NPS-OR-92-012

(2)

NAVAL POSTGRADUATE SCHOOL

Monterey, California

MOMENTS AND SIGNAL PROCESSING

Proceedings of the Conference held at the Naval Postgraduate School March 30-31 1992, Monterey, CA

Edited by:

P. Purdue, Naval Postgraduate School, Monterey, CA
H. Solomon, Stanford University, Stanford, CA

26 August 1992

Prepared for:
National Security Agency
Office of Naval Research

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was prepared for and funded by the National Security Agency and the Office of Naval Research. Reproduction of all or part of this report is authorized.

This report was prepared by:

PETER PURDUE
Chairman and Professor
Operations Research Department

Reviewed by:

PETER PURDUE
Chairman and Professor
Operations Research Department

Released by:

PAUL J. MARTO
Dean of Research

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NPSOR-92-012	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) OR	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION National Security Agency 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER RWSPD
8c ADDRESS (City, State, and ZIP Code) Ft George Meade, MD 20755-6000	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

MOMENTS AND SIGNAL PROCESSING

12 PERSONAL AUTHOR(S)
P. Purdue, H. Solomon

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
TECHNICAL REPORT	FROM Jan 92 TO Aug 92	920826	295

16 SUPPLEMENTARY NOTATION The views expressed in this report are those of the authors and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Statistical Moments, Signal Processing

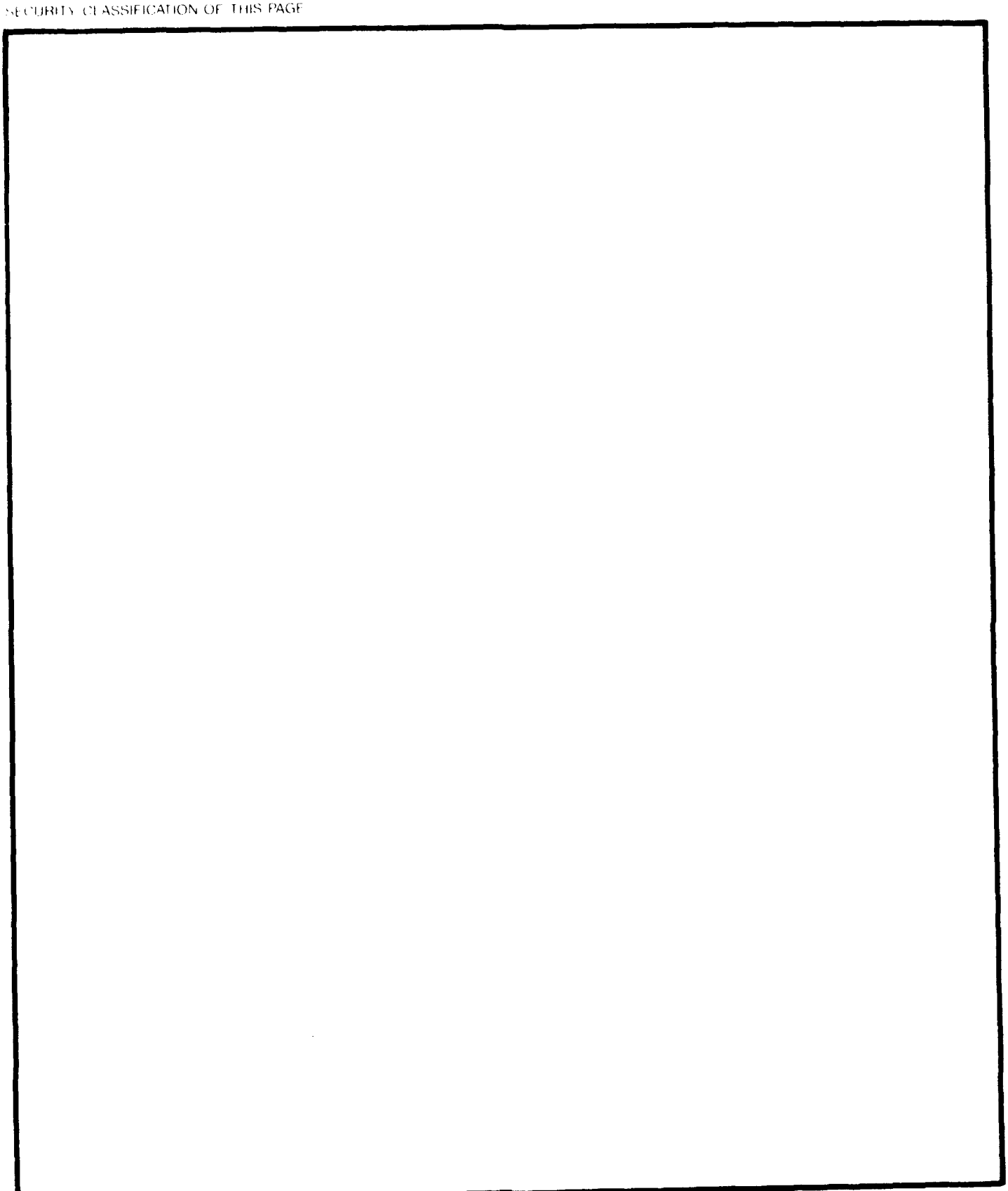
19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Proceedings of conference on Moments and Signal Processing held at the Naval Postgraduate School March 30-31, 1992, Monterey, CA.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Peter Purdue, Chairman and Professor OR Dept	22b TELEPHONE (Include Area Code) 408-646-2381 / 22c OFFICE SYMBOL OR

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 UNCLASSIFIED



CONTENTS

In recent years, moments and their uses have been investigated by mathematicians, statisticians, and engineers. In 1987, the American Mathematical Society sponsored a short course on "Moments in Mathematics" at its meeting in San Antonio, Texas. This led to a volume containing the six papers delivered there. The volume was published by the Society in its Short Course Series as Volume 37 in its *Proceedings of Symposia in Applied Mathematics.*

Recently, Dr. James Maar of the National Security Agency noted a number of problems in signal processing in which moments of distributions were important and yet statisticians and signal processor scientists were unaware of what had been accomplished by each other. He initiated discussions with Professor Peter Purdue of the Operations Research Department of the Naval Postgraduate School and Professor Herbert Solomon of the Statistics Department at Stanford University about developing a conference in which moments and signal processing and their interaction would be featured. Professor Purdue and Professor Solomon agreed to explore this idea and they developed and co-chaired a Conference on Moments and Signal Processing which was held at the Naval Postgraduate School on March 30–31, 1992. The Proceedings herein resulted from that conference.

The Conference developed around eight speakers whose interests include moments and statistics, signal processing, and interactions between the two. Professors Jerry Mendel and Max Nikias came from the signal processing community; Professors Satish Iyengar and Michael Stephens came from the statistical community. The remaining four, Professors David Brillinger, Ken-Shin Lii, Bruce Lindsay, and Ed Wegman, came at the subject in different shadings emanating from the central core of the Conference.

The Conference was supported substantively by the National Security Agency and partially by the Office of Naval Research. Many thanks are due to these agencies. A number of government scientists from the Department of Defense and a limited number of general community attendees participated in the Conference. This led to a lively audience of 40 to 50 participants over the two day period.

It is hoped that the wide availability of the papers in this report will lead to more communication between the two communities and of course within each group.

ADAPTIVE BLIND EQUALIZATION[1]

Yuanjie Chen and Chrysostomos L. Nikias

Department of Electrical Engineering - Systems
Signal and Image Processing Institute
University of Southern California
Los Angeles, CA 90089-2564

[1]This work was supported in part by the Office of Naval Research under contract N00014-92-J-1034 and the National Science under grant MIP-9206829

ABSTRACT

This tutorial paper is focused on two topics, namely: (i) to describe systematic methodologies for selecting nonlinear transformations for blind equalization algorithms (and thus new types of cumulants), and (ii) to give an overview of the existing blind equalization algorithms and point out their strengths as well as weaknesses. It is shown in this paper that all blind equalization algorithms belong in one of the following three categories, depending where the nonlinear transformation is being applied on the data: (i) the Bussgang algorithms, where the nonlinearity is in the output of the adaptive equalization filter; (ii) the polyspectra (or Higher-Order Spectra) algorithms, where the nonlinearity is in the input of the adaptive equalization filter; and (iii) the algorithms where the nonlinearity is inside the adaptive filter, *i.e.*, the nonlinear filter or neural network. We describe methodologies for selecting nonlinear transformations based on various optimality criteria such as MSE or MAP. We illustrate that such existing algorithms as Sato, Benveniste-Goursat, Godard or CMA, Stop-and-Go and Donoho are indeed special cases of the Bussgang family of techniques when the nonlinearity is memoryless. We present results that demonstrate the polyspectra-based algorithms exhibit faster convergence rate than Bussgang algorithms. However, this improved performance is at the expense of more computations per iteration. We also show that blind equalizers based on nonlinear filters or neural networks are more suited for channels that have nonlinear distortions.

The Godard or CMA algorithm is probably the most widely used blind equalizer in digital communications today due to its simplicity, low complexity and constant modulus property. Its main drawbacks, however, are slow convergence and no guarantee for global convergence starting from arbitrary initial guess. We present a new method for blind equalization, the CRIMNO algorithm (*i.e.*, criterion with memory nonlinearity), which is shown to have the same advantages as Godard (simplicity, low complexity, constant modulus property) and yet guaranteeing much faster convergence. The CRIMNO algorithm is flexible enough to address blind deconvolution problems when the input sequence is colored.

1 INTRODUCTION

Blind deconvolution or equalization is a signal processing procedure that recovers the input sequence applied to a linear time-invariant nonminimum phase system from its output only. Blind equalization algorithms are essentially adaptive filtering algorithms designed in such a way that they do not need the external supply of a desired response to generate the error signal in the output of the adaptive filter. In other words, the adaptive algorithm is "blind" to the desired response. However, the algorithm itself generates the desired response by applying a nonlinear transformation on sequences involved in the adaptation process. All blind equalization algorithms belong to one of the following three categories, depending where the nonlinear transformation is being applied on the data:

- The Bussgang algorithms, where the nonlinearity is in the **output** of the adaptive equalization filter;
- The Polyspectra (or Higher-Order Spectra) algorithms, where the nonlinearity is in the **input** of the adaptive equalization filter;
- The algorithms where the nonlinearity is **inside** the adaptive filter; *i.e.*, the filter is nonlinear (*e.g.* Volterra) or neural network.

The purpose of this paper is to provide an overview of the existing blind equalization algorithms and to discuss their advantages and limitations. Conventional equalization and carrier recovery techniques used in multilevel digital communication systems usually require an initial training period, during which a known data sequence (*i.e.*, training sequence) is transmitted [43], [45]. An alternative effective approach to this problem is to utilize blind equalizers which do not require any known training sequence during the startup period.

The paper describes systematic methodologies for selecting the nonlinearity based on various optimality criteria, such as maximum likelihood (ML), mean-square error (MSE) or maximum a posteriori (MAP). As an example, it is illustrated that such existing algorithms as Sato [46], [47] Benveniste-Goursat [5], [6] Godard or CMA [22], [50] and Stop-and-Go [41] are indeed special cases of the family of Bussgang techniques where the nonlinearity is memoryless [3], [4]. It is demonstrated that the polyspectra-based algorithms exhibit faster convergence rate than the Bussgang algorithms. However, this improved performance is at the expense of more computational complexity. On the other hand, blind equalizers based on nonlinear filters are well suited for channels that have nonlinear distortions [39], [40].

The Godard algorithm is probably the most widely used blind equalizer in digital communications today due to its simplicity, low computational complexity, and constant modulus property. Its main drawbacks, however, is slow convergence and no guarantee for global convergence (convergence starting from arbitrary initial guess). The paper describes the development of the CRIMNO algorithm (*i.e.*, criterion with memory nonlinearity) which is shown to have the same advantages as Godard algorithm (simplicity, low complexity, constant modulus property) and yet guaranteeing much faster convergence [12], [13]. Extension of the CRIMNO algorithm to the case of colored input signals is also presented.

The polyspectra-based adaptive blind equalization algorithms are also described in the paper. In particular, the Tricepstrum Equalization Algorithm (TEA) [24], the Power Cepstrum and Tricoherence Equalization Algorithm (POTEA) [7], and the Cross-Tricepstrum Equalization Algorithm (CTEA) [8] are presented, as well as their advantages and limitations. It is shown that these algorithms perform simultaneous identification and equalization of a nonminimum phase communication channel from its output only. Simulations with PAM and QAM signals

demonstrate the effectiveness of the polyspectra-based algorithms.

Finally, the paper provides an overview of the neural network based adaptive equalization algorithms either with or without a training sequence [11], [20], [26], [27], [39], [40], [49].

2 DEFINITION OF BLIND EQUALIZATION PROBLEM

Let us consider the discrete-time linear transmission channel whose impulse response $\{f(i)\}$ is unknown and possibly time-varying. The input data $\{x(i)\}$ are assumed to be independent and identically distributed (i.i.d.) random variables, with non-Gaussian probability density function. Let us also assume, without loss of generality, that the sequence $\{x(i)\}$ has mean $E\{x(i)\} = 0$ and variance $E\{|x(i)|^2\} = Q_x$. If $x(i)$ is real, we may drop the magnitude function and simply write $E\{x^2(i)\}$. Initially, noise is not taken into account in the output of the channel. From Figure 2.1, it follows that the model we consider is

$$\begin{aligned} y(i) &= f(i) * x(i) \\ &= \sum_k f(k)\, x(i-k) \end{aligned} \tag{2.1}$$

where "$*$" denotes linear convolution and $\{y(i)\}$ is the received sequence. The problem is to reconstruct (or restore) the input sequence $\{x(i)\}$ from the received sequence $\{y(i)\}$ or, equivalently, to identify the inverse filter (equalizer) $\{u(i)\}$ for the channel.

From Figure 2.1, we see that the output sequence $\{\tilde{x}(i)\}$ of the equalizer is given by

$$\tilde{x}(i) = u(i) * y(i)$$

$$= u(i) * (f(i) * x(i))$$

$$= u(i) * f(i) * x(i). \tag{2.2}$$

So, to achieve

$$\tilde{x}(i) = x(i-D)e^{j\theta} \tag{2.3}$$

where D is a constant delay and θ is a constant phase shift, it is required that

$$u(i) * f(i) = \delta(i-D) \tag{2.4}$$

where

$$\delta(i) = \begin{cases} (1)(e^{j\theta}), & i = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Performing the Fourier transform on (2.4), we obtain

$$U(w) \cdot F(w) = e^{j(\theta - wD)}. \tag{2.5}$$

In other words, the objective of the equalizer is to achieve a transfer function

$$U(w) = \frac{1}{F(w)} e^{j(\theta - wD)}. \tag{2.6}$$

In general, D and θ are unknown. However, the constant delay D does not affect the reconstruction of the original input sequence $\{x(i)\}$. The constant phase shift θ can be removed by a carry recovery technique. As such, in the sequel, it will be assumed that $D = 0$ and $\theta = 0$.

Blind equalization schemes may be classified into three categories; *i.e.*, those which utilize

nonlinearities in the **output** of the adaptive equalization filter, those which place the nonlinearity in the **input** of the adaptive equalization filter, and those which utilize adaptive nonlinear equalization filters. The Bussgang equalization algorithms with memoryless or memory nonlinearity belong to the first category whereas the higher-order cumulant-based equalizers (TEA, POTEA, etc.) belong to the second category, as they perform **memory** nonlinear transformation on the input data of the equalization filter. Blind equalizers based on nonlinear filters, such as the Volterra filter or neural networks, belong to the third category. Figures 2.2 (a)-(c) illustrate the block diagrams of the aforementioned three families of blind equalizers.

3 PERFORMANCE MEASURES FOR ALGORITHM EVALUATION

Four different performance measures are usually considered in simulation experiments for the testing of the blind equalization algorithms: the time-average squared error ($\mathrm{E_{ASE}}$), the transitional symbol error rate (SER), the residual intersymbol interference (ISI) and the discrete eye patterns [43], [44]. They are defined as follows.

Time-Average Squared Error($\mathrm{E_{ASE}}$ or MSE)

At iteration (i), the mean square error in the output of the equalizer is defined as :

$$\mathrm{E_{ASE}} = \frac{1}{N}\sum_{i=1}^{N}|x(i-D)-\tilde{x}(i)|^2 \tag{3.1}$$

where $\tilde{x}(i)$ is the output of the equalizer at iteration (i) and $x(i-D)$ is the corresponding true value. Note that the delay D, which is introduced by the channel and the equalizer, does not affect the recovery of the original information $\{x(i)\}$. However, it must be taken into account in

the calculation of MSE (i). The MSE (i) gives a measure of both the noise and residual ISI at the output of the equalizer.

Transitional Symbol Error Rate (SER)

The SER indicates the percentage of wrongly detected symbols in consecutive intervals of 500 symbols, *i.e.*,

$$\text{SER} = \frac{\#\text{of wrong detections in 500 symbols}}{500} \tag{3.2}$$

Residual ISI

The residual ISI in the output of equalizer is defined as follows. Let $\{f(i)\}$ be the channel impulse response and $\{u(i)\}$ the equalizer tap coefficients at iteration (i). Let $s(i) = f(i) * u(i)$, then

$$\text{ISI}(i) = \frac{\sum_i |s(i)|^2 - \max\{|s(i)|^2\}}{\max\{|s(i)|^2\}} \tag{3.3}$$

Physically, this indicates the amount of ISI present at the output of the equalizer due to imperfect equalization.

Discrete eye patterns

Discrete eye patterns (or equalized signal constellation) consist of all possible values of the output of the equalizer, $\tilde{x}(i)$, at iteration (i), **drawn in two-dimensional space**. We say that the eye pattern is open whenever the ideal decoding thresholds are easily distinguishable between neighboring equalized states.

In our simulations, all performance measures were calculated for many independent signal and noise realizations. For the E_{ASE}, time averaging over 100 samples were performed for each realization. The eye pattern at iteration (i) was obtained by drawing the output of equalizer for all

independent realizations and for a specific number of samples (for each realization) symmetrically located around (i).

4 ALGORITHMS WITH NONLINEARITY IN THE OUTPUT OF THE EQUALIZATION FILTER

Let us assume that a guess for the impulse response of the inverse filter (equalizer), $u_g(i)$ has been selected. Then,

$$u_g(i) * f(i) = \delta(i) + \epsilon(i) \tag{4.1}$$

where $\epsilon(i)$ accounts for the difference (error) between our guess $u_g(i)$ and the actual values of $u(i)$. If we convolve the initial guess of the inverse filter, $\{u_g(i)\}$, with the received sequence, $\{y(i)\}$, we obtain

$$\begin{aligned} \tilde{x}(i) &= y(i) * u_g(i) \\ &= x(i) * f(i) * u_g(i). \end{aligned} \tag{4.2}$$

Combining (4.2) with(4.1), we obtain

$$\begin{aligned} \tilde{x} &= x(i) * (\delta(i) + \epsilon(i)) \\ &= [x(i) * \delta(i)] + [x(i) * \epsilon(i)] \\ &= x(i) + n(i) \end{aligned} \tag{4.3}$$

where

$$n(i) = x(i) * \epsilon(i) \tag{4.4}$$

is the "convolutional noise", namely, the residual ISI arising from the difference between our guess $u_g(i)$ and the actual inverse filter $u(i)$.

Our problem now is to utilize the deconvolved sequence $\{\tilde{x}(i)\}$ to find the "best" estimate of $\{\tilde{x}(i)\}$; namely, $\{d(i)\}$. Note that in adaptive-filter literature $d(i)$ is used to represent the desired response [25]. Two criteria are employed to determine the "best" estimate of $x(i)$ from the given $\tilde{x}(i)$. These are the mean-square error (MSE) and maximum a posteriori (MAP).

Since the transmitted sequence $x(i)$ has a non-Gaussian probability density function, the MSE and MAP estimates are nonlinear transformations of $\tilde{x}(i)$. In general, the "best" estimate $d(i)$ is given by [3], [4], [23], [54].

$$d(i) = g[\tilde{x}(i)] \qquad \text{(memoryless)}$$

or

$$d(i) = g[\tilde{x}(i), \tilde{x}(i-1), \ldots, \tilde{x}(i-m)] \qquad (m\text{th} - \text{order memory}) \tag{4.5}$$

where $g[\cdot]$ is a nonlinear function with or without memory. The $d(i)$ is fed back into the adaptive equalization filter as shown in Figure 4.1. From this figure, it is also apparent that the nonlinear function $g[\cdot]$ is in the **output** of the equalization filter.

4.1 Optimum Selection of Nonlinearities

4.1.1 Nonlinearities with MSE Estimates

In summary, a well treated classical estimation problem is as follows:

$$\tilde{x}(i) \;=\; x(i) + n(i) \tag{4.6}$$

where

(i) $n(i)$ is Gaussian. Note that if $\epsilon(i)$ in (4.4) is long enough, the central limit theorem makes the Gaussianity assumption for $n(i)$ reasonable.

(ii) $\{x(i)\}$ are independent, identically distributed (i.i.d.) and in general non-Gaussian. The *pdf* of $x(i)$ is known; in digital communications the $\{x(i)\}$ are usually equi-probable discrete signal points.

(iii) $x(i)$ and $n(i)$ are assumed independent.

Given the $\tilde{x}(i)$, we seek the MSE estimate of $x(i)$, namely, $d_{\text{mse}}(i)$.

From Van Trees [52, p. 58], it follows that the best MSE estimate of $\{x(i)\}$ given $\{\tilde{x}(i)\}$ is the **mean of the a posteriori density**, *i.e.*,

$$\begin{aligned} d_{\text{mse}}(i) &= \int_{-\infty}^{+\infty} dx\; x P_{x/\tilde{x}}(x/\tilde{x}) \\ &= E\{x(i)/\tilde{x}(i)\}. \end{aligned} \tag{4.7}$$

where $P_{x/\tilde{x}}(x/\tilde{x}) = \frac{P_{\tilde{x}/x}(\tilde{x}/x)\cdot P_x(x)}{P_{\tilde{x}}(\tilde{x})}$ is the a posteriori density; $P_{x/\tilde{x}}(x/\tilde{x})$ is Gaussian, $N(x(i), Q_n)$, with Q_n being the variance of $\{n(i)\}$; the a priori density $P_x(x)$ is the *pdf* of $x(i)$, and $P_{x/\tilde{x}}(\tilde{x})$ behaves as a normalization constant in the integral of (4.7).

If $x(i)$ is zero-mean Gaussian with variance Q_x; *i.e.*, $P_x(x)$ is $N(0, Q_x)$, (4.7) reduces to

$$d_{\text{mse}}(i) = \frac{Q_x}{Q_x + Q_n}\tilde{x}(i) \tag{4.8}$$

which, in turn, implies that $g[\tilde{x}(i)]$ is a **linear** function. However, when $P_x(x)$ is non-Gaussian, the integral (4.7) can not be reduced to a simple expression and $g[\cdot]$ will be a nonlinear function. In the sequel, we show $d_{\text{mse}}(i)$ versus $\tilde{x}(i)$ when *pdf* $P_x(x)$ is uniform and Laplace.

Uniform Distribution

The a priori *pdf* is given by

$$P_x(x) = \begin{cases} \frac{1}{2\lambda} & -\lambda \leq x \leq \lambda \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

Consequently, the a posteriori *pdf* takes the form

$$P_{x/\tilde{x}}(x/\tilde{x}) = \begin{cases} \frac{A_1(x,\tilde{x})}{B_1(\tilde{x})} & -\lambda \leq x \leq \lambda \\ 0, & \text{otherwise.} \end{cases} \tag{4.10}$$

where

$$A_1(x,\tilde{x}) = \frac{1}{2\lambda}\frac{1}{\sqrt{2\pi Q_n}}\exp\left[-\frac{(x-\tilde{x})^2}{2Q_n}\right]$$

$$B_1(\tilde{x}) = \int_{-l}^{\lambda} A_1(x)dx.$$

Substituting (4.10) into (4.7), we obtain $d_{\text{mse}}(i)$ as a function of $\tilde{x}$. However, this relationship is not easy to express analytically and is obtained by numerical integration as shown in Figure 4.2.

Laplace Distribution

The a priori density is given by

$$P_x(x) = \frac{\lambda}{2}\exp[-\lambda|x|] \tag{4.11}$$

and thus the a posteriori density takes the form

$$P_{x/\tilde{x}}(x/\tilde{x}) = \frac{A_2(x,\tilde{x})}{B_2(\tilde{x})} \tag{4.12}$$

where

$$\begin{aligned} A_2(x,\tilde{x}) &= \frac{\lambda}{2}\cdot\exp\left[-\lambda|x|\cdot\frac{1}{\sqrt{2\pi Q_n}}\exp[-\frac{(x-\tilde{x})^2}{2Q_n}\right] \\ B_2(\tilde{x}) &= \int_{-\infty}^{+\infty} A_2(x)dx. \end{aligned}$$

Combining (4.12) with (4.7) and using numerical integration we obtain d_{mse} vs $\tilde{x}$ as shown in Figure 4.3.

4.1.2 Nonlinearities with MAP Estimates

In this section we treat the estimation problem

$$\tilde{x}(i) = x(i) + n(i)$$

where $n(i)$ is Gaussian and $x(i)$ is *i.i.d.* non-Gaussian. However, we seek MAP estimate of $x(i)$, namely $d_{\text{map}}(i)$ when $n(i)$ is white or colored, or correlated with $x(i)$. The colored noise case,

as well as the case of correlated noise with $x(i)$, will result into a memory nonlinear relationship between d_{map} and $\tilde{x}(i)$; i.e., $d_{\text{map}}(i) = g[\tilde{x}(i), \tilde{x}(i-1), \ldots, \tilde{x}(i-m)]$. If $x(i)$ is Gaussian *i.i.d.* and $n(i)$ is white Gaussian, independent from $x(i)$, then the $d_{\text{map}}(i)$ is identical to $d_{\text{mse}}(i)$ and is given by (4.8).

If we denote $\underline{x} = [x(i), x(i-1), \ldots, x(1)]$ and $\underline{\tilde{x}} = [\tilde{x}(i), \tilde{x}(i-1), \ldots, \tilde{x}(1)]$, then a posteriori *pdf* is given by Van Trees [p. 58]

$$P_{\underline{x}/\underline{\tilde{x}}}(\underline{x}/\underline{\tilde{x}}) = \frac{P_{\underline{x}}(\underline{x}) \cdot P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x})}{P(\underline{\tilde{x}})} \tag{4.13}$$

and the MAP estimate, $\underline{d}_{\text{map}}$, of $\underline{x}$ given $\underline{\tilde{x}}$ is the value of $\underline{x}$ which maximizes $\ell(\underline{x})$, where

$$\ell(\underline{x}) = \ell n P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x}) + \ell n P_{\underline{x}}(\underline{x}). \tag{4.14}$$

where the denominator of (4.13) does not contribute to the maximization of $\ell(\underline{x})$.

CASE I: White Gaussian Noise

In this case the $n(i)$ is white, Gaussian $N(0, Q_n)$, and independent of $x(i)$. It is also assumed that $\{x(i)\}$ are *i.i.d.* and non-Gaussian. Consequently, joint *pdfs* are expressed as products of marginal *pdfs* and the MAP estimate at each iteration $\{i\}$, $d_{\text{map}}(i)$, is obtained by maximizing

$$\ell(x(i)) = \ell n P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x}) + \ell n P_{\underline{x}}(\underline{x}).$$

That is to say that the **estimation problem is decoupled** and the resulting relationship $d_{\text{map}}(i)$ vs $\tilde{x}(i)$, is **memoryless**.

The following memoryless nonlinearities can be derived.

(i) **Uniform Distribution** (4.9)

$$d_{\text{map}}(i) = \begin{cases} -\lambda, & \tilde{x}(i) < -\lambda \\ \tilde{x}(i), & -\lambda \le \tilde{x}(i) \le \lambda \\ \lambda, & \tilde{x}(i) > \lambda \end{cases} \tag{4.15}$$

Note that d_{map} does not depend on Q_n.

(ii) **Laplace Distribution** (4.11)

$$d_{\text{map}}(i) = \begin{cases} \tilde{x}(i) + \lambda Q_n, & \tilde{x}(i) < -\lambda Q_n \\ 0, & -\lambda Q_n \le \tilde{x}(i) \le \lambda Q_n \\ \tilde{x}(i) - \lambda Q_n, & \tilde{x}(i) > \lambda Q_n. \end{cases} \tag{4.16}$$

Here the MAP estimate depends on Q_n. For the symmetric uniform and Laplace a priori distributions the resulting a posteriori *pdf*, $P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x})$, is asymmetric.

Figures 4.4 and 4.5 illustrate the MAP memoryless nonlinearities.

CASE II: Colored Gaussian Noise

In this case we assume that $n(i)$ is colored Gaussian $N(0, \underline{R})$ where $\underline{R}$ is $m \times m$ correlation matrix. On the other hand, $\{n(i)\}$. Based on these assumptions, the numerator of (4.13) is

$$P_{\underline{x}}(\underline{x}) \cdot P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x}) = \left[\prod_{i=1}^{m} P_x(x(i))\right] \cdot P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x}) \tag{4.17}$$

where

$$P_{\underline{\tilde{x}}/\underline{x}}(\underline{\tilde{x}}/\underline{x}) = \left(\frac{1}{2\pi|R|}\right)^{\frac{m}{2}} \exp\left[-\frac{1}{2}(\underline{\tilde{x}} - \underline{x})^T \underline{R}^{-1}(\underline{\tilde{x}} - \underline{x})\right]$$

and

$$\prod_{i=1}^{m} P_x(x(i)) \;=\; [P_X(x)]^m.$$

For mathematical tractability, we consider the case $m = 2$ and derive the memory nonlinear relationships $\underline{d}_{\text{map}}(i)$ vs $\underline{\tilde{x}}$.

For $m = 2$ the correlation matrix takes the form

$$\underline{R} = Q_n \cdot \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad |\rho| < 1. \tag{4.18}$$

For simplicity, we also define the following vectors

$$\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \;\triangleq\; \begin{pmatrix} \tilde{x}(i) \\ \tilde{x}(i-1) \end{pmatrix} \;=\; \underline{\tilde{x}}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;\triangleq\; \begin{pmatrix} x(i) \\ x(i-1) \end{pmatrix} \;=\; \underline{x}\,. \tag{4.19}$$

(i) Uniform Distribution (4.9)

Maximizing (4.17) is equivalent here to minimizing

$$J \;=\; (\underline{\tilde{x}} - \underline{x})^T \underline{R}^{-1} (\underline{\tilde{x}} - \underline{x}) \tag{4.20}$$

with the restrictions $-\lambda \le x_1 \le \lambda$, $-\lambda \le x_2 \le \lambda$. Hence, we seek a point in the area $X_2 = \{(x_1, x_2) : -\lambda \le x_1 \le \lambda,\ -\lambda \le x_2 \le \lambda\}$ such that J is minimized. Differentiating J

with respect to x_1 and x_2 and setting the derivative to zero we obtain

$$\begin{aligned}(\tilde{x}_1 - x_1) - \rho(\tilde{x}_2 - x_2) &= 0 \\ (\tilde{x}_2 - x_2) - \rho(\tilde{x}_1 - x_1) &= 0.\end{aligned} \tag{4.21}$$

From (4.21), it is apparent that if $\underline{\tilde{x}} \epsilon X_2$, that is $-\lambda \leq \tilde{x}_1 \leq \lambda$ and $-\lambda \leq \tilde{x}_2 \leq \lambda$, then

$$\underline{d}_{1\text{map}} = \begin{pmatrix} d_{1\text{map}} \\ d_{2\text{map}} \end{pmatrix} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \quad \text{for} \quad \tilde{x} \epsilon X_2 \tag{4.22}$$

when $\underline{\tilde{x}}$ is outside X_2, the minimum is achieved on the boundary of X_2. That is

$$\begin{aligned} d_{1\text{map}} &= k \cdot \lambda \cdot \text{sgn}[\tilde{x}_1] + (1-k) f_c[\tilde{x}_1 - \rho(\tilde{x}_2 - \lambda \text{sgn}[\tilde{x}_2])] \\ d_{2\text{map}} &= (1-k) \cdot \lambda \cdot \text{sgn}[\tilde{x}_2] + k \cdot f_c[\tilde{x}_2 - \rho(\tilde{x}_1 - \lambda \text{sgn}[\tilde{x}_1])] \\ &\text{for} \quad \underline{\tilde{x}} \notin X_2 \end{aligned} \tag{4.23}$$

where

$$f_c(x) = \begin{cases} \lambda, & x > \lambda \\ x, & |x| \leq \lambda \\ -\lambda, & x < -\lambda. \end{cases} \tag{4.24}$$

(ii) Laplace Distribution (4.11)

To obtain the MAP estimate is equivalent to minimize

$$J = \lambda|x_1| + \lambda|x_2| + \frac{1}{2}[(\underline{\tilde{x}} - \underline{x})^T \underline{R}^{-1} (\underline{\tilde{x}} - \underline{x})]. \tag{4.25}$$

The necessary conditions are

$$\begin{aligned} \lambda \mathrm{sgn}[x_1] + c(x_1 - \tilde{x}_1) - c\rho(x_2 - \tilde{x}_2) &= 0 \\ \lambda \mathrm{sgn}[x_2] + c(x_2 - \tilde{x}_2) - c\rho(x_1 - \tilde{x}_1) &= 0. \end{aligned} \tag{4.26}$$

where $c = \frac{1}{Q_n(1-\rho^2)}$. Clearly, (4.26) is a nonlinear system of equations. Two special cases are the following: 1) when $-\lambda/c \leq \tilde{x}_1 - \rho\tilde{x}_2,\ \tilde{x}_2 - \rho\tilde{x}_1 \leq \lambda/c$, then $d_{1\mathrm{map}} = 0$, and 2) when $\rho = 0$, the problem reduces to the case of white Gaussian noise.

