

Spectral Estimation for Sensor Arrays

by

Stephen W. Lang

S.B., Massachusetts Institute of Technology (1977)

M.S., Massachusetts Institute of Technology (1979)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August, 1981

© Stephen W. Lang, 1981

The author hereby grants to M.I.T. permission to reproduce and
to distribute copies of this thesis document in whole or in part.

Signature of Author _____
Department of Electrical Engineering and Computer Science
August 14, 1981

Certified by _____
James H. McClellan
Thesis Supervisor

Accepted by _____
Chairman, Departmental Graduate Committee

Archives

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

NOV 13 1981

LIBRARIES

Spectral Estimation for Sensor Arrays

by

Stephen W. Lang

Submitted to the Department of Electrical Engineering and Computer Science on August 14, 1981 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

The array processing problem is briefly discussed and an abstract spectral estimation problem is formulated. This abstract problem involves the estimation of a multi-dimensional frequency-wavevector power spectrum given certain measurements of the correlation function and knowledge of the spectral support.

The pursuit of correlation-matching spectral estimates leads to the extendibility question: does there exist any positive spectrum on the spectral support which matches a given set of correlation samples? In answering this question, the necessary mathematical framework is developed with which to analyze specific spectral estimation algorithms and to design algorithms for their computation.

This framework is exploited in the extension of two spectral estimation techniques from the time series case, for which they were originally formulated, to the more general array processing case. Pisarenko's method models the spectrum as a sum of impulses plus a noise component. Its computation in the time series case requires the solution of a eigenvalue problem; more generally it is shown to require the solution of a finite-dimensional linear optimization problem. The maximum entropy method (MEM) models the spectral density function as the inverse of a positive polynomial. Its computation in the time series case requires the solution of a system of linear equations; more generally it is shown to require the solution of a finite-dimensional convex optimization problem.

Algorithms are developed for the solution of the optimization problems involved in the computation of Pisarenko's estimate and the MEM estimate. Several spectral estimation examples using these algorithms are presented.

Thesis Supervisor: James H. McClellan

Title: Associate Professor of Electrical Engineering

To my parents

Acknowledgements

I thank James McClellan for pointing out the work of Brad Dickinson, which eventually led to this thesis, for many valuable discussions, and for his guidance of this work. I thank Arthur Baggeroer, Webster Dove, David Harris, and Bruce Musicus for technical discussions which particularly aided my work and I thank all the members of the Digital Signal Processing Group at M.I.T. for making my stay here pleasant and rewarding. Finally, I gratefully acknowledge the financial support given to me by the Hertz Foundation during my graduate years.

Contents

Abstract 2

Acknowledgements 4

Table of Contents 5

List of Figures 7

List of Tables 8

Chapter 1: Introduction 9

Chapter 2: The Array Processing Problem 14

Chapter 3: Extendibility 19

3.1 Introduction 19

3.2 Spectral Supports and Coarrays 20

3.3 Mathematical Terminology 21

3.4 Characterizations of Extendibility 23

3.5 Boundary and Interior 26

3.6 Power Spectral Density Functions 27

3.7 Sampling of the Spectral Support 28

Chapter 4: Pisarenko's Method 30

4.1 Introduction 30

4.2 Pisarenko's Method for Sensor Arrays 31

4.3 The Computation of Pisarenko's Estimate 35

Chapter 5: The Maximum Entropy Method 38

5.1 Introduction 38

5.2 The Maximum Entropy Spectrum 38

5.3 Dual Optimization Problems 40

5.4 Application of the Method of Steepest Descent 44

5.5 Comparison to Other Algorithms 46

Chapter 6: Spectral Estimation Examples 48

6.1 Introduction 48

6.2 The Importance of a Proper Spectral Support 49

6.3 Choosing the Correct Spectral Estimation Method 53

6.4 A Two-Dimensional MEM Spectrum 57

Chapter 7: Summary 60

Appendix A: The Extension Theorem for Spectral Density Functions 63

Appendix B: The Representation Theorem 66

Appendix C: The Uniqueness of Pisarenko's Estimate 69

Appendix D: The Existence of the MEM Estimate 71

References 74

Figures

Fig. 2.1: A three dish array. 17

Fig. 2.2: Spectral support for an array of dish antennas. 18

Fig. 3.1: E and P for $K = [-\pi, \pi]$ and $\Delta = \{0, \pm 1\}$. 23

Fig. 3.2: E and P for $K = [-\pi, \frac{\pi}{2}]$ and $\Delta = \{0, \pm 1\}$. 25

Fig. 3.3: Approximation of a spectral support by sampling. 29

Fig. 4.1: Decomposition of a vector \mathbf{r} into a vector \mathbf{r}' on the boundary of E plus a multiple of a given vector \mathbf{n} . 31

Fig. 5.1: The hyperplane $(\mathbf{r}, \mathbf{p}) = \nu(K)$ is tangent to the surface $H(\mathbf{p}) = H(\mathbf{p}')$ at the point $\mathbf{p} = \mathbf{p}'$. 41

Fig. 6.1: Six sensor linear non-uniform array and the positive portion of its coarray. 50

Fig. 6.2: Original spectral density function, whose correlation function, sampled on the coarray, produced the correlation vector. 50

Fig. 6.3: MEM spectral estimate for a spectral support of $[-\pi, \pi]$. 52

Fig. 6.4: MEM spectral estimate for a spectral support of $[-\frac{5\pi}{4}, \frac{5\pi}{4}]$. 52

Fig. 6.5: Four sensor array and its coarray. 54

Fig. 6.6: Spectral support for two-dimensional waves. 54

Fig. 6.7: MEM spectral estimate of two point sources in isotropic noise. 55

Fig. 6.8: Array and coarray for 2-D example. 58

Fig. 6.9: Spectrum from which correlation vector was derived. 59

Fig. 6.10: MEM spectral estimate. 59

Fig. A.1: Two open convex sets with the same closure are identical. 65

Fig. B.1: E for $K = [-\pi, \frac{\pi}{2}]$ and $\Delta = \{0, \pm 1\}$. 68

Tables

Table 6.1: Pisarenko's estimate of two point sources in isotropic noise. 56

Table 6.2: Pisarenko's estimate of three point sources in isotropic noise. 57

Chapter 1

Introduction

Just as the power spectrum of a stationary time series describes a distribution of power in frequency, the frequency-wavevector power spectrum of a homogeneous and stationary wavefield describes a distribution of power in wavevector and frequency, or equivalently, in propagation direction and frequency. The frequency-wavevector spectrum, or information which can be derived from it, is important in many applications areas such as radio-astronomy, radar, and sonar. Hence its estimation, from data provided by sensor arrays, is of great practical interest.

Chapter 2 contains a more detailed description of wavefields and sensor arrays and motivates the spectral estimation problem. The representation of a power spectrum as a measure and as a spectral density function is discussed. Chapter 2 introduces the *coarray*, the set of vector separations and time lags for which correlation samples are available, and the *spectral support*, the region of frequency-wavevector space containing power to which the sensors are sensitive. It concludes with the formulation of an abstract problem: the estimation of a power spectrum given only that it is positive on the spectral support, zero outside, and has certain known correlations for separations in the coarray. Although simpler than many problems encountered in practice, this abstraction retains the key features which distinguish the array problem from the problem of time series power spectral estimation; the multidimensionality of the spectral support and of the frequency variable, and the non-uniformity of the coarray. The formulation is a compromise; hopefully it is

specific enough to result in techniques of practical usefulness in array processing, and general enough so that the results will be useful in other areas as well.

Given this problem formulation, it is natural to consider spectral estimates which match the known information: spectral estimates which are positive, zero outside the spectral support, and which match the measured correlations. The pursuit of such *correlation-matching* spectral estimates raises two important questions. The first, and more fundamental, question concerns the existence of any such estimate. This *extendibility* problem has deep historical roots (Stewart 1976) and was recently raised by Dickinson (1980) with reference to the maximum entropy spectral estimation method. The extendibility problem is explored in chapter 3. Extendible sets of correlation measurements are characterized, for power spectra represented as measures and as spectral density functions. The effect which sampling of the spectral support has on extendibility is considered. In answering the extendibility question, the necessary mathematical framework is developed with which to analyze specific spectral estimation methods and to design algorithms for their computation.

The second question raised is that of uniqueness, is there a unique correlation-matching spectral estimate, and, if not, how can a specific one be chosen? In fact, except in very special cases, a unique estimate does not exist and the task of a spectral estimation method becomes the selection of one out of an ensemble of spectra satisfying the correlation, positivity, and spectral support constraints. Two spectral estimation methods, originally developed for time series analysis, are extended to the array problem. Chapter 4 concerns Pisarenko's method (Pisarenko 1973), which involves modelling the correlation measurements as a sum of two components. One, a noise component of known spectral shape but unknown amplitude, is made as large as possible without

making the other component non-extendible. The spectral estimate resulting from Pisarenko's method is revealed to solve a linear optimization problem. A solution to this optimization problem will always exist if the correlation measurements are extendible. In fact, Pisarenko's method is shown to be intimately related to the extendibility question and an algorithm for the computation of Pisarenko's estimate easily serves as an extendibility test. Although the estimate happens to be unique in the time series case, it is shown that it is not always unique in the more general array setting. The computation of Pisarenko's estimate, which in the time series case involves the solution of an eigenvalue problem, is shown, more generally, to involve the solution of a linear program.

Chapter 5 concerns the maximum entropy method (Burg 1975). Of the feasible spectral density functions, the maximum entropy method (MEM) selects that with the largest entropy as the spectral estimate. Given an additional constraint on the spectral support and coarray, easily met in practice, the MEM spectral estimate is shown to exist and be unique. Further, it is derived as the solution to certain convex optimization problems. While, in the time series case, the computation of the MEM estimate essentially involves only the solution of a set of linear equations, this no longer holds true in the more general array problem. However, the application of standard techniques to the convex optimization problems result in reliable computational algorithms. These are compared to algorithms previously proposed for the computation of the multi-dimensional MEM estimate.

Several spectral estimation examples are presented in chapter 6. These examples provide an opportunity to review some of the ideas presented previously, to expand upon them in specific contexts, and to discuss some details of

computation and implementation. Chapter 7 contains a summary of the research and raises some questions which may prove to be fruitful areas for future research.

In the process of exploring the array processing problem, several interesting themes are developed. The first involves the application of simple ideas from functional analysis to the spectral estimation problem. The set of correlation measurements are seen to naturally inhabit a real vector space, therefore functional analysis, the study of such spaces, provides the necessary mathematical tools. Functional analysis is, of course, the language of approximation theory; those familiar with filter design, long a problem of interest in digital signal processing, will find much they recognize here. The computation of Pisarenko's and the MEM estimate rely upon the solution of optimization problems. Iterative optimization algorithms must be considered, therefore, not just as aids for the design of, but as components of signal processing systems. Modern optimization theory, which can also be considered a branch of functional analysis, provides the basic techniques for the design of reliable computational algorithms. Some knowledge of functional analysis and of optimization theory is assumed on the part of the reader, although only the most basic ideas are made use of. The two books by Luenberger (1969, 1973) are appropriate references for those wishing to familiarize themselves with these topics.

Simple measure theoretic ideas provide the means for treating many different spectral supports in a unified manner. In particular, the discrete spectral supports necessary in practical computational algorithms are handled by the same theory as the spectral supports, consisting of an infinite number of points, which they are intended to approximate. Only the most basic ideas from measure theory are needed; again the reader is assumed to be familiar with

them. The book by Royden (1968) is an appropriate reference.

Chapter 2

The Array Processing Problem

Imagine a multi-dimensional homogeneous medium supporting scalar propagating waves and containing a group of sensors, called an *array*. The medium in the vicinity of the array is completely described by the scalar wave equation

$$\begin{aligned}\nabla^2 u(x, t) &= \ddot{u}(x, t) \\ u: \mathbb{R}^D \times \mathbb{R} &\rightarrow \mathbb{C}.\end{aligned}\tag{2.1}$$

The wavefield will be assumed to be homogeneous and stationary so that its second order statistics are described by a correlation function, or equivalently, by a power spectrum (Baggeroer 1976),

$$r(\delta, \tau) = E[u^*(x, t)u(x + \delta, t + \tau)] = \int e^{j(k \cdot \delta + \omega \tau)} d\mu.\tag{2.2}$$

The power spectrum is represented here by a positive measure, μ , which assigns a non-negative power to each measurable subset of frequency-wavevector space.

The use of a positive measure to represent a power spectrum provides the flexibility needed to deal with a range of spectral supports in a unified manner and to easily handle spectra which contain impulses: finite power at a single wavevector. It is more common to represent a power spectrum by means of a spectral density function. In the representation of a power spectrum by a spectral density function, there is some fixed measure ν which defines the integral of a function. The power spectrum is then represented as a positive function $S(k, \omega)$.

$$r(\delta, \tau) = \int e^{j(k \cdot \delta + \omega \tau)} S(k, \omega) d\nu.\tag{2.3}$$

Given a power spectral density function S , it is possible to define a corresponding positive measure by requiring the measure of a set B to equal the integral of the spectral density function over B :

$$\mu(B) = \int_B d\mu = \int_B S(k, \omega) d\nu. \quad (2.4)$$

Although more common, the representation of power spectra as power spectral density functions is slightly less general and somewhat more cumbersome than the representation in terms of measures. Both will be used in succeeding chapters.

However represented, the spectrum describes the distribution of incident power in temporal frequency and spatial wavevector or, equivalently, spatial bearing. It is important in many applications areas. In radio-astronomy, the formation of an image can be described as the estimation of a power spectrum. The detection and bearing estimation of targets in radar and sonar can be based upon information contained in a power spectral estimate.

The sensors each produce a time function which is the wavefield u sampled in space. The collection of time functions produced by all the sensors will be called the array output or the array response. The array output is to be processed so as to provide an estimate of the frequency-wavenumber spectrum. The stochastic character of the wavefield invariably leads to random variations of any spectral estimate based on the array output. To combat this effect, spectral estimates are often based on stable statistics derived from the array output. A common example of such a statistic is a correlation estimate calculated by multiplying the output of one sensor and a time delayed version of a second sensor and averaging over time. This process results in an estimate of the correlation function at a temporal lag corresponding to the delay time and

a spatial separation which is the vector distance between the two sensors. The averaging process results in statistical stability of the correlation estimates and therefore in statistical stability of any spectral estimate based on these correlation estimates. Estimates or samples of the correlations are only available for a finite set of inter-sensor separations, the *coarray* (Haubrich 1968).

$$\Delta = \{x - y : \text{sensor locations } x, y\} \quad (2.5)$$

Besides samples of the correlation function, two other pieces of information about the spectrum are available: the spectrum is known to be *positive* and it is confined to a finite region of frequency-wavevector space, the *spectral support*. Outside of this support the spectrum is assumed to be zero. The spectral support arises naturally in several ways. Actual sensors have a finite temporal bandwidth; they will be relatively insensitive to power outside of this bandwidth. If the wavefield has finite temporal bandwidth, a result perhaps of finite sensor bandwidth, and satisfies the homogeneous dispersion relation then it has natural cutoff wavenumbers. Also, if the source of the wavefield has a known angular extent or if the sensors are directional then there is limited angular beamwidth which translates to a cutoff wavenumber for a particular temporal frequency. All of these effects can be modelled by assuming that no power is present outside of a certain region of frequency-wavevector space. A known spectral support, based on the physics of the problem, constitutes important prior information which can be brought to bear on the spectral estimation problem. A simple example is provided by a sensor array composed of uniformly oriented dish antennas.

