = \frac{1}{2}\left[J'_n{}^2(\alpha) + \left(1 - \frac{n^2}{\alpha 2}\right) J^2{}_n(\alpha)\right] \tag{D.12}$$

Now if α and are different roots of $J_n(x) = 0$ we can write

$$\int_0^1 x\, J_n(\alpha x)\, J_n(\beta x)\, dx = 0,\ \alpha \neq \beta \tag{D.13}$$

and thus $J_n(\alpha x)$ and $J_n(\beta x)$ are orthogonal with respect to the weighting function x.

Having established orthogonality, a series expansion of an arbitrary function can be written in terms of Bessel functions with the form

$$f(x) = \sum_{m=1}^{\infty} C_m\, J_n(\lambda_m x) \tag{D.14}$$

where λ_1, λ_2, ... are the positive roots of $J_n(x) = 0$. The coefficients, C_m, are given by

$$C_m = \frac{2}{J_{n+1}{}^2(\lambda_m)} \int_0^1 xJ_n(\lambda_m x)\, f(x)dx \tag{D.15}$$

If we wish to expand f(t) over some arbitrary interval (o, a) the zero order Bessel series expansion becomes [5]

$$f(t) = \sum_{m=1}^{\infty} C_m J_o(\lambda_m t),\ o < t < a \tag{D.16}$$

with the coefficients, C_m, calculated from

$$C_m = \frac{2 \int_0^a t\, f(t)\, J_o(g_m t)dt}{a^2 [J_1(\lambda_m a)]^2} \qquad (D.17)$$

and λ_m, m = 1, 2, ... are the ascending order positive roots of $Jo(a\lambda_m)$= 0. The integral in the numberator of (D.17) is the finite Hankel transform; recent formulations of fast Hankel transforms based upon fast Fourier transform (FFT) algorithms now makes computation of (D.16) and (D.17) feasible. Expansion of a speech or seismic signal using Bessel functions will result in a feature set having different properties than that obtainable via Fourier techniques. Several unique properties of the Hankel transform (and consequently a Fourier-Bessel series) will be presented in the next section.

Properties of the Hankel Transform

The Hankel transform is defined by

$$F(W) = \int_0^\infty r\, f(r)\, J_o(Wr)dr \qquad (D.18)$$

and the corresponding inversion formula is

$$f(r) = \int_0^\infty W\, F(W)\, J_o(rW)dW\ . \qquad (D.19)$$

If

$$f_1(r) \overset{h}{\leftrightarrow} F_1(W),\ f_2(r) \overset{h}{\leftrightarrow} F_2(W)$$

then

$$\int_0^\infty r\, f_1(r) f_2^*(r)dr = \int_0^\infty W F_1(W) F_2^*(W)dW \qquad (D.20)$$

which is Parseval's formula.

The differential formula is

$$\frac{d^2f(r)}{dr^2} + \frac{1}{r}\frac{df(r)}{dr} \overset{h}{\leftrightarrow} -W^2F(W) \quad . \tag{D.21}$$

Other properties include:

Similarity

$$f(ar) \overset{h}{\leftrightarrow} \frac{1}{a^2} F\left(\frac{W}{a}\right) \tag{D.22}$$

and

Addition

$$f(r) + g(r) \leftrightarrow F(W) + G(W) \quad . \tag{D.23}$$

A shift property does not exist as the Bessel function Kernel is shift variant under argument translation. The shift variant property may be advantageous in the speaker identification.

Bracewell (1964) contains a fairly complete table of Hankel transforms; additional transform pairs are provided by Gerardi (1959) along with a relationship between the Hankel transform and the Laplace transform.

Fast Hankel Transform Algorithms

Recently, algorithms for efficient numerical evaluation of the Hankel transform have appeared in the literature. These timely results enable applications of Fourier-Bessel series techniques to be investigated without undue computational burden; additionally the fast Hankel transform algorithms are based upon readily available FFT algorithms.

Perhaps the earliest fast Hankel transform algorithm was published by Siegman. Siegman's technique is based upon what might be called a Gardner transform (1979) which transforms the shift variant Bessel Kernel to a kernel that is shift invariant. The new integral becomes a

cross correlation integral which then is computed via an FFT. A disadvantage of this technique is that the original function under transformation requires sampling at exponentially spaced points; non-linear sampling does not generally match the physical realities of data collection techniques and thus is awkward to implement.