4.2 The Bussgang Algorithms

Fig. 4.1 illustrates the Bussgang adaptive blind equalization algorithms when an LMS type or stochastic gradient algorithm [53] is used for the adaptation of the equalizer coefficients, and the nonlinearity $g^{(i)}[\cdot]$ is memoryless [3], [4], [23]. The following equations, consistent with the block diagram of Fig. 4.1, describe the Bussgang family of algorithms:

$$\begin{aligned} \underline{u}(i) &= [u_1(i), \ldots, u_N(i)]^T && \text{equalizer taps} \\ \underline{u}(0) &= [0, \ldots, 1, \ldots, 0]^T && \text{initial tap values} \\ \underline{y}(i) &= [y(i), \ldots, y(i-N+1)]^T && \text{input to the equalizer block of data} \\ i &= 0, 1, 2, \ldots && \text{iteration index} \\ \tilde{x}(i) &= \underline{u}^H(i)\underline{y}(i) && \text{equalizer output or reconstructed sequence} \\ d(i) &= g^{(i)}[\tilde{x}(i)] = g^{(i)}[\underline{u}^H(i)\underline{y}(i)] && \text{output of nonlinearity} \\ e(i) &= d(i) - \tilde{x}(i) && \text{error sequence} \\ \underline{u}(i+1) &= \underline{u}(i) + \mu\underline{y}(i) \cdot e^*(i) && \text{LMS-type adaptation} \end{aligned} \tag{4.27}$$

4.2.1 Convergence Rate and Properties

From (4.27) and Figure 4.1, it is apparent that the **output sequence** of the **nonlinear function**, $d(i)$, **"plays the role" of the desired response or the training sequence.** It is also apparent that the Bussgang technique is simple to implement and understand, and it may be viewed as a minor modification of the original LMS algorithm (the desired response of the original LMS adaptation is a memoryless transformation of the transversal filter output). As such, it is expected that the technique will have convergence that will depend on the eigenvalue spread of the autocorrelation matrix of the received data $\{y(i)\}$.

From (4.27), the LMS adaptation equation for the equalizer coefficients is given by

$$\underline{u}(i+1) = \underline{u}(i) + \mu \underline{y}(i)\, e^*(i) \tag{4.28}$$

If we obtain the expected value (ensemble averaging) of (4.28), we have

$$\begin{aligned} E\{\underline{u}(i+1)\} &= E\{\underline{u}(i)\} + \mu E\left\{\underline{y}(i)\left(g^{(i)^*}[\tilde{x}(i)] - \tilde{x}^*(i)\right)\right\} \\ &= E\{\underline{u}(i)\} + \mu E\left\{\underline{y}(i) g^{(i)^*}[\tilde{x}(i)]\right\} - \mu E\{\underline{y}(i)\tilde{x}^*(i)\}. \end{aligned} \tag{4.29}$$

The adaptive algorithm converges in the mean when

$$E\left\{\underline{y}(i) g^{(i)^*}[\tilde{x}(i)]\right\} = E\{\underline{y}(i)\tilde{x}^*(i)\} \qquad \text{(equilibrium)}$$

and it converges in the mean-square when

$$E\left\{\underline{u}^H(i)\underline{y}(i) g^{(i)^*}[\tilde{x}(i)]\right\} = E\{\underline{u}^H(i)\underline{y}(i)\tilde{x}^*(i)\}$$

$$E\left\{\tilde{x}(i)g^{(i)^*}[\tilde{x}(i)]\right\} = E\{\tilde{x}(i)\tilde{x}^*(\cdot)\}. \tag{4.30}$$

Thus, it is required that the equalizer output $\tilde{x}(i)$ be Bussgang at equilibrium. Note that identity (4.30) states that the autocorrelation of $\tilde{x}(i)$ (right-hand side) equals the cross correlation between $\tilde{x}(i)$ and a nonlinear transformation of $\tilde{x}(i)$ (left-hand side). Processes which satisfy property (4.30) are said to be Bussgang [10]. In summary, the adaptive Bussgang techniques converge when the equalizer output sequence, $\{\tilde{x}(i)\}$, becomes Bussgang (necessary condition).

A stochastic gradient algorithm (steepest descent) essentially minimizes iteratively a performance index $J(i) = E\{G[\tilde{x}(i)]\}$ with respect to the equalizer coefficients $\underline{u}(i)$. A more general form of the equalizer taps adaptation equation (4.28) is [25]

$$\underline{u}(i+1) = \underline{u}(i) - \mu\nabla_u J(i) \tag{4.31}$$

where $\nabla_u J(i)$ is the gradient of $J(i)$. Differentiating $J(i)$ by using the composite function rule, we obtain

$$\begin{aligned}\nabla_u J(i) &= -E\{\nabla_u[\tilde{x}(i)]\cdot\nabla_{\tilde{x}}[G(\tilde{x}(i))]\} \\ &= -E\{\underline{y}(i)\cdot\nabla_{\tilde{x}}[G(\tilde{x}(i))]\}\end{aligned} \tag{4.32}$$

By dropping the expectation operation, *i.e.*, by using a single-point unbiased estimate, we obtain

$$\hat{\nabla}_u J(i) \;=\; -\underline{y}(i)c^*(i) \tag{4.33}$$

where

$$\begin{aligned} c^*(i) &= \nabla_{\tilde{x}}[G(\tilde{x}(i))] \\ &= g^{(i)^*}[\tilde{x}(i)] - \tilde{x}^*(i) \end{aligned} \tag{4.34}$$

Equation (4.3.4) shows the relationship between the nonlinear function $g^{(i)}[\cdot]$ used in the Bussgang Techniques with the nonlinear cost function $G[\cdot]$ which defines the performance index, $J[\cdot]$.

Example for one dimensional modulation (PAM)

The first blind equalization algorithm was introduced by Sato in 1975 [47] for PAM signals. He chose the simple nonlinear function

$$g(\tilde{x}) \;=\; \gamma \mathrm{sgn}[\tilde{x}] \tag{4.35}$$

where γ is a gain parameter which must be chosen to satisfy the Bussgang property (4.30) i.e.,

$$E\{\hat{x}(i) \cdot \gamma \mathrm{sgn}[\tilde{x}(i)]\} \;=\; E\{|\tilde{x}(i)|^2\}$$

or

$$\gamma \;=\; E\{\tilde{x}(i)|^2\} \,/\, E\{|\tilde{x}(i)|\}. \tag{4.36}$$

We could also write Sato's algorithm in terms of

$$G(\tilde{x}) = \frac{1}{2}\tilde{x}^2 - \gamma|\tilde{x}| \tag{4.37}$$

4.2.2 Extension to QAM modulation

The extension of Bussgang algorithms to two-dimensional constellations (QAM) is somewhat straightforward [3], [4]. In the case of two independent quadrature carriers, the conditional mean estimate of an equivalent complex transmitted symbol x given the complex observation $\tilde{x} = \tilde{x}_R + j\tilde{x}_I$ can be written as

$$d = E\{x/\tilde{x}\} = g[\tilde{x}_R] + jg[\tilde{x}_I]. \tag{4.38}$$

We keep the notation simple by omitting (i). For example, the Sato nonlinearity for QAM signals takes the form [47].

$$g(\tilde{x}) = \gamma \text{csgn}(\tilde{x}) = \gamma\{\text{sgn}[\tilde{x}_R] + j\ \text{sgn}[\tilde{x}_I]\}. \tag{4.39}$$

It is clear that real and imaginary parts of the data can be estimated separately. The complex data equivalent of the adaptive Bussgang Techniques is described in (4.27), but with

$$g^{(i)}[\tilde{x}(i)] \triangleq g^{(i)}[\tilde{x}_R(i)] + j\ g^{(i)}[\tilde{x}_I(i)]. \tag{4.40}$$

Consequently, the error sequence is

$$e(i) = \{g^{(i)}[\tilde{x}_R(i)] - \tilde{x}_R(i)\} + j\left\{g^{(i)}[\tilde{x}_I(i)] - \tilde{x}_I(i)\right\}. \tag{4.41}$$

For example, the "Stop-and-Go" algorithm introduced by Picchi and Prati [41] is an adaptive Bussgang technique with the following nonlinearity

$$\begin{aligned} g[\tilde{x}(i)] &= \tilde{x}(i) + \frac{1}{2}A\hat{x}(i) - \frac{1}{2}A\tilde{x}(i) \\ &\quad + \frac{1}{2}B\hat{x}^*(i) - \frac{1}{2}B\tilde{x}^*(i) \end{aligned} \tag{4.42}$$

where $\tilde{x}(i)$ is defined as the quantizer (slicer) output in Figure 4.1 and (A, B) is a pair of integers taking values $(2,0)$ or $(1,1)$ or $(1,-1)$ or $(0,0)$. The values of (A, B) are generally different at each iteration, and how they are chosen is described later in this section.

Another example of a Bussgang technique is the **heuristic modification** of the Sato algorithm suggested by Benveniste and Goursat [5], [6]. In this case, the nonlinear function takes the form

$$\begin{aligned} g[\tilde{x}(i)] &= \tilde{x}(i) + k_1\hat{x}(i) - k_1\tilde{x}(i) + \\ &\quad k_2|\hat{x}(i) - \tilde{x}(i)| \cdot [\gamma \mathrm{csgn}[\tilde{x}(i)] - \tilde{x}(i)] \end{aligned}$$

or

$$\begin{aligned} g[\tilde{x}(i)] &= \tilde{x}(i) + |\hat{x}(i) - \tilde{x}(i)| \quad \{k_1 e^{j\arg[\hat{x}(i) - \tilde{x}(i)]} + \\ &\quad k_2[\gamma \mathrm{csgn}[\tilde{x}(i)] - \tilde{x}(i)]\} \end{aligned} \tag{4.43}$$

where k_2, k_2 are constants. From (4.38) we observe that the Benveniste-Goursat error function

may be seen as a weighted sum of the Decision Directed (DD) [43] and Sato errors. On the other hand the "Stop-and-Go" error function (4.37) is the weighted sum of the DD error and its conjugate. The weights of the two algorithms, however, are chosen in a completely different manner.

4.2.3 Unknown Carrier Phase: The Constant Modulus Property

Equation (4.33) can be written in polar coordinates as

$$d = E\left\{x/\tilde{x}\right\} = r\, e^{j\theta}\,. \tag{4.44}$$

If **we assume that all rotated constellations are equally likely**, since the carrier phase is unknown, then the conditional mean d in (4.39) has the same argument as $\tilde{x}$, and is given by

$$d = \hat{g}[|\tilde{x}|] \cdot e^{j\ \arg(\tilde{x})} \tag{4.45}$$

where $\hat{g}[\cdot]$ is a nonlinear function and $|\tilde{x}| = \sqrt{\tilde{x}_R^2 + \tilde{x}_I^2}$, $\arg(\tilde{x}) = \arctan[\tilde{x}_I/\tilde{x}_R]$. Combining (4.39) with (4.40) we obtain [3], [4], [23]

$$\begin{aligned} e(i) &= d(i) - \tilde{x}(i) \\ &= \hat{g}[|\tilde{x}(i)|]e^{j\arg[\tilde{x}(i)]} - \tilde{x}(i) \\ &= \tilde{x}(i)[\frac{g[|\tilde{x}(i)|]}{|\tilde{x}(i)|} - 1]. \end{aligned} \tag{4.46}$$

Hence, the error term is independent of any fixed phase rotation of the signal constellation. Equation (4.27) also represents the Bussgang technique for the case of unknown carrier phase,

provided we substitute $e(i)$ in (4.27) by $e(i)$ of (4.41).

Example: The Godard (or CMA) Algorithm [22], [50]

Under the assumption that all rotated constellations are equally likely, Godard [22] suggested that $\hat{g}[|\tilde{x}|]$ in (4.41) be chosen as

$$\hat{g}[|\tilde{x}|] \;=\; |\tilde{x}| + R_p|\tilde{x}|^{p-1} - |\tilde{x}|^{2p-1} \tag{4.47}$$

where R_p is a real constant. As we shall see this form has some very nice properties. Special cases of (4.42) include

$$\hat{g}[|\tilde{x}|] \;=\; (1 + R_2)|\tilde{x}| - |\tilde{x}|^3 \qquad (p = 2)$$

and

$$\hat{g}([|\tilde{x}|]) \;=\; R_1 \qquad (p = 1).$$

The parameter R_p is a gain constant which has to be chosen according to (4.30). Since

$$g[\tilde{x}(i)] \;=\; \frac{\tilde{x}(i)\hat{g}[|\tilde{x}(i)|]}{|\tilde{x}(i)|} \tag{4.48}$$

combining (4.43) with (4 30), we obtain

$$E\{|\tilde{x}(i)|^2 + R_p|\tilde{x}(i)|^p - |\tilde{x}(i)|^{2p}\} \;=\; E\{|\tilde{x}(i)|^2\}$$

or

$$R_p \;=\; \frac{E\{|\tilde{x}(i)|^{2p}\}}{E\{|\tilde{x}(i)|^p\}} \tag{4.49}$$

At perfect equalization, $\tilde{x}(i) = x(i)e^{j\theta}$ (assuming time delay $D = 0$), and thus

$$R_p = \frac{m_{2p}}{m_p}, \quad \text{where} \quad m_p = E\{|\tilde{x}(i)|^p\}.$$

Combining (4.34) and 4.43), we obtain the Godard performance index nonlinearity, namely,

$$G(\tilde{x}(i)) = \frac{1}{2p}(|\tilde{x}(i)|^p - R_p) \tag{4.50}$$

Fig. 4.6 summarizes the nonlinear functions of the Bussgang iterative techniques.

4.2.4 The Sato and Benveniste-Goursat Algorithms

Sato [46] introduced the first blind equalization scheme in 1975 by introducing the sign nonlinearity to generate the desired response of the adaptive scheme shown in Figure 4.1, *i.e.*, $d(i) = \gamma \text{ sgn } [\tilde{x}(i)]$. In 1986, Sato [47] extended his 1-D PAM algorithm to the multidimensional blind equalization problem where all transmitted signals become vector processes and all impulse responses (channel and equalizer) are square matrices. The extension, however, is straightforward. For example, in the two-dimensional case of QAM signals the "sign" nonlinearity becomes the "complex sign" defined by (4.34). The error signal of the Sato algorithm

$$e_s(i) = \gamma \text{ cgn } [\tilde{x}(i)] - \tilde{x}(i) \tag{4.51}$$

is very noisy around the solution unless the transmitted sequence $x(i)$ takes only the values ± 1. In other words, although $e_s(i)$ is zero-mean at the solution, it has a large variance. On the other hand, the Decision Directed (DD) error signal $e_D(i) = \tilde{x}(i) - \hat{x}(i)$ (see Figure 4.6) [33], though not robost for blind equalizers, enjoys the property of being identically zero at the solution. Hence,

Benveniste-Goursat [5] suggested the idea of combining (heuristically) both error signals in the form of a weighted averaging as follows

$$e_{BG}(i) = k_1\, e_D(i) + k_2\, e_S(i)\,|e_D(i)| \tag{4.52}$$

where k_1, k_2 are constants. The rationale behind the error expression (4.47) is the following. Before the eye of the equalizer opens, $|e_D(i)|$ is large and thus the Sato error $e_S(i)$ contributes to the proper direction. At the opening of the eye and thereafter $|e_D(i)|$ becomes small and the DD mode of the error $e_{BG}(i)$ takes over to speed up convergence and to achieve faster rate than the original Sato algorithm with $e_S(i)$. It is no wonder, therefore, that in our simulation experience we have seen the Benveniste-Goursat (BG) algorithm exhibiting initially very slow convergence.

A faster convergence rate has been observed only after the eye opens. The Benveniste-Goursat algorithm may be seen as the Sato algorithm that switches automatically to a DD one when the eye of the equalizer opens. The extension of the Benveniste-Goursat algorithm to a Decision Feedback Equalization (DFE) implementation [2] was given by Macchi et al. [32].

4.2.5 The Godard and Donoho (or Shalvi-Weinstein) Algorithms

The basic motivation behind the development of Godard's algorithm introduced in 1980 [22] was to find a cost function that characterizes the amount of ISI at the equalizer output independently of the carrier phase. Since the input sequence $x(i)$ is *i.i.d.*, the cost function that satisfies the aforementioned conditions is

$$J^{(p)} \;=\; E\left\{(|\tilde{x}(i)\,|^p - |\,x(i)|^p)^q\right\}, \tag{4.53}$$

which depends on the input sequence, For $p = 2$, and $q = 2$, $J^{(2)}$ takes the form

$$J^{(2)} = E\{|\tilde{x}(i)|^4 + |x(i)|^4 - 2|\tilde{x}(i)|^2|x(i)|^2\} \tag{4.54}$$

where we assume that $E\{x^2(i)\} = 0$. However, (4.48) or (4.49) can not be used in practice because $\{x(i)\}$ is inaccessible. To avoid this difficulty, Godard [22] suggested the use of a dispersion function

$$D^{(p)} = E\{(|\tilde{x}(i)|^p - R_p)^q\} \tag{4.55}$$

which was shown to behave like the cost function $J^{(p)}$ and yet it is independent of the input sequence. Note that R_p is defined by (4.44). Assuming $p = 2$, $q = 2$, (4.49) and (4.50) can be written as [22]

$$\begin{aligned} J^{(2)} &= J_1 + J_2 + \\ &\left\{4(E\{|x(i)|^2\})^2 \cdot |f(0)|^2 - 2(E\{|x(i)|^2\})^2\right\} \cdot \sum_k{}' |f(k)|^2 \end{aligned} \tag{4.56}$$

and

$$\begin{aligned} D^{(2)} &= J_1 + J_2 + \\ &\left\{4(E\{|x(i)|^2\})^2 \cdot |f(0)|^2 - 2E\{|x(i)|^4\}\right\} \cdot \left\{\sum_k{}' |f(k)|^2 + R_2^2 - E\{|x(i)|^4\}\right\} \end{aligned} \tag{4.57}$$

where $\sum_k'$ is taken for $k \neq 0$ and

$$J_1 = E\{|x(i)|^4\}\,(1 - |f(0)^2) + E\{|x(i)|^4\} \cdot \sum_k{}' |f(k)|^4,$$

$$J_2 = 2(E\{|x(i)|^2\})^2 \cdot \left\{ \left(\sum_k^{\prime} |f(k)|^2 \right)^2 - \sum_k^{\prime} |f(k)|^4 \right\}. \quad (4.58)$$

Comparing (4.51) with (4.52), we see that for $D^{(2)}$ to be similar to $J^{(2)}$, the following inequality must be satisfied:

$$4(E\{|x(i)|^2\})^2 \, |f(0)|^2 - 2E\{|x(i)|^4\} > 0$$

or

$$|f(0)|^2 > \frac{E\{|x(i)|^4\}}{2(E\{|x(i)|^2\})^2}. \quad (4.59)$$

Godard suggests (4.53) and $f(i) = 0$ for $i \neq 0$ as a way of initializing his algorithm.

Based on what has been reported in literature [50] and on our simulation experience, the Godard algorithm has always converged to a minimum that opens the eye when Godard's initialization procedure is being followed. The Godard algorithm is summarized in (4.27) and Fig. 4.6. Its convergence for $p = 2$ is better than $p = 1$. In addition, Godard noted that convergence improves when the step size μ is divided by 2 at each 10,000 iterations [22]. The Constant Modulus Algorithm (CMA), suggested independently by Treichler and Agee in 1983 [50], is the Godard algorithm for $p = 2$ and $R_2 = 1$. Ding et al. [15] reported that the Godard-type algorithms exhibit local (not global) undesirable minima.

Shalvi and Weinstein recently introduced [48] a blind equalization scheme based on the idea of matching the kurtosis measures between the transmitted sequence $\{x(i)\}$ and the reconstructed sequence $\{\tilde{x}(i)\}$ at the output of the equalizer. The kurtosis of the input complex sequence $x(i)$, is defined by

$$K(x(i)) = E\{|x(i)|^4\} - 2E^2\{|x(i)|^2\} - |E\{x^2(i)|\}|^2 \quad (4.60)$$

which is zero for complex Gaussian random variables. The important point made in [48] is that if $E\{|\tilde{x}(i)|^2\} = E\{|x(i)|^2\}$, then (1) $|K(\tilde{x}(i))| \leq |K(x(i))|$ and (2) $|K(\tilde{x}(i))| = |K(x(i))|$ if perfect equalization is achieved. Thus, the problem is to maximize the magnitude of the kurtosis measure $|K(\tilde{x}(i))|$ in the output of the equalizer at each iteration subject to the constraint $E\{|\tilde{x}(i)|^2\} = E\{|x(i)|^2\}$. One of the special cases of the Shalvi-Weinstein algorithm is the original Godard algorithm. It has recently been recently reported that the Shalvi-Weinstein algorithm was originally introduced by Donoho [16] for real-valued signals and that the algorithm's convergence is only guaranteed for infinite-length equalization filters.

4.2.6 The Stop-and-Go and Decision-Directed Algorithms

The basic idea behind the Stop-and-Go algorithm, which was proposed by Picchi and Prati [41] in 1987, is to retain the advantages of simplicity and fast convergence (in open eye-pattern conditions) of the Decision directed (DD) algorithm [33] while attempting to improve its blind convergence capabilities.

The adaptation error $e_D(i)$ used in the DD algorithm is [33]

$$e_D(i) = \hat{x}(i) - \tilde{x}(i) \tag{4.61}$$

where $\tilde{x}(i)$ is the output of the equalizer and $\hat{x}(i)$ the output of the threshold detector. Assuming that the equalizer initial tap setting corresponds to a closed eye-pattern, $e_D(i)$ will be large most of the time due to the large number of incorrect decisions $\hat{x}(i)$. Consequently, the DD algorithm cannot converge in closed eye-pattern conditions.

In the Stop-and-Go algorithm, Picchi and Prati proposed the use of the error sequence

$$e(i) = \frac{1}{2}\{A(i)e_D(i) + B(i)e_D^*(i)\} \tag{4.62}$$

where

$$\begin{aligned} A(i) &= I_R(i) + I_I(i) \\ B(i) &= I_R(i) - I_I(i) \end{aligned}$$

and

$$I_R(i) = \begin{cases} 1, & \text{if} \quad \text{sgn}[e_D(i)]_R = \text{sgn}[e_S(i)]_R \\ 0, & \text{otherwise} \end{cases}$$

$$I_I(i) = \begin{cases} 1, & \text{if} \quad \text{sgn}[e_D(i)]_I = \text{sgn}[e_S(i)]_I \\ 0, & \text{otherwise.} \end{cases}$$

Note that $e_S(i)$ is the Sato error given by (4.46).

From the foregoing, it is clear that the Stop-and-Go algorithm is essentially the DD algorithm when the eye is open. It is mostly during closed eye-pattern conditions that the Stop-and-Go adaptation rule takes place. Also, it is clear that the Benveniste-Goursat and Stop-and-Go algorithms have different convergence properties when the eye-pattern is closed and similar convergence properties when the eye is open. The modifications of this algorithms have been proposed to incorporate joint equalization and carrier recovery, decision feedback equalization [1] as well as fractionally spaced equalization [21], [45].

4.3 The CRIMNO Algorithm

Although the Bussgang algorithms are different from each other, as we have seen, they perform only memoryless nonlinear transformations on the equalizer outputs to generate the desired response. This, in turn, implies that the cost functions they attempt to minimize with respect to the equalizer coefficients are also memoryless. These algorithms do not explicitly employ the fact that *the transmitted data are statistically independent*, which is the essence of the new criterion we introduce in this section. Since statistical independence of the transmitted data involves more than one data symbols, this results in a memory nonlinear transformation on the equalizer outputs and thus a memory nonlinear cost function.

4.3.1 Criterion with Memory Nonlinearity

As we have seen, Godard solves the blind equalization problem by proposing a cost function which is independent of the transmitted data, and yet reaches its global minimum at perfect equalization. The Godard cost function (also known as the constant modulus algorithm (CMA) [22] is given by (4.50) and (4.44).

Note that only the expected value of some function of the *current* equalizer output appears in Godard's cost function. Therefore, the Godard criterion only makes use of the probability distribution of the transmitted data. It does not explicitly use the fact that the *transmitted data are statistically independent.*

Assume that perfect equalization is achievable and consider the situation where perfect equalization has indeed been achieved. That is

$$\tilde{x}(i) = x(i - D)$$

where d is some positive number, which accounts for the delay. Since the transmitted data $x(i)$ are statistically independent from each other, so are the equalizer outputs $\tilde{x}(i)$ at perfect equalization. In addition, for most transmitted data constellations, the mean of transmitted data $x(i)$ is zero. Therefore, at perfect equalization , we have

$$E\{\tilde{x}(i)\tilde{x}^*(i-l)\} = E\{x(i-D)x^*(i-l-D)\} = E\{x(i-D)\} \cdot E\{x^*(i-l-D)\} = 0$$

By making use of this property and combining it with Godard's criterion, we obtain a new criterion, called criterion with memory nonlinearity (CRIMNO), which is the minimization of the following cost function:

$$M^{(p)} = w_0 E\left(|\tilde{x}(i)|^p - R_p\right)^2 + w_1 \left|E\{\tilde{x}(i)\tilde{x}^*(i-1)\}\right|^2 + \cdots + w_M |E\{\tilde{x}(i)\tilde{x}^*(i-M)\}|^2. \quad (4.63)$$

The rationale behind the CRIMNO is that since each term reaches its global minimum at perfect equalization, by appropriately combining them, we can increase the convergence speed of the corresponding CRIMNO algorithm [12], [13]. This is clearly demonstrated in the simulations section.

Remarks:

1. Memory nonlinearity: the CRIMNO cost function depends not only on the *current* equalizer output, but also on the *previous* equalizer outputs. As such, it results to a criterion with memory nonlinearity. The parameter M determines the size of memory.

2. Generalization of the Godard criterion: when $w_0 = 1$, $w_i = 0$ for $i \neq 0$, the CRIMNO cost function reduces to the Godard cost function. Therefore, the CRIMNO criterion may be seen as a generalization of the Godard criterion.

3. Constant Modulus Property: the CRIMNO criterion preserves the constant modulus property inherent in Godard.

4.3.2 CRIMNO Blind Equalization Algorithm

Define the equalizer coefficient vector $\underline{u}(i) \triangleq [u_1(i), \cdots, u_N(i)]^T$, and the received signal vector $\underline{y}(i) \triangleq [y(i), \cdots, y(i-N+1)]^T$, where N is the length of the equalizer. Then the equalizer outputs are

$$\tilde{x}(i-l) = \underline{y}^T(i-l) \cdot \underline{u}(i), \; l = 0, 1, \cdots, M, \tag{4.64}$$

where superscript T denotes transposition of a vector.

Differentiating the cost function $M^{(2)}$ with respect to the equalizer coefficient vector $\underline{u}(i)$, we obtain [12]

$$\begin{aligned}\frac{\partial M^{(2)}}{\partial \underline{u}(i)} &= 4w_0 E[\underline{y}^*(i)\tilde{x}(i)(|\tilde{x}(i)|^2 - R_2)] \\ &+2w_1[E(\underline{y}^*(i-1)\tilde{x}(i))E(\tilde{x}^*(i)\tilde{x}(i-1)) + E(\underline{y}^*(i)\tilde{x}(i-1))E(\tilde{x}(i)\tilde{x}^*(i-1))] \\ &+\cdots \\ &+2w_M[E(\underline{y}^*(i-M)\tilde{x}(i))E(\tilde{x}^*(i)\tilde{x}(i-M)) + E(\underline{y}^*(i)\tilde{x}(i-M))E(\tilde{x}(i)\tilde{x}^*(i-M))]. \end{aligned} \tag{4.65}$$

By using the steepest descent method to search for the minimum point, we obtain

$$\underline{u}(i+1) = \underline{u}(i) - \alpha \cdot \{4w_0 E[\underline{y}^*(i)\tilde{x}(i)(|\tilde{x}(i)|^2 - R_2]$$

$$+2w_1[E(\underline{y}^*(i-1)\tilde{x}(i))E(\tilde{x}^*(i)\tilde{x}(i-1)) + E(\underline{y}^*(i)\tilde{x}(i-1))E(\tilde{x}(i)\tilde{x}^*(i-1))]$$

$$+\cdots$$

$$+2w_M[E(\underline{y}^*(i-M)\tilde{x}(i))E(\tilde{x}^*(i)\tilde{x}(i-M)) + E(\underline{y}^*(i)\tilde{x}(i-M))E(\tilde{x}(i)\tilde{x}^*(i-M))] \quad (4.66)$$

where

$$\underline{u}(i) \triangleq [u_1(i), \ldots, u_N(i)]^T$$

In (4.6), the expectation are the ensemble averages taken with respect to transmitted data $x(i)$ while the channel impulse response $f(i)$ and the equalizer coefficients $u(i)$ are treated as fixed.

If we use single point estimates for the ensemble averages, we obtain the stochastic gradient CRIMNO algorithm:

$$\begin{aligned}
\underline{u}(i+1) &= \underline{u}(i) - \alpha[4w_0\underline{y}^*(i)\tilde{x}(i)(|\tilde{x}(i)|^2 - R_2) + 2w_1(\underline{y}^*(i-1)\tilde{x}(i)|\tilde{x}(i-1)|^2 + \underline{y}^*(i-1)\tilde{x}(i-1)|\tilde{x}(i)|^2) \\
&\quad + \cdots + 2w_M(\underline{y}^*(i)\tilde{x}(i)|\tilde{x}(i-M)|^2 + \underline{y}^*(i-M)\tilde{x}(i-M)|\tilde{x}(i)|^2)] \\
&= \underline{u}(i) - \alpha[\underline{y}^*(i)\tilde{x}(i) * (4w_0|\tilde{x}(i)|^2 + 2w_1|\tilde{x}(i-1)|^2 + \cdots + 2w_M|\tilde{x}(i-M)|^2 - 4w_0R_2) \\
&\quad + 2w_1\underline{y}^*(i-1)\tilde{x}(i-1)|\tilde{x}(i)|^2 + \cdots + 2w_M\underline{y}^*(i-M)\tilde{x}(i-M)|\tilde{x}(i)|^2]. \qquad (4.67)
\end{aligned}$$

Note that at each iteration, all equalizer outputs $\tilde{x}(i-l), l = 0, 1, \cdots, M$ are recalculated using current (most recent) equalizer coefficient vector $\underline{u}(i)$ via $\tilde{x}(i-l) = \underline{y}^T(i-l)\underline{u}(i)$. This requires a lot of computations. If, instead of using the current equalizer coefficient vector $\underline{u}(i)$, we use the delayed equalizer coefficient vector $\underline{u}(i-l)$ to calculate $\tilde{x}(i-l)$. Note that (for small step-size, which is required for the stability of stochastic gradient-type algorithm, the difference between $\underline{u}(i)$ and $\underline{u}(i-l)$ is negligible. Then at each iteration we will need to calculate only one equalizer

output $\tilde{x}(i)$ using the current equalizer coefficient vector $\underline{u}(i)$.