Example 2.1: A three dish array.

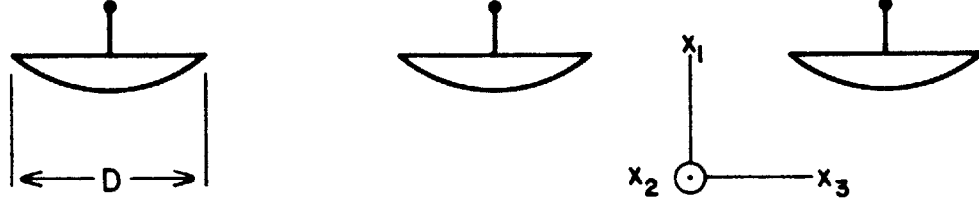


Fig. 2.1: A three dish array

A dish antenna of diameter D has a cutoff wavenumber of roughly

$$\sqrt{k_2^2 + k_3^2} \leq 0.61 \frac{\lambda}{D}.$$

Assuming that the wavefield satisfies the dispersion relation for a homogeneous, non-dispersive medium and that the receivers are only sensitive to a particular temporal ω_0 , suggests that the support for the spectral estimate should be the region of frequency-wavevector space which satisfies the dispersion relation and to which the sensors are sensitive. This region is defined by the three equations:

$$\begin{aligned} \omega &= \omega_0, \\ k_1^2 + k_2^2 + k_3^2 &= \left[\frac{2\pi}{\lambda} \right]^2, \\ \sqrt{k_2^2 + k_3^2} &\leq 0.61 \frac{\lambda}{D}. \end{aligned}$$

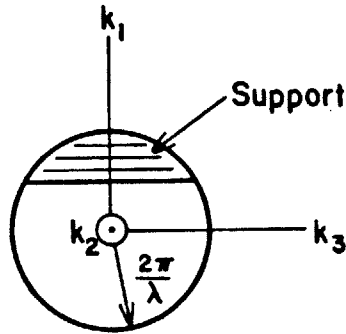


Fig. 2.2: Spectral support for an array of dish antennas.

In many applications much more data is available in the time dimension than in the space dimension. In these cases it is convenient to separate out the time variable by Fourier analyzing the time series output of each sensor and then doing a separate wavevector spectral estimate for each temporal frequency by using the Fourier coefficients as data for a wavevector spectral estimator. Thus the estimation problem is formulated for complex data even though physical wavefields are real valued. Fortunately conventional Fourier analysis is often satisfactory when data is abundant, as well as being implicit in the narrow-band character of many sensors. Where limited data in the time dimension makes the above approach impractical and wide-band sensor arrays are available the full problem may be treated by including the temporal variables τ and ω in the vectors δ and k . It shall be assumed that one of these two approaches has been taken; henceforth the temporal variables τ and ω will be dropped.

Chapter 3

Extendibility

3.1 Introduction

A simple model of the array processing problem was constructed in the last chapter: given certain correlation measurements and a spectral support, produce a spectral estimate. It is natural to use the known information about the spectrum to constrain the spectral estimate. Knowledge of certain correlations can be used by constraining the spectral estimate to have these same correlations. Knowledge that the spectrum is positive and zero outside the spectral support can be used by constraining the spectral estimate to be positive and zero outside the spectral support. Such spectral estimates are thus consistent with the known information; they are called *correlation-matching* spectral estimates. Two of these estimates, Pisarenko's method and the maximum entropy method will be discussed in later chapters.

The pursuit of correlation-matching spectral estimates raises a fundamental existence question. Given a finite collection of measured correlations and a spectral support, does there exist *any* correlation-matching spectral estimate? If such a spectral estimate exists, the measured correlations are said to be *extendible*; the correlation function associated with any correlation-matching spectral estimate is a suitable extension of the correlation measurements to all spatial separations. After some necessary mathematical definitions, the existence question is answered by characterizing the set of extendible correlation measurements. The mathematical framework developed to answer this question will prove useful in analyzing spectral estimation techniques and in

designing algorithms for their computation.

3.2 Spectral Supports and Coarrays

It is first necessary to define more carefully a spectral support and a coarray. The spectral support K is assumed to be a compact subset of \mathbf{R}^D , i.e. K is closed and bounded. This assumption is similar to the bandlimitedness assumption of time series analysis. Some such assumption is necessary because of the finite amount of data on which the spectral estimate is to be based. Assuming that K is compact leads to a certain technical advantage: a continuous function on a compact set attains its infimum and supremum. The assumption of this property will prove useful and leads to no practical disadvantage. In fact, it will be turned to advantage by using the spectral support as a device by which to incorporate certain prior knowledge into the spectral estimation problem in a simple manner. As discussed in the previous chapter, knowledge of dispersion relations and sensor directionality and frequency response can be used to construct an appropriate spectral support. The incorporation of this additional prior information into the spectral estimation problem is expected to result in improved spectral estimates.

In the last section, it was assumed that the coarray Δ , the set of spatial differences at which the correlation function is sampled, is the difference set of some set of sensor locations. Although this definition for the coarray arises naturally in the array processing problem, it can be relaxed somewhat. It will only be assumed that Δ is a finite subset of \mathbf{R}^D with the properties:

- (i) $0 \in \Delta$;
- (ii) if $\delta \in \Delta$, then $-\delta \in \Delta$;
- (iii) $\{e^{jk \cdot \delta} : \delta \in \Delta\}$ is a set of linearly independent functions on K .

Condition (i) implies knowledge of $\tau(0)$, the total power in the spectrum. Condition (ii) reflects the fact that the correlation function is conjugate-symmetric, thus if $\tau(\delta)$ is known, so too is $\tau(-\delta)$. Condition (iii) is only slightly more subtle than the first two. Linear independence means that if $p(\delta)$ are complex constants, not all zero, such that $p(\delta) = p^*(-\delta)$ then the real-valued function

$$\sum_{\delta \in \Delta} p(\delta) e^{-jk \cdot \delta} \quad (3.1)$$

is not identically zero on K . A correlation sample is a measurement on the spectrum, and this condition guarantees that these measurements are independent; each measurement gives new information about the spectrum.

If $D > 1$ then the spectral estimation problem is *multi-dimensional*. If $D = 1$, $K = [-\pi, \pi]$ and $\Delta = \{0, \pm 1, \dots, \pm M\}$ then the spectral estimation problem is that of the familiar *time series case* and the extendibility question reduces to the famous trigonometric moment problem (Grenander and Szegő 1958).

3.3 Mathematical Terminology

The spectral support and the coarray give rise to a few other important mathematical entities. Conjugate-symmetric complex-valued functions on Δ will play a central role in this work. A *conjugate-symmetric* function f on Δ is one for which $f(-\delta) = f^*(\delta)$ for all $\delta \in \Delta$. Correlation samples, from which spectral estimates are to be made, are such functions. Because of this symmetry, many of the expressions to follow are real-valued even though, for the sake of simplicity, they have been written in a form which suggests that they might be complex-valued. The coarray Δ has $2M+1$ elements, and so a conjugate-symmetric function on Δ is characterized by $2M+1$ independent real numbers.

Thus a conjugate-symmetric function on Δ may be thought of as a vector in \mathbb{R}^{2M+1} . A vector space over the real numbers is chosen because it is only multiplication by a *real* number which sends a correlation function into another correlation function. This representation of conjugate-symmetric functions on Δ as vectors hints at the importance of functional analysis, the study of linear vector spaces, to the spectral estimation problem.

Since $\{e^{jk \cdot \delta} : \delta \in \Delta\}$ is a linearly independent set of functions on K , it follows that each vector p in \mathbb{R}^{2M+1} can be uniquely associated with a real-valued Δ -polynomial $P(k)$ on K through the relation

$$P(k) = \sum_{\delta \in \Delta} p(\delta) e^{-jk \cdot \delta}. \quad (3.2)$$

p shall be termed *positive* if $P(k) \geq 0$ on K . P shall denote the set of those vectors associated with positive Δ -polynomials. From the compactness of K , it can be shown that P is a closed convex cone with its vertex at the origin. A cone with vertex at the origin is a set such that if f is in the set, then so is αf for all $\alpha \geq 0$ (Luenberger 1969). Cones are important kinds of sets in the spectral estimation problem because it is only multiplication by *positive* real numbers that sends a correlation function into another correlation function.

The inner product between a vector r of correlation samples and a vector p of polynomial coefficients shall be defined as

$$(r, p) = \sum_{\delta \in \Delta} r^*(\delta) p(\delta). \quad (3.3)$$

It should be noted that this is not the standard inner product in \mathbb{R}^{2M+1} , although a simple scaling of the zeroth correlation would make it so.

3.4 Characterizations of Extendibility

Let E denote the set of extendible correlation vectors. That is, $r \in E$ if

$$r(\delta) = \int_K e^{jk \cdot \delta} d\mu \quad (3.4)$$

for some positive measure μ on K . From the properties of the integral, it follows that E is a closed convex cone. Furthermore, a section through E at $r(0) = 1$:

$$E' = \{r \in E : r(0) = 1\} \quad (3.5)$$

is the convex hull of the compact set

$$A = \{r : r(\delta) = e^{jk \cdot \delta}, k \in K\}. \quad (3.6)$$

Thus E is the closed convex cone, with vertex at the origin, generated by A .

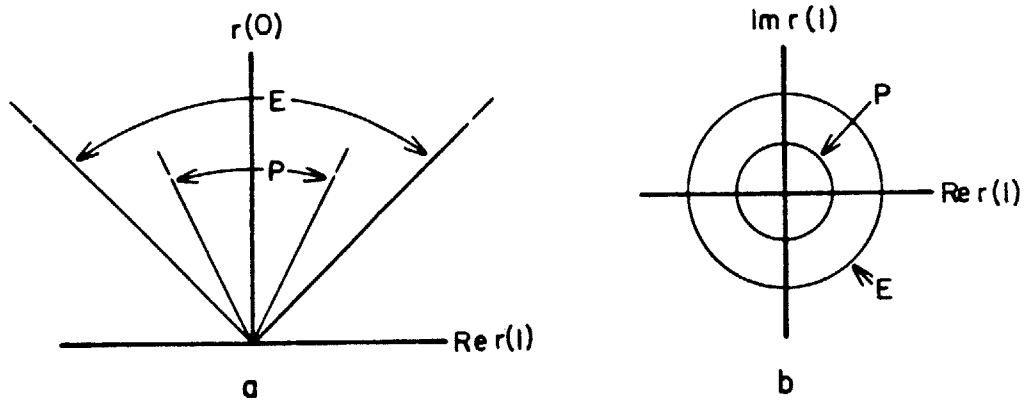


Fig. 3.1: E and P for $K = [-\pi, \pi]$ and $\Delta = \{0, \pm 1\}$. (a) shows a section through E and P at $\text{Im } r(1) = 0$ and (b) shows a section through E and P at $r(0) = 1$.

This characterization of extendible correlations is similar to that given originally by Carathéodory in 1907 for the trigonometric moment problem (Stewart 1976). It is important in that it relates the set of extendible correlation vectors directly to the spectral support and the coarray via fundamental properties of the integral. It thus gives a clear geometric picture of extendibility, and will be

useful in proofs.

A characterization of extendibility which is more useful in the development of spectral estimation methods results from expressing E as the intersection of all the closed half-spaces containing it (Luenberger 1969). This characterization involves duality, since half-spaces are defined by linear functionals, elements of the dual space. A closed half space is defined by a vector p and a real number c as

$$\{r: (r, p) \geq c\}. \quad (3.7)$$

To determine the particular half-spaces containing E , it is only necessary to consider those correlation vectors which generate E , positive multiples of vectors in the set A . p and c define a closed half-space containing E if and only if, for every $r \in A$, $(\alpha r, p) \geq c$ for every $\alpha \geq 0$. Hence $(\alpha r, p) = \alpha P(k) \geq c$ for every $k \in K$ and $\alpha \geq 0$. Since α may be made arbitrarily large, it must be true that $P(k) \geq 0$. The smallest half-space containing E for such a p corresponds to choosing $c = 0$. Thus

$$E = \bigcap_{p \in P} \{r: (r, p) \geq 0\}, \quad (3.8)$$

hence

The extension theorem: r is extendible if and only if $(r, p) \geq 0$ for all positive p .

Positive polynomials thus occur naturally in the extendibility problem, since they define the supporting hyperplanes of the set E of extendible correlation vectors. The extension theorem simply states that E and P are positive conjugate cones (Luenberger 1969).

Although the incorporation of a spectral support into the problem is new,

essentially the same characterization of extendibility through the use of positive polynomials has been used by Calderón and Pepinsky (1952), and Rudin (1963).

Fig. 3.2 shows the dependence of E on the spectral support. There are two ways of looking at this dependence. E is the convex cone generated by A ; because the spectral support has been reduced, E is smaller than in Fig. 3.1. That is the direct way; the indirect way involves constraints. The set K constrains the set P via the positivity condition and the set P constrains the set E via the extendibility theorem. Thus when K shrinks, P grows, and E shrinks.

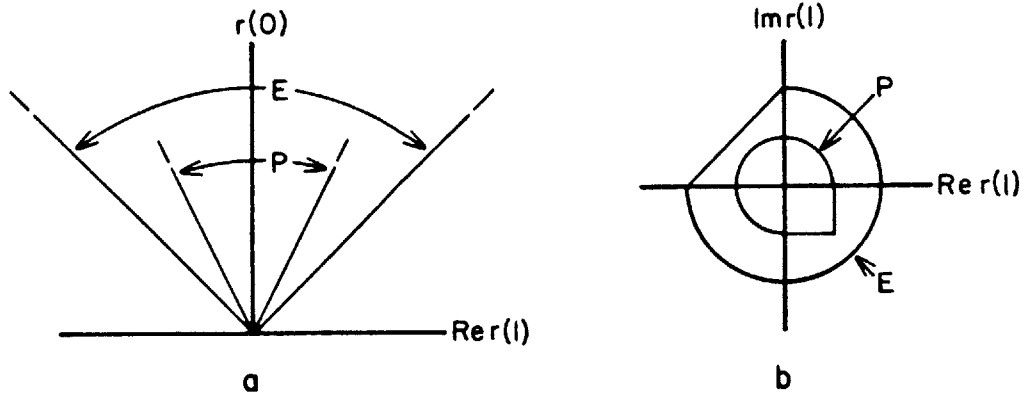


Fig. 3.2: E and P for $K = [-\pi, \frac{\pi}{2}]$ and $\Delta = \{0, \pm 1\}$. (a) shows a section through E and P at $\text{Im } r(1) = 0$ and (b) shows a section through E and P at $r(0) = 1$.