A similar algorithm was published by Johansen and Sorensen (1979) which mapped the Hankel transform integral to a convolution integral. Again, however, this mandates exponential sampling of the function under transformation.

Oppenheim, Frisk, and Martinez (1978) have proposed an algorithm based upon the two dimensional Fourier transform "projection-slice" theorem (Papoulis, 1977). Although this approach is also FFT based, transformation of a data series in one dimension requires a relatively complex interpolation step prior to applying the FFT. If the data naturally arises from a two dimensional circular symmetric pattern on a rectangular grid, however, this approach is well formulated.

A unique approach has been proposed by Cavanagh and Cook (1979) that involves expansion of the function under transformation into a set of Gaussian-Laguerre polynomials that have known analytic transforms. Their algorithm, however, suffers convergence problems and will not be considered further.

An algorithm allowing uniform sampling of the time series to be transformed has been published recently by Candel (1981, 1982). Since most time series are sampled equidistantly in time, this algorithm is very attractive. Additionally, the algorithm is simple to implement and is based upon computation of an FFT. As Candel's method was chosen to develop the results presented later in this report, the computation

procedure will be briefly summarized.

Reproducing (D.18) with slightly different notation to emphasize the functional dependence on time we have

$$F(W) = \int_0^\infty t\, f(t) J_o(Wt) dt \ . \tag{D.24}$$

Substituting an integral form of the zero order Bessel function

$$Jo(W) = \frac{1}{\pi} \int_{-1}^{1} \frac{e^{jxu}}{(1 - u^2)^{y1/2}} du \tag{D.25}$$

results in

$$F(W) = \frac{1}{\pi} \int_0^{\pi/2} \phi(r\cos\theta) d \tag{D.26}$$

where

$$\phi(y) = \int_{-\infty}^{\infty} t\, f(t) e^{jyt}\, dt \ . \tag{D.27}$$

(y) in (D.27) is seen to be a Fourier transform of the product t f(t) and can easily be evaluated via FFT techniques. F(W) in (D.26) is then formed by sampling $\phi(y)$ with an $r\cos\theta$ spacing. $\phi(y)$ need not be interpolated, however, as nearest neighbor selection is sufficiently accurate provided the FFT length is chosen to provide adequate resolution.

With f(t) discretized as

$$f(n) = f(n\Delta t),\ n = 0, 1, \ldots, N - 1 \ , \tag{D.28}$$

F(W) discretized as

$$F(1) = \frac{F(1\Delta W)}{(\Delta t)^2} \ , \tag{D.29}$$

and with the sampling contraint

$$(\Delta W)\ (\Delta t) = \frac{2\pi}{N} \tag{D.30}$$

the algorithm follows directly.

The computation steps are summarized below:

1.) $\phi(k) = \sum_{n=0}^{N-1} f(n) e^{j \frac{2\pi}{N} nk}$

2.) $\Theta_i = (\frac{\pi}{N})\ (\frac{N}{2} - i + \frac{1}{2})$, $i = 1, 2, \ldots, N/2$

3.) $K(i,l)$ = Integer part $[l \cos \Theta_i + 1/2]$

4.) $F(l) = \frac{1}{N} \sum_{i=1}^{N/2} \phi[k(i,l)]$, $l = 0, 1, \ldots, N/2$

Candel's algorithms will now be re-written in pseudo-code to highlight a computation bottle neck:

```
φ(k) = FFT{n · f(n)}
 Do l = 0, N/2
     sum = 0.
     Do i = 1, N/2
         Θi = (π/N)(N/2 - i + 1/2)
         K(i,l) = int [l cos(Θi) + 1/2]
         sum = sum + [K(i,l)]
     end do
        F(l) = (1/N) · sum
     end do
```

Note that computation of $F(l)$, $l = 0, 1, 2, \ldots, N/2$ requires a nested "do loop" with the inner loop executing (N/2) squared times. Obviously, this is unacceptable, since such a structure negates the savings gained by recasting the Hankel transform integral into a Fourier integral suitable for implementation via FFT techniques.