4.3.3 Adaptive Weight CRIMNO Algorithm

The shape of the cost function depends on the choice of weight w_l. So does the performance of the CRIMNO algorithm. Here, we describe an ad hoc way of adjusting the weights on-line in the blind equalization process.

The basic idea is to estimate the values of all terms in the CRIMNO cost function over a block of data and then set the weights used in the next block proportional to the deviations of the corresponding terms from their ideal values at perfect equalization. The rationale behind this scheme is that if one term in the criterion has a large deviation from its ideal value, then in the next block the weight associated with it will be set equal to a large value, and consequently, the gradient-descent method will bring it down quickly.

To elaborate on this idea, we rewrite the CRIMNO cost function as

$$M^{(p)} = w_0 J_0 + w_1 J_1 + \cdots + w_M J_M, \tag{4.68}$$

where

$$\begin{aligned} J_0 &= E(|\tilde{x}(i)|^p - R_p)^2 \\ J_l &= |E(\tilde{x}(i)\tilde{x}^*(i-l)|^2 \; 1 \le l \le M. \end{aligned} \tag{4.69}$$

Define the deviation of the lth term $D(J_l)$ by

$$D(J_l) \triangleq |J_l - J_l^{(o)}|, \tag{4.70}$$

where $J_l^{(o)}$ is the value of J_l at perfect equalization ($J_l^{(o)} = 0,\ l = 1, \cdots, M$). Then the weights are adjusted using the following formulae:

$$\begin{aligned} w_0 &= \begin{cases} \gamma_0 D(J_0) & \gamma_0 D(J_0) < \lambda \\ \lambda & \gamma_0 D(J_0) \geq \lambda \end{cases} \\ w_i &= \begin{cases} \gamma D(J_l) & \gamma D(J_l) < \lambda \\ \lambda & \gamma D(J_l) \geq \lambda \end{cases} \end{aligned} \tag{4.71}$$

where $\lambda_0 > 0$ is the scaling constant for the first term, $\gamma > 0$ is the scaling constant for the other terms in the CRIMNO cost function, and λ is a constraint on the maximum value of the weights to guarantee the stability of the algorithm.

The CRIMNO algorithm with weights adjusted in this way is called adaptive weight CRIMNO algorithm. Some in-depth comments are provided below:

1. When the deviations of all terms vary proportionally, the adaptive weight scheme becomes an adaptive step-size algorithm. Moreover, the adaptation is done automatically. So when the algorithm converges, then weights decrease to zero. Hence, the adaptive weight CRIMNO algorithm acquires as a byproduct the decreasing step-size, which has been proven to be an optimal strategy for equalization [51].

2. For the adaptive weight CRIMNO algorithm, the shape of the cost function is changing.

The local minima of the cost function are also changing. Thus, what is local minimum of the cost function at one iteration may not be at the next iteration. However, **whatever the change of the weights, the global minimum does not change, and it always corresponds to perfect equalization.**

3. The adaptive weight CRIMNO algorithm tends to move out of a local minimum of the cost function quickly, if the cost function has local minima and the algorithm gets trapped in one of them. This is based on the following arguments. In the adaptive weight CRIMNO algorithm, the equalizer coefficient increment, $\Delta\underline{u}(i+1) = \underline{u}(i+1) - \underline{u}(i)$ is a random vector, the variance of which determines how fast the algorithm will move out of a local minimum. The variance of the equalizer coefficient increment depends on the step-size α, gradient $\frac{\partial J_l}{\partial \underline{u}(i)}$ and the weights w_l (proportional to $D(J_l)$). The step-size and gradient are the same with the fixed weight CRIMNO algorithm; we thus concentrate on the third one: w_l, or equivalently $D(J_l)$. At a global minimum of the cost function, $D(J_l)$ are all small, thus, the variance of the equalizer coefficient increment is small. Therefore, the algorithm will remain near the global minimum. However, that is not the case with a local minimum. In that case, $D(J_l)$ will be large, therefore, the variance of the equalizer coefficient increment will be large (relative ot the case at the global minimum), and the algorithm will move out of that minimum quickly. Moreover, the larger the deviation $D(J_l)$, the more quickly the algorithm will move out of the local minimum.

4. Blocks of data are used to estimate $\{J_l\}$. The block length should be sufficiently long to make the variances of the estimates small, but not long enough to make the weight update fall behind.

4.3.4 CRIMNO Extensions

In this section, the CRIMNO ideas, *i.e.*, memory nonlinearity, are extended to the following cases: (1) the case of correlated inputs; (2) the case when higher-order correlation terms [38] are utilized.

Colored CRIMNO

One of the key assumptions in the CRIMNO criterion is that the transmitted data are independent and identically distributed (*i.i.d*). However, in practice, this may not be true for QAM signals. Usually, in order to overcome the phase ambiguity caused by the squaring loop for carrier recovery, differential encoding techniques are used, which correlate the input data when the source symbols are not equiprobable. Since the operations of differential encoding are known, the autocorrelations of the input data can be derived. In the case where the autocorrelations of the input data are known a priori, the CRIMNO criterion can be modified as follows:

$$M_c^{(p)} = w_0 E(|\tilde{x}(i)|^p - R_p)^2 + w_1 |E(\tilde{x}(i)\tilde{x}^*(i-1) - \beta_1|^2 + \cdots + w_M |E(\tilde{x}(i)\tilde{x}^*(i-M)) - \beta_M|^2 \quad (4.72)$$

where $\beta_l \triangleq E(x(i)x^*(i-l))$ are **the known autocorrelations of the transmitted data.**

Higher-Order Correlation CRIMNO

Here, a criterion which exploits the higher-order correlations, such as the fourth-order statistics of the equalizer output, is given below:

$$\begin{aligned} M_h^{(p)} = {} & w_0 E(|\tilde{x}(i)|^p - R_p)^2 + \sum_l w_l |E(\tilde{x}(i)\tilde{x}^*(i-l))|^2 \\ & + \sum_{j,k,l \text{ all different}} v_{jkl} |E(\tilde{x}(i)\tilde{x}^*(i-j)\tilde{x}(i-k)\tilde{x}^*(i-l)|^2 \end{aligned} \quad (4.73)$$

The performance of both (4.73) and (4.74) criteria needs to be investigated.

4.3.5 Computer Simulation

Computer simulations have been conducted to compare the performance of the adaptive weight CRIMNO algorithm with that of the Godard (or CMA) algorithm. Fig. 4.6 shows the performance of the adaptive weight CRIMNO algorithm, compared with that of the Godard algorithm under the different step-sizes, including the optimum one: We see that the performance of the adaptive weight CRIMNO algorithm is better than or approaches that of the Godard algorithm with optimum step-size. Fig. 4.7 shows the performance of the adaptive weight CRIMNO algorithm for different memory sizes ($M = 2.4.6$). Fig. 4.8 shows that the corresponding eye-patterns at iteration 20000. We see that the larger the memory size M, the better the performance of the adaptive weight CRIMNO algorithm. Table 4.2 lists the computational complexity of the CRIMNO algorithm, the adaptive weight CRIMNO algorithm, and the Godard algorithm. We see that there is only a little increase in computational complexity. Therefore, the performance improvement is achieved at the expense of little increase in computational complexity.

5 ALGORITHMS WITH NONLINEARITY IN THE INPUT OF THE EQUALIZATION FILTER

The Polyspectra Based Techniques

Another class of blind equalization algorithms are those algorithms which are based on higher-order cumulants or polyspectra [36], such as the tricepstrum equalization algorithm (TEA) [24], the power cepstrum and tricoherence equalization algorithm (POTEA) [7], and the cross-tricepstrum equalization algorithm (CTEA) [8]. All these algorithms perform nonlinear transfor-

mation on the input of the equalization filter. This nonlinear transformation, *e.g.* the generation of the higher-order cumulants or polyspectra of the received data, is a memory nonlinear transformation, because it employs both the present and the past values of the received data. The use of the higher-order statistics of the received data is necessary for blind equalization, since the correct phase information about the channel can not be extracted from only the second-order statistics of the received data [14], [29], [34], [35], [37], [42].

5.1 Definitions and Properties: Cumulants and Higher Order Spectra

The readers are assumed to be somewhat familiar with the basic material of higher-order spectra. However, some important properties which will be used in the subsequent sections are given.

5.1.1 Definitions

1. Definition of Cumulants:

   Given a set of n real random variables $\{x_1, x_2, \cdots, x_n\}$, their nth joint cumulants of order is defined as

$$L(x_1, x_2, \cdots, x_n) \triangleq (-j)^n \frac{\partial^n ln\Phi(v_1, v_2, \cdots, v_n)}{\partial v_1 \partial v_2 \cdots \partial v_n}\Bigg| v_1 = v_2 = \cdots = v_n = 0 \qquad (5.1)$$

   where

$$\Phi(v_1, v_2, \cdots, v_n) = E\{\exp\, j(v_1 x_1 + \cdots + v_n x_n)\}. \qquad (5.2)$$

   Given a real stationary random sequence $\{x(i)\}$ with zero mean, $E\{x(i)\} = 0$, then the nth-order cumulant of the random sequence depends only on the time difference and is

defined as

$$L_x(\tau_1, \tau_2, \cdots, \tau_{n-1}) \triangleq (-j)^n \frac{\partial^n ln\Phi_x(v_1, v_2, \cdots, v_n)}{\partial v_1 \partial v_2 \cdots \partial v_n}\Bigg| v_1 = v_2 = \cdots = v_n = 0 \qquad (5.3)$$

where $\tau_1, \tau_2, \cdots, \tau_{n-1}$ are integers and

$$\Phi_x(v_1, v_2, \cdots, v_n) = E\{\exp j(v_1 x(i) + v_2 x(i+\tau_1) + \cdots + v_n x(i+\tau_{n-1}))\} \qquad (5.4)$$

Given a set of real jointly stationary random sequences $\{x_k(i)\}$, $k = 1, 2, \cdots, n$ with zero mean, $E\{x_k(i)\} = 0$, then the nth-order cross-cumulant of the sequences depends only on the time difference and is defined as

$$L_{x,1,2,\cdots,n}(\tau_1, \tau_2, \cdots, \tau_{n-1}) \triangleq (-j)^n \frac{\partial^n ln\Phi_{x,1,2,\cdots,n}(v_1, v_2, \cdots, v_n)}{\partial v_1 \partial v_2 \cdots \partial v_n}\Bigg| v_1 = v_2 = \cdots = v_n = 0 \qquad (5.5)$$

where $\tau_1, \tau_2, \cdots, \tau_{n-1}$ are integers and

$$\Phi_{x,1,2,\cdots,n}(v_1, v_2, \cdots, v_n) = E\{\exp j(v_1 x_1(i) + v_2 x_2(i+\tau_1) + \cdots + v_n x_n(i+\tau_{n-1}))\}. \qquad (5.6)$$

2. Definitions of Higher-Order Spectra.

Higher-order spectra are defined to be the Z-transforms of the corresponding cumulants [34], [38]. Specifically, a nth-order spectrum of a real stationary zero mean random sequence $\{x(i)\}$ is just the $(n-1)$-dimensional Fourier transform of the nth-order cumulant $L_x(\tau_1, \tau_2, \cdots, \tau_{n-1})$ of the random sequence. That is

$$S_x(z_1, z_2 \cdots, z_{n-1}) \triangleq \sum_{\tau_1,\tau_2,\cdots,\tau_{n-1}} L_x(\tau_1, \tau_2, \cdots, \tau_{n-1}) \prod_{l=1}^{n-1} Z_l^{-\tau_l}. \tag{5.7}$$

When $n = 2, 3, 4$ the corresponding spectrum is called power spectrum, bispectrum, and trispectrum, respectively.

A nth-order cross-spectrum of a set of real stationary zero mean random sequences $\{x_k(i)\}$, $k = 1, 2, \cdots, n$, is defined as the $(n-1)$ dimensional Z-transform of the nth-order cumulant $L_{x,1,2,\cdots,n}(\tau_1, \tau_2, \cdots \tau_{n-1})$ of the random sequence, that is

$$S_{x,1,2,\cdots,n}(z_1, z_2, \cdots, z_{n-1}) \triangleq \sum_{\tau_1,\tau_2,\cdots\tau_{n-1}} L_{x,1,2,\cdots,n}(\tau_1, \tau_2, \cdots, \tau_{n-1}) \prod_{l=1}^{n-1} z_l^{-\tau_l}. \tag{5.8}$$

3. Definitions of coherence.

Coherence is defined as the higher-order spectrum normalized by the power spectrum. Specifically, a nth-order coherence of a real stationary zero mean random sequence $x(i)$ is defined as

$$R_x(z_1, z_2, \cdots, z_{n-1}) \triangleq \frac{S_x(z_1, z_2, \cdots, z_{n-1})}{[S_x(z_1)S_x(z_2)\cdots S_x(z_{n-1})S_x(\prod_{l=1}^{n-1} z_l^{-1})]^{\frac{1}{2}}} \tag{5.9}$$

An alternative definition for the nth-order coherence, which is equivalent to the above definitions, is

$$R_x(z_1, z_2, \cdots, z_{n-1}) \triangleq \left[\frac{S_x(z_1, z_2, \cdots, z_{n-1})}{S_x(z_1^{-1}, z_2^{-1}, \cdots, z_{n-1}^{-1})}\right]^{\frac{1}{2}} \tag{5.10}$$

4. Definitions of Cepstrum of Higher-Order Spectrum

The cepstrum is defined as the inverse Z-transform of the log function of the spectrum. Specifically, a cepstrum for the nth-order spectrum of a real stationary zero mean random

sequence $\{x(i)\}$ is defined as

$$c_x(\tau_1, \tau_2, \cdots, \tau_{n-1}) \triangleq Z^{-1}\left[ln\ S_x(z_1, z_2, \cdots, z_{n-1})\right] \tag{5.11}$$

A cepstrum for the nth-order cross spectrum of a set of real stationary zero mean random sequence $\{x(i)\}$, $i = 1, 2, \cdots, n$, is defined as

$$c_{x,1,2,\cdots,n}(\tau_1, \tau_2, \cdots \tau_{n-1}) \triangleq Z^{-1}\left[ln\ S_{x,1,2,\cdots,n}(z_1, z_2, \cdots, z_{n-1})\right] \tag{5.12}$$

When $n = 2, 3, 4$, the corresponding cepstrum is called power cepstrum, bicepstrum and tricepstrum, respectively.

5.1.2 Properties

Some important properties of cumulants are shown below.

1. If $x_1, x_2, \cdots, x_n$ can be divided into two or more groups which are statistically independent, then the cumulant $L(x_1, x_2, \cdots, x_n)$ is zero.

   Specifically, if $\{x(i)\}$ are an independent, identically distributed random variables, the nth-order cumulant of the sequence $\{x(i)\}$ is

$$L_x(\tau_1, \tau_2, \cdots, \tau_{n-1}) = \gamma\delta(\tau_1)\delta(\tau_2)\cdots\delta(\tau_{n-1}) \tag{5.13}$$

2. Cumulants of higher order ($n \geq 3$) are zero for Gaussian processes.

3. If $\{x(i)\}$ and $\{y(i)\}$ are statistically independent random sequences and, $z(i) = x(i) + y(i)$, then

$$L_z(\tau_1, \tau_2, \cdots, \tau_{n-1}) = L_x(\tau_1, \tau_2, \cdots, \tau_{n-1}) + L_y(\tau_1, \tau_2, \cdots, \tau_{n-1}). \tag{5.14}$$

5.2 Tricepstrum Equalization Algorithm (TEA)

5.2.1 Problem Formulations

We assume that the received sequence after being demodulated, low-pass filtered and synchronously sampled (at rate $\frac{1}{T}$) can be written as:

$$y(i) = z(i) + w(i) = \sum_{k=-L_1}^{L_2} f(k)x(i-k) + w(i) \tag{5.15}$$

where the nonminimum phase equivalent channel impulse response $\{f(i)\}$ accounts for the transmitter filter, non-ideal channel (or multipath propagation), and receiver filter impulse response; the input data sequence $\{x(i)\}$ is generally complex, non-Gaussian, white, *i.i.d.*, with $E\{x(i)\} = 0$, $E\{x(i)^3\} = 0$ and $E\{x(i)^4\} - 3[E\{x(i)^2\}]^2 = \gamma_x \neq 0$; for example $\{x(i)\}$ could be a multi-level symmetric PAM sequence or the complex baseband equivalent sequence of a symmetric QAM signal; the additive noise $\{w(i)\}$ is zero-mean, Gaussian, generally complex and statistically independent from $\{x(i)\}$; we also assume that the channel transfer function $F(z)$ (Z-transform of $\{f(i)\}$) admits the factorization [24]

$$F(z) = A \cdot I(z^{-1}) \cdot O(z) \tag{5.16}$$

the factor $I(z^{-1}) = \frac{\prod_{k=1}^{L_3}(1-a_k z^{-1})}{\prod_{k=1}^{L_4}(1-c_k z^{-1})}, |a_k| < 1, |c_k| < 1$, is a minimum phase polynomial, *i.e.*, with zeros and poles inside the unit circle. The factor $O(z) = \prod_{k=1}^{L_2}(1 - b_k z), |b_k| < 1$ is a maximum

phase polynomial, *i.e.*, with zeros outside the unit circle. The parameter A is a constant gain factor. Finally, the sequence $\{y(i)\}$ is the input to the blind equalizer.

5.2.2 Relations of Tricepstrum of the Linear Filter Output

The input to the channel, $x(i)$, is a non-Gaussian i.i.d. random sequence, thus

$$S_x(z_1, z_2, z_3) = \gamma_x. \tag{5.17}$$

The trispectrum of the output, $y(i)$, of the channel (linear filter) is

$$\begin{aligned} S_y(z_1, z_2, z_3) &= \gamma_x F(z_1)F(z_2)F(z_3)F(z_1^{-1}z_2^{-1}z_3^{-1}) \\ &= \gamma_x \cdot A^4 \cdot I(z_1^{-1})I(z_2^{-1}) \cdot I(z_3^{-1}) \cdot I(z_1, z_2, z_3)\, O(z_1) \cdot O(z_2) \cdot O(z_3) \cdot O(z_1^{-1}z_2^{-1}z_3^{-1}) \end{aligned} \tag{5.18}$$

Taking the logarithm of $S_y(z_1, z_2, z_3)$ and then the inverse Z-transform, after some manipulation, we obtain [24]

$$c_y(m,n,l) = \frac{1}{2}\begin{cases} \log(\gamma_x A^4) & m=n=l=0 \\ -\frac{1}{m}A^{(m)} & m>0, n=l=0 \\ -\frac{1}{n}A^{(n)} & n>0, m=l=0 \\ -\frac{1}{l}A^{(l)} & l>0, m=n=0 \\ \frac{1}{m}B^{(-m)} & m<0, n=l=0 \\ \frac{1}{n}B^{(-n)} & n<0, m=l=0 \\ \frac{1}{l}B^{(-l)} & l<0, m=n=0 \\ -\frac{1}{n}B^{(n)} & m=n=l>0 \\ \frac{l}{n}A^{(n)} & m=n=l<0 \\ 0 & \text{otherwise} \end{cases} \tag{5.19}$$

where, $A^{(I)}, B^{(J)}$ are the minimum and maximum phase differential cepstrum parameters of the system, corresponding to $I(z^{-1})$ and $O(z)$, respectively. They are defined as follows:

$$A^{(I)} \stackrel{def}{=} \sum_{k=1}^{L_3} a_k^I - \sum_{k=1}^{L_4} c_k^I \qquad B^{(J)} \stackrel{def}{=} \sum_{k=1}^{L_2} b_k^J \tag{5.20}$$

In addition, the following identity holds between the fourth-order cumulants $L_y(m,n,l)$ and the tricepstrum $c_y(m,n,l)$:

$$\sum_{J=1}^{\infty}\Big\{A^{(J)}[L_y(m-J,n,l) - L_y(m+J,n+J,l+J)]\Big\} + \sum_{J=1}^{\infty}\Big\{B^{(J)}[L_y(m-J,n-J,l-J) - L_y(m+J,n,l)]\Big\} = -m \cdot L_y(m,n,l) \tag{5.21}$$

where we define,

$$J \cdot c.(J,0,0) = \begin{cases} -A^{(J)}, & J = 1, \ldots \infty \\ B^{(-J)}, & J = -1, \ldots - \infty. \end{cases}$$

$A^{(J)}, B^{(J)}, J = 1, 2, \ldots$ are the minimum and maximum phase cepstral coefficients respectively, which are related to the zeros $F(z)$. However, in practice, the summation terms in (5.21) can be approximated by arbitrarily large but finite values because $A^{(J)}$ and $B^{(J)}$ decay exponentially as J increases.

In practice the fourth-order cumulants $L_y(\cdot)$ in (5.21) need to be substituted by their estimates $\hat{L}_y(\cdot)$ obtained from a finite length window of the received samples $\{y(i)\}$.

The TEA algorithm, uses (5.21) in order to form an overdetermined system of equations, *i.e.*, we have more equations than unknowns. Then, TEA solves this overdetermined system of equations, adaptively, using an LMS adaptation algorithm. At each iteration an estimate of the cepstral parameters $\{A^{(I)}\}$ and $\{B^{(J)}\}$ is computed. The coefficients of the equalizer are calculated for $\{A^{(I)}\}$ and $\{B^{(J)}\}$ by means of the iterative formulas.

5.2.3 TEA Algorithm

Let:

$\{y(i)\}$:	The received zero-mean synchronously sampled communication signal.
N_1, N_2:	Lengths of minimum and maximum phase components of the equalizer.
p, q:	Lengths of minimum and maximum phase cepstral parameters.
$\hat{M}_y^{(i)}(m,n,l)$:	Estimated fourth-order moments of $\{y(i)\}$ at iteration (i).
$\hat{R}_y^{(i)}(j)$:	Estimated second-order moments of $\{y(i)\}$ at iteration (i).
$\hat{L}_y^{(i)}(m,n,l)$:	Estimated fourth-order cumulants of $\{y(i)\}$ at iteration (i).

Symmetric PAM or QAM Signaling:

In general, for 1-D (*e.g.* PAM) or 2-D (*e.g.* QAM) signaling with symmetric constellations:

$$\hat{L}_y^{(i)}(m,n,l) = \hat{M}_y^{(i)}(m,n,l) - \hat{R}_y^{(i)}(m)\cdot\hat{R}_y^{(i)}(n-l) - \hat{R}_y^{(i)}(n)\hat{R}_y^{(i)}(l-m) - \hat{R}_y^{(i)}(l)\hat{R}_y^{(i)}(n-m) \quad (5.22)$$

For symmetric square ($L \times L$) QAM constellations:

$$\hat{L}_y^{(i)}(m,n,l) = \hat{M}_y^{(i)}(m,n,l) \quad (5.23)$$

and $A_{(i)}^{(I)}, B_{(i)}^{(J)}$ are the minimum and maximum phase differential cepstrum parameters at iteration (i) respectively. L_1 and L_2 are the orders of the minimum phase and maximum phase components of the FIR channel, respectively. Note that, $\{a_i\}$, $|a_i| < 1$ and $\{\frac{1}{b_i}\}$, $|b_i| < 1$ are the zeros of the minimum and maximum phase components of the FIR channel, respectively.

$\{u(i)\}$: The coefficients of the equalizer at iteration (i).

$\{\tilde{x}(i)\}$: The coefficients of the equalizer at iteration (i).

At iteration (i): $i = 1, 2, \ldots$

Step 1 Estimate adaptively the $L_y^{(i)}(m,n,l)$, $-M \leq m,n,l \leq M$, from finite length window of $\{y(k)\}$ as described below. M should be sufficiently large so that $L_y(m,n,l) \simeq 0$ for $|m|$, $|n|$, $|l| > M$. Assuming that at iteration (0) we have received the time samples $\{y(1), \ldots y(I_{\text{lag}})\}$ we proceed as follows:

Stationary Case with Growing Rectangular Window

$$\hat{M}_y^{(i)}(m,n,l) = (1-\eta(i))\cdot\hat{M}_y^{(i-1)}(m,n,l) + \eta(i)\cdot y(S_4^i)y(S_4^i+m)y(S_4^i+n)y(S_4^i+l) \quad (5.24)$$

$$\hat{R}_y^{(i)}(j) = (1-\eta(i))\cdot\hat{R}_y^{(i-1)}(j) + \eta(i)\cdot y(S_2^i)y(S_2^i+j) \quad (5.25)$$

where, $\eta(i) = \frac{1}{i+I_{\text{lag}}}$, $S_2^i = \min(i+I_{\text{lag}}, i+I_{\text{lag}}-m, i+I_{\text{lag}}-n, i+I_{\text{lag}}-l)$, $S_2^i = \min(i+$

$I_{\text{lag}}, i + I_{\text{lag}} - j)$. Finally substitute (5.24) and (5.25) into (5.22) or (5.23).

Nonstationary Case

First Way:

$$\begin{aligned} &\text{for} \quad i \leq K \quad \text{use (5.24), (5.25) with} \quad \eta(i) = \frac{1}{i + I_{\text{lag}}} \\ &\text{for} \quad i > K \quad \text{use (5.24), (5.25) with} \quad \eta(i) = \eta = \text{fixed} \end{aligned} \tag{5.26}$$

η should have a small value $(0 < \eta < 1)$, for example $\eta = 0.01$.

Second Way: (for symmetric L^2- QAM signaling)

Since in this case the second-order moment $R_y(j) = 0$, we can use $M_y(m, n, l)$ with a forgetting factor $w, 0 < w < 1$ as follows. (S_4^i is as before):

$$(i+I_{\text{lag}})\cdot \hat{M}_y^{(i)}(m,n,l) = w\cdot(i-1+I_{\text{lag}})\cdot \hat{M}_y^{(i-1)}(m,n,l)+y(S_4^i)y(S_4^i+m)y(S_4^i+n)y(S_4^i+l) \tag{5.27}$$

and substitute $(i + I_{\text{lag}}) \cdot \hat{M}_y^{(i)}(m, n, l)$ for $\hat{L}_y^{(i)}(m, n, l)$ everywhere.

Third Way:

Formulas (5.24) and (5.25) could be used in nonstationary environments by reinitializing the algorithm after certain number of iteration or when a channel change is detected,

Remarks:

- By using the symmetry properties of fourth-order cumulants only $\frac{(2M+1)^3}{24}$ cumulants need to be calculated.
- The assumption that I_{lag} data have been received at iteration (0) avoids ill conditioning of the matrices of the system given in Step 3. It causes a delay to I_{lag} at the input of the

equalizer.

Step 2

Select p, q arbitrarily large so that $A^{(I)} \simeq 0$ and $B^{(J)} \simeq 0$ for $I > p$ and $J > q$. For example, $C = 10^{-4}$ (very small constant)

$$A^{(I)} \simeq 0 \quad \text{for} \quad I > p = int\left[\log \frac{C}{\alpha}\right]$$
$$B^{(J)} \simeq 0 \quad \text{for} \quad J > p = int\left[\log \frac{C}{\beta}\right] \tag{5.28}$$

where, $int[\cdot]$ denotes integer part and $max|a_i| < \alpha < 1$, $max|b_i| < \beta < 1$.

Define: $w = max(p, q), z \leq \frac{w}{2}, s \leq z$.

Step 3

Using the relation:

$$\sum_{I=1}^{p} \left\{ A_{(i)}^{(I)} \left[\hat{L}_y^{(i)}(m - I, n, l) - \hat{L}_y^{(i)}(m + I, n + I, l + I) \right] \right\} +$$
$$\sum_{J=1}^{q} \left\{ B_{(i)}^{(J)} \left[\hat{L}_y^{(i)}(m - J, n - J, l - J) - \hat{L}_y^{(i)}(m + J, n, l) \right] \right\} = -m \cdot \hat{L}_y^{(i)}(m, n, l) \tag{5.29}$$

with $m = -w, \ldots, -1, 1, \ldots, w, n = -z, \ldots, 0, \ldots, z$ and $l = -s, \ldots, 0, \ldots, s$ to form the overdetermined system of equations:

$$\hat{P}(i) \cdot \hat{a}(i) = \hat{p}(i) \qquad i = 0, 1, 2, \ldots \tag{5.30}$$

where $\hat{P}(i)$ is $[N_p \times (p + q)]$ (where $N_p = 2w \times (2z + 1) \times (2s + 1)$) matrix with entries of the form $\{\hat{L}_y^{(i)}(m, n, l) - \hat{L}_y^{(i)}(\sigma, \tau, \lambda)\}$; $\hat{a}(i) = [\hat{A}_{(i)}^{(1)}, \ldots, \hat{A}_{(i)}^{(p)}, \hat{B}_{(i)}^{(1)}, \ldots, \hat{B}_{(i)}^{(q)}]^T$ (T denotes transpose)

is the $(p+q) \times 1$ vector of unknown cepstral parameters; $\hat{p}(i)$ is the $N_p \times 1$ vector with entries of the form $\{-m \cdot \hat{L}_y(m,n,l)\}$.

Step 4

Assume that initially $\hat{a}(0) = [0,\ldots,0]^T$. Update $\hat{a}(i) = [\hat{A}(1),\ldots,\hat{A}_{(i)}^{(p)}, \hat{B}(1),\ldots,\hat{B}_{(i)}^{(q)}]^T$ as follows

$$\hat{a}(i+1) = \hat{a}(i) + \mu(1) \cdot \hat{P}^H(i) \cdot \hat{e}(i), \tag{5.31}$$

$$\hat{e}(i+1) = \hat{p}(i) - \hat{P}(i) \cdot \hat{a}(i), \quad 0 < \mu(i) < 2/tr\{\hat{P}^H(i) \cdot \hat{P}(i)\} \tag{5.32}$$

Step 5

Calculate the equalizer normalized coefficients. Initialize $\hat{i}_{inv}(i,0) = \hat{o}_{inv}(i,0) = 1$ and the estimate:

$$\hat{i}_{inv}(i,k) = \frac{1}{k} \sum_{n=2}^{k+1} [\hat{A}_{(i)}^{(n-1)}] \cdot \hat{i}_{inv}(i, k-n+1)$$

$$k = 1,\ldots,N_1 \tag{5.33}$$

$$\hat{o}_{inv}(i,k) = \frac{1}{k} \sum_{n=k+1} [-\hat{B}_{(i)}^{(1-n)}] \cdot \hat{o}_{inv}(i, k-n+1)$$

$$k = -1,\ldots,-N_2 \tag{5.34}$$

where (i) is the iteration index taking values $i = 1,2,3\ldots$ Then,

$$\hat{u}_{norm}(i,k) = \hat{i}_{inv}(i,k) * \hat{o}_{inv}(i,k), k = -N_2,\ldots,0,\ldots,N_1 \tag{5.35}$$

where $\{*\}$ denotes linear convolution.

Step 6

Estimate the gain factor $\hat{A}(i)$ as follows: In step (1) we have already calculated:

$$\hat{L}_y^{(i)}(0,0,0) \simeq \gamma_x \cdot \sum_k (f(k))^4$$
$$\hat{M}_2^{(i)}(0) \simeq Q_x \cdot \sum_k (f(k))^2 \tag{5.36}$$

where $Q_x = E\{(x(k))^2\}$, $\gamma_x = E\{(x(k))^4\} - 3 \cdot Q_x^2$ are known. Also:

$$\begin{aligned} \hat{i}(i,k) &= -\frac{1}{k} \sum_{n=2}^{k+1} \hat{A}_{(i)}^{(n-1)} \cdot \hat{i}(i,k-n+1), \quad k = 1,\ldots,p \\ \hat{o}(i,k) &= \frac{1}{k} \sum_{n=k+1}^{0} \hat{B}_{(i)}^{(1-n)} \cdot \hat{o}(i,k-n+1), \quad k = -1,\ldots,q \end{aligned} \tag{5.37}$$

and $\hat{f}(i,k) = \hat{i}(i,k) * \hat{o}(i,k)$, $\{*\}$ denotes convolution, $Q_{\hat{f}}(i) = \sum_k (\hat{f}(i,k))^2$, $\gamma_{\hat{f}}(i) = \sum_k (\hat{f}(i,k))^4$.