In the time series case, the extendibility theorem reduces to a test of the positive-definiteness of the correlation samples. Hence extendibility may be thought of as a more general analog of positive-definiteness.

Example 3.1: The time series case; $K = [-\pi, \pi]$, $\Delta = \{0, \pm 1, \dots, \pm M\}$.

In this case, the extendibility problem reduces to the trigonometric

moment problem (Grenander and Szegö 1958). Although not generally true, it follows in this case, from the fundamental theorem of algebra, that a positive polynomial may be factored as the squared magnitude of an M-th degree trigonometric polynomial:

$$P(k) = |A(k)|^2.$$

The inner product (r, p) becomes a Toeplitz form in the coefficients of the new polynomial

$$(r, p) = \sum_{i,j=0}^M a^*(i) r(i-j) a(j).$$

Thus the requirement that the inner product (r, p) be positive for all positive polynomials reduces to a requirement that the Toeplitz form corresponding to the correlation measurements be positive definite.

3.5 Boundary and Interior

Sometimes it is necessary to know more about a correlation vector than that it is extendible, or to know more about a polynomial than that it is positive. The discussion of Pisarenko's method in chapter 4, for example, involves vectors on the boundaries of E and P and the discussion of the maximum entropy method in chapter 5 involves vectors in the interiors.

The *boundary* of a closed set consists of those members which are arbitrarily close to some vector not in the set. The *interior* of a closed set consists of those members which are not on the boundary. The boundary and the interior of a finite dimensional set do not depend upon a particular choice of vector norm (Hoffman 1975). Moreover, since P and E are convex sets, they have interiors and boundaries which are particularly simple to characterize.

The boundary of P , denoted ∂P , consists of those positive polynomials which are zero for some $k \in K$. The interior of P , denoted P° , consists of those polynomials which are strictly positive on K .

Positive polynomials may be used to define the boundary and the interior of E . The boundary of E , denoted ∂E , consists of those extendible correlation vectors which make a zero inner product with some non-zero positive polynomial. The interior of E , denoted E° , consists of those correlation vectors which make strictly positive inner products with every positive polynomial.

3.6 Power Spectral Density Functions

Many spectral estimation methods represent the power spectrum, not as a measure, but as a spectral density function. This leads to a modification of the extendibility problem: given a fixed finite measure ν which defines the integral

$$\tau(\delta) = \int_K S(k) e^{jk \cdot \delta} d\nu, \quad \delta \in \Delta, \quad (3.9)$$

which correlation vectors τ can be derived from some bounded, strictly positive function S ? Under one additional constraint on ν , easily satisfied in practice, it can be shown that vectors which can be represented in this fashion are exactly those vectors in the interior of E .

The extension theorem for spectral density functions: If every neighborhood of every point in K contains a set of non-zero measure, then there exists a positive function $S(k)$ which is bounded and bounded away from zero such that

$$\tau(\delta) = \int_K S(k) e^{jk \cdot \delta} d\nu, \quad \delta \in \Delta,$$

if and only if

$$(\tau, p) > 0$$

for all positive p .

The proof of this theorem is contained in Appendix A. The ability to represent correlation vectors in E° by spectral density functions which are bounded away from zero will be of use in the proof of the existence of the MEM spectral estimate.

3.7 Sampling of the Spectral Support

A finite computational algorithm can only evaluate a function at a finite number of points. In practice, therefore, the only spectral supports which can be dealt with are those composed of a finite number of points:

$$K = \{k_i \in \mathbb{R}^D : i = 0, \dots, N-1\}. \quad (3.10)$$

Many spectral supports of interest contain an infinite number of points. In designing computational algorithms for spectral estimation on these spectral supports, the support must be approximated by one having a finite number of points. A measure μ on a support consisting of a finite number of points is completely characterized by its value $\mu(k_i)$ on each point. Thus the inverse Fourier integral reduces to a finite sum:

$$\int_K e^{jk \cdot \delta} d\mu = \sum_{i=0}^{N-1} e^{jk_i \cdot \delta} \mu(k_i). \quad (3.11)$$

Similarly, for spectral density functions,

$$\int_K e^{jk \cdot \delta} S(k) d\nu = \sum_{i=0}^{N-1} e^{jk_i \cdot \delta} S(k_i) \nu(k_i). \quad (3.12)$$

The measure ν can be considered to define a quadrature rule for integrals over the spectral support.

From the definitions of extendible correlation vectors and of positive poly-

nomials, it can be seen that if a new spectral support is formed by choosing a finite number of points out of some original spectral support then the new set E is a convex polytope inscribed within the original set E and the new set P is a convex polytope circumscribed about the original set P . Hence the new E is smaller than the original E and the new P is larger than the original P . By sampling the original spectral support sufficiently densely, these polytopes can be made to approximate the original sets to arbitrary precision. For example, Fig. 3.3 shows the effect of approximating the spectral support $[-\pi, \pi]$ by the four samples $\{0, \pm \frac{\pi}{2}, \pi\}$ for $\Delta = \{0, \pm 1\}$. The original E and P cones have a circular cross-section, as in Fig. 3.1. The cones corresponding to the sampled support have a square cross-section. The boundaries of the new and old cones intersect at vectors corresponding to the sample points.

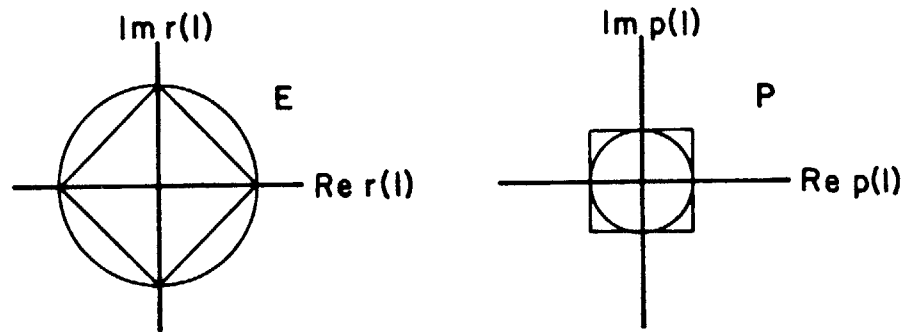


Fig. 3.3: Approximation of a spectral support by sampling; a section at $\tau(0) = 1$.

Chapter 4

Pisarenko's Method

4.1 Introduction

Pisarenko (1973) described a time series spectral estimation method in which the spectrum is modelled as the sum of line components plus a white noise component. If the white noise component is made as large as possible, he showed that the position and amplitudes of the lines needed to match the measured correlations are uniquely determined. In this section, Pisarenko's method will be derived in the more general array setting and for a more general noise component. The relationship of Pisarenko's method to the extendibility question will be demonstrated.

The extended Pisarenko's estimate will be derived as the solution of an optimization problem involving the minimization of a linear functional over a convex region defined by linear constraints. A solution to this optimization problem always exists, but it may not be unique. A dual optimization problem is derived which, in the time series case, leads to the familiar interpretation of Pisarenko's method as the design of a constrained least squares smoothing filter.

Algorithms for the computation of Pisarenko's method are discussed. A primal optimization problem is written, for the case of a spectral support composed of a finite number of points, as a standard form linear program. The application of the simplex method to the solution of this primal linear program is discussed. A dual linear program is presented which is related to the dual optimization problem developed before. The possible existence of

computational algorithms faster than the simplex method is discussed.

4.2 Pisarenko's Method for Sensor Arrays

The basis of Pisarenko's method is the unique decomposition of a correlation vector \mathbf{r} into a multiple of a given noise correlation vector \mathbf{n} , in the interior of E , plus a remainder \mathbf{r}' , on the boundary of E :

$$\mathbf{r} = \mathbf{r}' + \alpha \mathbf{n}. \quad (4.1)$$

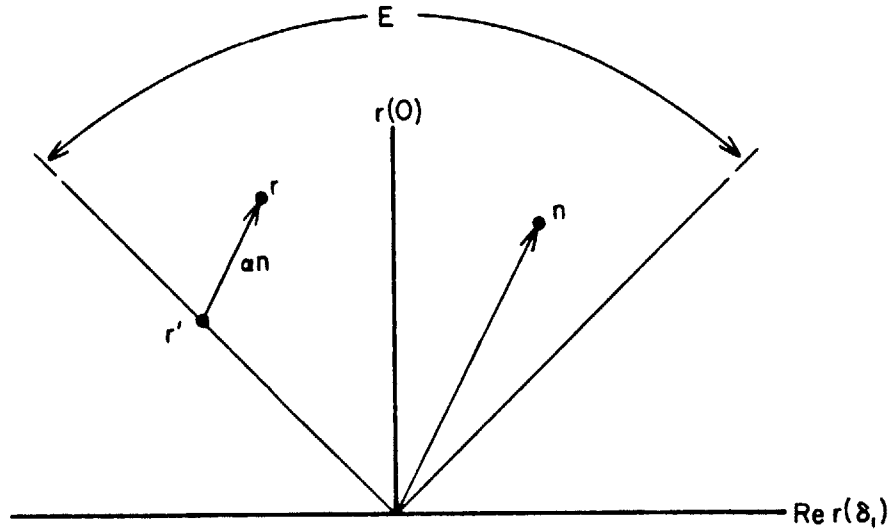


Fig. 4.1: Decomposition of a vector \mathbf{r} into a vector \mathbf{r}' on the boundary of E plus a multiple of a given vector \mathbf{n} .

The assumption that \mathbf{n} is in the interior of E implies that such a decomposition of an arbitrary vector \mathbf{r} exists and is unique. Consider the one-parameter family of correlation vectors

$$\mathbf{r}_c = \mathbf{r} - c\mathbf{n}. \quad (4.2)$$

For c sufficiently positive, \mathbf{r}_c must be non-extendible, since the assumption that $\mathbf{n} \in E^\circ$ implies that $n(0) > 0$. For c sufficiently negative, $\mathbf{r}_c = |c|[\mathbf{n} + \frac{1}{|c|}\mathbf{r}]$ must

be extendible, since the assumption that $n \in E^\circ$ implies that E contains a neighborhood of n . The convexity of E implies that there is some greatest number α such that $r' = r - \alpha n$ is extendible. Since it is arbitrarily close to some non-extendible vector, r' must be on the boundary of E . Furthermore, since $\alpha \geq 0$ if and only if r is extendible, this decomposition of r can also be used as an extendibility test.

This unique decomposition of r can be formulated as a primal spectral optimization problem. α is the largest number such that the remainder $r' = r - \alpha n$ is extendible. Thus

$$\alpha = \max_{\mu \geq 0} \frac{1}{n(0)} \left[r(0) - \int_K d\mu \right] \quad (4.3)$$

such that

$$r(\delta) = \int_K e^{jk \cdot \delta} d\mu + \frac{1}{n(0)} \left[r(0) - \int_K d\mu \right] n(\delta), \quad \delta \in \Delta, \delta \neq 0.$$

The maximum is attained for some positive measure μ' corresponding to the remainder r' . Since n is extendible, it corresponds to some positive measure μ_n . Hence

$$r(\delta) = \int_K e^{jk \cdot \delta} [d\mu' + \alpha d\mu_n]. \quad (4.4)$$

If $\alpha \geq 0$ then $\mu' + \alpha \mu_n$ is a positive measure which matches the correlation measurements and which has the largest possible noise component.

Polynomials can be used to derive some further information about the remainder r' and its spectral representation μ' . r' is on the boundary of E , hence it makes a zero inner product with some non-zero positive polynomial:

$$(r', p') = \int_K P'(k) d\mu' = 0. \quad (4.5)$$

It follows that $\mu'(B) = 0$ for any set B which is disjoint from the zero set of $P'(k)$. This suggests the final step in the derivation of Pisarenko's method, which is the association of the remainder τ' with an impulsive spectrum. The fact that the objective functional of the primal optimization problem is not strictly convex hints that the solution μ' will not be unique in general. One special case in which μ' is uniquely determined is the time series case considered by Pisarenko.

Example 4.1: The time series case: $K = [-\pi, \pi]$, $\Delta = \{0, \pm 1, \dots, \pm M\}$.

Since τ' is on the boundary of E , there is some non-zero positive p' such that $(\tau', p') = 0$. Every such p' and every spectral representation μ' of τ' must satisfy $(\tau', p') = \int_K P'(k) d\mu' = 0$. As in example 3.1, $P'(k) = |A(k)|^2$ for some M -th degree trigonometric polynomial $A(k)$. $A(k)$, and hence $P'(k)$, can be zero at no more than M points. The spectrum μ' , therefore, must be a sum of impulses at these points. Since it is possible to construct a positive polynomial which is zero at $N \leq M$ arbitrarily selected points, it follows that τ' has a unique spectral representation as a sum of impulses at the common zeros of all positive polynomials p' such that $(\tau', p') = 0$.

For the solution, μ' , to the primal optimization problem (4.3) to be unique for every correlation vector τ , every correlation vector τ' in ∂E must have a unique spectral representation. Example 4.1 shows that, in the time series case, every such τ' has a unique spectral representation in terms of no more than M impulses. More generally, the extension theorem combined with Carathéodory's theorem (Cheney 1966) shows that there is at least one spectral representation of τ' as a sum of no more than $2M$ impulses.

The representation theorem: If τ' is on the boundary of E, then for some $2M$ non-negative $a(i)$ and some $k_i \in K$:

$$\tau'(\delta) = \sum_{i=1}^{2M} a(i) e^{jk_i \cdot \delta}. \quad (4.6)$$

(See appendix B for proof.)

This representation, and thus the solution to the primal optimization problem (4.3) may not be unique. This uniqueness problem is discussed in appendix C, where it is shown that, in many multi-dimensional cases of interest, vectors exist on the boundary of E with more than one spectral representation. Although the existence of such vectors is shown it is not clear how common they may be.