Candel proposes a simple solution based upon an asymptotic series expansion of the Bessel function kernel. For $x >> 1$, $J_0(x)$ can be expressed

$$J_0(x) \simeq (2/(\pi x))^{1/2} \cos(x - \pi/4) \tag{D.31}$$

and now the Hankel transform may be written

$$F(W) = \int_0^\infty tf(t)(2/\pi tW)^{1/2} \cos(tW - \pi/4)dt \ . \tag{D.32}$$

By defining

$$f(t) = f(-t)e^{i\pi/2}, \ t \leq 0$$

the asymptotic Hankel transform may be expressed as a Fourier integral

$$F(W) = \frac{1}{2} \int_{-\infty}^{\infty} \left(\frac{2}{\pi W}\right)^{1/2} f(t)|t|^{1/2} e^{-jWt}dt \ ; \tag{D.33}$$

now a fast Fourier transform algorithm can be used to evaluate F(W). With $f(n) = f(n \cdot \Delta t)$, $F(l) = F(l\,\Delta w)/(t)^2$, and $(\Delta t) \cdot (\Delta W) = 2\pi/N$, F(W) can be estimated by:

$$F(l) = \frac{N^{1/2}}{2!l^{1/2}} \sum_{n=0}^{N-1} |n| \ f(n)e^{-j\pi/4} \ e^{j\frac{2\pi}{N}ln} \ . \tag{D.34}$$

Therefore, F(l) may be computed by performing a fast Fourier transform on the new sequence $[\,|n|^{1/2} \ f(n)e^{-j\pi/4}]$. The resulting FFT is then scaled by $N^{1/2}/(2\pi l^{1/2})$ and the calculations are completed. Since a single FFT is required with no frequency interpolation necessary this algorithm is very efficient. Unfortunately, the asymptotic expansion is inaccurate for values near the origin so a combination algorithm is required; that is, the very efficient algorithm is used for large argument values and the inefficient interpolation algorithm is used for small argument values. The crossover point is dependent upon the func-

tion under transformation, but hopefully relative few values need calculation via the "slow" fast Hankel transform (frequency interpolation method). Further study of Fourier-Bessel expansion techniques would include an investigation into convergence rates of the asymptotic method, but that has not been pursued in this appendix.

Simulation Results

A computer simulation was developed in order to evaluate the potential use of Fourier-Bessel series expansion coefficients as a feature set in speaker identification algorithms or well log signal decomposition. The simulation is set up such that several test functions are available for transformation for coefficient accuracy checking as well as a short segments of digitized speech for comparison against its spectrogram. Additionally, the transformed function can be regenerated using the computed F-B coefficients and a normalized error metric computed. Also available is synthetically generated acoustic well log data obtained from the Amoco Corporation. No attempt has been made to minimize execution speed other than utilizing an FFT based fast Hankel transform (Candel's algorithm 1981).

As demonstrated previously in the discussion of Candel's fast Hankel transform, although the time domain samples are equally spaced, computation of values in the transform domain requires $r\cos\theta$ type sampling of the FFT output. To avoid complex interpolation schemes, the FFT is zero padded to such a length that the resulting frequency resolution permits nearest neighbor selection of frequency samples. In order to quantify somewhat the FFT length required for accurate coefficient generation several test functions were expanded in a F-B series then

reconstructed with the calculated coefficients to form an estimated of the original function. With the estimate formed, denoted $\hat{f}(n)$, an error measure was computed as

$$E = \frac{1}{M} \sum_{n=1}^{N} [f(n) - \hat{f}(n)]^2 \qquad (D.35)$$

with M equal to the number of coefficients selected to construct f(n) and N equal to the time series length. For comparison purposes the standard Fourier series coefficients were also computed and then used to generate an estimate of the transformed function. Functions transformed for this test were:

1.) $f(t) = e^{j\omega_o t}$ (chirp signal)

2.) $f(t) = \sin(t)$

3.) $f(t) = J_o(t)$

4.) $f(t) = e^{-t}$

5.) $f(t) = 4\alpha\ t\ e^{-\alpha t} \sin(\omega_o t)$

In addition to the above five test functions, an acoustic well log synthetic trace and a short segment of speech were expanded in a F-B series.

Table 1. Summarizes the relative error magnitude as a result of computing the F-B and Fourier series expansion for the five test functions and two sets of real data. The series expansion coefficicents were calculated with a 512 point FFT, 256 test data points, and using 256 coefficients to reconstruct the original time series.