Then (the sign of $\frac{1}{\hat{A}(i)}$ cannot be identified):

For L-PAM Signaling:

$$\left| \frac{1}{\hat{A}(i)} \right| \simeq \left(\frac{Q_x \cdot Q_{\hat{f}}(i)}{\hat{M}_2^{(i)}(0)} \right)^{\frac{1}{2}} \tag{5.38}$$

For L^2-QAM Signaling:

$$\frac{1}{\hat{A}(i)} \simeq \left(\frac{\gamma_x \gamma_{\hat{f}}(i)}{\hat{L}_y^{(i)}(0,0,0)} \right)^{\frac{1}{4}} = |\gamma_x|^{\frac{1}{4}} \cdot e^{j\frac{\pi}{4}} \cdot \left(\frac{\gamma_{\hat{f}}(i)}{\hat{L}_y^{(i)}(0,0,0)} \right)^{\frac{1}{4}} \tag{5.39}$$

since $\gamma_x < 0$ for equi-probable L^2-QAM signaling.

Step 7

Let, $\underline{y}(i) = [y(i+N_2),\ldots,y(i-N_1)]^T$ and $[\underline{\hat{u}}_{norm}(i)] = [\hat{u}_{norm}(i,-N_2),\ldots,\hat{u}_{norm}(i,N_1)]^T$. Fi-

nally, the output of the TEA equalizer is:

$$\tilde{x}(i) = \frac{1}{\hat{A}(i)} \cdot [\underline{\hat{u}}_{norm}(i)]^T \cdot \underline{y}(i) \tag{5.40}$$

While most of the Bussgang blind equalization algorithms, which are based on non-MSE cost function minimization, have not been shown to be globally convergent and cases of their mis-convergence have been encountered, the TEA algorithm, designed as described above, is a more reliable alternative, as it guarantees convergence.

Remarks:

1. Since Gaussian noise is suppressed in the fourth-order cumulant domain, the identification of the channel response does not take into account the observation noise. Consequently, the proposed equalizers work under the zero-forcing (ZF) constraint. For the same reason, we expect that the identification of the channel will be satisfactory even in low signal to noise (SNR) conditions.

2. The ability of the tricepstrum method to identify separately the maximum and minimum phase components of the channel makes possible the design and implementation of different equalization structures.

3. In the recursive formulas (5.37) we used the following properties that relate time impulse responses with cepstrum coefficients: (i) a channel and its inverse have opposite in sign cepstrum coefficients, (ii) the cepstrum coefficients of the convolution of two minimum phase or two maximum phase sequences, are equal to the sum of the corresponding cepstrum coefficients of the individual sequences and (iii) two finite impulse response (FIR) sequences with

conjugate roots have also conjugate cepstrum coefficients. These become unique features of the TEA equalizer when is compared with other equalization schemes.

4. The described algorithm is based only on the statistics of the received sequence $\{y(i)\}$ and does not take into account the decisions $\{\hat{x}(i)\}$ at the output of the equalizer. Consequently wrong decisions (and thus error propagation effects) do not affect the convergence of the proposed equalization schemes.

5. Instead of using the LMS algorithm to solve adaptively the system of equations (5.30), one may employ a Recursive Least-Squares (RLS) algorithm [25] which will have a faster convergence at the expense of even more computations.

5.2.4 Power Cepstrum and Tricoherence Equalization Algorithm (POTEA) [7]

5.2.5 Relations of Power Cepstrum and Tricoherence of the Linear Filter Output

The problem is as formulated in Section 5.2.1, the channel output $y(i)$ is the convolution of the non-Gaussian *i.i.d.* random sequence $x(i)$ with the channel impulse response $f(i)$ plus some noise. The cepstrum of the power spectrum of the channel output $y(i)$, can be shown after some algebra to be equal to [7].

$$c_{p_y}(m) = \begin{cases} ln|A_2| & m = 0 \\ -\frac{1}{m}[A^{*(m)} + B^{(m)}] & m > 0 \\ \frac{1}{m}[A^{(-m)} + B^{*(-m)}] & m < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.41}$$

where $A^{(k)}$, $B^{(k)}$ are the minimum and maximum phase cepstral coefficients of $F(z)$. These are :

$$\begin{aligned} A^{(k)} &= \sum_{i=1}^{L_3} a_i^k - \sum_{i=1}^{L_4} c_i^k \\ B^{(k)} &= \sum_{i=1}^{L_2} b_i^k, \end{aligned} \tag{5.42}$$

where $\{a_i\}$ and $\{b_i\}$ are the zeros of $F(z)$ inside and outside of the unit circle respectively.

Remarks:

1. $A^{(k)}$, $B^{(k)}$ decay exponentially and thus their length can be truncated in practice at $k = p$, so that $A^{(p)}$, $B^{(p)}$ are arbitrarily small.

2. **If the channel $F(z)$ has cepstral coefficients $A^{(k)}$, $B^{(k)}$, its inverse filter, $U(z)$, has cepstral coefficients $-A^{(k)}$, $-B^{(k)}$.** It is also shown in [7] that if we define $S^{(k)} \triangleq A^{(k)} + B^{*(k)}$ and $r_y(k) \triangleq E\{y(i+k)y^*(i)\}$, then the following relations holds:

$$\sum_{k=1}^{p} S^{*(k)}[-r_y(m-k)] + \sum_{k=1}^{p} S^{(k)}[r_y(m+k)] = m r_y(m), \quad m = 1, \cdots 2p \tag{5.43}$$

where p is some integer, the choice of which is discussed in [24]. Now let us consider the cepstrum of the tricoherence.

$$R_y(z_1, z_2, z_3) = \left[\frac{S_y(z_1, z_2, z_3)}{S_y^*(z_1^{-1}, z_2^{-1}, z_3^{-1})} \right]^{\frac{1}{2}} \tag{5.44}$$

It has been shown that the trispectrum of the received data satisfies:

$$S_y(z_1, z_2, z_3) = \gamma_x F^*(z_1^{-1}) F(z_2) F^*(z_3^{-1}) F(z_1^{-1} z_2^{-1} z_3^{-1}) \tag{5.45}$$

Therefore,

$$R_y(z_1, z_2, z_3) = \left[\frac{F^*(z_1^{-1}) F(z_2) F^*(z_3^{-1}) F(z_1^{-1} z_2^{-1} z_3^{-1})}{F(z_1^{-1}) F^*(z_2) F(z_3^{-1}) F^*(z_1 z_2 z_3)} \right]^{\frac{1}{2}} \tag{5.46}$$

After some algebra, we obtain

$$R_y(m,n,l) = \frac{1}{2} \begin{cases} ln|A_1| & m=0, n=0, l=0 \\ -\frac{1}{m}[A^{*(m)} - B^{(m)}] & m>0, n=0, l=0 \\ -\frac{1}{m}[A^{*(-m)} - B^{*(-m)}] & m<0, n=0, l=0 \\ -\frac{1}{n}[A^{*(n)} - B^{*(n)}] & m=0, n>0, l=0 \\ -\frac{1}{n}[A^{*(-n)} - B^{*(-n)}] & m=0, n>0, l=0 \\ \frac{1}{m}[A^{*(m)} - B^{(m)}] & m=n=l>0 \\ \frac{1}{m}[A^{(-m)} - B^{*(-m)}] & m=n=l<0 \\ -\frac{1}{l}[A^{*(l)} - B^{(l)}] & m=0, n=0, l>0 \\ -\frac{1}{l}[A^{(-l)} - B^{*(-l)}] & m=0, n=0, l<0 \\ 0 & \text{otherwise} \end{cases} \tag{5.47}$$

Taking the logarithm of both sides of (5.44), we obtain,

$$R_y(z_1, z_2, z_3) = \frac{1}{2} \left[lnS_y(z_1, z_2, z_3) - lnS_y^*(z_1^{-1}, z_2^{-1}, z_3^{-1}) \right] \tag{5.48}$$

Differentiating with respect to Z_1 and performing inverse Z-transform, we obtain

$$\begin{aligned} &2L_y(m,n,l) * L_y^*(-m,-n,-l) * [-mR_y(m,n,l)] \\ &= L_y^*(-m,-n,-l) * [-mL_y(m,n,l)] + L_y(m,n,l) * [mL_y^*(m,n,l)] \end{aligned} \tag{5.49}$$

By defining the following functions:

$$\begin{aligned}\theta_1(m,n,l) &\triangleq L_y^*(-m,-n,-l) * L_y(m,n,l) \\ \theta_2(m,n,l) &\triangleq L_y^*(-m,-n,-l) * mL_y(m,n,l)\end{aligned} \tag{5.50}$$

are combining (5.49) and (5.50), we obtain:

$$2\theta_1(m,n,l) * [mR_y(m,n,l)] = \theta_2(m,n,l) + \theta_2^*(-m,-n,-l) \tag{5.51}$$

Defining $D^{(k)} = A^{(k)} - B^{*(k)}$ and combining (5.47), we obtain:

$$\begin{aligned}&\sum_{k-=1}^{p} D^{*(k)}[\theta_1(m-k,n-k,l-k) - \theta_1(m-k,n,l)] \\ &\quad + \sum_{k-=1}^{p} D^{(k)}[\theta_1(m+k,n+k,l+k) - \theta_1(m+k,n,l)] \\ &\quad = \theta_2(m,n,l) + \theta_2^*(-m,-n,-l)\end{aligned} \tag{5.52}$$

A rule of thumb is to define $w = p$, $z \leq w/2$, $h \leq z$ and then take $m = -w,\ldots,-1,1,\ldots,w$, $n = -z,\ldots z$, $l = -h,\ldots,h$ to form a linear overdetermined system to equations.

5.3 The POTEA Algorithm

In this section the POTEA algorithm is given in detail.

Let

N_1, N_2: Lengths of minimum and maximum phase components of the equalizer.

p: Length minimum and maximum phase cepstral parameters,

At iteration $i = 1, 2, \ldots$.

Step 1 Estimate adaptively the $L_y^{(1)}(m,n,l)$ for $-M_1 \le m,n,l \le M_1$, and $r_y^{(i)}(m)$ for $-M_2 \le m \le M_2$ from a finite length window of $\{y(n)\}$, and then generate the following functions:

$$\begin{aligned}\theta_1^{(i)}(m,n,l) &= L_y^{*(i)}(-m,-n,-l) * L_y^{(i)}(m,n,l)\\ \theta_2^{(i)}(m,n,l) &= L_y^{*(i)}(-m,-n,-l) * mL_y^{(i)}(m,n,l)\end{aligned}$$

Step 2 Choose p arbitrarily such that $A^{(p+1)} \approx 0$, $B^{(p+1)} \approx 0$ and define $w = p$, $z \le \frac{w}{2}$, $h \le z$.

Step 3 Form the equations

$$\sum_{k=1}^{p} S^{*(k)}[-r_y(m-k)] + \sum_{k=1}^{p} S^{(k)}[r_y(m+k)] = mr_y(m), \quad m = 1,\cdots,2p \tag{5.53}$$

where $S^{(k)} = A^{(k)} + B^{*(k)}$, $k = 1,\ldots,p$.

$$\begin{aligned}&\sum_{k=1}^{p} D^{*(k)}[\theta_1(m-k,n-k,l-k) - \theta_1(m-k,n,l)]\\ &+\sum_{k=1}^{p} D^{(k)}[\theta_1(m+k,n+k,l+k) - \theta_1(m+k,n,l)]\\ &\qquad = \theta_2(m,n,l) + \theta_2^*(-m,-n,-l)\end{aligned} \tag{5.54}$$

and the following system of equations

$$\hat{P}\hat{a} = \hat{p} \tag{5.55}$$

$$\hat{Q}\hat{b} = \hat{q} \tag{5.56}$$

where the matrices $\hat{P}, \hat{a}, \hat{p}, \hat{Q}, \hat{b}$ and $\hat{q}$ are defined above.

Step 4 Solve adaptively the above systems employing LMS-type adaptation as follows:

$$\hat{a}(i+1) = \hat{a}(i) + \mu(i)\hat{P}^H(i)\hat{e}(i) \tag{5.57}$$

$$\hat{b}(i+1) = \hat{b}(i) + \mu'(i)\hat{Q}^H(i)\hat{e}'(i) \tag{5.58}$$

where

$$\begin{aligned} \hat{e}(i) &= \hat{p}(i) - \hat{P}(i)\hat{a}(i) \\ \hat{e}'(i) &= \hat{q}(i) - \hat{Q}(i)\hat{b}(i) \\ 0 &< \mu(i) < \frac{2}{tr(\hat{P}^H\hat{P})} \\ 0 &< \mu'(i) < \frac{2}{tr(\hat{Q}^H\hat{Q})} \end{aligned}$$

The algorithm at instant i minimizes the mean square error:

$$\begin{aligned} \hat{J}(i) &= E\{\hat{e}^H(i)\hat{e}(i)\} \\ \hat{J}'(i) &= E\{\hat{e}'^H(i)\hat{e}'(i)\} \end{aligned}$$

Step 5 Calculate $A^{(k)}$ and $B^{(k)}$ as follows:

$$A^{(k)} = \frac{S^{(k)} + D^{(k)}}{2}, B^{(k)} = (\frac{S^{(k)} - D^{(k)}}{2})^* \tag{5.59}$$

Step 6 Calculate

$$\hat{i}_{eq}(i,k) = \frac{1}{k}\sum_{n=2}^{k+1}[A_{(i)}^{(n-1)}]\hat{i}_{eq}(i,k-n+1), k = 1,\ldots,N_1 \tag{5.60}$$

$$\hat{o}_{eq}(i,k) = \frac{1}{k}\sum_{n=k+1}^{0}[-B_{(i)}^{(1-n)}]\hat{o}_{eq}(i,k-n+1), k=-1,\ldots,-N_2 \qquad (5.61)$$

with initialization : $\hat{i}_{eq}(i,0) = \hat{o}_{eq}(i,0) = 1$. The normalized ($A = 1$) estimate $\hat{u}_{norm}(i,k)$ at iteration (i) is given by:

$$\hat{u}_{norm}(i,k) = \hat{i}_{eq}(i,k) * \hat{o}_{eq}(i,k) \qquad (5.62)$$

Step 7 Estimate gain factor $\hat{A}(i)$

Step 8 The reconstructed transmitted sequence at iteration(i) is:

$$\tilde{x}(i) = \frac{1}{\hat{A}(i)}\sum_{k=-N_2}^{N_1}\hat{u}_{norm}(i,k)y(i-k) \qquad (5.63)$$

Computational Complexity

In this section the computational complexity of POTEA is presented and compared with the computational complexity of TEA.

PAM

POTEA: $\frac{3(2M+1)^3}{8} + 3(2M+1) + 2p(N_p+p+1) + \frac{N^2+8N+3}{4} + (4M)^3\log_2 4M$

TEA: $\frac{3(2M+1)^3}{8} + 3(2M+1) + (p+q)(2N_p+1) + \frac{N^2+8N+3}{4}$

QAM

POTEA: $4[\frac{3(2M+1)^3}{8} + 2(2M+1) + 2p(2N_p+4p+2) + \frac{N^2+8N+3}{4} + (4M)^3\log_2 4M]$

TEA: $4[\frac{(2M+1)^3}{6} + (p+q)(2N_p+1) + \frac{N^2+8N+3}{4}]$

5.4 Cross-Tricepstrum Equalization Algorithm (CTEA) [8]

5.4.1 Problem Formulations

Assume we have n measurements at each time index k, $y_i(k), i = 1, 2, \ldots n$, where

$$y_i(k) = f_i(k) * x(k) + n_i(k) \tag{5.64}$$

(shown in Figure 5.1 for $n = 4$) and

1. $f_i(k)$ is the impulse response of a discrete time linear time invariant system,
2. $x(k)$ is a non-Gaussian, nth order white process with cumulant $\gamma_x \neq 0$,
3. $n_i(k)$ is zero-mean additive noise, with $n_i(k)$ independent of $n_j(k)$ for $i \neq j$ and independent of $x(k)$. No assumptions are made about *pdf* for whiteness (in time) of $n_i(k)$.

We also assume that each impulse response $h_i(k)$ is stable with no zeros on the unit circle and that its Z transform $F_i(z)$ can be written as [8]

$$F_i(z) = A_i I_i(z^{-1}) O_i(z) \tag{5.65}$$

where the A_i are gain constants, the r_i are integer linear phase factors,

$$I_i(z^{-1}) = \frac{\prod_{j=1}^{L_{i3}}(1 - a_{ij}z^{-1})}{\prod_{j=1}^{L_{i4}}(1 - c_{ij}z^{-1})}$$

is the minimum phase component and

$$O_i(z) = \prod_{j=1}^{L_{i2}}(1 - b_{ij}z)$$

is the maximum phase component, with zeros a_{ij} and poles c_{ij} inside and zeros b_{ij} outside the unit circle (*i.e.* $|a_{ij}| < 1, |b_{ij}| < 1$, and $|c_{ij}| < 1$).

5.4.2 Relation of Cross-Tricepstrum of the Linear Filter Output

With the above assumptions, the nth-order cross-spectrum of the $y_i(k)$ can be written as

$$S_{y,1,2,\cdots,n}(z_1, z_2, \cdots, z_{n-1}) = \gamma_x F_1(z_1) F_2(z_2) \cdots F_{n-1}(z_{n-1}) F_n(\prod_{i=1}^{n-1} z_i^{-1}) \tag{5.66}$$

Taking the logarithm and performing inverse Z-transform on both sides, we obtain after some algebra the following results:

$$c_{y,1,2,\cdots,n}(m_1, m_2, \cdots, m_{n-1}) = \begin{cases} ln\gamma_x & m_1 = m_2 = \ldots = m_{n-1} = 0, \\ -(1/m_i)A_i(m_i) & m_i > 0, m_j = 0, j \leq i, \\ & i = 1, 2, \ldots, n-1, \\ (1/m_i)B_i(-m_i) & m_i < 0, m_j = 0, j \neq i, \\ & i = 1, 2, \ldots, n-1, \\ -(1/m_n)A_n(-m_n) & m_1 = m_2 = \ldots = m_{n-1} < 0, \\ -(1/m_n)B_n(m_n) & m_1 = m_2 = \ldots = m_{n-1} > 0, \\ 0 & \text{otherwise} \end{cases} \tag{5.67}$$

with

$$\begin{aligned} A_i(k) &\triangleq \sum_{j=1}^{L_{i3}} (a_{ij})^k - \sum_{j=1}^{L_{i4}} (c_{ij})^k \\ B_i(k) &\triangleq \sum_{j=1}^{L_{i2}} (b_{ij})^k. \end{aligned} \tag{5.68}$$

This results means that the n-th order cross-cepstrum is non-zero on n lines only in its domain and that on each of these lines we find the complex cepstrum of a zero-linear phase, scaled version of *one* of the n impulse responses.

Now, to develop a least squares solution for the A_i and B_i, we take first partial derivatives of the logarithm of (5.66), independently with respect to each of its variables, followed by inverse $\mathcal{Z}$ transforms. Letting $S_{y,1,2,\ldots,n}(m_1, m_2, \ldots, m_{n-1})$ denote the n-th order cross cumulants of the y_i, we get the following $n-1$ equations relating the cross cumulants to the cepstral coefficients:

$$\begin{aligned} & S_{y,1,2,\ldots n}(m_1, m_2, \ldots, m_n) * (m_i \; c_{y,1,2,\ldots,n}(m_1, m_2, \ldots, m_{n-1})) \\ & \quad = -m_i \; S_{y,1,2,\ldots n}(m_1, m_2, \ldots, m_{n-1}) \end{aligned}$$

for $i = 1, 2, \ldots n-1$. Each equations involves an $(n-1)$ dimensional convolution. However, plugging in (5.67) reduces each equation to a single finite summation:

$$\begin{aligned} & \sum_{k=1}^{\infty} A_i(k) S_{y,1,2,\ldots,n}(t_1, t_2, \ldots, t_{n-1}) - B_i(k) S_{y,1,2,\ldots,n}(u_1, u_2, \ldots, u_{n-1}) \\ & \quad - A_n(k) S_{y,1,2,\ldots,n}(m_i + k, m_i + k, \ldots, m_i + k) \\ & \quad + B_n(k) S_{y,1,2,\ldots,n}(m_i - k, m_i - k, \ldots, m_i - k) \\ & \quad = -m_i S_{y,1,2,\ldots,n}(m_1, m_2, \ldots, m_{n-1}) \end{aligned} \tag{5.69}$$

where

$$\begin{aligned} t_i &= m_i - k \\ u_i &= m_i + k \\ t_j &= u_j = m_j \quad j \neq i. \end{aligned}$$

From equation (5.68) the sums in (5.69) decay, so we can truncate them to p_i and q_i for the terms involving A_i and B_i respectively (see [8]) and rewrite (5.69) as a finite dimensional vector dot product equation. Writing $M > p_i + q_i + p_n + q_n$ equations at M points in the $n-1$ dimensional domain of $S_{y,1,2,\ldots n}$ we can form the overdetermined system

$$\underline{R}_{in} \cdot \underline{C}_{in} = \underline{\gamma}_{in} \tag{5.70}$$

5.4.3 Cross-TEA (CTEA) Algorithm

In this section we describe the CTEA algorithm for blind equalization of QAM signals with four receivers. The algorithm has two stages at each iteration:

1. Channel identification and deconvolution
2. Combining by use of a decision rule

Channel Identification and Deconvolution

Step 1. Estimate the cross-cumulants and kurtoses of the received data recursively.

Step 2. Form the systems of equations (5.70) and solve each system in turn to get the cepstral coefficients for each channel[1]

Step 3. From the results of the previous step, estimate the forward and inverse channel impulse responses up to a desired length.

Step 4. From the estimated forward impulse response and the kurtoses, estimate the gains $A_i^{(j)}$ for each channel.

[1]The cepstral coefficients for channel four can be estimated from the solution of one of the three systems or an average of all three.

Step 5. With the estimated inverse response, $f_{i,\text{inv}}^{(j)}(k)$, and the estimated gain for each channel, deconvolve to estimate the input symbol as

$$\hat{x}_i(j) = \frac{1}{A_i^{(j)}} y_i(j) * f_{i,\text{inv}}^{(j)}(k)$$

Combining Decision Rules

As illustrated in Figure 5.1, from the four estimates $\tilde{x}_i(j)$ we need to form a single quantized decisions $\hat{x}(j)$. We describe here an optimal combining rule in the case of a perfect equalizer, as well as three sub-optimal schemes, arithmetic mean, majority rule, and median (which for $n = 4$ channels is equivalent to α-trimmed mean with $\alpha = 1$).

Optimal Decision for the Perfect Equalizer [8]

We consider the following assumptions:

1. $x(k)$ is complex and uniformly distributed,
2. $u_i(k)$ is the perfect equalizer for $f_i(k)$, *i.e.* $f_i(k) * u_i(k) = \delta(k)$, and
3. $n_i(k)$ are zero-mean, complex Gaussian variables with known variance σ_i^2 and are independent across channels.,

Since we will do symbol by symbol detection, we will drop the time index k for simplicity. With these assumptions,

$$\tilde{x}_i = x + n_i * u_i \triangleq x + \tilde{n}_i.$$

Therefore, the conditional probability density of $\hat{x}$ given X, $p(\hat{x}|x)$, is complex Gaussian with mean x and variance

$$\tilde{\sigma}_i^2 = \sigma_i^2 \sum_k |u_i(k)|^2.$$

Since the noise in each channel is independent, the maximum likelihood estimate $\hat{x}$ of x given the four observations $\tilde{x}_i$ (assuming x to be from a continuous distribution) is

$$\begin{aligned} \hat{x}_R &= \frac{\frac{1}{4}\sum_i \tilde{\sigma}_i^{-2}\tilde{x}_{i,R}}{\sum_i \tilde{\sigma}_i^{-2}} \\ \hat{x}_I &= \frac{\frac{1}{4}\sum_i \tilde{\sigma}_i^{-2}\tilde{x}_{i,I}}{\sum_i \tilde{\sigma}_i^{-2}} \end{aligned}$$

where the subscript R and I denote real and imaginary parts respectively. Note that if the noise has the same variance in all channels then this result reduces to the arithmetic man. If, on the other hand, we assume that x belongs to a known discrete set $\mathcal{D}$ then we need to find $\hat{x} \in \mathcal{D}$ which satisfies

$$\min_{\hat{x}\in\mathcal{D}} \sum_i \tilde{\sigma}_i^{-2}|\tilde{x}_i - \hat{x}|^2$$

or equivalently

$$\min_{\hat{x}\in\mathcal{D}} \sum_i \tilde{\sigma}_i^{-2}\left(|\hat{x}|^2 - 2(\hat{x}_R\tilde{x}_{i,R} + \tilde{x}_I\hat{x}_{i,I})\right).$$

Of course the assumptions of perfect equalization and known noise variance are not realistic in practice so we describe below three sub-optimal combining rules which we tested in our simulations.

Arithmetic Mean

Step 1. Form a soft decision statistic

$$\tilde{x}(j) = \frac{1}{4}\sum_{i=1}^{4} \tilde{x}_i(j).$$

(If information is available about the relative quality of the channels then a weighted mean could be used.)

Step 2. Put $\tilde{x}(j)$ through a decision device to get $\hat{x}(j)$.

Majority Rule

Step 1. Put each estimate through a decision device to form four decision statistics $\hat{x}_i(j)$.

Step 2. If there is a plurality among the $\hat{x}_i(j)$ in one region of the decision space then that is the decision. If there is a tie (all four different or two votes for each of two decisions) use a tie-breaking procedure. One method would be to pick the decision region that has the smallest average squared decision error. For example, if $\hat{x}_i(j) = \hat{x}_2(j) \neq \hat{x}_3(j) = \hat{x}_4(j)$:

$$\text{Let } d_1 = \sum_{i=1}^{2} |\hat{x}_i(j) - \tilde{x}_i(j)|^2$$

$$\text{Let } d_2 = \sum_{i=3}^{4} |\hat{x}_i(j) - \tilde{x}_i(j)|^2$$

Then

$$\text{Choose} \quad \hat{x}_1(j) \quad d_1 \leq d_2$$

$$\hat{x}_2(j) \quad d_2 > d_1.$$

Median

Step 1. Order the real and imaginary parts of the $\tilde{x}_i(j)$ separately.

Step 2. Set

$$\text{REAL}\{\tilde{x}(j)\} = \text{median}\{\text{REAL}\{\tilde{x}_i(j)\}\}$$

$$\text{IMAG}\{\tilde{x}(j)\} = \text{median}\{\text{IMAG}\{\tilde{x}_i(j)\}\}$$

Step 3. Put $\tilde{x}(j)$ through the decision to get $\hat{x}(j)$.

5.5 Computer Simulations

Computer simulations has been employed to compare the performance of the blind equalization algorithms. The performance metric used are those in Sections 2. And the following issues are addressed.

5.5.1 TEA vs. Bussgang-type Algorithms

Fig. 5.2-5.4 show the performance of the TEA algorithm, compared with that of Bussgang-type algorithms, such as Godard, Benveniste-Goursat, Stop-and-Go algorithms. We see that the TEA algorithm opens the eye much faster than the Bussgang-type algorithms. This performance improvement is achieved at the expense of larger computational complexity.

5.5.2 POTEA vs. TEA

Fig. 5.5-5.6 show the performance of the POTEA algorithm, compared with that of TEA. We see that the POTEA algorithm converges faster than the TEA algorithm. The performance improvement is achieved at the expense of further increase in computational complexity.

5.5.3 CTEA vs. TEA

Fig. 5.7-5.8 show the performance of the CTEA algorithm compared with that of TEA algorithm. We see that the CTEA algorithm converges faster than the TEA algorithm for some channels. The performance improvement is achieved at the expense of further increase in computational complexity.

6 ALGORITHM WITH NONLINEARITY INSIDE THE EQUALIZATION FILTER

Still another class of bind equalization algorithms are those algorithms which use Volterra filters [9], [10] or neural networks [20], [26], [27]. This class of algorithms perform nonlinear operations inside the equalization filter. It is therefore also be able to correctly extract the phase information of the unknown channel from its output only. In this section, we will concentrate on those algorithms based on neural network.

6.1 Review of Equalization Techniques Based on Neural Networks

Equalization is a technique which is used to combat the intersymbol interference caused by non-ideal channels. Usually, equalizers are implemented using linear transversal filters [17], [18], [30], [31]. However, when the unknown channel has deep spectral nulls or some severe nonlinear distortions, such as phase jitter and frequency offset, linear equalizers are not powerful enough to compensate all of these. That is why nonlinear filters, such as those implemented by Volterra filter or neural network, come in and play an important role.

Neural Networks (NNWs) are mathematical models of theorized mind and brain activities. The fundamental idea of NNWs is to organize many simple identical processing elements into

layers to perform more sophisticated tasks. The properties of NNWs include: massive parallelism; high computation rates; great capability for non-linear problems, continuous adaptation; inherent fault tolerance and ease for VLSI implementation, etc. All these properties make NNWs attractive to various applications. Several neural network based algorithms have been proposed for equalization problems.

1 Multi-Layer Perceptron

The multi-layer Perceptron (MLP) [39], [40] is one of the most widely used implementations of NNWs. It comprises a number of nodes which are arranged in layers, as shown in Figure 6.1. A node receives a number of inputs $x_1, x_2, \cdots, x_n$, which are then multiplied by a set of weights $w_1, w_2, \cdots, w_n$ and the resultant values are summed up. A constant v is added to this weighted sum of inputs, known as the node threshold, and the output of the node is obtained by evaluating a nonlinear (sigmoid) function, $f(.)$, which is called activation function.

The architecture of a perceptron can be described by a sequence of integers $n_0, n_2, \cdots, n_k$. where n_0 is the dimension of the input to the network, and the number of nodes in each successive layer, ordered from input to output, is $n_1, n_2, \cdots, n_k$. In this notation, the MLP produces a nonlinear mapping $g = R^{n_0} \rightarrow R^{n_k}$.

The updating of the connection coefficients of the MLP is done iteratively by using back-propagation (BP) algorithm with the following formula:

$$(w_{i+1}, v_{i+1}) = (w_i, v_i) + \Delta_i \tag{6.1}$$

and

$$\Delta_i = -(\beta, \eta) \cdot \frac{de^2}{d(w_i, v_i)} + \alpha \cdot \Delta_{i-1}. \tag{6.2}$$

2 Self-Organizing Feature Maps

The topology by self-organizing feature map (SOM), which is introduced by Kohonen [26], [27] consists of two layers of nodes, referred to as input layer and output layer, which are fully connected with different connection weights. The inputs to the SOM can be any continuous values, whereas each of the output-layer node represent a pattern class that the input vector may belong to. That means the outputs of SOMs are discrete values, and therefore, the SOM is sometimes also referred to as learning vector quantizer.

The SOM works iteratively as follows. First, find the set of connection coefficients W_g which is the closest to the input vector A_k,

$$\| A_k - W_g \| = \min_{j=1}^{p} \| A_k - W_j \| . \tag{6.3}$$

Second, perform the following quantization of the output-layer node:

$$b_g = \begin{cases} 1, & \text{if } \| A_k - W_g \| = \min \| A_k - W_j \| \\ 0, & \text{otherwise.} \end{cases} \tag{6.4}$$

and then move W_g closer to A_k using the equation

$$\Delta W_{ij} = \begin{cases} \alpha(k) \cdot [a_i^k - W_{ig}], & j = g \\ \beta(k) \cdot [a_i^k - W_{ij}], & j \in N_r, j \neq g \\ 0, & j \notin N_r, \end{cases} \tag{6.5}$$

where N_r is the topological neighborhood of the winning node b_g which consists b_g itself and its direct neighbors up to the depth $1, 2, \cdots$, and $\alpha(k)$ and $\beta(k)$ are the learning rate at time k.