Given τ , if α and the locations of the impulses in the solution μ' could be determined, then the impulse amplitudes could be calculated simply by solving a set of linear equations. A dual optimization problem (4.8) will now be derived which gives α and p' such that $(\tau', p') = 0$. Then, if τ' has a unique spectral representation, the line locations can then be determined from the zeros of $P(k)$. From the extendibility theorem,

$$(\tau', p) = (\tau - \alpha n, p) = (n, p)[(\tau, p)/(n, p) - \alpha] \geq 0. \quad (4.7)$$

Since $n \in E^\circ$ and $\tau' \in \partial E$, it follows that $(n, p) > 0$ and $(\tau', p) \geq 0$ for all $p \in P$. Furthermore, since $(\tau', p') = 0$ for some $p' \in P$, it follows that

$$\alpha = \min (\tau, p) \quad (4.8)$$

over the set

$$\{p \in P: (n, p) = 1\},$$

and that the minimum is attained at p' . In the time series case, this dual prob-

lem reduces to the same eigenvector problem derived by Pisarenko and leads to the interpretation of Pisarenko's method as determining a constrained least squares smoothing filter.

Example 4.2: The time series case: $K = [-\pi, \pi]$, $\Delta = \{0, \pm 1, \dots, \pm M\}$.

As in example 3.1,

$$(\tau, p) = \sum_{i,j=0}^M a^*(i) \tau(i-j) a(j).$$

Furthermore, if n corresponds to white noise,

$$(n, p) = p(0) = \sum_{i=0}^M |a(i)|^2.$$

Thus the dual optimization problem reduces to finding the eigenvector of the Toeplitz matrix associated with τ corresponding to the smallest eigenvalue. If there are several such eigenvectors, the impulses in μ' are located at the common zeros of the corresponding polynomials. Any normalized eigenvector corresponding to the minimal eigenvalue gives the coefficients of a smoothing filter, the sum of whose squared magnitudes is constrained to be one, which gives the least output power when fed an input process whose correlations are described by τ (Sullivan, Frost, and Treichler 1978).

4.3 The Computation of Pisarenko's Estimate

In the design of algorithms to compute Pisarenko's estimate, one is necessarily concerned with spectral supports composed of a finite number of points:

$$K = \{k_i \in \mathbb{R}^D : i = 0, \dots, N-1\}. \quad (4.9)$$

A measure μ on such a support is completely characterized by its value $\mu(k_i)$ on each point. The primal problem (4.3) can be written as the standard form

linear program

$$\min_{\mu \geq 0} \sum_{i=0}^{N-1} \mu(k_i) \quad (4.10)$$

such that

$$\sum_{i=0}^{N-1} \left[e^{jk_i \cdot \delta} - \frac{n(\delta)}{n(0)} \right] \mu(k_i) = \tau(\delta) - \frac{\tau(0)n(\delta)}{n(0)}, \quad \delta \in \Delta, \delta \neq 0 \quad (4.11)$$

with N variables and $2M$ constraints. The fundamental theorem of linear programming (Luenberger 1973) is equivalent to the representation theorem in this case. Given that a solution exists to this linear program, as shown in the previous section, the fundamental theorem guarantees a solution in which no more than $2M$ of the $\mu(k_i)$'s are non-zero, a so-called *basic* solution.

The (unsymmetric) dual linear program can be shown (Luenberger 1973) to be

$$\max_{\lambda} \sum_{\delta \neq 0} \lambda(\delta) \left[\tau^*(\delta) - \frac{\tau(0)n^*(\delta)}{n(0)} \right] \quad (4.12)$$

such that

$$\sum_{\delta \neq 0} \lambda(\delta) \left[e^{-jk_i \cdot \delta} - \frac{n^*(\delta)}{n(0)} \right] \leq 1, \quad i = 0, \dots, N-1. \quad (4.13)$$

The dual optimization problem derived in the last section can be derived from this linear program by defining

$$p(\delta) = -\frac{1}{n(0)} \lambda(\delta), \quad \delta \neq 0, \quad (4.14)$$

and by introducing the additional variable $p(0)$ and adjoining the additional constraint equation

$$(n, p) = 1. \quad (4.15)$$

The primal problem can be solved using the *simplex method* (Luenberger 1973). The application of the simplex method to the primal problem results in essentially the same computational algorithm as the application of the (single) *exchange method* to the dual problem (Stiefel 1960). By incorporating a technique to avoid cycling, such as that due to Charnes (1952), an algorithm can be obtained which is guaranteed to converge to an optimal solution in a finite (typically $O(M)$) number of steps.

The problem of Chebyshev approximation is related to the computation of Pisarenko's estimate; it also can be formulated as the minimization of a linear functional over a convex space defined by linear inequality constraints (Cheney 1966). It also has been solved using the simplex (single exchange) method. However, for the particular problem of the Chebyshev approximation of continuous functions by polynomials in one variable, a computational method exists which is significantly faster than the simplex method, a *multiple exchange* method due to Remes. Although attempts have been made to extend this method to more general problems (Harris and Mersereau 1977), the resulting algorithms are not well understood; in particular, there is no proof of convergence. Recent work in this area (Blatt 1978) is promising, however, and the development of a reliable multiple-exchange method for Chebyshev approximation would have important consequences for the related problem of the computation of Pisarenko's estimate.

Chapter 5

The Maximum Entropy Method

5.1 Introduction

In maximum entropy method (MEM), a unique spectral estimate is obtained by optimizing a convex functional over a collection of spectral density functions which satisfy the correlation-matching constraints. The positivity constraint is implicitly enforced by the selection of a functional which penalizes low values of the spectral density function.

The maximum entropy method is developed in the array processing setting. Conditions are posed which guarantee the existence and uniqueness of the MEM spectral estimate and for which the estimate can be characterized as having a particular parametric form. These conditions involve certain constraints on the coarray and the spectral support which are easily met in practice. The calculation of the MEM estimate is reduced to the solution of either of two finite dimensional convex optimization problems. The application of standard optimization techniques to these problems results in iterative computational algorithms which are guaranteed to converge. The resulting algorithms are compared to those previously proposed.

5.2 The Maximum Entropy Spectrum

MEM is so named because it can be derived (Burg 1975) as the solution of a constrained entropy maximization problem. Consider optimizing the entropy functional

$$H = \int_{\mathbf{k}} \ln S(\mathbf{k}) d\nu \quad (5.1)$$

over the set of all continuous and strictly positive functions of k which satisfy the correlation-matching constraints

$$r(\delta) = \int_K S(k) e^{jk \cdot \delta} d\nu, \quad \delta \in \Delta. \quad (5.2)$$

Integrals over K are defined by some measure ν , as described in section 3.6. If an optimum S' exists of the required form then a Lagrange multiplier theorem (Luenberger 1969, p. 188) implies that

$$S'(k) = \frac{1}{P'(k)} \quad (5.3)$$

for some strictly positive Δ -polynomial P' . Conversely, if there exists a positive Δ -polynomial P' whose inverse $S'(k)$ is a spectral density function which satisfies the correlation-matching constraints, then the convexity of H guarantees the uniqueness and optimality of S' .

The existence proof presented in appendix D, therefore, merely demonstrates that if r is a correlation vector in the interior of E , then there exists a polynomial p' , in the interior of P , such that $\frac{1}{P'(k)}$ is a spectral density function which satisfies the correlation-matching constraints. The basis of the proof is to consider (5.2, 5.3) as a mapping from P° into E° , which must be shown to be invertible. It is easily seen that this mapping is continuous and one-to-one. Therefore, if it can be shown that points on the boundary of P are mapped into points on the boundary of E , the mapping must be onto, hence invertible.

Two conditions are assumed in the proof of this theorem. First it is assumed that every neighborhood of every point in K contains a set of non-zero measure. This is the same condition posed in section 3.6, under which correlation vectors in E° can be associated with spectral density functions which are bounded and bounded away from zero. Second, it is assumed that if $p_n \in P^\circ$ is a

sequence converging to $p \in \partial P$, then

$$\int_K \frac{d\nu}{P_n(k)} \rightarrow \infty. \quad (5.4)$$

It follows that if r_n is the correlation vector corresponding to the spectral density function $\frac{1}{P_n(k)}$, then

$$\left(\frac{r_n}{r_n(0)}, p_n \right) = \left[\int_K \frac{d\nu}{P_n(k)} \right]^{-1} \rightarrow 0. \quad (5.5)$$

Thus the sequence of normalized correlation vectors $\frac{1}{r_n(0)} r_n$ approaches ∂E .

This shows that points on the boundary of P are mapped into points on the boundary of E . Although there are some simple spectral supports which do not satisfy these two conditions, the supports can easily be modified to do so. These conditions, therefore, are not very restrictive in practice.

5.3 Dual Optimization Problems

In this section, it is shown that the polynomial corresponding to the MEM estimate can be obtained as the solution to either of two finite-dimensional optimization problems, both of which require the optimization of a convex functional over a convex set. These well-behaved optimization problems can be successfully attacked with a number of standard algorithms.

The correlation vector, as well as the entropy, may be thought of as functions of a polynomial $p \in P^n$:

$$r_p(\delta) = \int_K \frac{e^{jk \cdot \delta}}{P(k)} d\nu, \quad (5.6)$$

$$H(p) = - \int_K \ln P(k) d\nu. \quad (5.7)$$

Taking the gradient of H with respect to the polynomial coefficients results in

$$\frac{\partial H}{\partial p(\delta)} = - \int_K \frac{e^{-jk \cdot \delta}}{P(k)} d\nu = -r_p^*(\delta). \quad (5.8)$$

The polynomial p' corresponding to the MEM spectrum makes a known inner product with the correlation vector:

$$(\tau, p') = \sum_{\delta \in \Delta} \tau^*(\delta) p'(\delta) = \int_K \frac{P'(k)}{P(k)} d\nu = \nu(K). \quad (5.9)$$

Therefore, the hyperplane $(\tau, p) = \nu(K)$ is tangent to the surface $H(p) = H(p')$ at the point $p = p'$, where the correlation matching constraints are satisfied.

This situation is illustrated in Fig. 5.1.

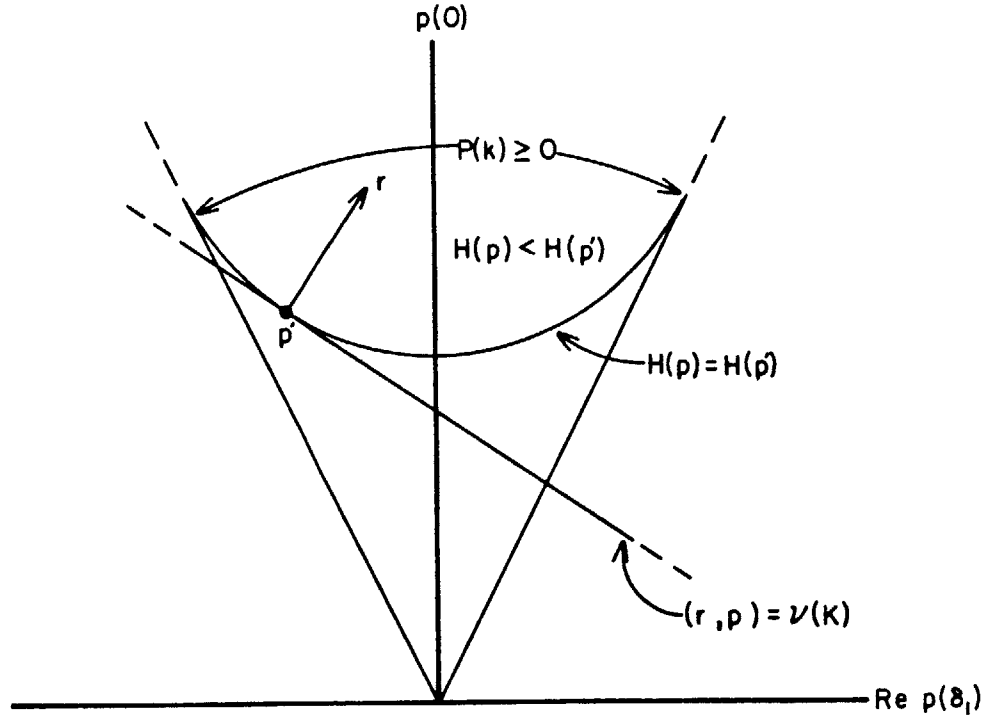


Fig. 5.1: The hyperplane $(\tau, p) = \nu(K)$ is tangent to the surface $H(p) = H(p')$ at the point $p = p'$.

$H(\mathbf{p})$ is a strictly convex function of \mathbf{p} . It follows that surfaces of constant entropy are convex, and that the intersection point \mathbf{p}' may be obtained as the solution to either of two optimization problems, which are dual to the primal entropy maximization problem (5.1, 5.2).

Of all points on the hyperplane $(\mathbf{r}, \mathbf{p}) = \nu(K)$, the minimum entropy is attained at $\mathbf{p} = \mathbf{p}'$:

$$H(\mathbf{p}') = \min H(\mathbf{p}) \text{ such that } (\mathbf{r}, \mathbf{p}) = \nu(K). \quad (5.10)$$

Of all points in the convex set $H(\mathbf{p}) \leq H(\mathbf{p}')$, the minimum projection onto \mathbf{r} is attained at $\mathbf{p} = \mathbf{p}'$:

$$\nu(K) = \min (\mathbf{r}, \mathbf{p}) \text{ such that } H(\mathbf{p}) \leq H(\mathbf{p}'). \quad (5.11)$$

The first optimization problem (5.10) is amenable to solution by any of a number of standard techniques, such as the method of steepest descent or a quasi-Newton method. It will form the basis for a computational algorithm to be discussed in the next section.

In the second optimization problem (5.11), it is not necessary to know $H(\mathbf{p}')$. Since the objective functional is linear, the optimum will be attained on the boundary of the constraint set. It can be shown that the solution of (5.11) simply scales with $H(\mathbf{p}')$, thus it may be solved with the equality constraint $H(\mathbf{p}) = 0$, giving \mathbf{q} . It is easily verified that the scaling yields

$$\mathbf{p}' = \frac{\nu(K)}{(\mathbf{r}, \mathbf{q})} \mathbf{q} \quad (5.12)$$

and that

$$H(\mathbf{p}') = \ln \left[\frac{(\mathbf{r}, \mathbf{q})}{\nu(K)} \right]. \quad (5.13)$$

In the time-series case, the second optimization problem leads to the usual

interpretation of the MEM estimate as a least-squares predictor.

Example 5.1: MEM in the time-series case: $K = [-\pi, \pi]$, $\Delta = \{0, \pm 1, \dots, \pm M\}$.