Relative error entries in Table 1 were calculated by the following sequence of steps:

1.) Compute 256 Fourier-Besel series coefficients and 256 Fourier series coefficients utilizing 256 samples of the test signal (a 512 point FFT was used to compute the Fourier coefficients and also to implement the FHT algorithm)

2.) Reconstruct the test signal by performing a finite summation of the respective series using 256 coefficients calculated in step 1.; namely,

a) Fourier series given by

$$f(t) \simeq \sum_{n=-128}^{128} C_n e^{jn\omega_o t}$$

b) Fourier-Bessel series given by

$$f(t) \simeq \sum_{n=1}^{256} C_n J_o(\lambda_n t)$$

where λ_n, n=1,2,...,256 are the positive roots of Jo(t) = 0.

3.) Calculate the error measure using Equation (D.35).

TABLE IV

COMPARISON OF FOURIER-BESSEL
SERIES AND FOURIER SERIES

	Fourier-Bessel	Fourier Series
Signal	Error	Error
Linear FM	1.00	.0001
$J_0(t)$	.005	.0000
$te^{-t}\sin(\omega t)$	.005	.004
$e^{-t/2}$	.001	.0003
sin (t)	1.05	.0000
clean speech	.13	.003
Acoustic Log	.047	.0000

The Fourier-Series expansion wins this type of comparison hands down, however it must be pointed out that this evaluation is not entirely fair to the Fourier-Bessel series. For example, computation of a fast Hankel transform requires "interpolation" of the FFT samples (nearest neighbor rule) while the Fourier series coefficients coincide exactly with the FFT samples. Thus, the only errors reflected in the Fourier series expansion are due to truncation, but the Fourier-Bessel series suffers in addition to truncation errors, inaccuracies from "interpolating" the FFT samples. The only apparent solution to this shortcoming of a F-B series expansion is to improve the FFT resolution by increasing the number of samples (zero padding). By way of illustration consider the acoustic well log data; as can be seen from Table 1, E = .047. Increasing the FFT length to 8192 and still retaining 256 coefficients reduced the error to .0004. An FFT length of 8192 with 500 coefficients computed further reduced the error to .0001. As the FFT length is becoming rather long a more fruitful approach may be to combine a more accurate interpolation scheme with a shorter FFT (say 512 or 1024 point FFT).

The goal at this stage, however, is not to solve the algorithm complexity problem, but rather to look for ways in which the Fourier-Bessel expansion leads to a feature set that is somehow "better" than say a feature set based on Fourier coefficients. For example, a signal exhibiting both amplitude and frequency modulation characteristics (such as speech or an acoustic well log) may be more compactly represented by Bessel function basis vectors rather than by pure sinusoids. Also, it is possible that the F-B coefficients in some sense better capture the fundamental nature of the speech waveform; the shift <u>variant</u> property

may be desireable and possibly result in improved speaker identification/authentication probabilities.

A final point should be made concerning the data presented in Table 1. For the test function, $f(t) = J_0(t)$, the Fourier series coefficients produced an extremely accurate reconstruction of the function under transformation. A F-B series expansion resulted in a higher error, but the number of coefficients required was dramatically different. Regenerating f(t) = Jo(t) from Fourier coefficients required all 256 values to achieve the result; by contrast just one Fourier-Bessel coefficient is required to reconstruct the function. Admittedly, this is a pathelogical case in that any function decomposed into basis vectors of the same analytic form will produce a single coefficient. Indeed, expanding the test signal f(t) = sin(t) via Fourier series requires a single coefficient. Nevertheless, the point being made is that an unknown signal will be more efficiently (more information in fewer coefficients) represented if expanded in the set of basis functions that "resemble" itself. Obviously, much research remains to be done concering the potential use of Fourier-Bessel series for date compression before one can draw definite conclusions.

The error metric comparison between a Fourier series expansion and a Fourier-Bessel series expansion provides a "quick look", however a better method is to inspect some actual time series plots and the resulting series expansions. Additionally a plot of the reconstructed data gives one a very good "feel" for the relative accuracy of the series expansion coefficients.