6.2 The MLPs Equalization Algorithm for PAM and QAM Signals

The applications of MLP in equalization problems so far, have been limited to binary $\{0, 1\}$ or bipolar $\{-1, 1\}$ valued data and real valued channel models [11], [20], [49]. In this section, we introduce for the first time a **new implementation structure** of MLP which works well with L-PAM ($L > 2$) and N-QAM (N¿4) signals.

Looking into a MLP structure, we find out that it is the sigmoid function of the output layer nodes that confines the network outputs to the range $[-1, 1]$. In our equalization problem, the signals are equally spaced and symmetric with respect to either the original point of the coordinate, or to the x and y axes. Thus **we can just scale up the node function of the output layer by a constant factor C which is large enough to cover our maximum signal range**, e.g., [-15,15] for 16-PAM or 256-QAM signals. So, for the output layer, we have [30], [40]

$$f_M(x) = C \cdot \frac{1 - e^{\alpha x}}{1 + e^{\alpha x}}, \qquad (c \geq 1) \tag{6.6}$$

as the activation function. For the hidden layers, we still use the sigmoid function

$$f_i(x) = \frac{1 - e^{\alpha x}}{1 + e^{\alpha x}}. \tag{6.7}$$

The idea of adding another constant α comes form the thought that a smaller α, equivalently, a lower slope in Figure 6.2, would avo'd high vibration, and in turn, decrease the chance of

divergence in the course of weight adjustment.

For complex channel models and QAM signals, we use complex connection coefficients to get the weighted sum to which a complex threshold is added. Then the sigmoid functions of the real and the imaginary parts of the threshold added weighted sum are evaluated separately. Again, for the output layer nodes, the outputs are multiplied by a constant C. Using the steepest descent formula (Eq. 6.1, 6.2), we get the adaptation algorithm of our new MLP equalizer which is described in Table 6.1 [30], [40].

Simulation are conducted to examine the performance of MLP equalizers. The equalizer is implemented by the new MLP structure with only one output node. The input data to the system x_i are assumed to be independent of each other. The delayed input sequence x_{i-d}, where d is channel dependent, is used as the training sequence. The performance of MLP equalizers is evaluated by calculating the mean square error (MSE) $E[(x - \tilde{x})^2]$ and the average symbol error rate (SER) of the quantizer output. The eye pattern of equalizer outputs around certain number of iterations is shown in Figure 6.3.

Figure 6.4 illustrates the performance comparison between MLP and LMS-based linear transversal equalizer with the same number of inputs. The structure (the number of nodes in the hidden layer) of the MLP has been fine-tuned through experiment. The step size μ of the LMS-based equalizer is also optimized (the biggest value without causing divergence). From Fig. 6, it appears that the new structure of MLP works no much better, as a channel equalizer, than the simple linear adaptive equalizer. As a matter of fact, both methods end giving similar results.

7 CONCLUSIONS

The purpose of this paper is to provide a tutorial review of existing blind equalization algorithms for digital communications. Three families of techniques have been described, namely, the Bussgang techniques, the polyspectra-based techniques, and methods based on nonlinear equalization filters or neural networks. The complexity of the Bussgang techniques is approximately $2N$ multiplications per iteration, where N is the order of the linear equalization filter. On the other hand, the polyspectra-based techniques require approximately $\frac{1}{2}N^3$ multiplications per iteration. However, as it has been demonstrated in the paper, the polyspectra-based techniques achieve significantly faster convergence rate than the Bussgang techniques. Finally, it is pointed out in the paper that blind equalizers based on nonlinear filters or neural networks are better suited for equalization of channels with nonlinear distortions.

List of Figures

List of Tables

REFERENCES

[1] Austin, M.E., "Decision-Feedback Equalization for Digital Communication over Dispersive Channels," *IEEE Trans. Comm* , Vol. COM-30, pp. 2421-2433, November 1982.

[2] Belfiori, C.A. and J.H. Park, Dr., "Decision-Feedback Equalization," *Proc. IEEE*, Vol 67, pp: 1143-1156, August 1979.

[3] Bellini, S. "Bussgang Techniques for Blind Equalization," *Proc. IEEE-Globecom'86*, pp. 46.1.1-46.1.7.

[4] Bellini, S. and F. Rocca, "Blind Equalization: Polyspectra or Bussgang Technbiques," in Digital Communications, E. Biglieri and G. Prati Eds., Northholland, 1986, pp. 251-262.

[5] Benveniste, A. and M. Goursat, "Blind Equalizers," *IEEE Trans. on Communications*, Vol. COM-32, pp. 871-883, August 1984.

[6] Benveniste, S., M. Goursat and G. Ruget, "Robust Identification of Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications," *IEEE Trans. Automat. Contr.*, Vol. AC-25, No. 3, pp. 385-398, 1980.

[7] Bessios, A.G. and C.L. Nikias, "POTEA: The Power Cepstrum and Tricoherence Equalization Algorithm," *Proc. SPIE*, San Diego, CA July 1991.

[8] Brooks, D.H. and C.L. Nikias, "Cross-Bicepstrum and Cross-Tricepstrum Approaches to Multichannel Deconvolution," *Proc. Int. Signal Processing Workshop in Higher-Order Statistics*, Chamrousse, France, July 1991.

[9] Biglieri, S. Barberis, and M. Catena, "Analysis and Compensation of Nonlinearities in Digital Transmission Systems," *IEEE Sel. Areas in Comm.*, Vol. 6, No. 1, pp. 42-51,

January 1988.

[10] Biglieri, E., A. Gersho, R.D. Gitlin, and T.L. Lim, "Adaptive Cancellation of Nonlinear Symbol Interference for Voiceband Data Transmission," *IEEE Sel. Areas in Comm.*, Vol. SAC-2, No. 5, September 1984.

[10] Bussgang, J.J., "Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals", *MIT Technical Report*, No. 216, March 1952.

[11] Chen, S., G.J. Gibson, C.F.N. Cowan , and P.M. Grant, "Adaptive Equalization of Finite Non-Linear Channels Using Multilayer Perceptrons," *Signal Processing*, 20, pp. 107-109, 1990.

[12] Chen, Y., C.L. Nikias and J.G. Proakis, "CRIMNO: Criterion with Memory Nonlinearity for Blind Equalization," *Proc. Intern. Signal Processing Workshop in Higher-Order Statistics*, Chamrousse, France, July 1991.

[13] Chen, Y., C.L. Nikias and J.G. Proakis, "CRIMNO: Criterion with Memory Nonlinearity for Blind Equalization," *Proc. 25th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1991.

[14] Chiang, H.H. and C.L. Nikias, "Deconvolution and Identification of Nonminimum Phase FIR Systems Based on Cumulants," *IEEE Trans. Automatic Control*, Vol. AC-35, No. 1, January 1990.

[15] Ding, Z., R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson, "Existence and Avoidance of Ill-convergence of Godard Blind Equalizers in Data Communication Systems," 23rd Conference on Information Sciences and Systems, Baltimore, MD, March 1989.

[16] Donoho, D.L., "On Minimum entropy Deconvolution," in D.F. Findley, (ed.), *Applied Time Series Analysis, II*, Academic Press, NY, 1981.

[17] Forney, G.D., Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence on Intersymbol Interferences," *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 363-378, May 1972.

[18] Foschini, G.J., "Equalizing without Altering or Detecting Data," *AT&T Technical J.*, Vol. 64, No. 8, pp. 1885-1909, October 1985.

[19] Gersho, A. and T.L. Lim, "Adaptive Cancellation of Intersymbol interference for Data Transmission," *The Bell Syst. Tech. Journal*, pp. 1997-2021, November 1981.

[20] Gibson, G.J., S. Siu, and C.F.N. Cowan, "Application of Multilayer Perceptrons as Adaptive Channel Equalizers," *Proceedings IEEE Internatioal Conference ASSP*, Glasgow, Scotland, May 23-26, 1989, pp. 1183-1186.

[21] Gitlin, R.D. and S.B. Weinstein, "Fractionally-Spaced Equalization: An Improved Digital Transversal Equalizer," Bell Syst. Tech. J., Vol. 60, pp. 275-296, February 1981.

[22] Godard, D.N., "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. on Communications*, Vol. COM-28, pp. 1867-1875, November 1980.

[23] Godfrey, R. and F. Rocca, "Zero Memory Nonlinear Deconvolution," *Geophysical Prospecting*, Vol. 29, pp. 189-228, 1981.

[24] Hatzinakos, D. and C.L, Nikias, "Blind Equalization using a Tricepstrum Based Algorithm," *IEEE Trans. Comm.*, Vol. COM-39, May 1991.

[25] Haykin, S., Adaptive Filter Theory, Prentice Hall, Inc., Englewood Cliffs, NJ, 1991.

[26] Kohonen, T., K. Raivio, O. Simula, "Combining Linear Equalization and Self-Organizing Adaptation in Dynamic Discrete-Signal Detection," *IJCNN-90*, Vol. 1, San Diego, CA, 1990.

[27] Kohonen, T., "Self-Organization and Associative Memory," Series in Information Sciences, Vol. 8, Springer-Verlag, 3rd Edition, 1989.

[28] Kolmogorov, A.N., "On the Representation of Continuous Functions of Many Variables by Superposition of Continuous Functions of One Variable and Addition," *Dokl. Adad. Nauka USSR*, Vol. 144, pp. 953-956. 1957.

[29] Lii, K.S. and M. Rosenblatt, "Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non-Gaussian Linear Processes," *Ann. Statistics*, Vol. 10, pp. 1195-1208, 1982.

[30] Lucky, R.W., "Automatic Equalization for Digital Communications," *Bell Systems Tech. J.*, Vol. 44, pp. 547-588, 1965.

[31] Lucky, R.W., "Techniques for Adpative Equalization for Digital Communications," *Bell Systems Tech. J.*, Vol. 45, pp. 555-286.

[32] Macchi, O. and E. Eweda, "Convergence Analysis of Self-Adaptive Equalizers," *IEEE Trans. Inform. Theory*, Vol. IT-30, pp. 161-176, March 1984.

[33] Mazo, J.E., "Analysis of Decision-Directed Equalizer Convergence," *The Bell Systems Tech. J.*, Vol. 59, pp. 1857-1876, 1980.

[34] Mendel, J.M., "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proceedings IEEE*, Vol. 9, pp. 278-305, March 1991.

[35] Nikias, C.L., "ARMA Bispectrum Approach to Nonminimum Phase System Identification," *IEEE Trans. ASSP*, Vol. 36, No. 4, pp. 513-525, April 1988.

[36] Nikias, C.L., "Blind Deconvolution using Higher-Order Statistics," *Proc. International Workshop on Higher-Order Statistics*, France, July 1991.

[37] Nikias, C.L., and H.H. Chiang, "Higher-Order Spectrum Estimation via Noncausal Autoregressive Modeling and Deconvolution," *IEEE Transactions ASSP*,Vol. 36, No. 12, pp. 1911-1914, December 1988.

[38] Nikias, C.L. and M.R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework," *Proc. IEEE*, Vol. 75, No. 7, pp. 869-891, July 1987.

[39] Peng, M., C.L. Nikias and J. Proakis, "Adaptive Equalization for PAM and QAM Signals with Neural Networks," *Proc. 25th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1991.

[40] Peng, M., C.L. Nikias and J. Proakis, "Adaptive Equalization with Neural Networks : New Multi-layer Perception Structures and their Evaluation," *ICASSP'92*, San Francisco, CA, March 1992.

[41] Picchi, G. and G. Prati, "Blind Equalization and Carrier Recovery Using a 'Stop-and-Go' Decision-Directed Algorithm," *IEEE Trans. on Communications*, Vol. COM-35, No. 9, pp. 877-887, September 1987.

[42] Porat, B. and B. Friedlander, "A Blind SAG-SO-DFD-FS Equalizer," *Proc. of the IEEE Globecom'88*, pp. 30.4.1-30.4.5.

[43] Proakis, J.G., "Advances in Equalization for Intersymbol Interference," on *Advances in Communication Systems*, Vol. 4, A.J. Viterbi (ed.), Academic Press, New York, 1975.

[44] Proakis, J.G.,"Digital Communications," McGraw-Hill, New York, 1989, Second Edition.

[45] Qureshi, S.U.H. and G.D. Forney, Jr., "Performance Properties of a T/2 Equalizer," pp. 11.1.1-11.1.14, Los Angeles, CA, December 1977.

[46] Sato, Y., et al., "Blind Suppression of Time Dependency and its Extension to Multi-Dimensional Equalization," *Proc. of IEEE-ICC'86*, pp. 46.4.1-46.4.5.

[47] Sato, Y., "A Method for Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems," *IEEE Trans. on Communications*, Vol. COM-23, pp. 679-682, June 1975.

[48] Shalvi, O. and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," *IEEE Trans. Information Theory*, Vol. 36, pp. 312-321, 1990.

[49] Siu, S., G.J. Gibson, and C.F.N. Cowan, "Decision Feedback Equalization Using Neural Network Structures," *IEEE International Conference on Neural Networks*, London.

[50] Treichler, J.R. and B.G. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol ASSP-31, No. 2, pp. 459-471, April 1983.

[51] Ungerbeck, G., "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communication," *IBM J. Res. Develop.*, pp. 546-555, November 1972.

[52] Van Trees, H.L., Detection, Estimation and Modulation Theory, Part I, J. Wiley and Sons, Inc., New York, 1968.

[53] Widrow, B. and S.D. Stearns, Adaptive Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1985.

[54] Wiggins, R.A., "Minimum Entropy Deconvolution," *Geoexploration*, Vol. 16, pp.21-35, 1978.

Table 4.1 Nonlinear Functions of Bussgang Iterative Techniques.

$\underline{u}(i) = [u_1,(i), \cdots, u_N(i)]^T$ equalizer taps

$\underline{y}(i) = [y(i), \cdots, y(i-N+1)]^T$ input to the equalizer block of data

At iteration$\{i\}$, $i = 1, 2, \cdots$

$\tilde{x}(i) = \underline{u}^H(i)\ \underline{y}(i)$

$e(i) = g^{(i)}[\tilde{x}(i)] - \tilde{x}(i)$

$\underline{u}(i+1) = \underline{u}(i) + \mu\ \underline{y}(i)\ e^*(i)$

Algorithm	**Nonlinear function:** $g[\tilde{x}(i)] =$
LMS training mode	$x(i)$ (linear)
Decision Directed Mode	$\hat{x}(i)$
Sato	γ csgn $[\tilde{x}(i)]$
Benveniste-Goursat	$\tilde{x}(i) + k_1\,(\hat{x}(i) - \tilde{x}(i)) + k_2\lvert\hat{x}(i) - \tilde{x}(i)\rvert \cdot (\gamma\ \text{csgn}[\tilde{x}(i)] - \tilde{x}(i))$
Godard $p, q = 2$	$\frac{\tilde{x}(i)}{\lvert\tilde{x}(i)\rvert} \cdot \left\{\lvert\tilde{x}(i)\rvert + R_p\lvert\tilde{x}(i)\rvert^{p-1} - \lvert\tilde{x}(i)\rvert^{2p-1}\right\}$
Stop-and-Go	$\tilde{x}(i) + \frac{1}{2}A\,(\hat{x}(i) - \tilde{x}(i)) + \frac{1}{2}B\,(\hat{x}(i) - \tilde{x}(i))^*$ (A,B) = (2,0), (1,1), (1,-1) or (0,0), depending on the signs of DD and Sato errors

Table 4.2 Comparison of Computational Complexity

	Godard	CRIMNO (memory size M)		Adaptive Weight CRIMNO (memory size M)
		Version I	Version II	Version I
Real Multiplication	4N+5	4N+8M+5	MN+8M+4N+5	4N+10M+5

Table 6.1 Complex MLP adaptation algorithm.

1). Assign small random complex numbers to all the connections and thresholds.

2). Forward propagate inputs through the network:

$$\tilde{a}_{i+1,j} = \sum_{l=1}^{n_i} a_{i,l} \cdot w_{i,l,j}^{*} + v_{i,j} = \tilde{a}_{i+1,j}^{I} + j \cdot \tilde{a}_{i+1,j}^{Q},$$

$$\tilde{a}_{i+1,j} = f(\tilde{a}_{i+1,j}^{I}) + j \cdot f(\tilde{a}_{i+1,j}^{Q}),$$

where $i = 1, \cdots M$ (M is the number of layers), $f(\cdot)$ is the sigmoid function, and get the output,

$$\hat{x} = C \cdot a_{M1}.$$

3). Present the training signal to find the output error,

$$\tilde{e}_m = e_M^I \left[1 - (\hat{x}^I/C)^2\right]/C + j e_M^Q \left[1 - (\hat{x}^Q/C)^2\right]/C$$

where $e_M = x_{i-d} - \hat{x}$.

4). Find the backpropagation error,

$$\tilde{e}_{ij} = \underline{e}_{ij}^{I} \cdot [1 - (a_{ij}^{I})^2] + j \cdot \underline{e}_{ij}^{Q} \cdot [1 - (a_{ij}^{Q})^2],$$

where

$$\underline{e}_{i,j} = \sum_{l=1}^{n_{i+1}} w_{i,j,l} \cdot \tilde{e}_{i+1,l}.$$

5). Adjust connections and thresholds:

$$w_{i,j,k}(n+1) = w_{i,j,k}(n) + \eta \cdot \tilde{e}_{i+1,j}^{*}(n) \cdot a_{ij}(n),$$
$$v_{ij}(n+1) = v_{ij}(n) + \beta \cdot \tilde{e}_{ij}(n).$$

where "$*$" denotes conjugate operator. The momentum term can also be added.

6). Back to Step 2.

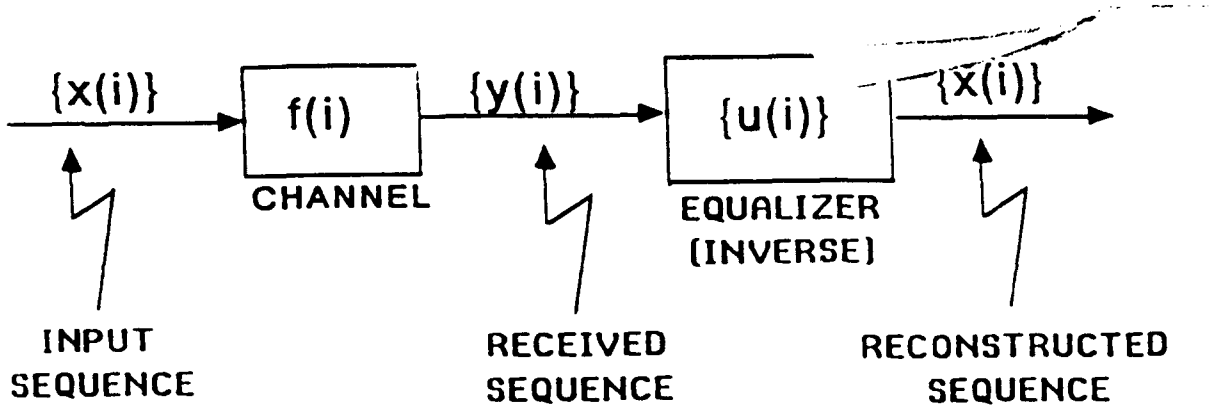


Figure 2.1 Block Diagram of Channel and Equalization Filter

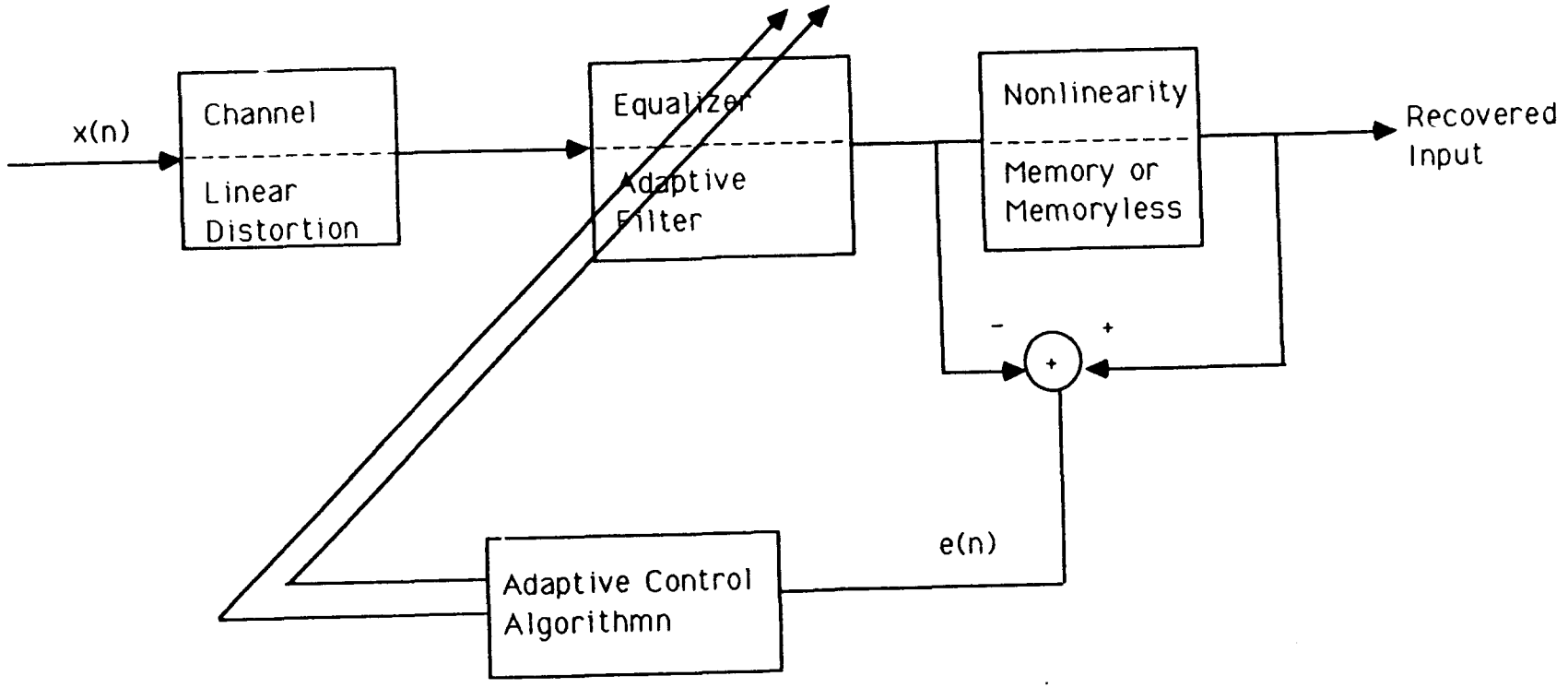


Figure 2. 2 (a) The Bussgang Algorithms: Nonlinearity is in the Output of Equalization Filter.

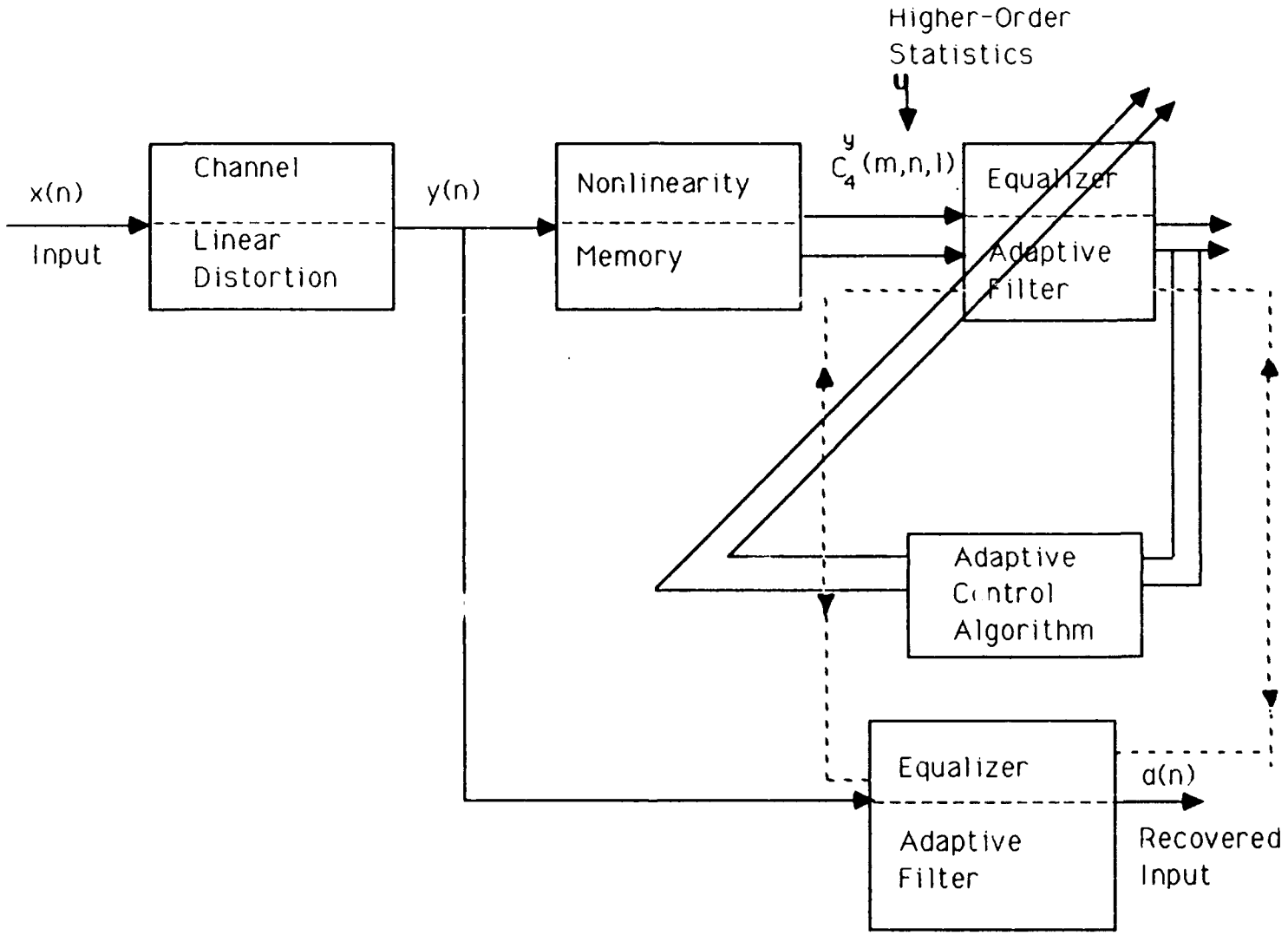


Figure 2. 2 (b) The Polyspectra Algorithms: Nonlinearity is in the Input of Equalization Filter.

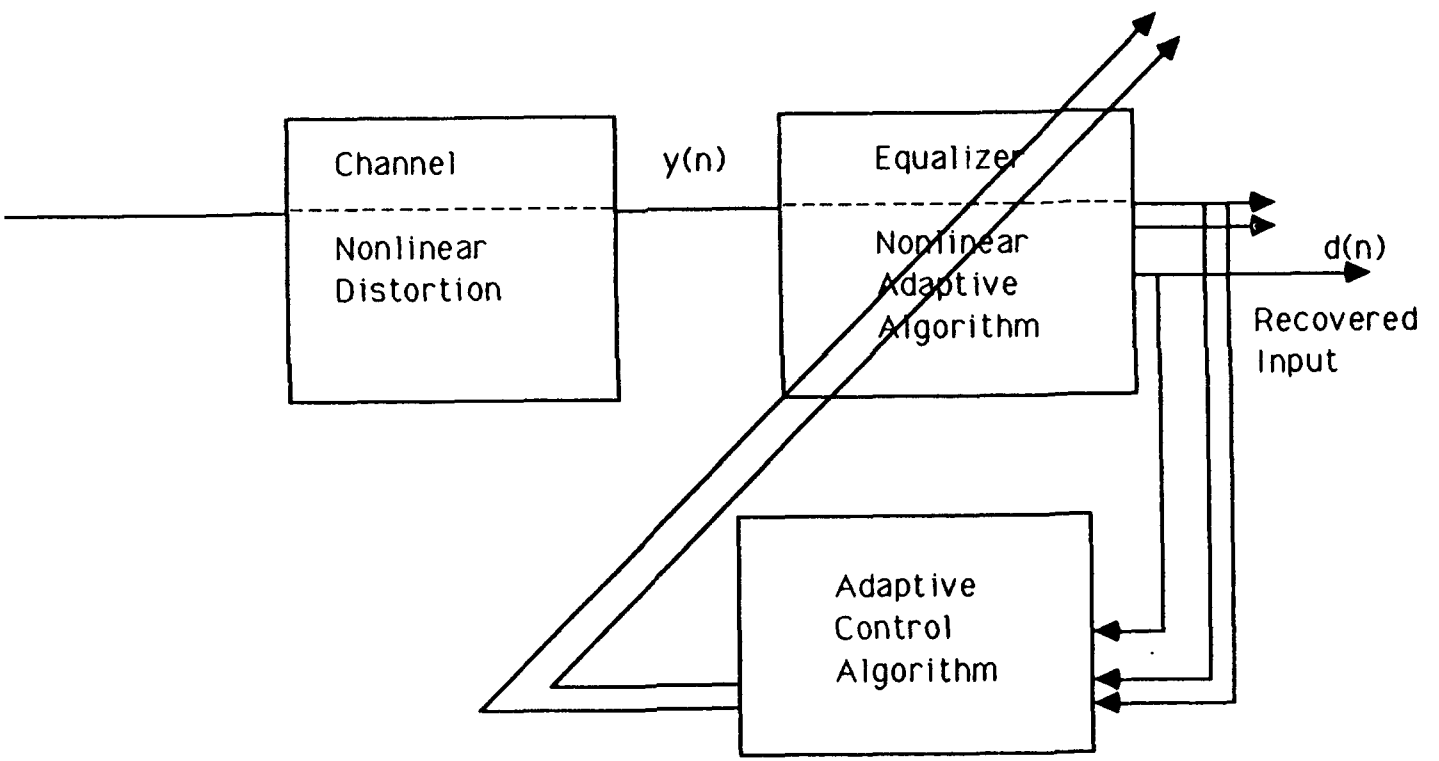


Figure 2. 2 (c) Blind Equalizers with Nonlinearity Inside the Equalization Filter.

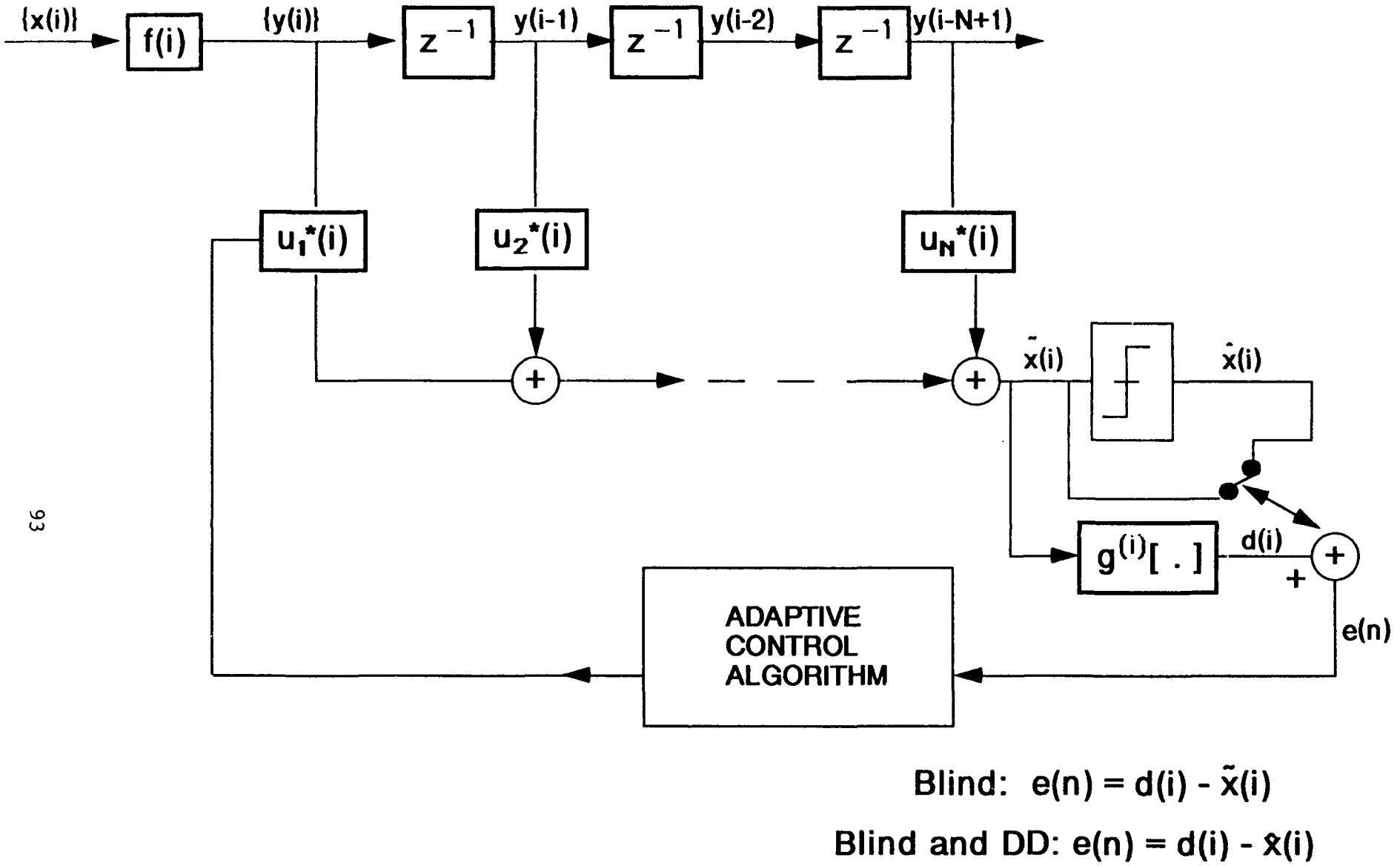


Figure 4.1 Linear Blind Equalization Filter with Nonlinearity in the Output.