$P(k) \in \mathcal{P}$ may be factored:

$$P(k) = |A(k)|^2 = c \prod_{i=1}^M |1 - z_i e^{-jk}|^2$$

where A is a M -th degree trigonometric polynomial with zeros at the locations z_i . In this case,

$$H(p) = \ln c$$

and

$$(r, p) = \int_{-\pi}^{\pi} S(k) P(k) \frac{dk}{2\pi}.$$

Therefore, minimizing (r, p) s.t. $H(p)=0$ is equivalent to minimizing the power out of the linear predictor, with transfer function $1 + \sum_{n=1}^M a(n)z^{-n}$, which is fed an input process whose correlation are described by r . This results in the familiar Toeplitz normal equations (Burg 1975)

$$\sum_{j=1}^M r(i-j)a(j) = -r(-i), \quad i = 1, \dots, M$$

$$(r, q) = r(0) + \sum_{j=1}^M r(-j)a(j)$$

and giving the MEM spectral estimate

$$S(k) = \frac{(r, q)}{|A(k)|^2}.$$

5.4 Application of the Method of Steepest Descent

The method of steepest descent is one of the oldest methods of minimization, as well as the basis for more advanced quasi-Newton methods (Luenberger 1973). This section discusses the application of steepest descent to the problem of obtaining the MEM spectral estimate, in particular the solution of

$$H(\mathbf{p}') = \min H(\mathbf{p}) \text{ such that } (\mathbf{r}, \mathbf{p}) = \nu(K). \quad (5.14)$$

The constraint $(\mathbf{r}, \mathbf{p}) = \nu(K)$ can be used to eliminate the parameter $\mathbf{p}(0)$ via the equation

$$\mathbf{p}(0) = \frac{1}{\mathbf{r}(0)} [\nu(K) - \sum_{\delta \neq 0} \mathbf{r}^*(\delta) \mathbf{p}(\delta)]. \quad (5.15)$$

The reduced problem is now an unconstrained minimization problem in the $2M$ real variables $\{\text{Re } \mathbf{p}(\delta), \text{Im } \mathbf{p}(\delta); \delta \in \Delta, \delta \neq 0\}$.

In the method of steepest descent, the iteration

$$\mathbf{p}_{i+1} = \mathbf{p}_i - \alpha_i \mathbf{g}_i \quad (5.16)$$

is used, where α_i is a non-negative scalar minimizing $H(\mathbf{p}_i - \alpha_i \mathbf{g}_i)$ and \mathbf{g}_i is the gradient vector:

$$\mathbf{g}_i(\delta) = -2[\mathbf{r}_{\mathbf{p}_i}(\delta) - \frac{\mathbf{r}_{\mathbf{p}_i}(0)}{\mathbf{r}(0)} \mathbf{r}(\delta)], \quad \delta \neq 0. \quad (5.17)$$

$\mathbf{g}_i(0)$ is determined from the constraint $(\mathbf{r}, \mathbf{p}_i) = \nu(K)$, thus

$$\mathbf{g}_i(0) = -\frac{1}{\mathbf{r}(0)} \sum_{\delta \in \Delta} \mathbf{r}^*(\delta) \mathbf{g}_i(\delta). \quad (5.18)$$

Thus, in the i -th iteration, the gradient \mathbf{g}_i is calculated. Since the correlations $\mathbf{r}_{\mathbf{p}_i}$, of the spectrum $\frac{1}{\mathbf{P}_i(k)}$, must be calculated to obtain the gradient, the difference between these correlations and the desired correlations, $\mathbf{r} - \mathbf{r}_{\mathbf{p}_i}$, can

be conveniently used as a stopping criterion. Then a line search is performed from the point p_i , in the direction $-g_i$, to a minimum of H . The minimum point is taken as p_{i+1} . Since H is convex, it is convenient to search for a zero of the derivative

$$\frac{dH(p_i - \alpha g_i)}{d\alpha} = \int_K \frac{G_i(k)}{P_i(k) - \alpha G_i(k)} d\nu \quad (5.19)$$

rather than a minimum of H . Note that $G_i(k)$ should be positive somewhere on K , since

$$0 = (r, g_i) = \int_K \frac{G_i(k)}{P'(k)} d\nu \quad (5.20)$$

and $\frac{1}{P'(k)} > 0$. Thus $\frac{dH}{d\alpha}$, which starts out negative, turns positive when $P_i(k) - \alpha G_i(k)$ becomes small in some region of K where G_i is positive. The line search is restricted to the interval $\alpha \in \left[0, \max_K \frac{G_i(k)}{P_i(k)}\right]$, within which a zero of $\frac{dH}{d\alpha}$ must lie. The convexity of H guarantees the convergence of this algorithm.

Of course, in actual algorithms, one is necessarily concerned with spectral supports composed of a finite number of points:

$$K = \{k_i \in \mathbb{R}^D : i = 0, \dots, N-1\}. \quad (5.21)$$

The measure ν , which defines a particular integral over K , is completely characterized by its values $\nu(k_i)$ on each point. The integral of a function over K reduces to a finite sum

$$\int_K F(k) d\nu = \sum_{i=0}^{N-1} F(k_i) \nu(k_i). \quad (5.22)$$

The measure ν then, merely defines a quadrature rule for integrals over K . As long as $\nu(k_i) > 0$, the conditions for the existence of the MEM estimate,

discussed in section 5.2, are trivially met.

5.5 Comparison to Other Algorithms

In his Ph.D. thesis (pp. 97), Burg discussed a general variational approach to the estimation of a function and he proposed a general, iterative, solution procedure. When his solution procedure is specialized to the MEM problem, it is seen to be an application of a version of Newton's Method to the problem:

$$\min [H(\mathbf{p}) + (\mathbf{r}, \mathbf{p})]. \quad (5.23)$$

This minimization problem can be shown to be dual to the original entropy maximization problem (5.1, 5.2). The two minimization problems (5.10, 5.11) derived in section 5.3 can be derived from (5.23) by holding one of the two terms fixed. This reduction in dimensionality is obtained by using the additional information that $(\mathbf{r}, \mathbf{p}') = \nu(K)$. The convexity of the objective functional in (5.23) guarantees the convergence of a suitably modified Newton's method (Luenberger 1973), although not of the version suggested by Burg. Specifically, a line search along the search direction must be done. Other authors have actually constructed iterative algorithms for the computation of the MEM estimate. Several of these can be shown to be the result of applying standard optimization techniques to (5.23).

Ong (1971) actually applied a version of Newton's method to (5.23) for the computation of two-dimensional MEM spectra. Again, lack of a line search means that the resulting algorithm is not guaranteed to converge.

The algorithm proposed by Woods (1976) is similar to the method of steepest descent applied to (5.23). Woods, deals with a normalized polynomial $q = c\mathbf{p}$, where $c = \frac{1}{p(0)}$, with the corresponding spectral density function

$\frac{c}{Q(k)}$. The resulting algorithm iteratively updates c and q in a manner similar to moving p in the direction of the negative gradient of (5.23). It is not exactly the same, however. Furthermore, required calculations are done only approximately. Either of these characteristics, as well as the absence of a line search along the search direction, create potential convergence problems.

More recently, Jain and Ranganath (1978) have applied a coordinate descent method to the solution of (5.23). They do implement a line search and the resulting algorithm is guaranteed to converge, although typically much more slowly than steepest descent (Luenberger 1973).

Two other distinctly different approaches to computing the MEM estimate have been taken. Wernecke and D'Addario (1977) have attacked the primal problem (5.1, 5.2). Although this allows them to trade off accuracy in correlation-matching for increases in entropy, it results in more computation because the primal problem is of much higher dimensionality than the dual problems, since typically $N \gg M$. Lim and Malik (1981) have used an alternating-projection type of algorithm which has not been proven to converge.

Chapter 6

Spectral Estimation Examples

6.1 Introduction

Three examples of spectral estimation using MEM and Pisarenko's method are presented in this chapter. These examples provide an opportunity to review some of the ideas presented in previous chapters, to expand upon them in a specific context, and to discuss some details of computation and implementation.

The MEM estimate was computed using the method of steepest descent, as discussed in section 5.4, for the first five iterations, in order to get close to the optimum. Succeeding iterations, if any, used a self-scaling quasi-Newton method due to Luenberger (1973), which converges significantly faster. The quasi-Newton method involves only few modifications to the the method of steepest descent, principally involving the accumulation of an approximate Hessian matrix and the use of this matrix to modify the gradient vector in the choice of a search direction. The stopping criterion used involved the Chebyshev norm of the difference between the correlation data and the correlations of the current model, considered as vectors in \mathbb{R}^{2M+1} . When the norm of this error was sufficiently small, less than $\tau(0) \times 10^{-2}$ for the examples to be presented, the iteration was stopped. This stopping criterion was arbitrarily selected. Tightening this error threshold did not, however, significantly change the spectral estimates in the following examples.

Pisarenko's estimate was computed using the simplex method to solve the primal linear program (4.10, 4.11), as described in section 4.3. The algorithm is

guaranteed to stop after a finite number of steps.

The desired spectral supports consisted of infinite numbers of points; they were approximated by supports containing a finite number of points, densely sampling the desired supports. The bulk of the time spent in the computation of both the MEM and Pisarenko's estimate was used in, essentially, evaluating polynomials over the spectral support. This required $O(MN)$ computations per iteration, for a spectral support consisting of N points. The algorithms were written in the C programming language and run under the UNIX* operating system on a PDP* 11/50. Because of the small address space of the computer used, sine and cosine tables had to be stored on disk instead of in main memory, which significantly slowed the algorithms. An exception was the simple 1-D MEM example of section 6.2, for which a simple recursion was used to calculate the necessary sines and cosines.

6.2 The Importance of a Proper Spectral Support

Fig. 6.1 shows a 6 sensor linear non-uniform array and its coarray, which consists of 15 non-zero $\pm\delta$ pairs, specifically all the integer multiples of ± 0.75 up to ± 12.75 except for ± 10.5 and ± 11.25 . Consider the correlation vector whose elements are given by

$$r(\delta) = \int_{-\pi}^{\pi} e^{jk\delta} S(k) \frac{dk}{2\pi}, \quad \delta \in \Delta, \quad (6.1)$$

where $S(k)$ is pictured in Fig. 6.2.

The spectral support, often known from dispersion relations or from sensor specifications, constitutes important prior information about the spectrum.

*UNIX is a trademark of Bell Laboratories.

*PDP is a trademark of Digital Equipment Corporation.

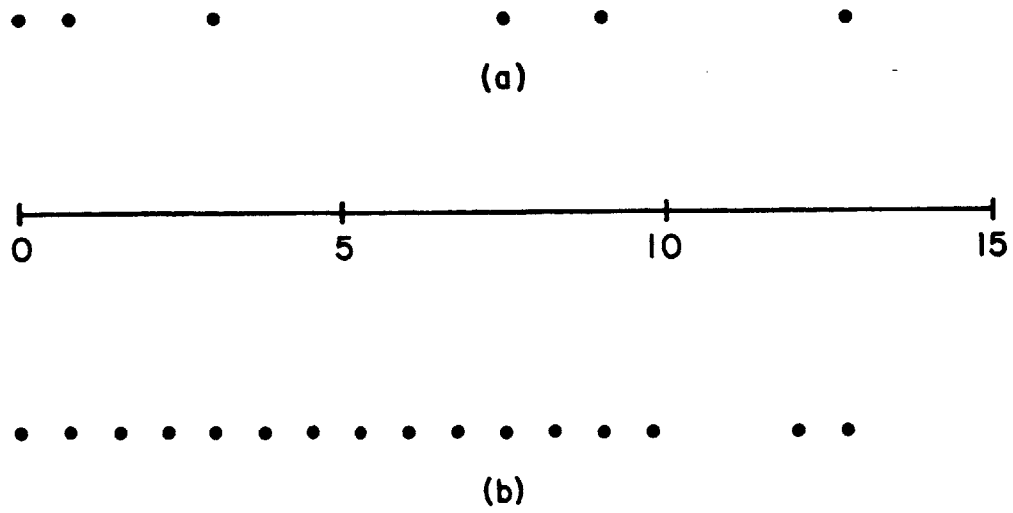


Fig. 6.1: Six sensor linear non-uniform array (a) and the positive portion of its coarray (b).

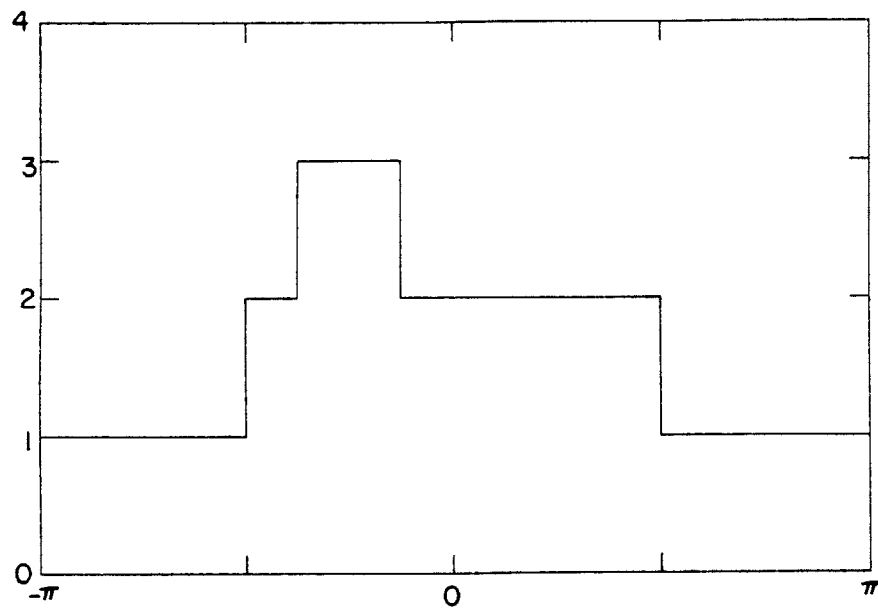


Fig. 6.2: Original spectral density function, whose correlation function, sampled on the coarray, produced the correlation vector.

This example concerns the effects of choosing the spectral support incorrectly.

Several spectral estimates, both MEM and Pisarenko's method, were computed for spectral supports approximating various intervals. All used a spectral support of 401 points, uniformly distributed across the interval to be approximated. For the MEM estimates, the measure ν corresponded to quadrature of the integrals by the trapezoidal rule; $\nu(k_i) = \frac{m(K)}{400}$, except for the endpoints, which were half of this value. Here $m(K)$ denotes the length of the interval K .

First the MEM spectrum was computed using the correct spectral support of $[-\pi, \pi]$. The result, shown in Fig. 6.3, is a reasonable facsimile of the original spectrum.

Next, the MEM spectrum was computed using a larger spectral support of $[-\frac{5\pi}{4}, \frac{5\pi}{4}]$. The result, shown in Fig. 6.4, has marked differences from the original spectrum. These differences can be explained by recalling that the MEM spectrum maximizes the entropy integral

$$\int_{-\frac{5\pi}{4}}^{\frac{5\pi}{4}} \ln S(k) \frac{dk}{2\pi} \quad (6.2)$$

subject to the correlation constraints. Since the original spectrum is zero outside of the region $[-\pi, \pi]$, it has an entropy of $-\infty$. MEM picks a spectrum which is positive everywhere in order to obtain a higher entropy; to do this while still satisfying the correlation matching constraints requires quite complicated behavior in the region $[-\pi, \pi]$. In fact, computing Pisarenko's estimate, for a noise vector corresponding to a constant spectral density level of 1 over the interval $[-\frac{5\pi}{4}, \frac{5\pi}{4}]$, resulted in an $\alpha = 3.3 \times 10^{-4}$. Thus τ is close to ∂E , the polynomial corresponding to the MEM estimate is close to ∂P , and high peaks in the MEM estimate are to be expected.