Two types of data were considered in this evaluation: a short sample of relatively clean speech (no noise added) and a synthetic

acoustic well log trace. The results of expanding the acoustic well log trace (Figure 77) into a Fourier-Bessel series (512 coefficients) can be viewed in Figure 78. A fair degree of data compression is available and the dynamic range has not increased. Significant research would be required, however, to determine if the Fourier-Bessel coefficients can be related to the geologic properties of the formation from the acoustic log was generated. As another example, a segment of speech (Figure 79) was also expanded into a Fourier-Bessel series. The coefficients are shown in Figure 80; whether or not these coefficients constitute a desirable feature set in the speaker identification problem is an open question.

Fourier-Bessel Decomposition Use

Given that the Fourier-Bessel series coefficients can be generated, a natural question is what use to make of them. One choice is to view the coefficents as a feature set to be input to any desired pattern recognition strategy and hope for improved performance.

We might also wonder how the coefficients vary with time; after all speech is quasi-stationary (or stationary over short time segments) and it is reasonable to expect that the F-B coefficients can put to good use in a display format similar to that of the spectrogram. To test this hypothesis a segment of speech signal could be broken up into short analysis windows (as with the short time Fourier transform), F-B coefficients calculated for each piece, and finally converted to a grey scale image for convenient viewing.

After the F-B coefficients are calculated for each analysis frame, they would be displayed as a function of time to form a two-dimensional

"image" quite analogous to the spectrogram.

One intriguing possible use of the F-B coefficient-time dispay that comes to mind is to subject this "image" to conventional 2-D signal processing algorithms and hope for improved results. Since the F-B coefficients are real the noisy phase problem upon reconstruction is avoided, which may be advantagous. The entire range of image processing algorithms developed over the past several decades would be available for exploitation to improve upon the speech characteristics.

It appears that the application of F-B series to speech processing, particularly speaker identification, bears further research. The shift variant property of the Hankel transform may prove valuable for non-stationary analysis and some indications exist that fewer coefficients may be required. Since the coefficients are real the speech can be directly reconstructed from its coefficient time index plot without need to retain phase components; this may prove to be of some use when conversion back to the time domain is desired. The performance of specific classification algorithms utilizing the F-B coefficients as a feature set needs to be evaluated.

One topic not mentioned at all in this appendix is that of window functions. It was felt that at this stage of the game a rectangular window was best so as not to "muddy the waters" with excessive test parameters to control. Proper choice of data smoothing windows can be expected to improve F-B series convergence at points of discontinuity, however, "proper choice" is probably an open question. Windows designed to suppress Fourier sidelobe leakage, for example, may no longer be the best choice; indeed, there may be undesireable and as of yet unforeseen side effects when computing a Fourier-Bessel expansion. One expects

though that optimum window choice for F-B coefficient generation would be a fruitful research endeavor.

Conclusions

The purpose of this appendix was to present the general signal decomposition problem in terms of an orthogonal series expansion. Focus was primarily held on the Fourier-Bessel series expansion with the Fourier series expansion utilized here and therefore comparison purposes. A fast Hankel transform algorithm was presented that allows the Fourier-Bessel series coefficients to be computed efficiently and hopefully new applications will now be investigated. The fast Hankel transform technique was illustrated with several test functions, segments of acoustic well log data, and a clean speech sample. Obviously, much research remains to be accomplished concerning the F-B series and now an efficient computational algorithm makes such research feasible.