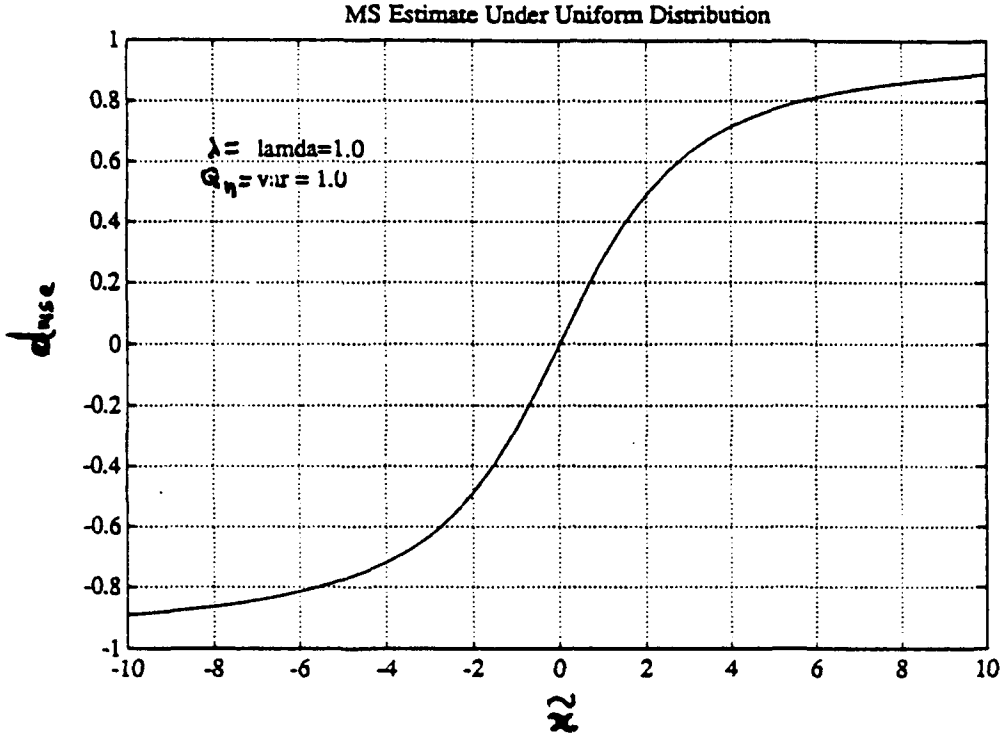


Figure 4.2 MS Estimate Under Uniform Distribution

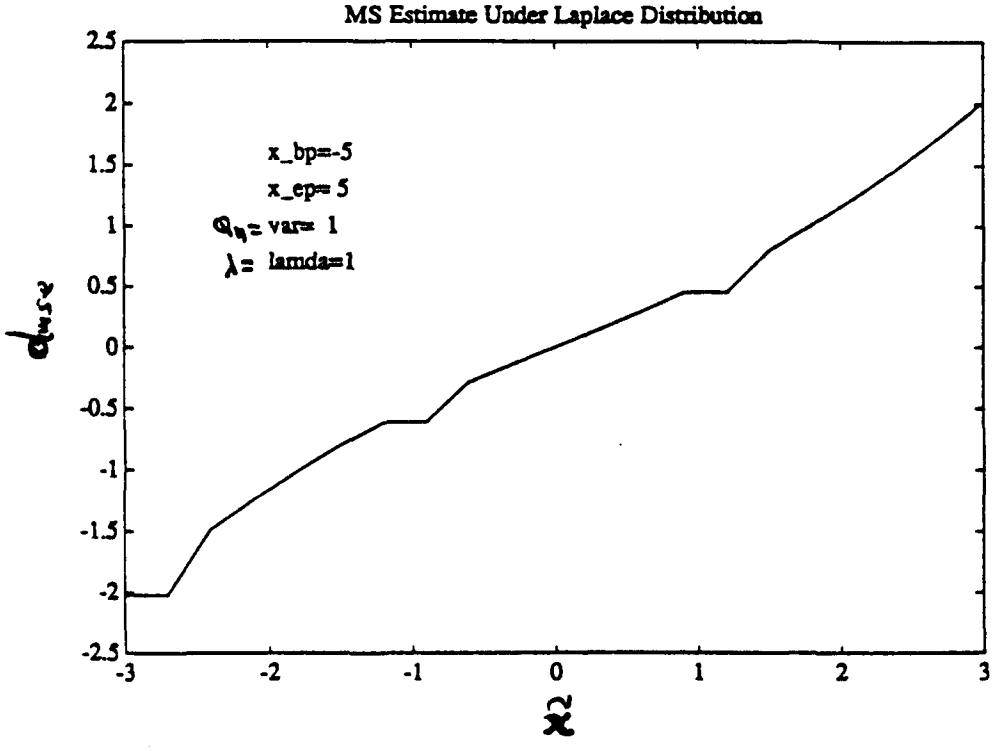


Figure 4.3 MS Estimate Under Laplace Distribution

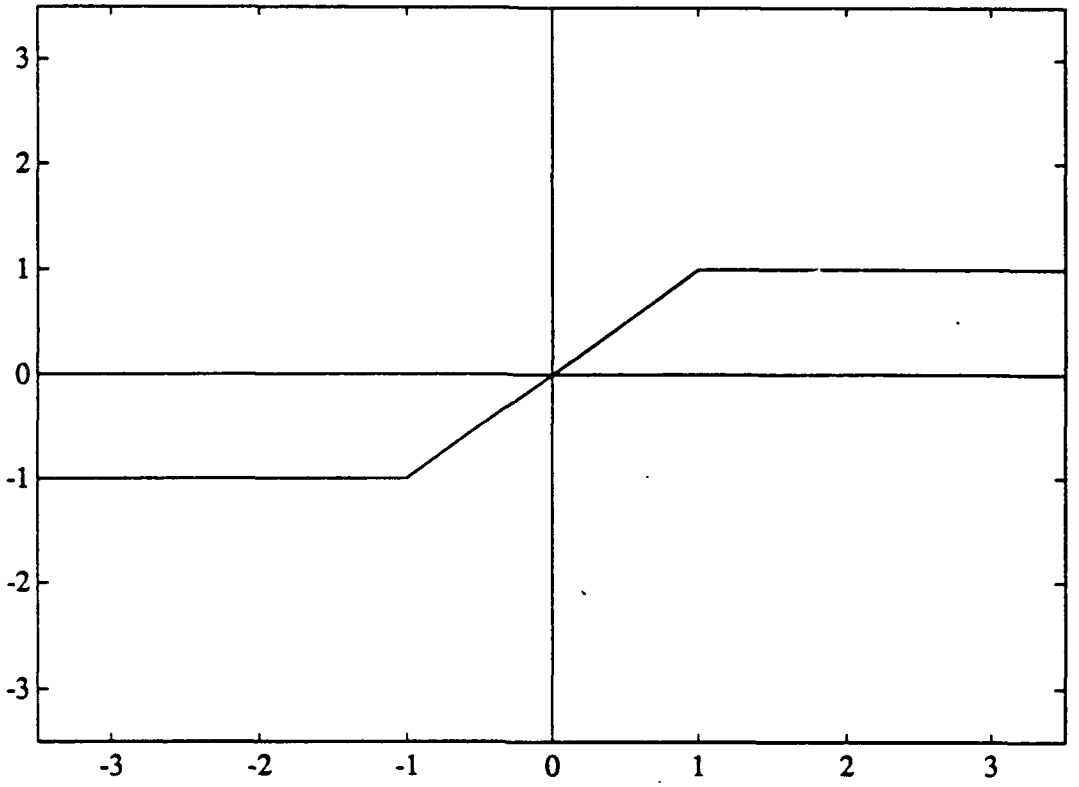


Figure 4.4 MAP Estimate Under Uniform Distribution

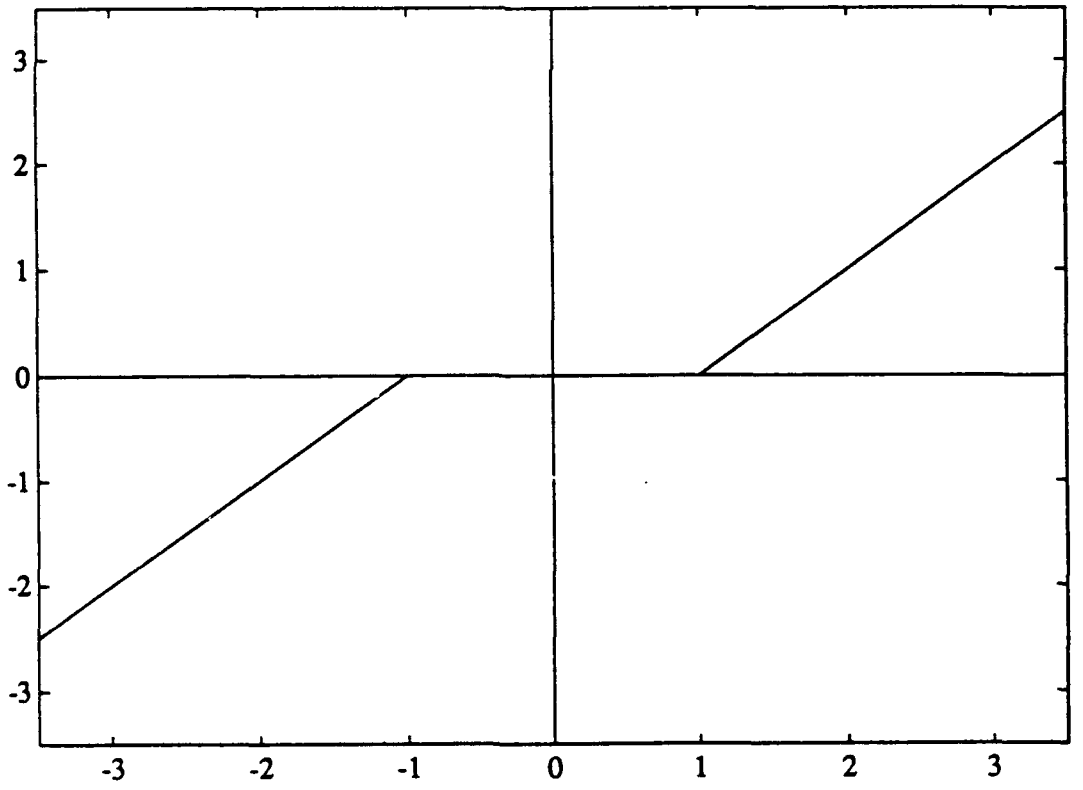
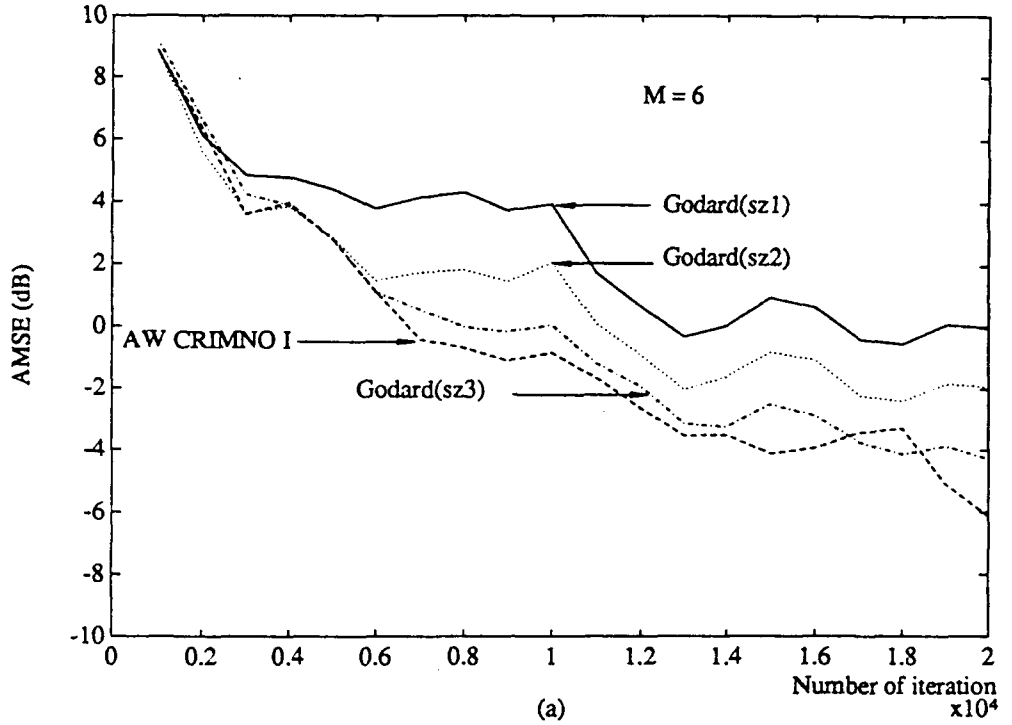


Figure 4.5 MAP Estimate Under Laplace Distribution



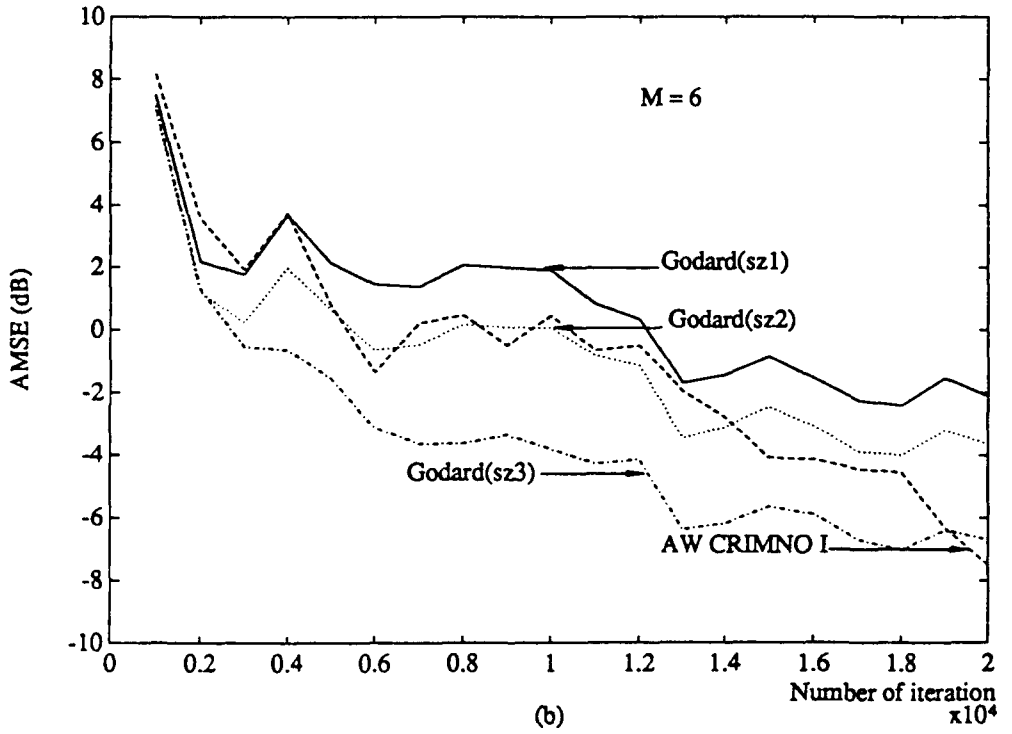
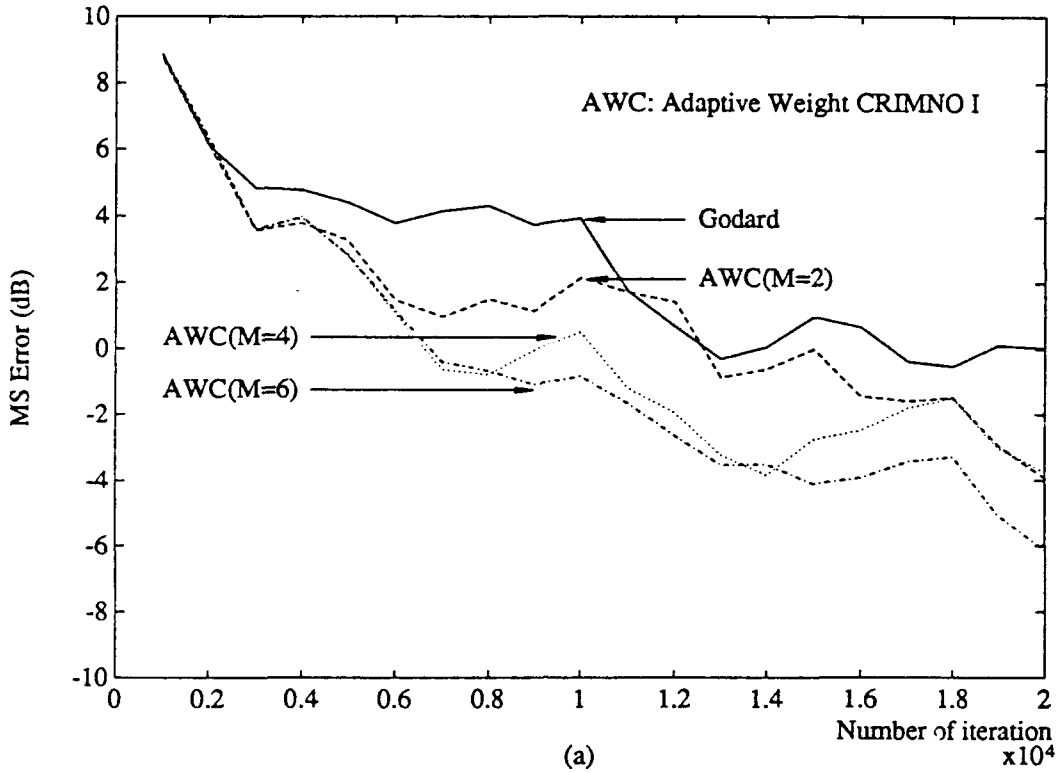
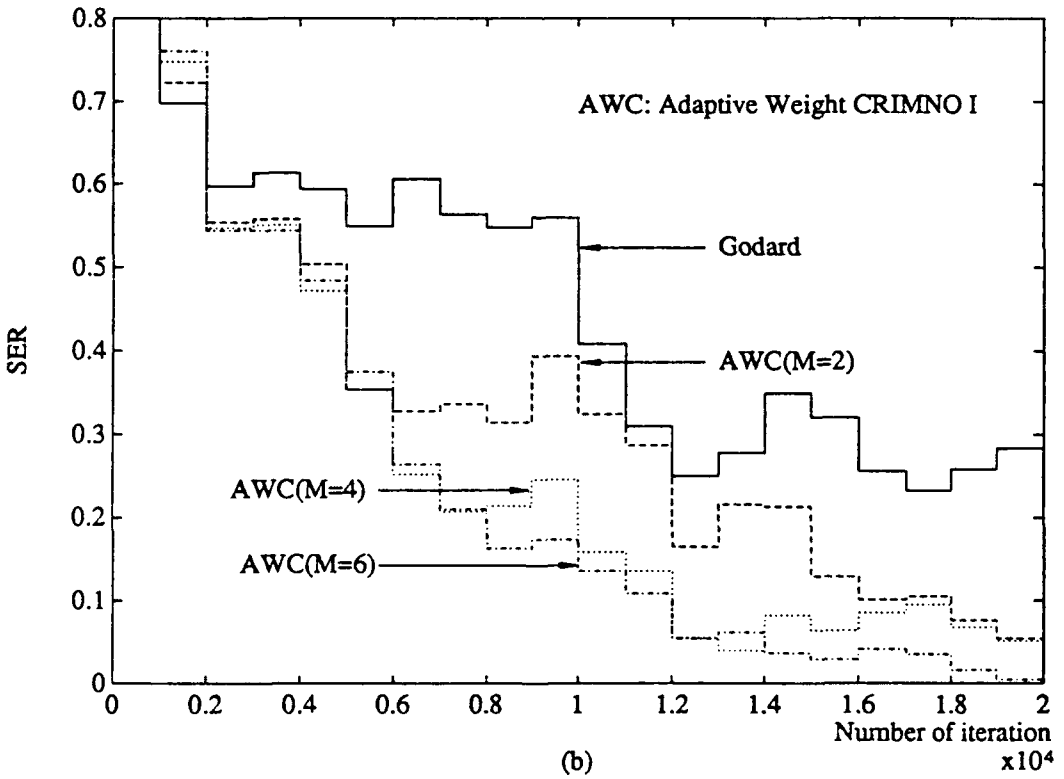


Figure 4.6 Comparison of the adaptive weight CRIMNO algorithm (szl) with Godard's algorithm of different step-size (sz3 is the optimum step-size): (a) the real channel; (b) the synthetic channel.


AWC: Adaptive Weight CRIMNO I
Godard
AWC(M=2)
AWC(M=4)
AWC(M=6)
MS Error (dB)
Number of iteration
x10^4

(a)


AWC: Adaptive Weight CRIMNO I
Godard
AWC(M=2)
AWC(M=4)
AWC(M=6)
SER
Number of iteration
x10^4

(b)

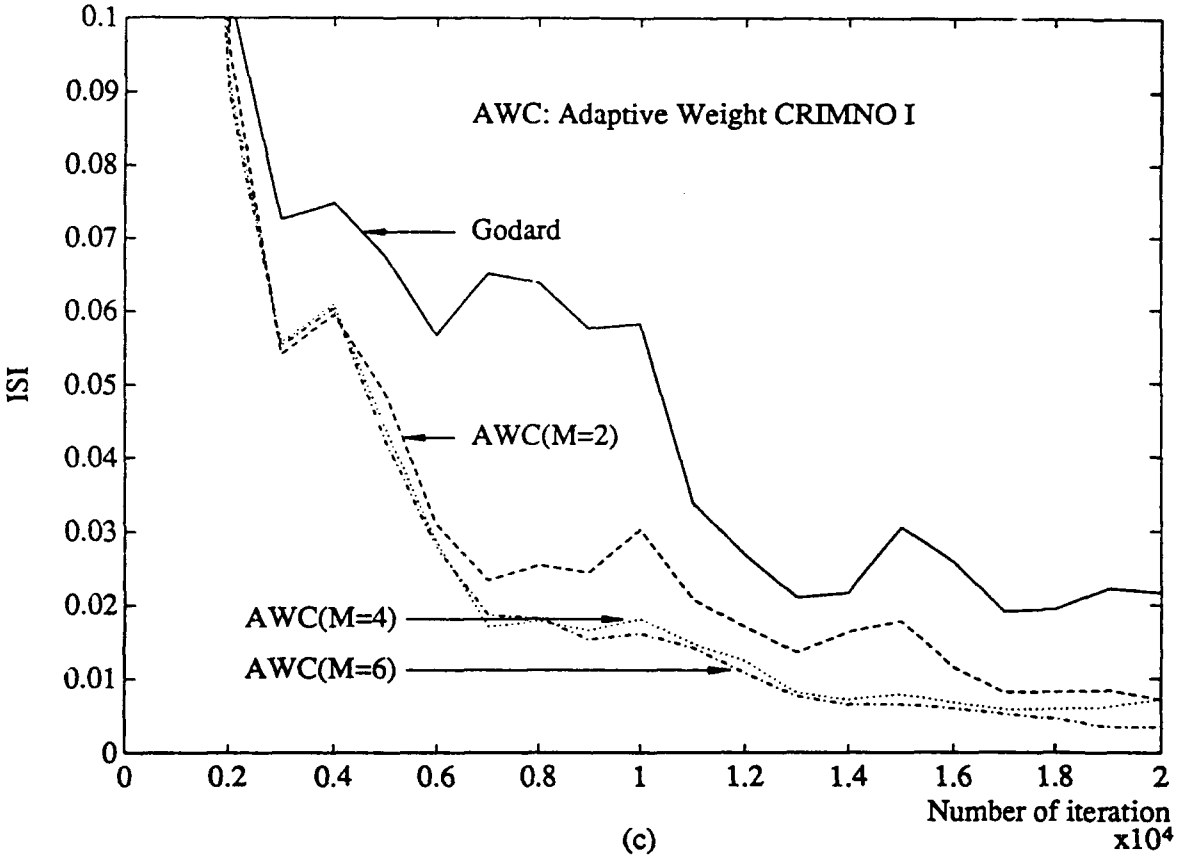


Figure 4.7 Effect of memory size M on the adaptive weight CRIMNO algorithm: (channel 2) (a) Mean square error; (b) Probability of error; (c) Intersymbol interference.

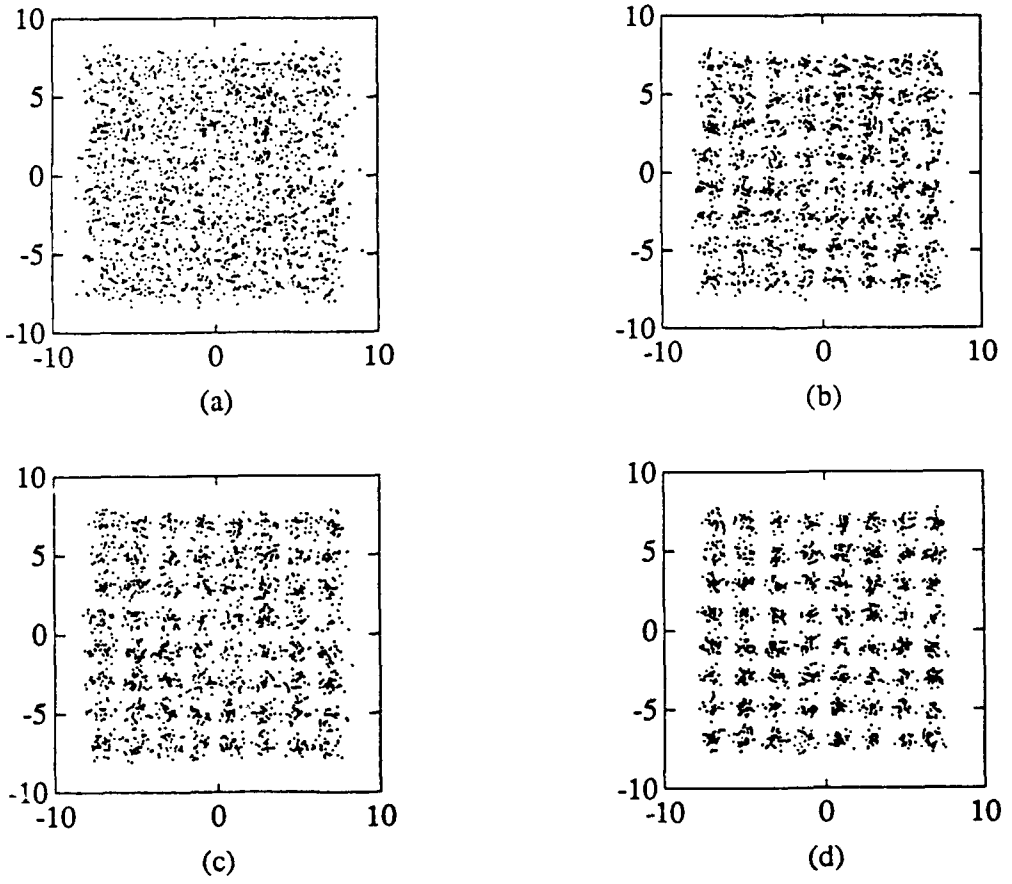


Figure 4.8 Eye pattern of adaptive weight CRIMNO algorithms with different memory size M at iteration 20000. (a) Godard; (b) Adaptive weight CRIMNO ($M=2$); (c) Adaptive weight CRIMNO ($M=4$); (d) Adaptive weight CRIMNO ($M=6$).

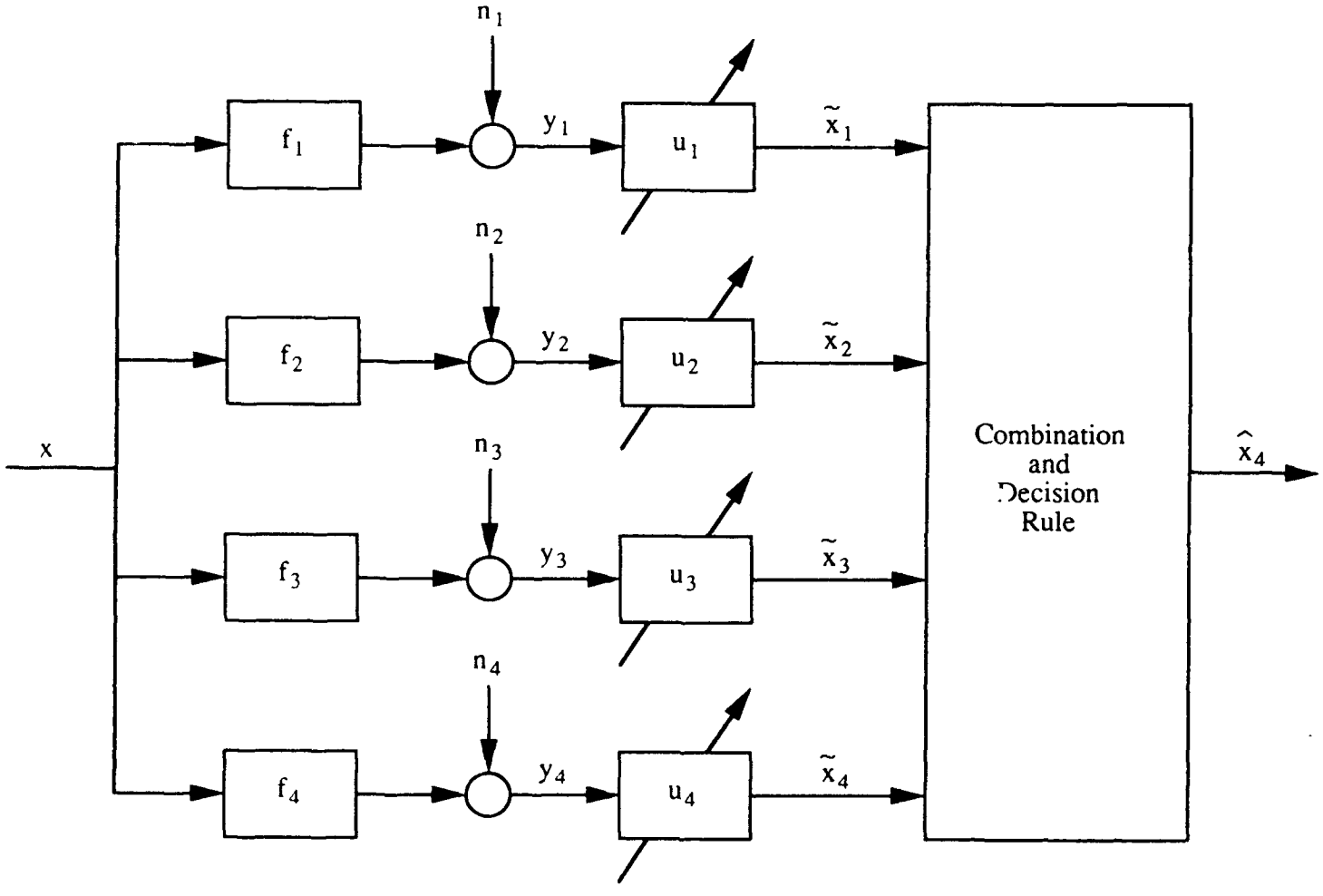


Figure 5.1 Diagram of four parallel equalized systems with additive noise.