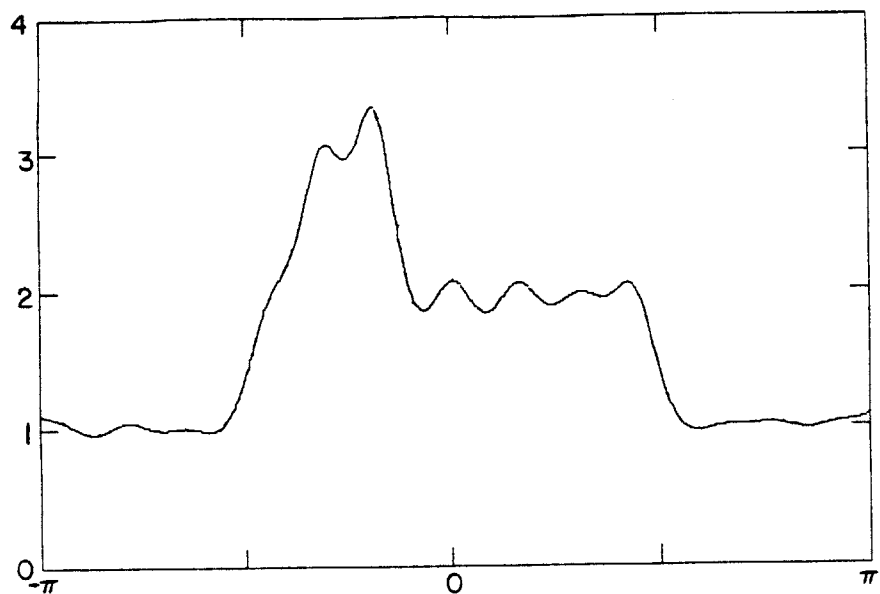


Fig. 6.3: MEM spectral estimate for a spectral support of $[-\pi, \pi]$. (5 iterations, 13 seconds CPU time.)

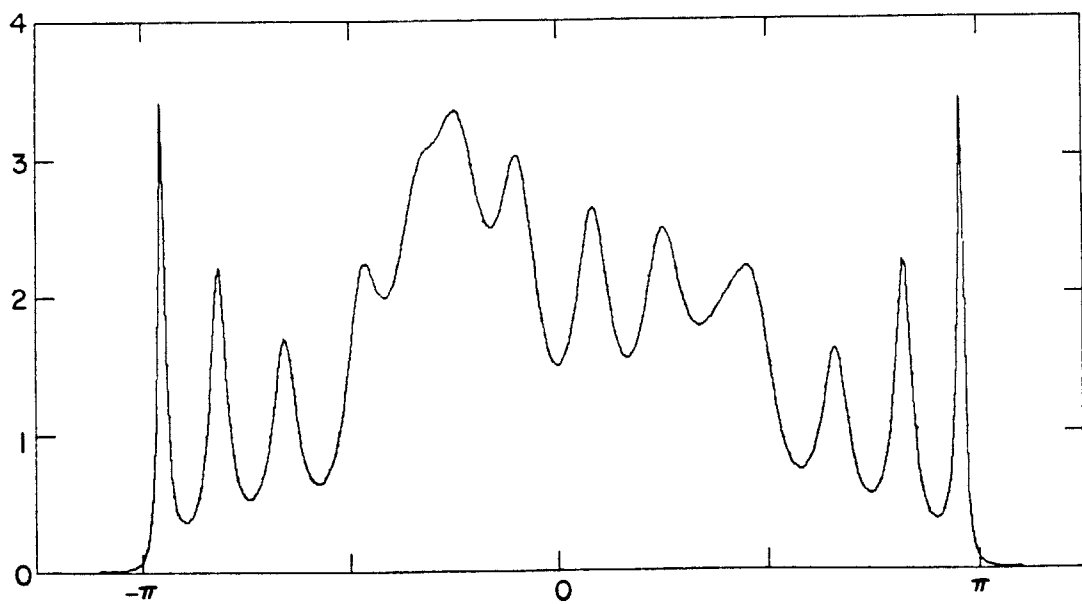


Fig. 6.4: MEM spectral estimate for a spectral support of $[-\frac{5\pi}{4}, \frac{5\pi}{4}]$. (36 iterations, 97 seconds CPU time.)

The introduction of significant regions into the spectral support where there is no power has done considerable violence to the MEM estimate. This behavior will be common to all spectral estimation methods which optimize functionals severely penalizing low spectral values in order to implicitly enforce the positivity constraints, as proposed by Burg (1975).

Finally, the spectral support was chosen too small, $[-\frac{3\pi}{4}, \frac{3\pi}{4}]$. When an attempt was made to compute the MEM estimate, a positive polynomial was discovered, on the ninth iteration, which makes a zero inner product with the correlation vector. Thus the correlation vector is outside of E° . To verify this, Pisarenko's estimate was computed, for a noise vector corresponding to a constant spectral density level of 1 over the interval $[-\frac{3\pi}{4}, \frac{3\pi}{4}]$. This resulted in an alpha of -1.7×10^5 , indicating that the correlation vector is not extendible. Unlike the poor behavior of MEM when the spectral support was made too large, this non-extendibility problem when the spectral support is made too small is fundamental; there is no positive spectrum on this too-small support which matches the correlations.

6.3 Choosing the Correct Spectral Estimation Method

Given the model of the array processing problem developed in chapter 2, the task of a spectral estimation method is simply to pick out one of the many positive spectra which are consistent with the correlation measurements and the known spectral support. There are many ways to do this, Pisarenko's method and MEM are just two. The selection of an estimation method is a way in which to incorporate additional prior knowledge into the problem. Consider the array and coarray pictured in Fig. 6.5 and the spectral support pictured in Fig.

6.6, consisting of the boundary of the circle of radius π , corresponding to two-dimensional waves in a homogeneous, non-dispersive medium. Further, consider the correlation vector corresponding to two impulses of unit power at 270° and at 275° plus noise of constant spectral density level 1.

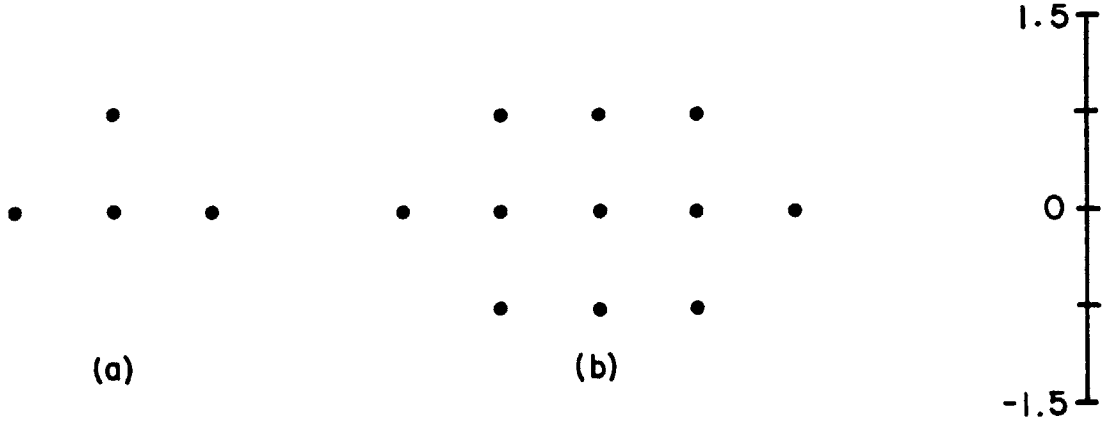


Fig. 6.5: Four sensor array (a) and its coarray (b).

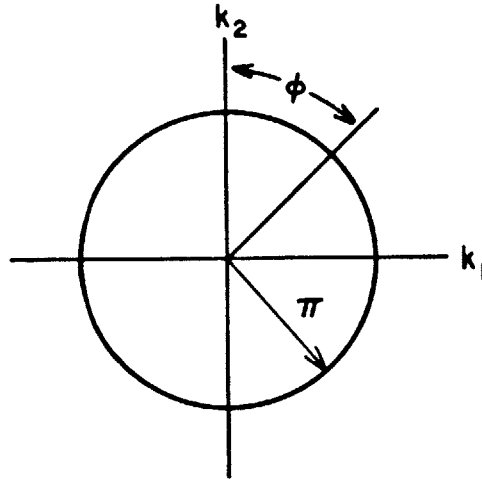


Fig. 6.6: Spectral support for two-dimensional waves.

Both MEM and Pisarenko spectral estimates were computed for spectral supports approximating the circle. Both used a spectral support of 360 points, one per degree. For the MEM estimates, the measure ν assigned equal weight to each sample point. Fig. 6.7 shows the MEM spectral estimate. Note that the two

peaks are not resolved. Table 6.1 shows the Pisarenko estimate, in the form of power at $2M$ points. Note that, except for the effects of machine rounding error the peaks are exactly resolved. Thus, if it is known that the spectrum consists of point sources in background noise of some known spectral shape, then Pisarenko's method is more appropriate than MEM because it effectively incorporates this prior knowledge into the estimation problem.

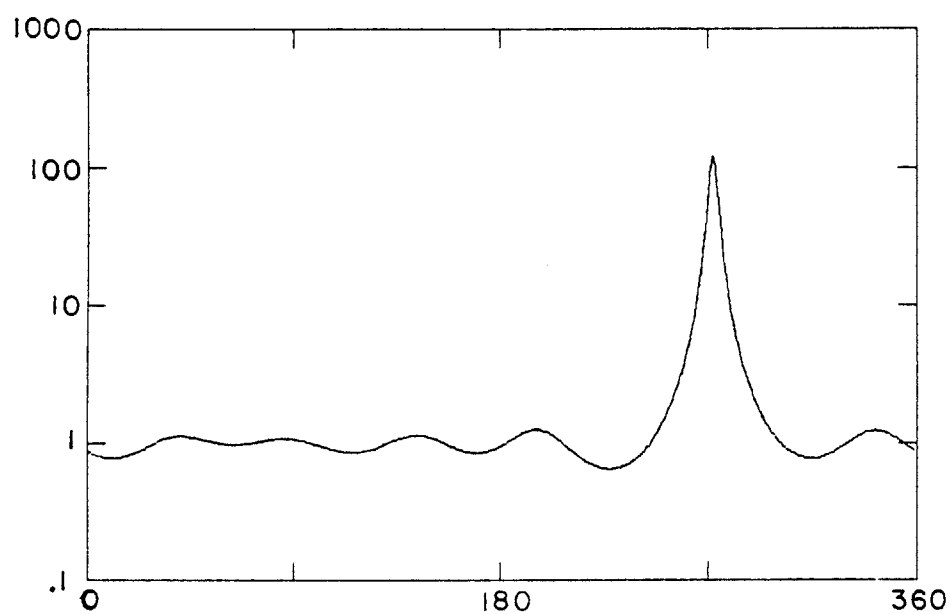


Fig. 6.7: MEM spectral estimate of two point sources in isotropic noise. (16 iterations, 33 seconds CPU time.)

$\alpha = 0.999999$	
φ	$\mu(\varphi)$
12	0.000000
13	0.000000
91	0.000000
167	0.000000
269	0.011591
270	0.968303
271	0.023218
275	0.992969
276	0.003917
305	0.000000

Table 6.1: Pisarenko's estimate of two point sources in isotropic noise. (105 iterations, 72 seconds CPU time.)

Consider the effect on Pisarenko's estimate of adding a point source, of unit power, at 180.5° , half-way between two sample points. In order to represent this point source by power at the sample points requires power at $2M$ sample points. These sample points do not include both 270° and 275° . When all three point sources are present, the spectra do not simply add. In fact the presence of the point source at 180.5° significantly disturbs the spectrum around 270° (Table 6.2).

$\alpha = 0.997803$	
φ	$\mu(\varphi)$
62	0.000429
63	0.000062
118	0.000150
180	0.500525
181	0.500090
270	0.806229
273	0.860291
277	0.245172
278	0.088951
359	0.000293

Table 6.2: Pisarenko's estimate of three point sources in isotropic noise. (55 iterations, 41 seconds CPU time.)

6.4 A Two-Dimensional MEM Spectrum

Consider the array and coarray in Fig. 6.8 and a spectral support, consisting of a circular disk of radius $2\sqrt{\pi}$. This spectral support can be considered as a limiting case of the spherical cap support of example 2.1, where the beamwidth of the dish antennas is very small.

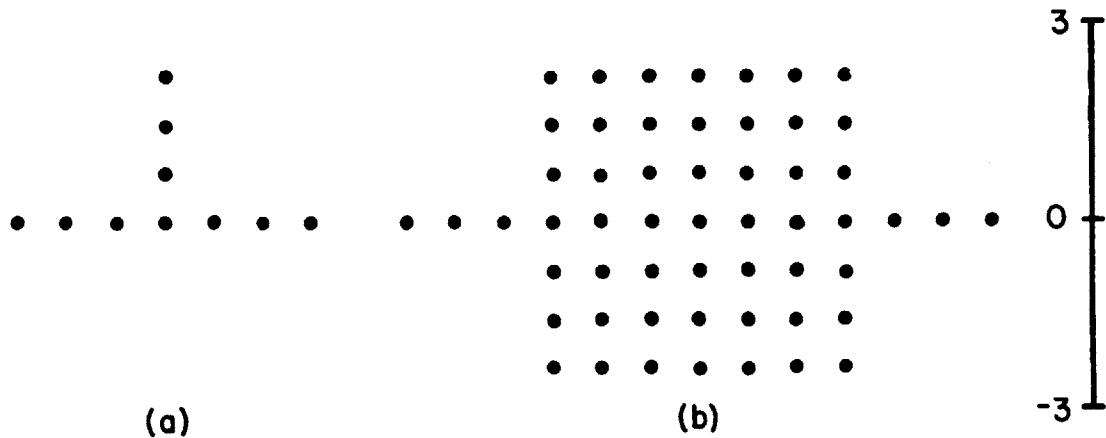


Fig. 6.8: Array (a) and coarray (b) for 2-D example.

Consider the correlation vector corresponding to the spectral density function in Fig. 6.9, which is of constant spectral level 1, except for a central circular portion, of radius $\sqrt{\pi}$, which is of spectral level 2. The spectral support was sampled in rectangular pattern, approximately 60 points from the center to the radius. If more sample points had been used, an entire spectral estimate could not have been kept in main memory at one time. Equal weight was assigned to each point.