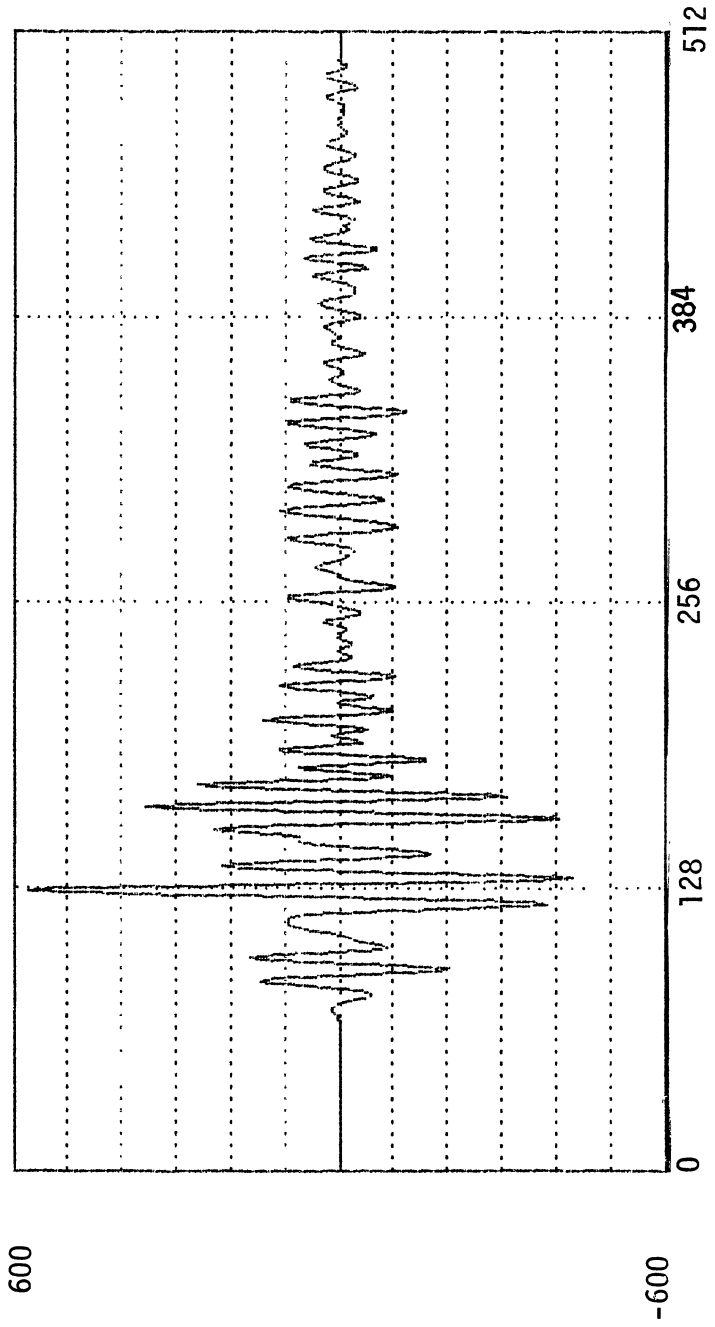

Figure 77. Synthetic Acoustic Well Log Trace

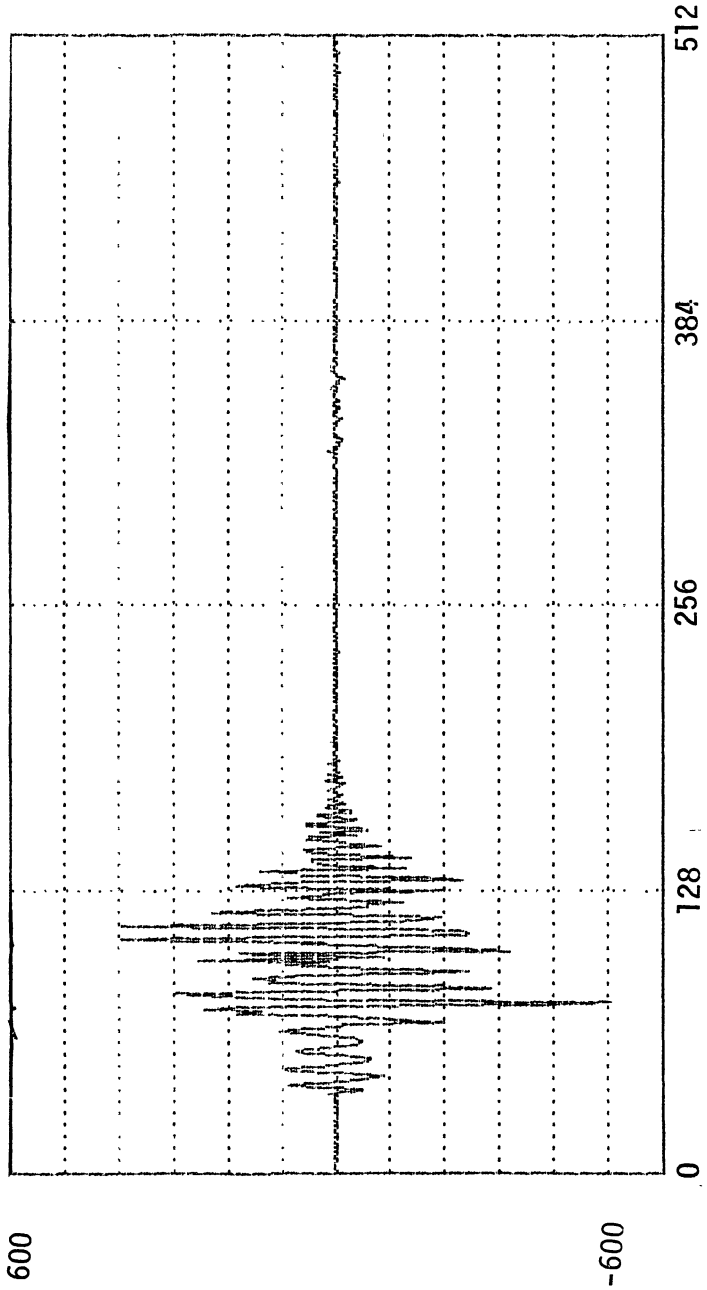

Figure 78. Fourier-Bessel Series Expansion of Synthetic Acoustic Well Log Data

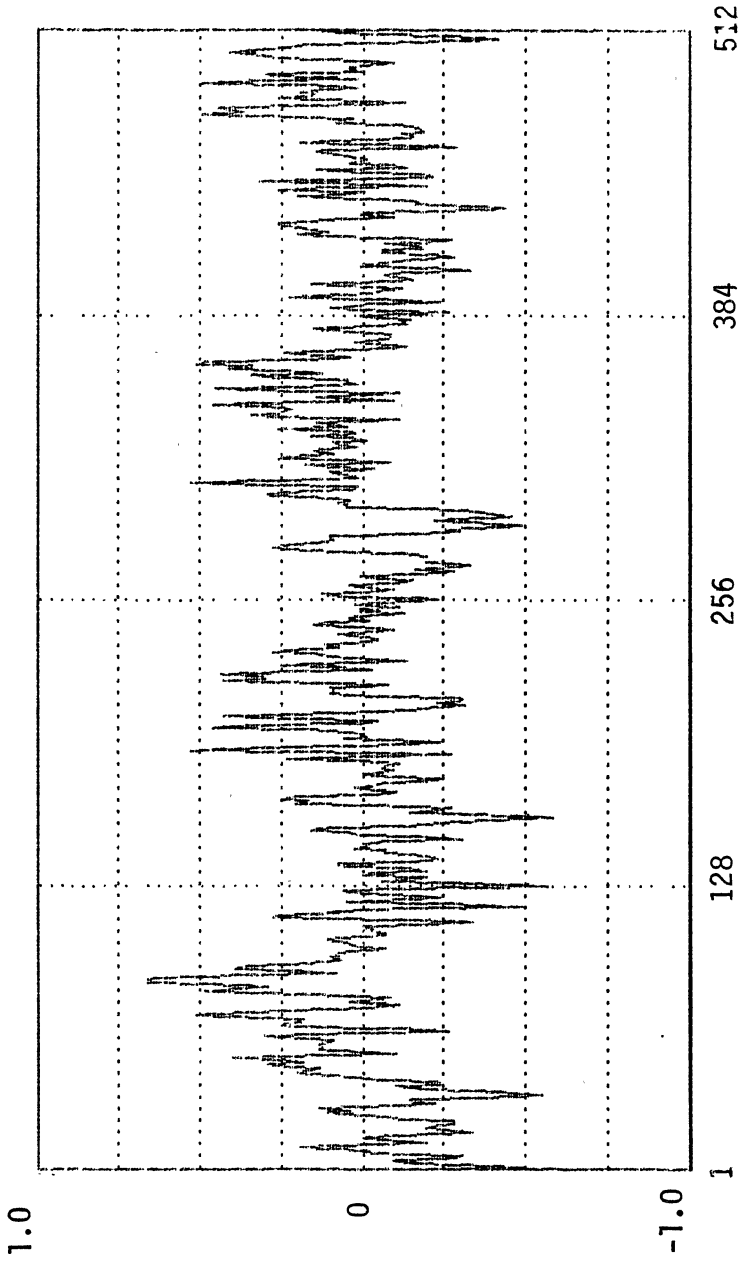

Figure 79. Speech Data

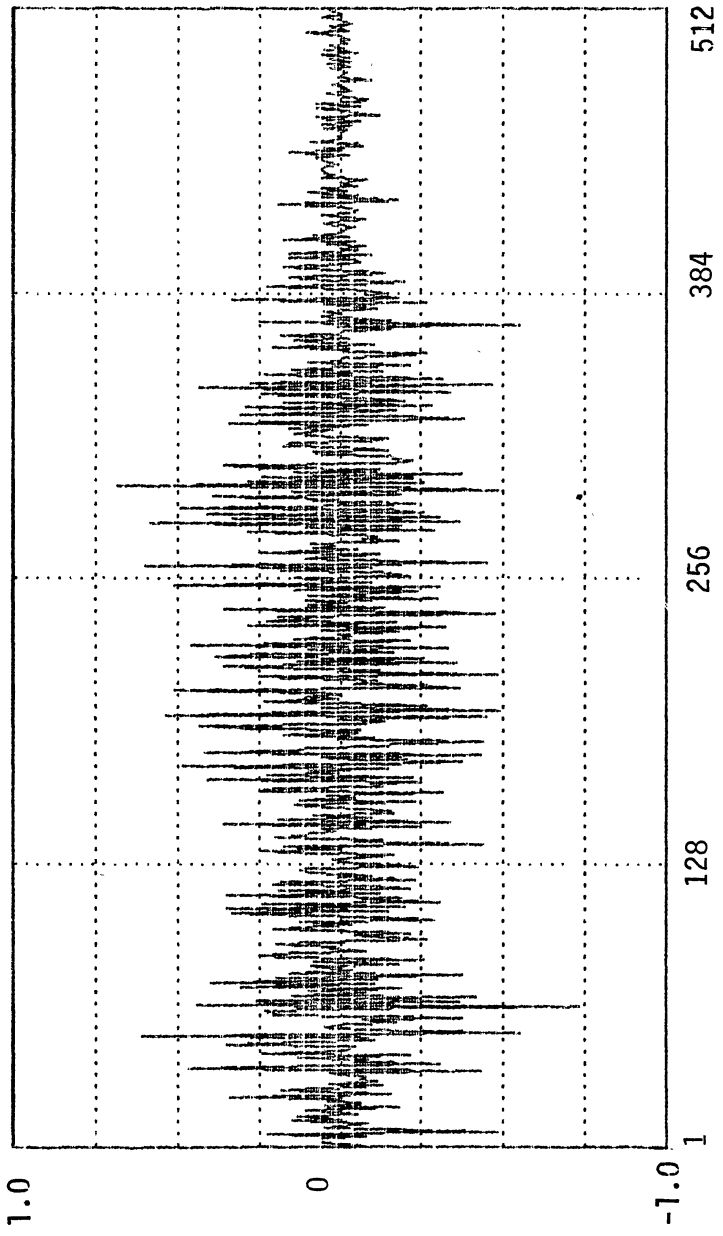

Figure 80. Fourier-Bessel Series Expansion of Speech Data

VITA

James Edwin Schroeder

Candidate for the Degree of

Doctor of Philosophy

Thesis: LINEAR PREDICTIVE SPECTRAL ANALYSIS VIA THE L_p NORM

Major Field: ELECTRICAL ENGINEERING

Biographical:

Personal Data: Born in Waverly, Iowa, August 21, 1947. Married to Sarah P. Lowe June 15, 1974.

Education: Graduated from Waverly-Shell Rock High School, Waverly, Iowa, in May, 1965; received Bachelor of Science degree in Electrical Engineering from the University of Iowa, in December, 1976; received Master of Science in Electrical Engineering from the University of Iowa, in July, 1978; enrolled in doctoral program at the University of Iowa, 1980-1982; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1985

Professional Experience: Computer Operator, Lutheran Mutual Life, 1965-1966; Electronics Technician, US Navy, 1966-1972; Graduate Teaching Assistant, University of Iowa, 1977-1978; Engineer Scientist, Rockwell-Collins Avionics, Cedar Rapids, Iowa, 1978-1980; Programmer Analyst, University of Iowa, 1980-1981; Graduate Teaching Assistant, University of Iowa, 1981-1982; Associate Prinicipal Engineer, Harris Corporation, Melbourne, Florida, 1982-1984; Lecturer, Oklahoma State University, 1984-1985.