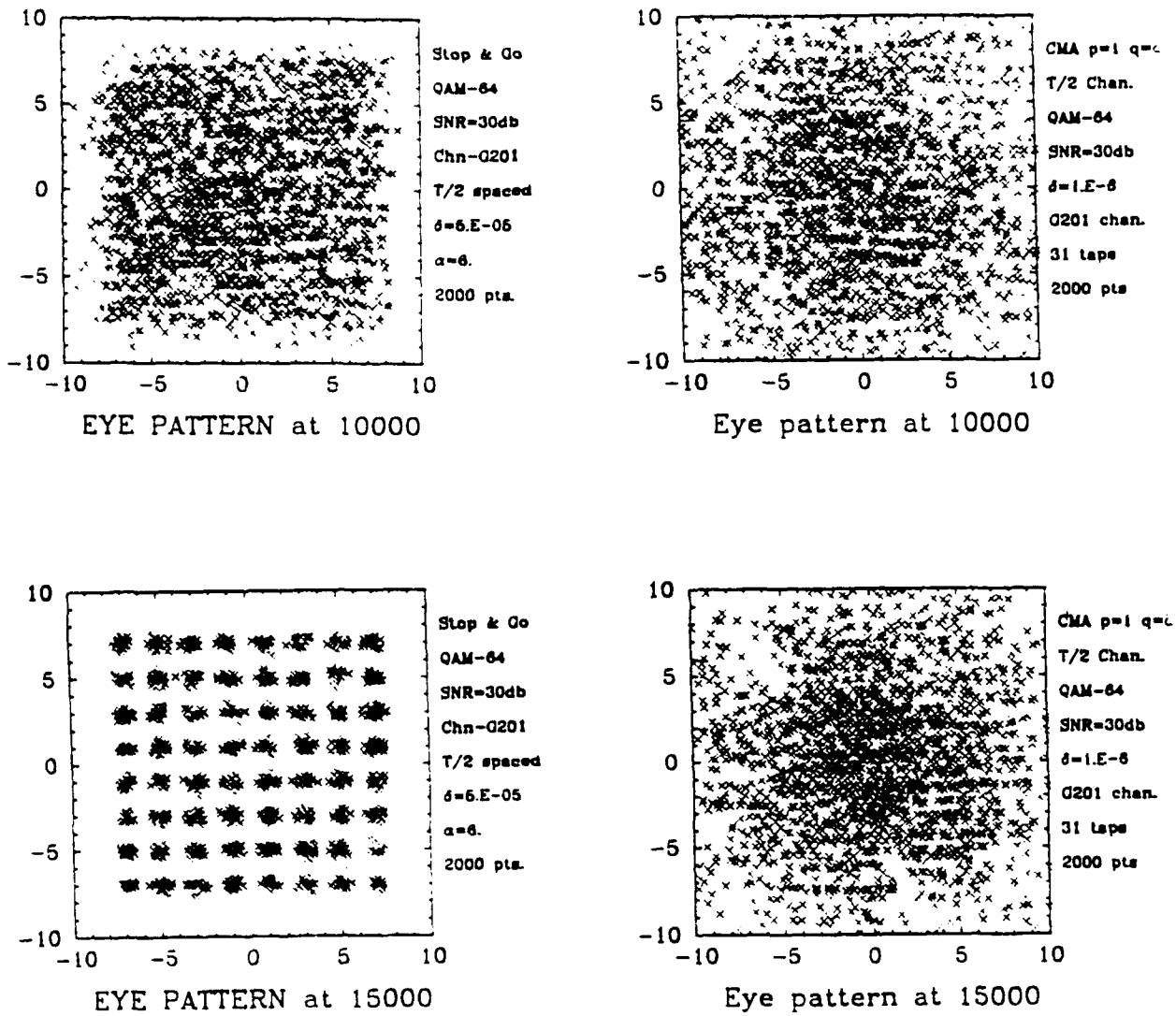


Figure 5.2 Channel G201 with QAM-64: Stop-and-Go and CMA (p=1, q=2) algorithms.

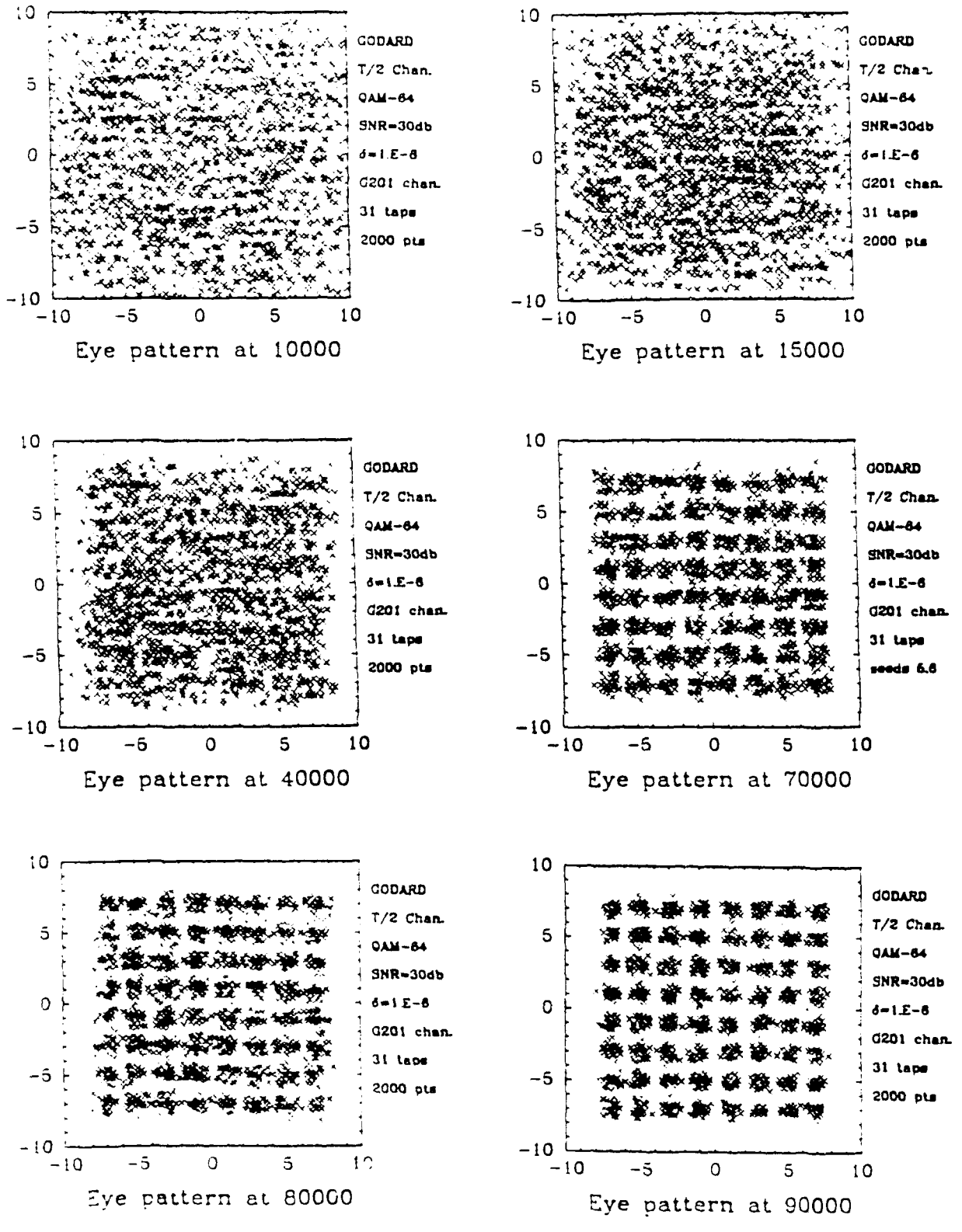


Figure 5.3 Channel G201 with QAM-64: Godard algorithms.

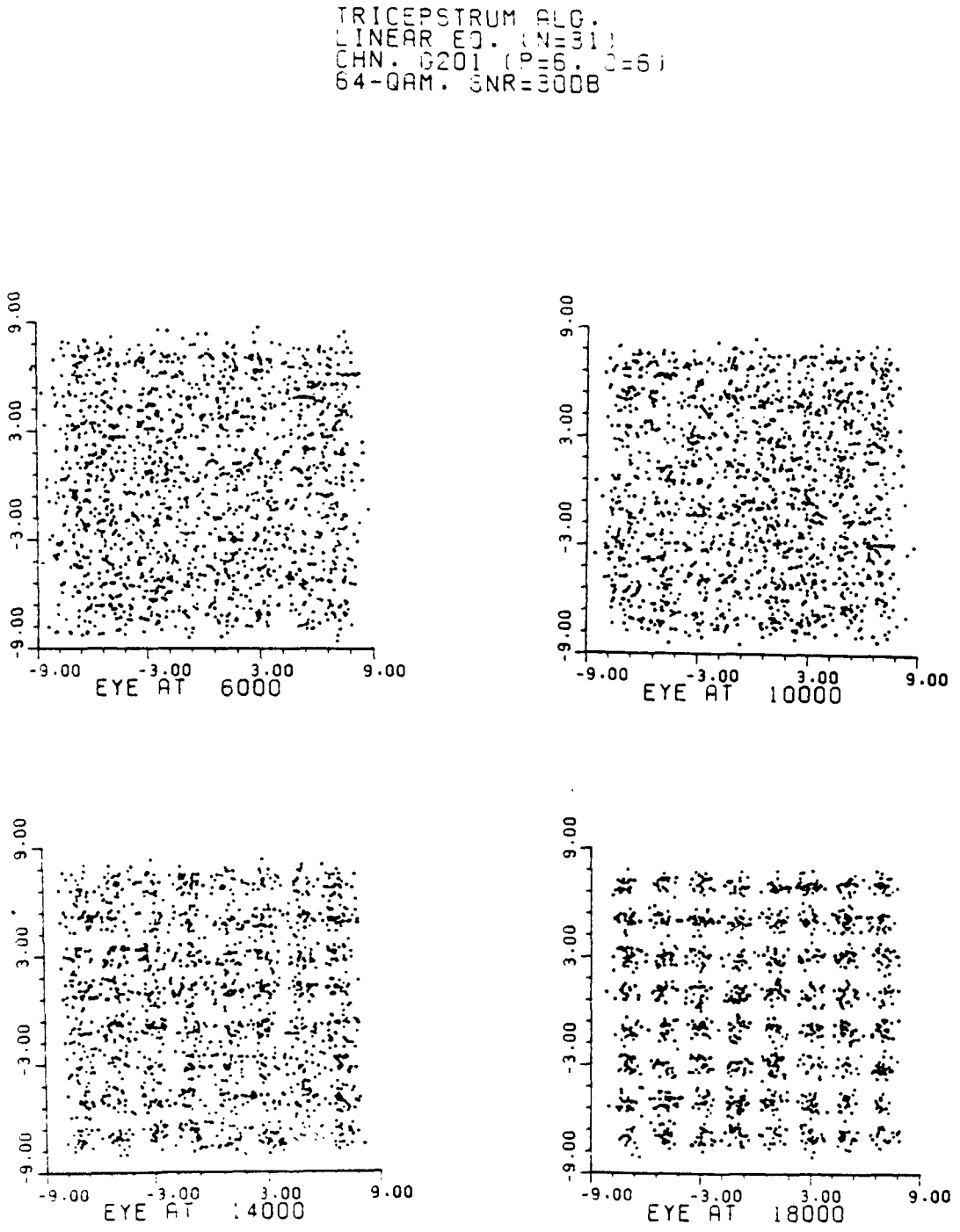


Figure 5.4 Channel G201 with QAM-64: TEA algorithms.

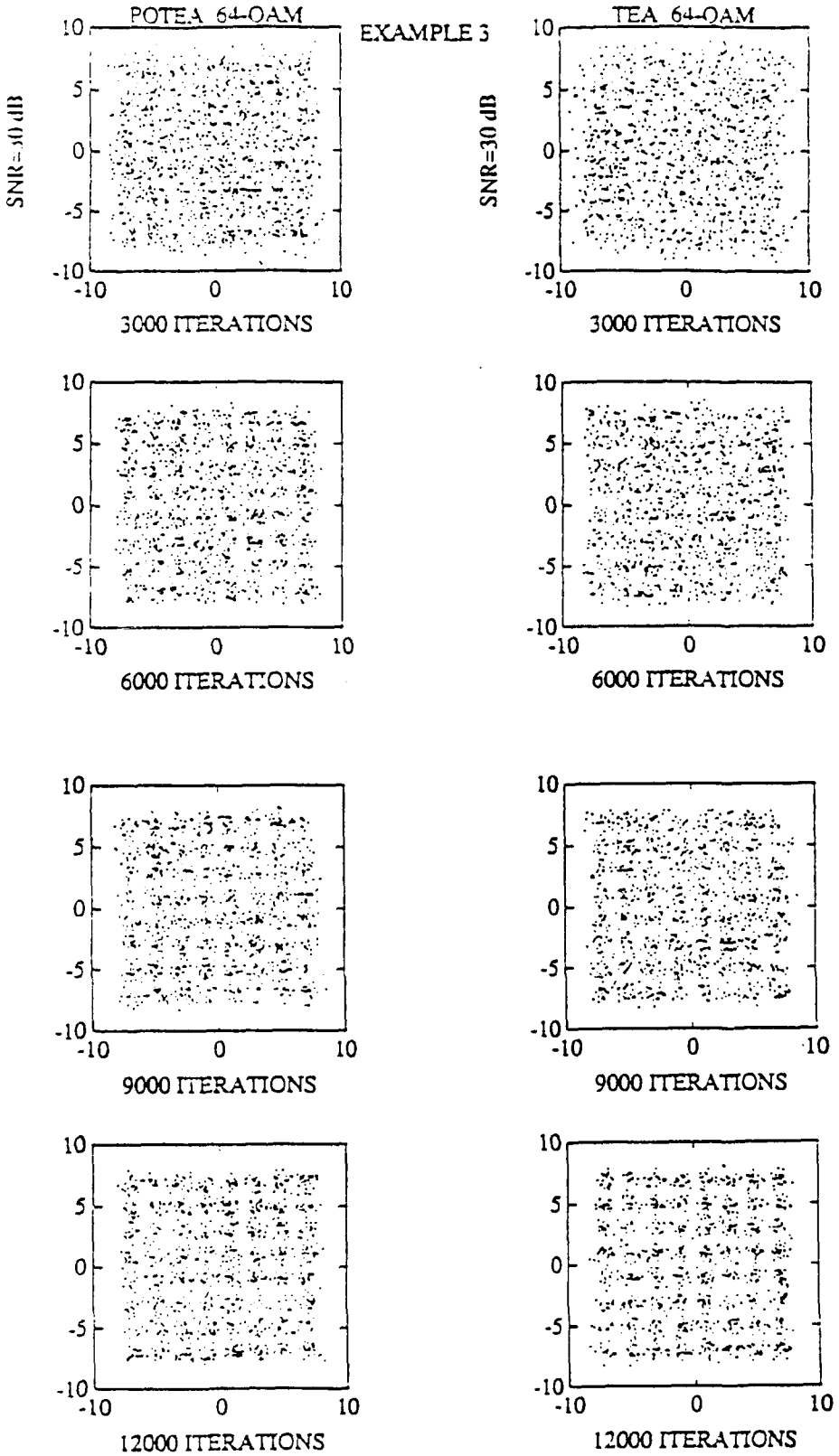


Figure 5.5 Eye-patterns of the Godard algorithm vs. those of the TEA algorithm.

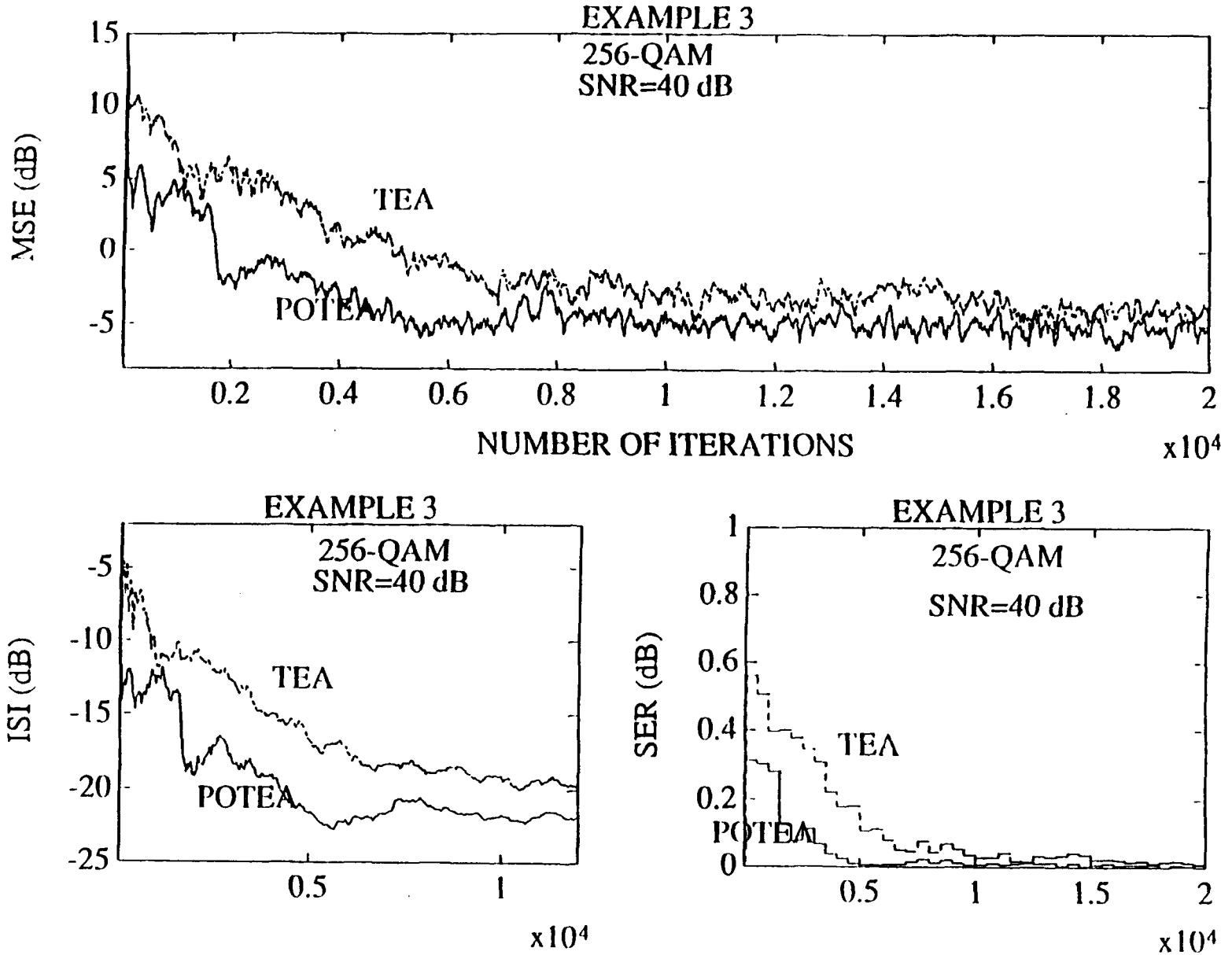


Figure 5.6 Performance comparison of the POTEA algorithm with the TEA algorithm.

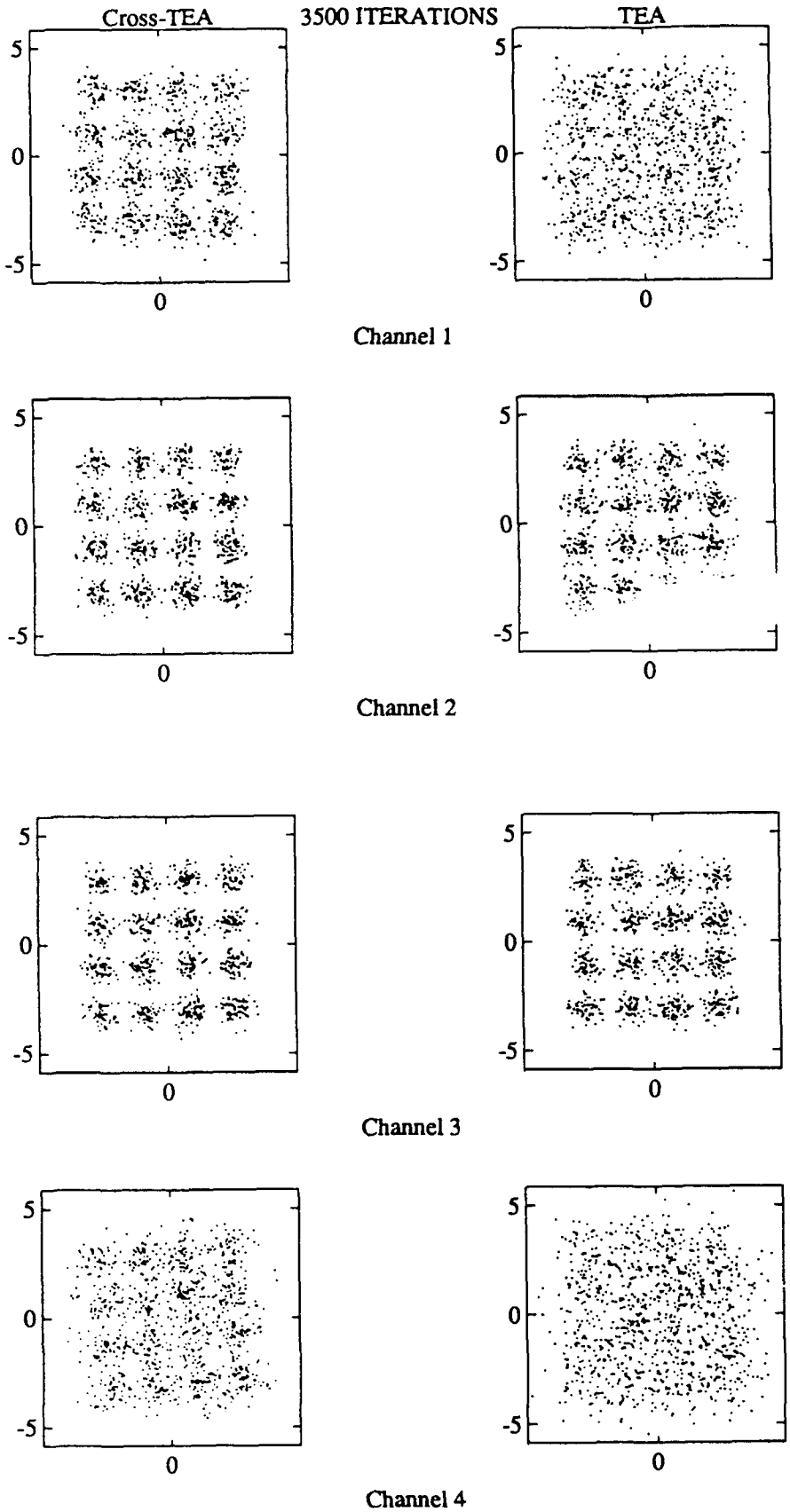


Figure 5.7 Eye Diagrams for CTEA and TEA at 3500 iterations..

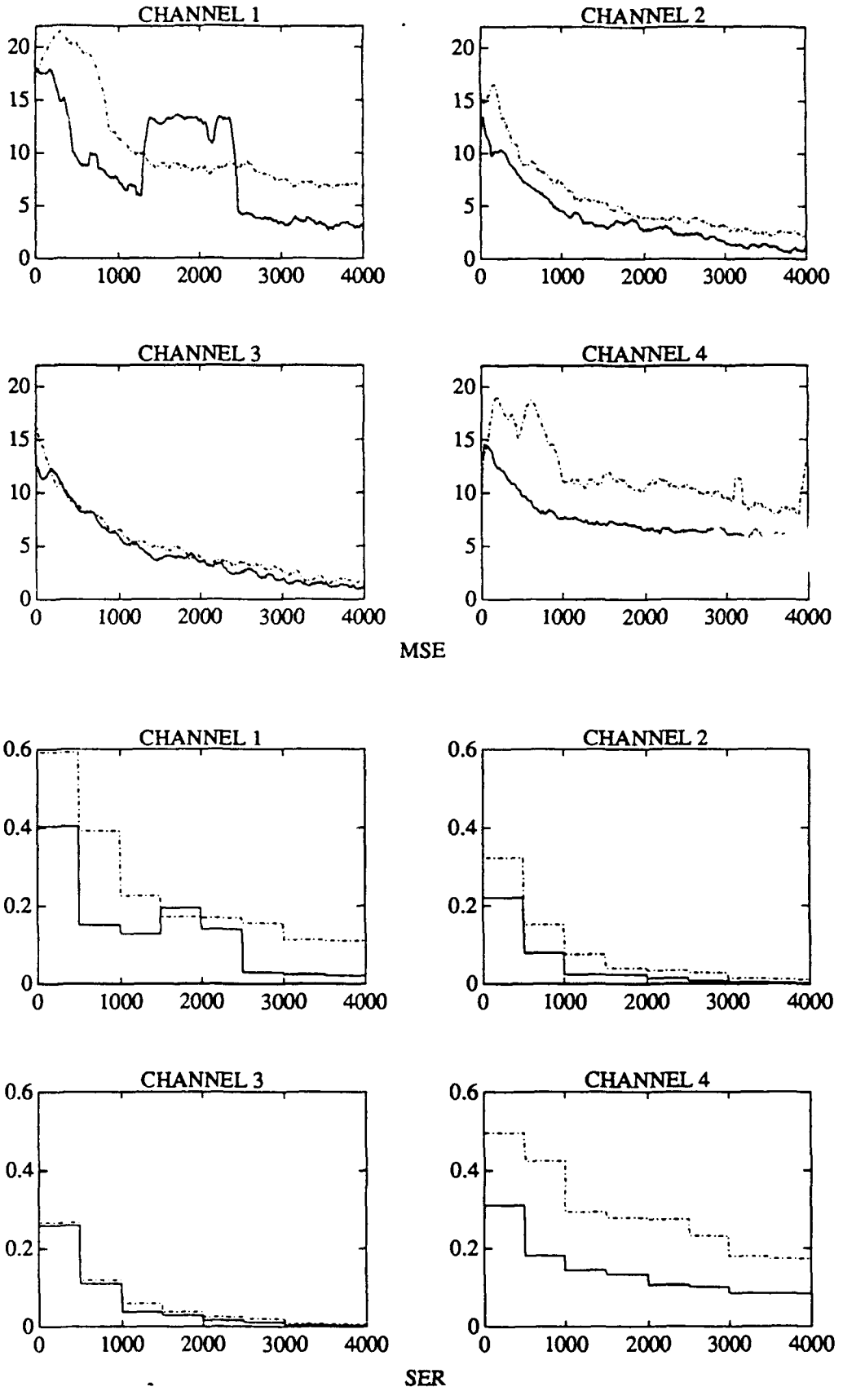


Figure 5.8 MSE and SER plots for all four channels, CTEA vs. TEA.

MOMENTS, CUMULANTS AND SOME APPLICATIONS TO STATIONARY RANDOM PROCESSES

BY DAVID R. BRILLINGER*

University of California, Berkeley

The paper ranges over some basic ideas concerning moments and cumulants, focusing on the case of random processes. Uses of moments and cumulants in developing large sample approximate distributions, in system identification and in inferring causal connections of a network of point processes are presented.

1. Introduction. Moments and cumulants find many uses in main stream statistics generally and with random processes particularly. Moments reflect the parameters of distributions and hence, as via the method of moments, may be used to estimate distributional parameters. Moments may be employed to develop approximations to the statistical distributions of quantities, such as sums in central limit theorems and associated expansions. Moments may be used to study the independence of variates. Moments unify diverse random processes, such as point processes and random fields, and diverse domains, such as the line or space-time.

2. Ordinary case. One can begin by asking: What is a moment? To provide an answer to this question, consider the case of the 0–1 valued

---

*Research partially supported by NSF Grant DMS-8900613

AMS 1980 subject classifications. Primary 62M10, 62M99.

Key words and phrases. Coherence, cumulant, moment, partial coherence, point process, system identification, time series.

variates X, Y, Z. For these variates

$$E\{XYZ\} = Prob\{X = 1, Y = 1, Z = 1\}$$

This provides an interpretation for a (third-order) moment in terms of a quantity having a primitive existence, namely a probability. Higher-order moments have a similar interpretation. One can proceed to general random variables, by noting that these may be approximated by step (or simple) functions, see eg. Feller (1966), page 107.

Next one can ask: What is a cumulant? One answer is to say that it is a combination of moments that vanishes when some subset of the variates is independent of the others. Suppose for example that X is independent of (Y, Z). The third order joint cumulant may be defined by

$$cum\{X, Y, Z\} = \tag{1}$$

$$E\{XYZ\} - E\{X\}E\{YZ\} - E\{Y\}E\{XZ\} - E\{Z\}E\{XY\} + 2E\{X\}E\{Y\}E\{Z\}$$

By substitution one quickly sees that this last expression vanishes in the case that X is independent of (Y, Z).

Expresion (1) gives one definition of a joint cumulant. An alternate way to proceed is to state that that cumulant is given by the coefficient of $i^3\alpha\beta\gamma$ in the Taylor expansion of

$$\log\left[E\{e^{i(\alpha X+\beta Y+\gamma Z)}\}\right]$$

supposing one exists.

Taking the log here converts factorizations into additivities and one sees immediately why the joint cumulants vanish in the case of independence.

Streitberg (1990) sets down a sequence of conditions that actually characterize a cumulant. These are:

1. Symmetry

$$cum\{X_1, X_2, \cdots\} = cum\{X_2, X_1, \cdots\}$$

2. Multilinearity

$$cum\{\alpha X_1, X_2, \cdots\} = \alpha cum\{X_1, X_2, \cdots\}$$

$$cum\{X_1+Y_1, X_2, \cdots\} = cum\{X_1, \cdots\} + cum\{Y_1, \cdots\}$$

3. Moment property, if the moments of **X** and **Y** are identical up to order k

$$cum\{\mathbf{X}\} = cum\{\mathbf{Y}\}$$

4. Normalization, in the expansion in terms of moments

$$cum\{X_1, \cdots, X_k\} = E\{X_1 \cdots X_k\} + \cdots$$

5. Interaction, if a subset is independent of the remainder

$$cum\{X_1, \cdots, X_k\} = 0$$

Cumulants provide a measure of Gaussianity. If the variate X is normal, then

$$cum_k\{X\} = 0 \tag{2}$$

for $k > 2$. (Here cum_k denotes the joint cumulant of X with itself k times.) Putting (2) together with the fact that the normal distribution is determined by its moments, provides a particularly brief proof of the central limit theorem. Namely suppose that $X_1, X_2, \cdots$ are independent and identically distributed with $E\{X\} = 0$ and $var\{X\} = 1$. Suppose all moments exist for X. Consider

$$S_n = (X_1 + \cdots + X_n)/\sqrt{n} \tag{3}$$

Then

$$cum_k\{S_n\} = n\ cum_k\{X\} / n^{\frac{k}{2}}$$

which tends to 0 for $k > 2$, as n tends to infinity, and in consequence S_n has a limiting normal distribution.