Fig. 6.10 is a contour plot of the resulting MEM spectral estimate. Although the correlations correspond to a circularly symmetric spectrum, the asymmetry of the coarray causes an asymmetry in the spectral estimate. In other respects, the estimate is a reasonable copy of the original spectrum, exhibiting a rise at radii near $\sqrt{\pi}$ and plateaus at larger and smaller radii.

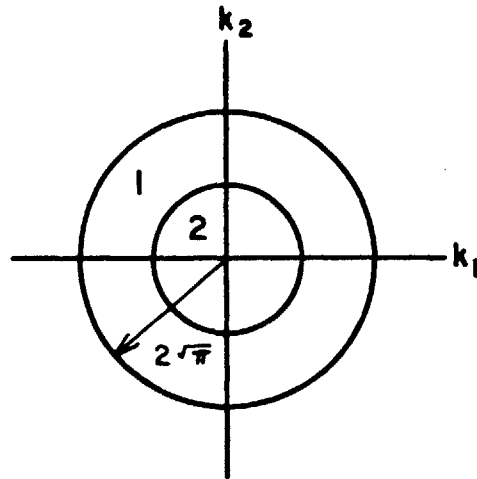


Fig. 6.9: Spectrum from which correlation vector was derived.

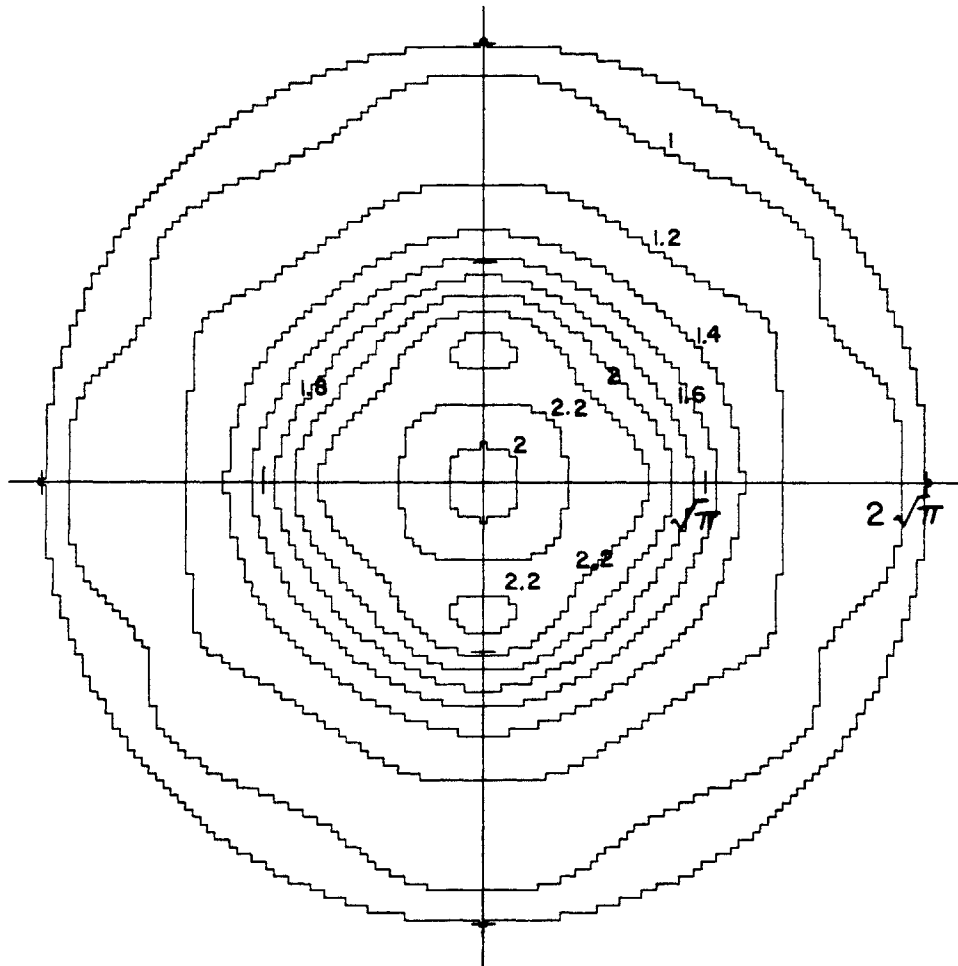


Fig. 6.10: MEM spectral estimate. (2 iterations, 733 seconds CPU time.)

Chapter 7

Summary

This thesis has been concerned with what is probably the simplest interesting problem in array processing; the estimation of a power spectrum with a known support, given certain samples of its correlation function. This problem retains the key ingredients of multi-dimensional spectra, non-uniformly sampled correlation functions, and arbitrary spectral supports. The investigation of correlation-matching spectral estimates led to the extendibility problem. This problem is called the trigonometric moment problem, in the time series case, and its solution involves consideration of the positive-definiteness of the correlation samples. Extendibility can therefore be considered as a generalization of positive-definiteness. Trigonometric polynomials, so common in the trigonometric moment problem, turned out to be a special case of Δ -polynomials, of central importance in the extendibility problem. The representation of power spectra, both as measures and as power spectral density functions, was considered.

Building on the theoretical framework developed in solving the extendibility problem, two spectral estimation methods were extended from the time series case to the array processing problem. Pisarenko's method was shown to be intimately related to the extendibility problem. Although it leads to a unique spectral estimate in the time-series case, it was shown not to do so in general. In the time series case, its computation involves the solution of a eigenvalue problem; its computation was shown to involve the solution of a linear optimization problem in general. Sufficient constraints on the spectral support and the

coarray were posed which ensure the existence of the MEM spectral estimate in its usual parametric form. Its computation, which in the time series case involves the solution of a system of linear equations, more generally involves the solution of a convex optimization problem.

Because the computation of both Pisarenko's and the MEM estimate involve the solution of optimization problems, iterative solution algorithms must be considered as components in, not just as design aids for, signal processing systems. Thus the design of reliable and fast algorithms is important. Because of the use of simple ideas from measure theory, the analytical techniques developed in this thesis also apply exactly to problems involving discrete spectral supports, for which computational algorithms can be devised. This allows the exact analysis of these algorithms, essential in the design of reliable computational algorithms. This formula has been used with great success in digital signal processing in the context of sampled data systems, principally because techniques exist to exactly analyze the behavior of the discrete time systems and hence to design reliable algorithms.

The relationship between the solutions to these discretized problems and the solutions to the original problems, involving spectral supports with a infinite number of points, has not been explored. However, it is only to be expected that the denser the sampling, the better the degree of approximation. Efforts to increase the sampling density without proportionally increasing the amount of computation, or to decrease the amount of computation for the same sampling density, should prove fruitful. In particular, the interpolation and adaptive grid techniques used in filter design should be useful. The computational algorithms discussed have been the result of applying standard optimization techniques to discretized optimization problems. New techniques, such as the

use of multiple exchange algorithms in the computation of Pisarenko's estimate, are of great interest.

Aside from just extending existing time series spectral estimation methods to the array processing case, the development of completely new methods is of interest. If a new method is to involve the optimization of some functional, the question of which functional to optimize is of interest. Further, there is the issue of implicit versus explicit positivity constraints. Burg points out that, in the absence of an explicit positivity constraint, spectral estimation methods which optimize some functional subject to correlation constraints result in parametric models simply related to Δ -polynomials. The use of explicit positivity constraints should result in more varied forms, such as Pisarenko's method, and should not exhibit the bad behavior of MEM, described in section 6.2, which results from the use of the functional being optimized to enforce the positivity constraint implicitly. More fundamentally, there is the question of how important is the positivity constraint. Constraints only result in improved spectral estimates if they would be violated, were they not imposed. The positivity constraint is an inequality constraint, while the correlation and spectral support constraints are both equality constraints. Thus one would expect the positivity constraint to be less useful than the other two.

Finally, there is the problem of enlarging the simple model of the array processing problem used in this thesis. In the time series case, direct data methods for spectral estimation are related to methods which assume knowledge of samples of the correlation function. One might hope that similar direct data methods could be developed for the array processing case.

Appendix A

The Extension Theorem for Spectral Density Functions

This appendix concerns the extension theorem for spectral density functions, discussed in section 3.6. This theorem identifies the interior of E with the set of correlation vectors which can be associated with spectral density functions which are bounded and bounded away from zero. It is assumed that every neighborhood of every point in K contains a set of non-zero measure. This condition guarantees that correlation vectors corresponding to impulses in K can be approximated by correlation vectors corresponding to bounded spectral density functions.

The extension theorem for spectral density functions: If every neighborhood of every point in K contains a set of non-zero measure, then there exists a positive function $S(k)$ which is bounded and bounded away from zero such that

$$r(\delta) = \int_K S(k) e^{jk \cdot \delta} d\nu, \quad \delta \in \Delta, \quad (\text{A.1})$$

if and only if

$$(r, p) > 0 \quad (\text{A.2})$$

for all positive p .

Proof: Define E_B as the set of correlation vectors corresponding to spectral density functions which are bounded and bounded away from zero:

$$E_B = \{r: r(\delta) = \int_K e^{jk \cdot \delta} S(k), b > S(k) > \varepsilon > 0 \text{ for some } b, \varepsilon\}. \quad (\text{A.3})$$

It follows immediately that E_B is convex. It will be shown that E_B is open and

that its closure is E . From these properties it will follow that $E_B = E^\circ$.

Consider the mapping from a bounded function $Q(k)$ into a vector τ defined by

$$\tau(\delta) = \int_K e^{jk \cdot \delta} Q(k) d\nu. \quad (A.4)$$

Because $\{e^{jk \cdot \delta} : \delta \in \Delta\}$ is a set of linearly independent functions on K and since every neighborhood of every point in K contains a set of non-zero measure, it follows that the image of the set of bounded Δ -polynomials

$$\{q \in \mathbb{R}^{2M+1} : |Q(k)| < \varepsilon\}, \quad (A.5)$$

under this mapping, is a neighborhood of 0. Therefore the image of

$$\{S(k) + Q(k) : |Q(k)| < \varepsilon\} \quad (A.6)$$

is a subset of E_B which is a neighborhood of τ . Hence E_B is open.

Consider the subsets of K

$$K_n = \{k \in K : \|k - k_0\| < \frac{1}{n}\}, \quad (A.7)$$

the corresponding functions in E_B

$$S_n(k) = \frac{\chi_{K_n}(k)}{\nu(K_n)} + \frac{1}{n}, \quad (A.8)$$

where $\chi_{K_n}(k)$ is the characteristic function of the set K_n , and the corresponding vectors

$$\tau_n(\delta) = \int_K e^{jk \cdot \delta} S_n(k) d\nu. \quad (A.9)$$

It can easily be shown that $\lim_n \tau_n = \tau_{k_0}$ where $\tau_{k_0}(\delta) = e^{jk_0 \cdot \delta}$. Thus τ_{k_0} is in the closure of E_B . It follows from Carathéodory's theorem (Cheney 1966) that every

$\tau \in E$ can be written as a positive sum of some $2M+1$ such τ_{k_i} . Since each τ_{k_i} is in the closure of E_B it follows that each $\tau \in E$ is also. Therefore the closure of E_B is E .

E° is the largest open set contained in E (Hoffman 1975). Suppose $\tau_1 \in E_B$ and that $\tau_2 \in E^\circ$ and that $\tau_2 \notin E_B$. Since both E_B and E° are open, E_B contains an open ball around τ_1 and E° contains an open ball around τ_2 .

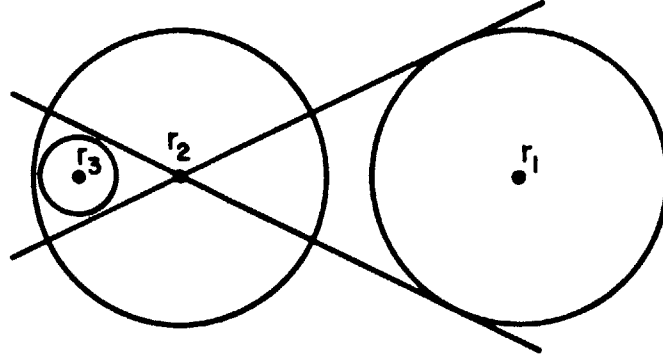


Fig. A.1: Two open convex sets with the same closure are identical.

The open ball around τ_3 must be entirely outside of E_B or else, by convexity, τ_2 would be in E_B . But then τ_3 is a vector in E which is not in the closure of E_B , a contradiction. Therefore there exists no such τ_2 and $E_B = E^\circ$.

Appendix B

The Representation Theorem

The representation theorem of section 4.2 is a simple extension of Carathéodory's theorem (Cheney 1966) for correlation vectors on the boundary of E , making use of the extension theorem. It is the generalization of "A theorem of C. Carathéodory" (Grenander and Szegö 1958, chapter 4) to multiple dimensions. In view of the chapter 4 derivation of Pisarenko's method as a linear program, the representation theorem may also be viewed as an extension of the fundamental theorem of linear programming (Luenberger 1973).

The representation theorem: If r' is on the boundary of E , then for some $2M$ non-negative $a(i)$ and some $k_i \in K$:

$$r'(\delta) = \sum_{i=1}^{2M} a(i) e^{jk_i \cdot \delta}. \quad (B.1)$$

Proof: Consider the compact convex set $E' = \{r \in E: r(0) = 1\} \subset \mathbb{R}^{2M}$. If $r \in E'$ then $r(\delta) = \int_K e^{jk \cdot \delta} d\mu$ for some positive μ such that $\int_K d\mu = 1$. Thus E' is in the closed convex hull of $A = \{r: r(\delta) = e^{jk \cdot \delta}, k \in K\}$. Since A is compact, by Carathéodory's theorem, E' is the convex hull of A and any element of E' can be expressed as a convex combination of $2M+1$ elements of A :

$$r = \sum_{i=1}^{2M+1} \vartheta_i r_i \quad (B.2)$$

with $\vartheta_i \geq 0$, $\sum_{i=1}^{2M+1} \vartheta_i = 1$, and $r_i \in A$. If one of the ϑ_i is zero, the proof is complete.

Otherwise, since r is on the boundary of E' , there is some non-zero $p \in P$ such that

$$0 = (\tau, p) = \sum_{i=1}^{2M+1} \vartheta_i (\tau_i, p). \quad (\text{B.3})$$

Thus, for each i , $(\tau_i, p) = 0$. The τ_i 's must be linearly dependent, so there are some $\beta_i \in \mathbb{R}$, not all zero, such that $\sum_{i=1}^{2M+1} \beta_i \tau_i = 0$. Let λ be the number with the smallest magnitude such that $\vartheta_i + \lambda \beta_i = 0$ for some i . Then

$$\tau = \sum_{i=1}^{2M+1} (\vartheta_i + \lambda \beta_i) \tau_i. \quad (\text{B.4})$$

One of the coefficients is zero, reducing this to a sum over only $2M$ terms. Recognizing that any element of E is a scaled version of an element of E' completes the proof.

Note that, in the times series case, τ' could be expressed as a sum of no more than M complex exponentials while this theorem only guarantees a representation in terms of $2M$ exponentials. This is not a deficiency in the proof, but a genuine feature of the problem, as the following one-dimensional example shows.

Example B.1: $K = [-\pi, \frac{\pi}{2}]$, $\Delta = \{0, \pm 1\}$.

Suppose that τ is on the straight portion of the boundary of E , as indicated in Fig. B.1. Clearly, τ has a unique representation as a convex sum of members of A in terms of the *two* correlation vectors corresponding to $k = \frac{\pi}{2}$ and $k = -\pi$,

$$\tau(\delta) = \frac{1}{2} e^{j\frac{\pi}{2}\delta} + \frac{1}{2} e^{-j\pi\delta}.$$

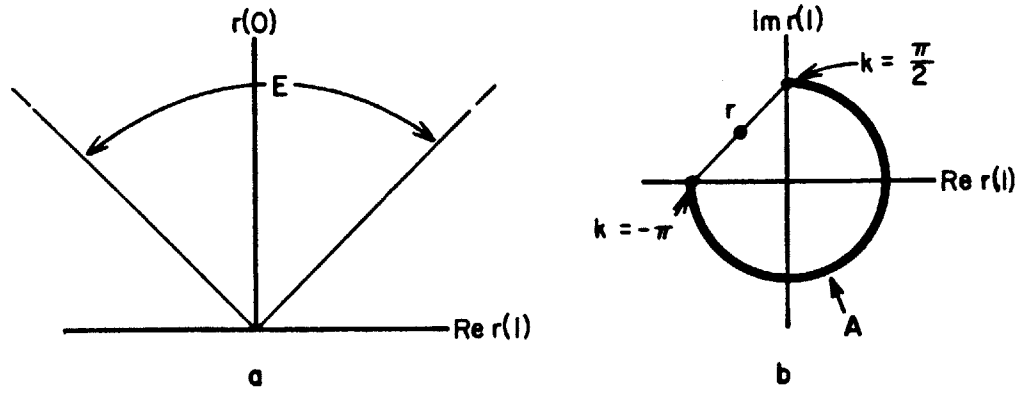


Fig. B.1: E for $K = [-\pi, \frac{\pi}{2}]$ and $\Delta = \{0, \pm 1\}$. (a) shows a section through E at $\text{Im } r(1) = 0$ and (b) shows a section through E at $r(0) = 1$.

Appendix C

The Uniqueness of Pisarenko's Estimate

As discussed in section 4.2, Pisarenko's estimate is unique if one and only one spectrum can be associated with each correlation vector on the boundary of E (∂E). Let r_k denote the correlation vector with elements $r_k(\delta) = e^{jk \cdot \delta}$ for some $k \in K$. Trivial uniqueness problems result if two distinct k 's give rise to the same r_k 's; it shall be assumed that this is not the case. Consider the set of correlation vectors corresponding to the zero set of some non-zero positive polynomial p ,

$$Z = \{r_k : (r_k, p) = P(k) = 0\}. \quad (C.1)$$

Any vector $r' \in E$ which makes a zero inner product with p can be expressed as a sum of positive multiples of vectors from the set Z . It follows that, if this set is linearly independent the representation is unique. Conversely, if this set is linearly dependent, then an r' on the boundary of E can be constructed which has more than one spectral representation. If the set is linearly dependent then there is a finite collection of non-zero real numbers $c(i)$ and $r_{k_i} \in Z$ such that

$$\sum_i c(i) r_{k_i} = 0. \quad (C.2)$$

Because $r_{k_i}(0) = 1$ for all i , there must be at least one $c(i)$ which is strictly positive and one which is strictly negative. Thus

$$r' = \sum_{c(i) > 0} c(i) r_{k_i} = \sum_{c(i) < 0} [-c(i)] r_{k_i} \quad (C.3)$$

is a non-zero correlation vector on the boundary of E with at least two spectral

representations.

Therefore Pisarenko's estimate is unique if and only if the set of correlation vectors corresponding to the zero set of each non-zero positive polynomial is linearly independent. In particular, for Pisarenko's estimate to be unique, no non-zero positive polynomial can have more than $2M$ zeros. This condition is similar to, though not quite as strict as the Haar condition (Rice 1964), which involves all polynomials, not just positive ones.

The factorability of polynomials, in the time series case, leads to a strong result. In the time series case, a non-zero positive polynomial can have no more than M zeros. Furthermore, a non-zero positive polynomial can be constructed with up to M arbitrarily located zeros. This implies (Example 4.1) that a correlation vector in ∂E has a unique spectral representation and that this spectrum is composed of M or fewer impulses. Furthermore, it implies that any spectrum composed of M or fewer impulses has a correlation vector in ∂E .

However, a simple example shows that Pisarenko's estimate is not guaranteed to be unique in most multi-dimensional situations. Consider the non-zero positive polynomial

$$P(k) = 1 - \cos(k \cdot \delta) \geq 0 \quad (C.4)$$

for some non-zero $\delta \in \Delta$. The zero set of $P(k)$ includes the hyperplane in K :

$$\{k : k \cdot \delta = 0\}. \quad (C.5)$$

Many spectral supports, of practical interest, intersect this hyperplane at an infinite number of points, implying the existence of some correlation vector on the boundary of E with a non-unique spectral representation. This non-uniqueness problem is similar to the non-uniqueness problem in multi-dimensional Chebyshev approximation (Rice 1969).

Appendix D

The Existence of the MEM Estimate

As discussed in section 5.2, MEM estimate can be considered as a mapping from the interior of P (P°) into the interior of E (E°). The problem is to show that this mapping is onto. This mapping idea was used by Woods (1976), however the proof given below is both more general and less complicated than his.

Theorem: If every neighborhood of every point in K contains a set of non-zero measure and if, for every $p \in \partial P$, $\int_K \frac{d\nu}{P(k)} = \infty$ then, for every $r \in E^\circ$, there is a $p \in P^\circ$ such that

$$r(\delta) = \int_K \frac{e^{jk \cdot \delta}}{P(k)} d\nu. \quad (D.1)$$

Proof: The condition that, for every $p \in \partial P$, $\int_K \frac{d\nu}{P(k)} = \infty$, means that if $p_n \in P^\circ$ is a sequence converging to $p \in \partial P$, then $\int_K \frac{d\nu}{P_n(k)} \rightarrow \infty$.

Two important facts derive from the assumption that every neighborhood of every point in K contains a set of non-zero measure. First, remembering that K is compact, it follows that

$$0 < \varepsilon_1 = \min_{p \in P} \int_K P(k) d\nu \quad \text{such that} \quad 1 = \max_K P(k). \quad (D.2)$$

Second, it follows that any vector r in E° can be associated with a spectral density function which is bounded and bounded away from zero:

$$0 < \varepsilon_2 < S(k). \quad (D.3)$$

Consider the mapping T from P° into E° defined by (D.1). T is a continuously differentiable mapping whose Jacobian is nowhere 0. Therefore, by the inverse mapping theorem (Hoffman 1975), $T(P^\circ)$ is open and contains a neighborhood of the correlation vector $n \in E^\circ$ corresponding to the spectral density function $N(k) = \varepsilon_2$. The proof now proceeds by contradiction. Assume that there is some $r \in E^\circ$ which is not in $T(P^\circ)$ and consider the one-parameter family of correlation vectors:

$$r_\alpha = (1-\alpha)n + \alpha r. \quad (D.4)$$

There is some smallest α , $0 < \alpha_\infty < 1$, such that $r_{\alpha_\infty} \notin T(P^\circ)$. Thus there exists a sequence $\alpha_n < \alpha_\infty$ and two corresponding sequences, $r_{\alpha_n} \in E^\circ$, and $p_n \in P^\circ$, such that $\alpha_n \rightarrow \alpha_\infty$, $r_{\alpha_n} \rightarrow r_{\alpha_\infty}$, and $r_{\alpha_n} = T(p_n)$. $P_n(k)$ is bounded uniformly in n , since

$$\begin{aligned} \nu(K) &= \int_K \frac{P_n(k)}{P_n(k)} d\nu = \int_K P_n(k) [(1-\alpha)N(k) + \alpha S(k)] d\nu \\ &> \varepsilon_2 \int_K P_n(k) d\nu > \varepsilon_1 \varepsilon_2 \max_K P_n(k). \end{aligned} \quad (D.5)$$

implies that

$$\max_K P_n(k) < \frac{\nu(K)}{\varepsilon_1 \varepsilon_2}. \quad (D.6)$$

Since

$$\nu(K) [\max_K P_n(k)]^2 \geq \int_K P_n(k)^2 d\nu = \sum_{\delta_1, \delta_2} p_n^*(\delta_1) g(\delta_1, \delta_2) p_n(\delta_2) \quad (D.7)$$

where

$$g(\delta_1, \delta_2) = \int_K e^{-j(\delta_1 - \delta_2) \cdot k} d\nu \quad (D.8)$$

is a positive-definite Grammian matrix, this bound on $P_n(k)$ also bounds p_n .

Since p_n is bounded, the Bolzano-Weierstrass theorem states that some subsequence of p_n converges, say to $p_\infty \in P$. The boundedness of the integrals

$$\int_K \frac{d\nu}{P_n(k)} = r_{a_n}(0) \leq n(0) + r(0) < \infty, \quad (D.9)$$

implies that $p_{a_\infty} \in P^\circ$. The continuity of T implies that $T(p_\infty) = r_{a_\infty}$, thus $r_{a_\infty} \in T(P^\circ)$, a contradiction. Therefore there exists no $r \in E^\circ$ which is not in $T(P^\circ)$.

A simple example can be given in which the condition that $\int_K \frac{d\nu}{P(k)} = \infty$ for every $p \in \partial P$ is violated. Consider the two-dimensional, square spectral support $[-\frac{\pi}{4}, \frac{\pi}{4}]^2$ and the positive Δ -polynomial $P(k) = \cos(k_1 + k_2)$. $P(k)$ is on the boundary of P , being zero at the two corners $k_1 = k_2 = \pm \frac{\pi}{4}$, and nowhere else. Unfortunately

$$\begin{aligned} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \frac{dk_1 dk_2}{P(k)} &= \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} \frac{dk_1 dk_2}{\cos(k_1 + k_2)} = 4 \int_0^{\frac{\pi}{2}} \frac{x dx}{\sin x} = \\ &8 \left[\frac{1}{1^2} - \frac{1}{3^2} + \frac{1}{5^2} - \frac{1}{7^2} + \dots \right] \end{aligned} \quad (D.10)$$

which is certainly finite. This example depends upon the fact that the spectral support has corners. If the corners were even slightly rounded then the conditions for the existence of the MEM estimate would be fulfilled.

References

- Baggeroer, A. B., "Space/Time Random Processes and Optimum Array Processing," Naval Undersea Center, San Diego, Ca., NUC TP 506, April 1976.
- Blackman, R. B. and J. W. Tukey, *The Measurement of Power Spectra*, Dover, 1959.
- Blatt, H. P., "A General Remes Algorithm in Real or Complex Normed Linear Spaces," *Multivariate Approximation*, D. C. Handscomb, ed., Academic Press, pp. 145-153, 1978.
- Burg, J. P., "Maximum Entropy Spectral Analysis," Ph.D. thesis, Stanford University, May 1975.
- Calderón, A. and R. Pepinsky, "On the Phases of Fourier Coefficients for Positive Real Functions," *Computing Methods and the Phase Problem in X-Ray Crystal Analysis*, The X-Ray Crystal Analysis Laboratory, Department of Physics, Pennsylvania State College, pp. 339-348, 1952.
- Charnes, A., "Optimality and Degeneracy in Linear Programming," *Econometrica*, vol. 20, pp. 160-170, 1952.
- Cheney, E. W., *Introduction to Approximation Theory*, McGraw-Hill, 1966.
- Dickinson, B. W., "Two-Dimensional Markov Spectrum Estimates Need Not Exist," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 120-121, Jan. 1980.
- Grenander, U. and G. Szegő, *Toeplitz Forms and Their Applications*, University of California Press, Berkeley and Los Angeles, 1958.

- Harris, D. B., and R. M. Mersereau, "A Comparison of Algorithms for Minimax Design of Two-Dimensional Linear Phase FIR Digital Filters," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. ASSP-25, pp. 492-500, December 1977.
- Haubrich, R. A., "Array Design," *Bulletin of the Seismological Society of America*, vol. 58, pp. 977-991, June 1968.
- Hoffman, K., *Analysis in Euclidean Space*, Prentice-Hall, 1975.
- Jain, A. K. and S. Ranganath, "Two Dimensional Spectral Estimation," *Proceedings of the RADC Spectrum Estimation Workshop*, pp. 151-157, May 1978.
- Lim, J. S. and N. A. Malik, "A New Algorithm for Two-Dimensional Maximum Entropy Power Spectrum Estimation", *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. ASSP-29, pp. 401-413, June 1981.
- Luenberger, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, 1973.
- Luenberger, D. G., *Optimization by Vector Space Methods*, Wiley, 1969.
- Ong, C., "An Investigation of Two New High-Resolution Two-Dimensional Spectral Estimate Techniques," *Special Report No. 1*, Long-Period Array Processing Development, Texas Instruments, Inc., April 1971.
- Pisarenko, V. F., "The Retrieval of Harmonics from a Covariance Function," *Geophys. J. R. Astr. Soc.*, vol. 33, pp. 347-366, 1973.

Rice, J. R., *The Approximation of Functions, vol. 1 - Linear Theory*, Addison Wesley, 1964.

Rice, J. R., *The Approximation of Functions, vol. 2 - Nonlinear and Multivariate Theory*, Addison Wesley, 1969.

Royden, H. L., *Real Analysis*, Macmillan, 1968.

Rudin, W., "The Extension Problem for Positive-Definite Functions," *Ill. J. Math.*, vol. 7, pp. 532-539, 1963.

Stiefel, E., "Note on Jordan Elimination, Linear Programming, and Tchebycheff Approximation," *Numerische Mathematik*, vol 2, pp. 1-17, 1960.

Stewart, J., "Positive Definite Functions and Generalizations, an Historical Survey," *Rocky Mountain Journal of Mathematics*, vol. 6, pp. 409-434, Summer 1976.

Sullivan, T. M., O. L. Frost, and J. R. Treichler, "High Resolution Signal Estimation," Argosystems, Inc., Sunnyvale California, June 1978.

Wernecke, S. J. and L. R. D'Addario, "Maximum Entropy Image Reconstruction," *IEEE Trans. Comput.*, vol. C-26, pp. 351-364, April 1977.

Woods, J. W., "Two-Dimensional Markov Spectral Estimation," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 552-559, Sept. 1976.