An error bound may be given for the degree of approximation of the distribution of a random variable by a normal, via bounds on the cumulants. In Rudzkis *et al. (1978)* the following result is developed. Consider a variate Y with mean 0 and variance 1. Suppose that

$$|cum_k\{Y\}| \leq \frac{H(k!)^{1+\nu}}{\Delta^{k-2}}$$

for some $\nu \geq 0$, $H \geq 1$, then in the interval $0 \leq u \leq \delta/H$

$$\sup_u |Prob\{Y < u\} - \Phi(u)| \leq \frac{18H}{\delta}$$

where

$$\delta = \frac{1}{7}\left[\frac{\sqrt{2}\Delta}{6}\right]^{1/(1+2\nu)}$$

In the case of a sum, such as (3), one can take $\Delta = \sqrt{n}$ for example.

3. Time series case. Consider a stationary time series $X(t)$ with domain $t = 0, \pm 1, \pm 2, \cdots$. If the $k-th$ moment exists, from the stationarity, the moment function

$$E\{X(t+u_1)\cdots X(t+u_{k-1})X(t)\}$$

will not depend on t, nor will the associated cumulant function

$$c_k(u_1, \cdots, u_{k-1})$$

$$= cum\{X(t+u_1), \cdots X(t+u_{k-1}), X(t)\} \tag{4}$$

The Fourier transforms of these $c_k(.)$ give the higher-order spectra of the series. These functions may be estimated given stretches of data.

It was indicated, by property 5 above, that a joint cumulant measures statistical dependence. This suggests formalizing the intuitive notation that values at a distance in time are not strongly dependent via

$$\sum_{u_1} \cdots \sum_{u_{k-1}} |c_k(u_1, \cdots, u_{k-1})| < \infty \tag{5}$$

for $k = 2, \cdots$. It is now direct to provide a central limit theorem for sums of values of a stationary time series. One has

$$cum_k \{\sum_1^T X(t) / \sqrt{T}\}$$

$$= \sum_{t_1} \cdots \sum_{t_k} c_k(t_1 - t_k, \cdots, t_{k-1} - t_k) / T^{k/2}$$

$$= \sum_t \left[\sum_{u_1} \cdots \sum_{u_{k-1}} c_k(u_1, \cdots, u_{k-1})\right] / T^{k/2}$$

$$\simeq \sum_u c_2(u) \qquad k = 2$$

and

$$\to 0 \qquad k > 2$$

following (5), giving the limit normal distribution.

Another aspect of the use of cumulants is that a calculus exists for manipulating polynomials in basic variates. Suppose that

$$Y = g(X_1, \cdots, X_L)$$
$$= \sum_{\mathbf{i}} \alpha_{i_1 \cdots i_L} X_1^{i_1} \cdots X_L^{i_L} \tag{6}$$

One has directly from (6) that

$$E\{Y^k\} = \sum_{\mathbf{m}} \beta_{m_1 \cdots m_L} E\{X_1^{m_1} \cdots X_L^{m_L}\}$$

but perhaps more usefully, there are rules due to Fisher, see Leonov and Shiryaev (1959), Speed (1983), providing an expression

$$cum_k\{Y\} = \sum_{\sigma} \gamma_\sigma \, cum\{X_j : j \in \sigma_1\} \cdots cum\{X_j : j \in \sigma_p\}$$

where $\sigma = (\sigma_1, \cdots, \sigma_p)$ is a partition of subscripts into blocks and the γ_σ are coefficients.

A time series analog of an expansion, like (6) for ordinary variates, is

provided by the Volterra expansion

$$Y(t) = a_0 + \sum_u a_1(t-u)X(u) + \sum_{u_1,u_2} a_2(t-u_1,t-u_2)X(u_1)X(u_2) + \cdots \quad (7)$$

Using the Cramer representation of the process, namely

$$X(t) = \int e^{it\lambda} dZ_X(\lambda)$$

(7) may be written

$$a_0 + \int e^{it\lambda} A_1(\lambda) dZ_X(\lambda) + \iint e^{it(\lambda_1+\lambda_2)} A_2(\lambda_1,\lambda_2) dZ_X(\lambda_1) dZ_X(\lambda_2) + \cdots$$

in terms of the Fourier transforms of the $a_1(.), a_2(.), \cdots$. This form often simplifies the development of particular analytic results.

Consideration now turns to the use of moments and cumulants in the identification of nonlinear systems. In the case of a polynomial system like (7), Lee and Schetzen (1965) develop estimates of the functions $a_1(.), a_2(.), \cdots$ via empirical moments of the form

$$\frac{1}{T} \sum_{t=0}^{T-1} X(t+u_1) \cdots X(t+u_k) Y(t)$$

for the case that the input, $X(.)$, is Gaussian white noise.

For the case of stationary Gaussian input and a quadratic system

$$Y(t) = a_0 + \sum_u a_1(t-u)X(u) + \sum_{u_1,u_2} a_2(t-u_1,t-u_2)X(u_1)X(u_2) + noise$$

Tick (1961) developed an estimation procedure as follows. Define the cross-spectrum and cross-bispectrum via

$$cum\{dZ_X(\lambda), dZ_Y(\mu)\} = \delta(\lambda+\mu) f_{XY}(\lambda) d\lambda d\mu$$

$$cum\{dZ_X(\lambda_1), dZ_X(\lambda_2), dZ_Y(\lambda_3)\} = \delta(\lambda_1+\lambda_2+\lambda_3) f_{XXY}(\lambda_1,\lambda_2) d\lambda_1 d\lambda_2 d\lambda_3$$

respectively. One has

$$f_{YX}(\lambda) = A_1(\lambda) f_{XX}(\lambda)$$

$$f_{XXY}(-\lambda_1,-\lambda_2) = 2A_2(-\lambda_1,-\lambda_2) f_{XX}(\lambda_1) f_{XX}(\lambda_2)$$

relations from which estimates of the transfer functions, A, may be

developed, based on estimates of the spectra that appear.

Another system that may be identified, in a like manner, takes the form, for input $X(.)$ and output $Y(.)$,

$$U(t) = \sum_u a(t-u)X(u)$$

$$V(t) = G[U(t)]$$

$$Y(t) = \mu + \sum_u b(t-u)V(u) + noise$$

i.e. involves an instantaneous nonlinearity, $G[.]$, and two linear filters. In the case that $X(.)$ is stationary Gaussian, one can develop the relationships

$$f_{YX}(\lambda) = L_1 A(\lambda)B(\lambda)f_{XX}(\lambda)$$

$$f_{XXY}(\lambda_1,\lambda_2) = L_2 A(-\lambda_1)A(-\lambda_2)B(-\lambda_1-\lambda_2)f_{XX}(\lambda_1)f_{XX}(\lambda_2)$$

where L_1, L_2 are constants. See Korenberg (1973) and Brillinger (1977). Estimates of the identifiable unknowns may be developed based on estimates of the spectra appearing.

4. Point process case. Consider isolated points, τ_k, scattered along the real line. Let $N(t)$ count the number in $(0,t]$ and $dN(t)$ the number in the small interval $(t,t+dt]$. Typically $dN(t)$ will be 0 or 1.

The $k-th$ order product density of the point process $N(.)$ is $p_k(.)$ given by

$$E\{dN(t_1) \cdots dN(t_k)\}$$

$$= Prob\{dN(t_1)=1, \cdots, dN(t_k)=1\}$$

$$= p_k(t_1, \cdots, t_k)dt_1 \cdots dt_k$$

for $t_1, \cdots, t_k$ distinct and $k = 1, 2, \cdots$. This relates to the moments

of the process as follows. Write $N^{(k)} = N(N-1) \cdots (N-k+1)$, then the $k-th$ factorial moment of $N(t)$ is

$$E\{N(t)^{(k)}\} = \int_0^t \cdots \int_0^t p_k(t_1, \cdots, t_k) dt_1 \cdots dt_k$$

The corresponding cumulant density is given by

$$cum\{dN(t_1), \cdots, dN(t_k)\} = q_k(t_1, \cdots, t_k) dt_1 \cdots dt_k$$

for $t_1, \cdots, t_k$ distinct. The $k-th$ factorial cumulant of $N(t)$ is now

$$\int_0^t \cdots \int_0^t q_k(t_1, \cdots, t_k) dt_1 \cdots dt_k$$

In the case of a Poisson process, the product densities will be given by

$$p_k(t_1, \cdots, t_k) = p(t_1) \cdots p(t_k)$$

with $p(t)$ the intensity of the process and the cumulant densities will vanish for $k > 1$.

As an example of the use of moments to derive an alternate limit theorem, suppose one has $N_1(.), \cdots, N_n(.)$ i.i.d.copies of a point process $N(.)$. Suppose they are superposed and rescaled to form the point process

$$M_n(t) = N_1(\frac{t}{n}) + \cdots + N_n(\frac{t}{n})$$

The $k-th$ factorial cumulant of this process is

$$\int_0^{t/n} \cdots \int_0^{t/n} n\, q_k(t_1, \cdots, t_k) dt_1 \cdots dt_k$$
$$\approx n(\frac{t}{n})^k q_k(0, \cdots, 0)$$

for large n, assuming continuity at 0. This cumulant tends to $tq_1(0)$ for $k = 1$ and to 0 for $k > 1$ and in consquence one has a Poisson limit for the variate $M_n(t)$.

5. Extensions. The preceding results and definitions extend quite directly to the cases of: a spatial process $X(x,y)$, a marked point process

$\sum_j M_j \ \delta(t-\tau_j)$, a hybrid process $X(\tau_j)$ and a line process, for example.

6. An example. In this section second-order moments and cumulants are employed to infer the causal connections amongst some contemporaneous point processes.

Consider the stationary bivariate point process (M, N) with points τ_k and γ_l respectively. In what follows an estimate of the product density of order 2 will be needed. The parameter is defined via

$$p_{MN}(u)\ dudt = E\{dM(t+u)dN(t)\}$$

$$= Prob\{dM(t+u) = 1,\ dN(t) = 1\}$$

This last suggests basing an estimate on the count

$$\#\{|\tau_k - \gamma_l - u| < \frac{h}{2}\} \tag{8}$$

for some small binwidth h. Details are given in Brillinger (1976). One result is that it appears more pertinent to graph the square root of the estimate. In the case that the processes M and N are independent, one will have $p_{MN}(u) = p_M p_N$, which possibility may be examined via the statistic (8).

The suggested estimate will be illustrated with some neurophysiological data. Concern in the experiment was with auditory paths in the brain of the cat. To collect data, microelectrodes were inserted with location tuned to sound response. Data was recorded when the neurons were firing spontaneously. Also responses were evoked experimentally by 200 msec. noise bursts, that were applied every 1000 msec., via speakers inserted in the ears. The firing times of 8 neurons were recorded. Figure 1 provides the data itself for 4 selected cells, 2 in the case with stimulation, 2 when the firing is spontaneous. Each horizontal line plots firings as a function

of time since stimulus initiation in a 1000 msec. time period. The stimulus was applied 505 times in these examples. In the stimulated case one notices vertical darkening corresponding to excess firing just after the stimulus has been applied. Neurophysiologists speak of locking. In the spontaneous case no locking is apparent. There is some evidence of non-stationarity in this case.

Figure 2 provides the square root of a multiple of (8). The horizontal dashed lines are ±2 standard errors about a horizontal line corresponding to independence in the stationary case. One infers that the cell pairs are associated in each case. However in the stimulated case one has to wonder if the apparent association of units 6 and 7 is not due to the fact that the cells are being stimulated at the same times.

Fourier techniques provide one means to address this concern. Write

$$d_M^T(\lambda) = \sum_k e^{-i\lambda\tau_k}$$
$$d_N^T(\lambda) = \sum_l e^{-i\lambda\gamma_l}$$

for the data $0 \leq \tau_k, \gamma_l < T$. For $\lambda \neq 0$ one has

$$E\{d_M^T(\lambda)\overline{d_N^T(\lambda)}\} \simeq 2\pi T\ f_{MN}(\lambda)$$

with $f_{MN}(.)$ the cross-spectrum given by

$$f_{MN}(\lambda) = \frac{1}{2\pi}\int e^{-i\lambda u} q_{MN}(u)\, du$$

A useful quantity for measuring the association of M and N may now be defined. It is the coherence,

$$|R_{MN}(\lambda)|^2 = |f_{MN}(\lambda)|^2 / f_{MM}(\lambda) f_{NN}(\lambda)$$

with the interpretation

$$\lim_{T\to\infty} |corr\{d_M^T(\lambda), d_N^T(\lambda)\}|^2$$

It satisfies $0 \leq |R_{MN}(\lambda)|^2 \leq 1$, with greater association corresponding to

values nearer 1. Figure 3 provides coherence estimates for the cell pairs of Figure 2. This evidence of association is in accord with that of Figure 2. The dashed horizontal line provides the 95% point of the null distribution of the coherence estimate.

To return to the driving question of how to "remove" the effects of the stimulus, one can consider the partial coherence. This has the interpretation

$$\lim_{T\to\infty} |corr\{d_M^T - \alpha d_S^T, d_N^T - \beta d_S^T\}|^2$$

with α, β regression coefficients and S referring to the process of stimulus times. Suppressing the dependence on λ the partial coherence is given by $|R_{MN|S}|^2$ where

$$R_{MN|S} = \frac{R_{MN} - R_{MS}R_{SN}}{\sqrt{(1-|R_{MS}|^2)(1-|R_{NS}|^2)}}$$

Figure 4 provides the estimated partial coherence of neurons 6 and 7 in the stimulated case. The level apparent in the top graph of Figure 3 has fallen off substantially suggesting that the association evidenced in Figures 2 and 3 is due to the stimulus.

For interests sake Figure 5 provides the coherence estimate for neurons 3 and 4 in the case of applied stimulation. One might wonder if they would become more strongly associated in the presence of stimulation. The results do not suggest that this has happened.

7. Conclusions. In summary, moments and cumulants may be employed to develop approximations to distributions, approximations such as the normal or the Poisson. They may be employed in system identification. They may be used to infer the "wiring" diagram of a collection of interacting point processes.

The approach presented is nonparametric, not based on special stochastic processes described by finite dimensional parameters. Brillinger (1991) provides a variety of references concerning the work pre 1980 on higher moments and spectra.

Acknowledgements. The neurophysiological data were provided by Alessandro Villa. Terry Speed mentioned the Streitberg (1990) result.

REFERENCES

BRILLINGER, D. R. (1976). Estimation of the second-order intensities of a bivariate stationary point process. J. Roy. Statist. Soc. B 38, 60-66.

BRILLINGER, D. R. (1977). The identification of a particular nonlinear time series system. Biometrika 64, 509-515.

BRILLINGER, D. R. (1991). Some history of the study of higher-order moments and spectra. Statistica Sinica 1, 465-476.

BRILLINGER, D. R. (1992). Nerve cell spike train data analysis: a progression of technique. J. Amer. Statist. Assoc. 87, June.

FELLER, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II. New York, Wiley.

KORENBERG, M. J. (1973). Cross-correlation analysis of neural cascades. In Proc. 10th Annual Rocky Mountain Biomed. Symp. 47-51. Instrument Society of America, Pitttsburgh.

LEE, Y. W. and SCHETZEN, M. (1965). Measurement of the Wiener kernels of a nonlinear system by crosscorrelation. Internat. J. Control 2, 237-254.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. Theor. Prob. Appl. 4, 319-329.

RUDZKIS, R., SAULIS, L. and STATULEVICIUS, V. (1978). A general lemma on probabilities of large deviations. Liet. Mat. Rink. 18, 99-116.

SPEED, T. P. (1983). Cumulants and partition lattices. Austral. J. Statist. 25, 378-388.

STREITBERG, B. (1990). Lancaster interactions revisited. Ann. Statist. 18, 1878-1885.

TICK, L. (1961). The estimation of transfer functions of quadratic systems. Technometrics 3, 563-567.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CA 94720

Figure Legends

Figure 1. Rastor plot of the firing times of 4 neurons in successive 1000 msec. periods. There are 505 horizontal lines of firing times.

Figure 2. The square root of a multiple of the quantity (6). Were the processes independent and stationary then about 5% of the values should lie outside the band defined by the two horizontal dashed lines.

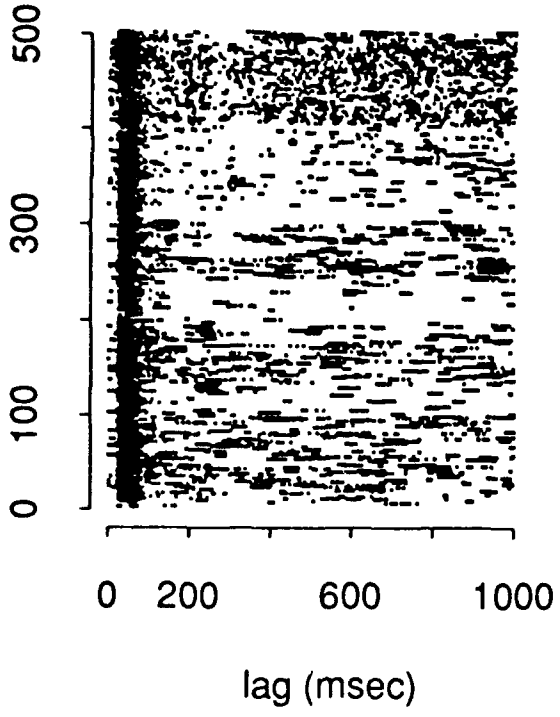
Figure 3. Estimated coherences of cells 6 and 7 in the stimulated case and 3 and 4 when the firing is spontaneous.

Figure 4. Estimated partial coherence of cells 6 and 7 "removing" the effect of the stimulus.

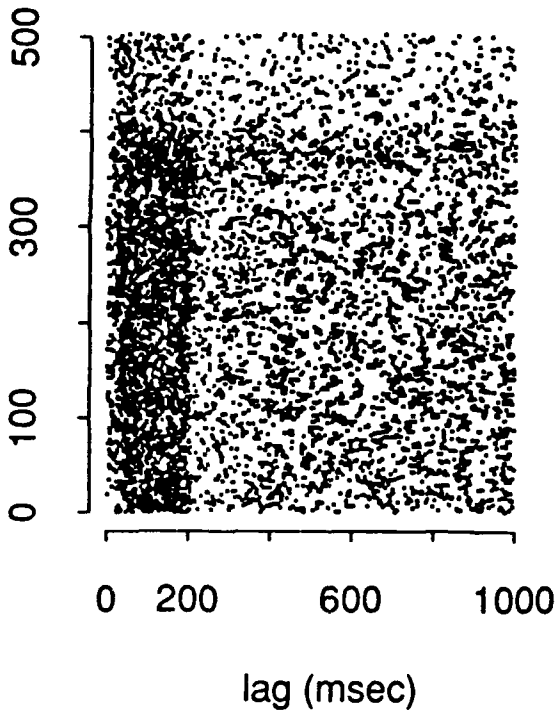
Figure 5. The estimated coherence of cells 3 and 4 in the case of stimulation.

Spike times following stimulus

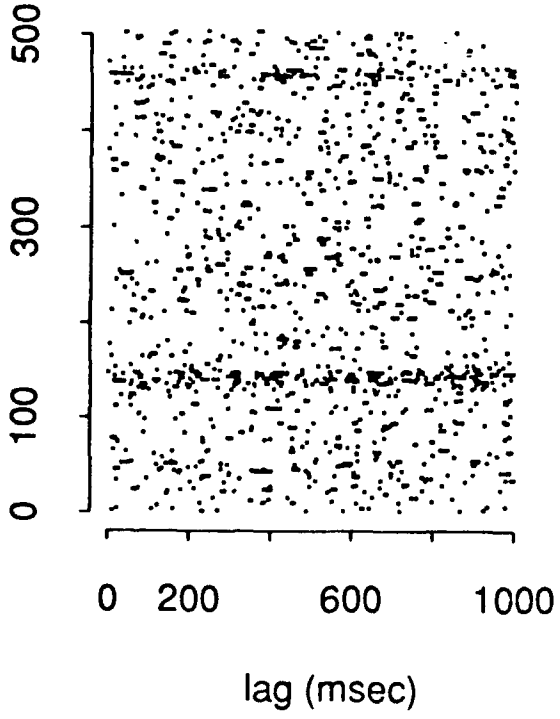
Stimlated unit 6



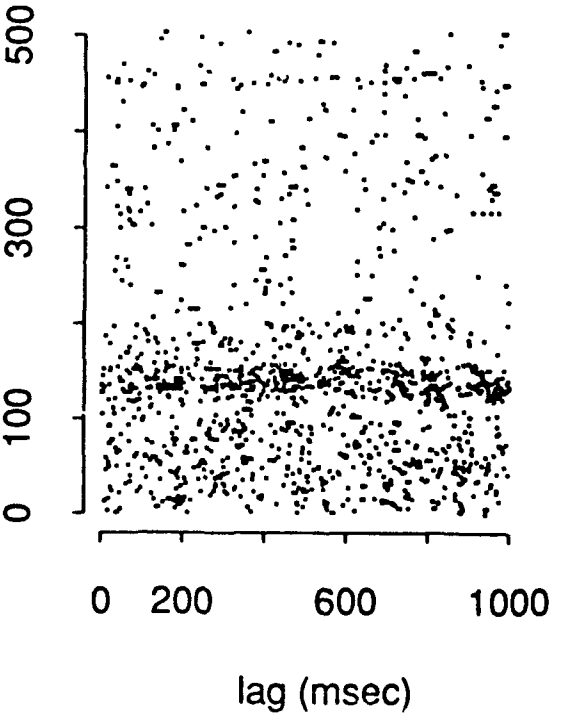
Stimlated unit 7



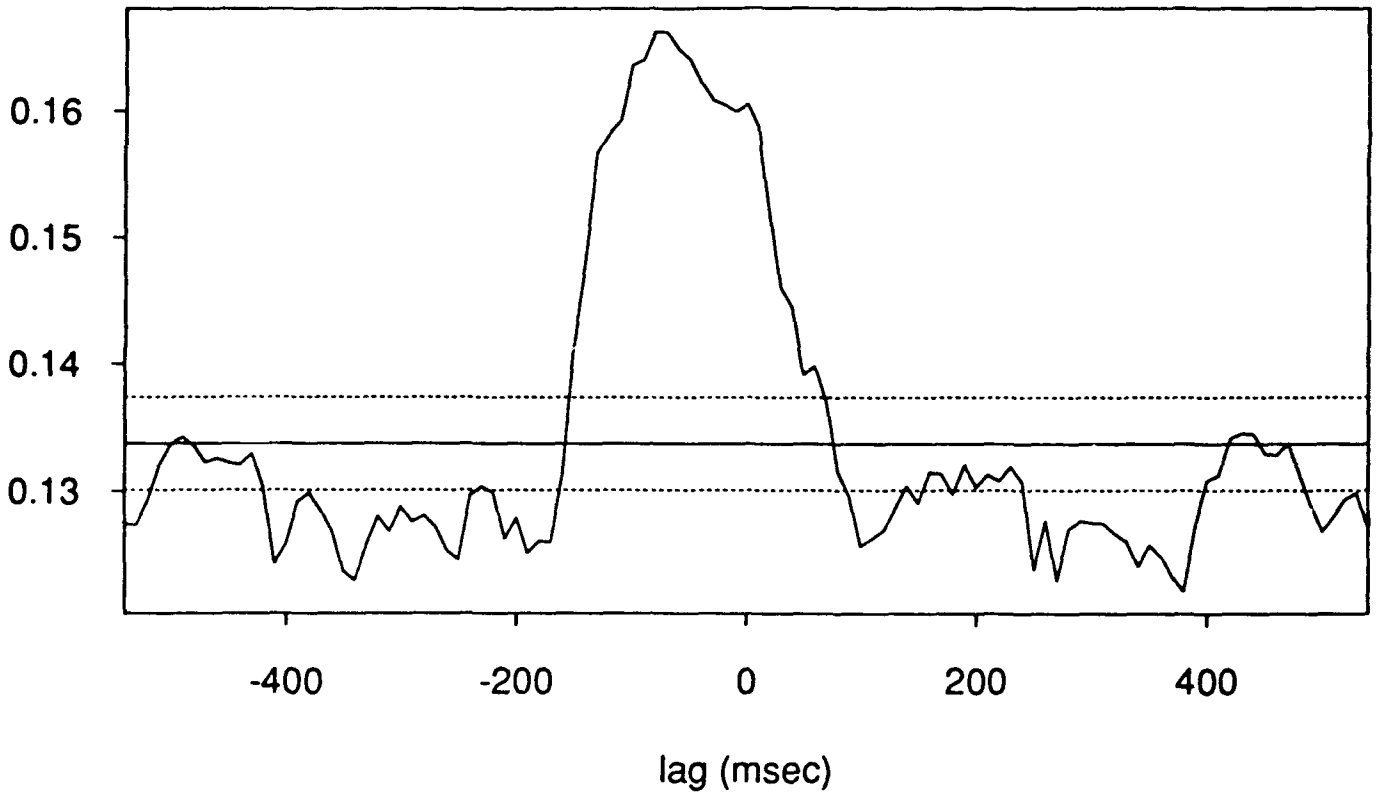
Spontaneous unit 3



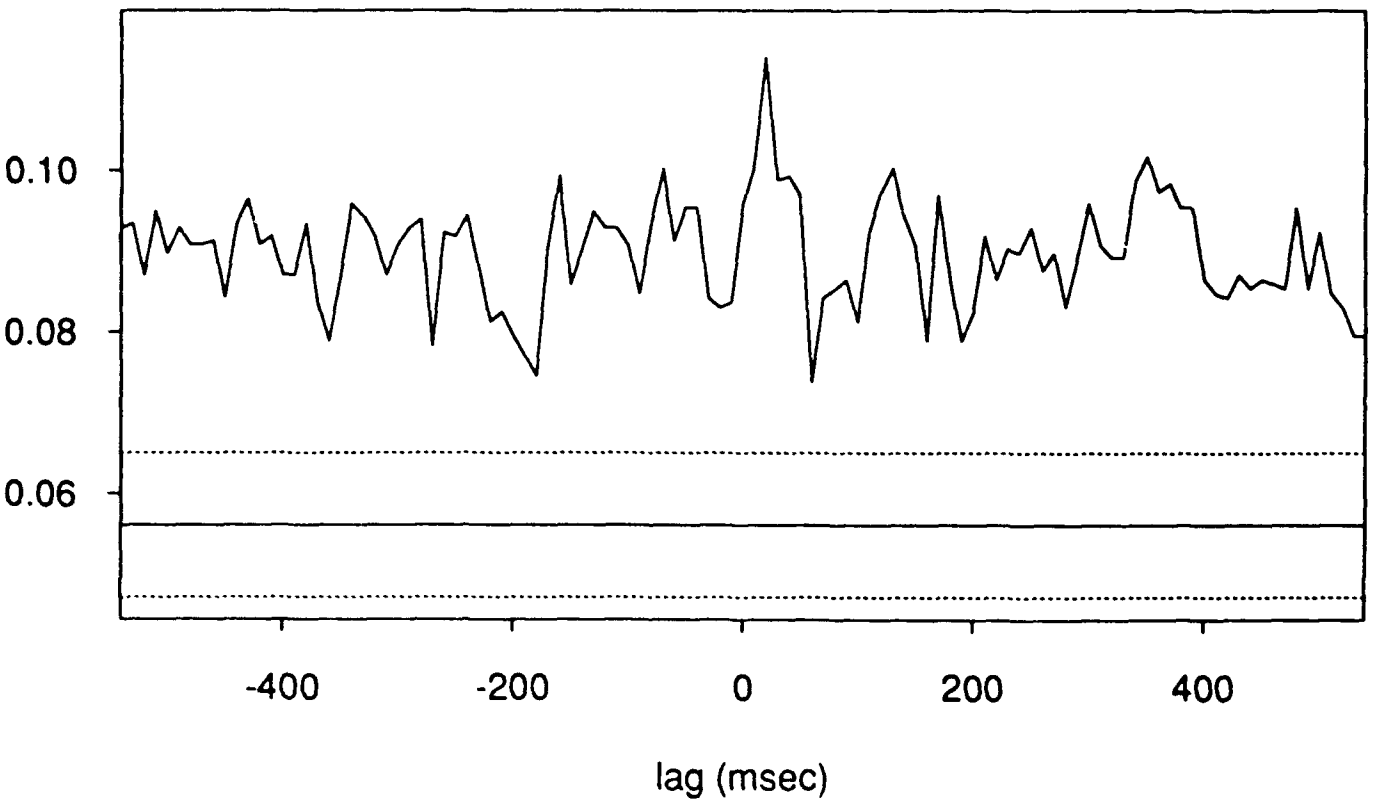
Spontaneous unit 4



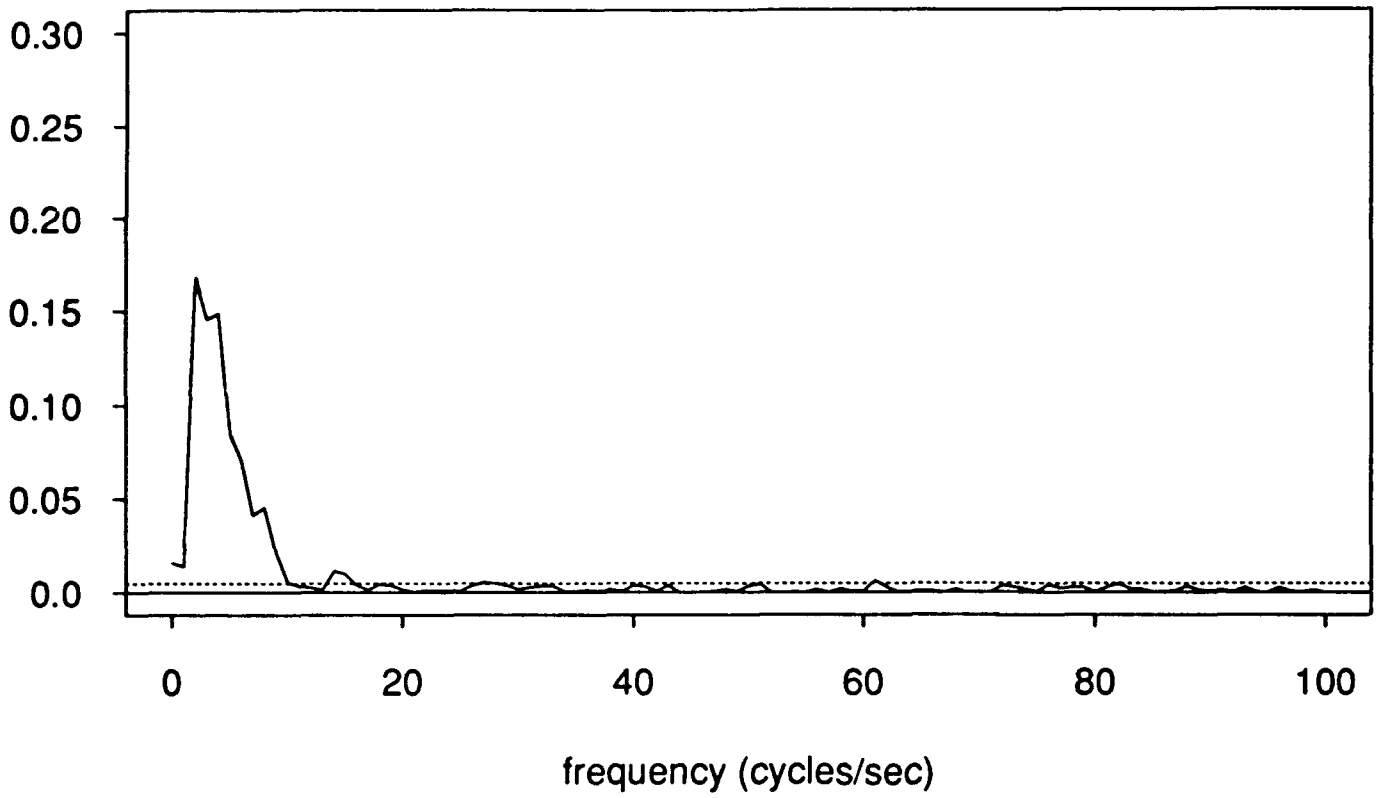
Stimulated units 6&7, sqrt(crossint)



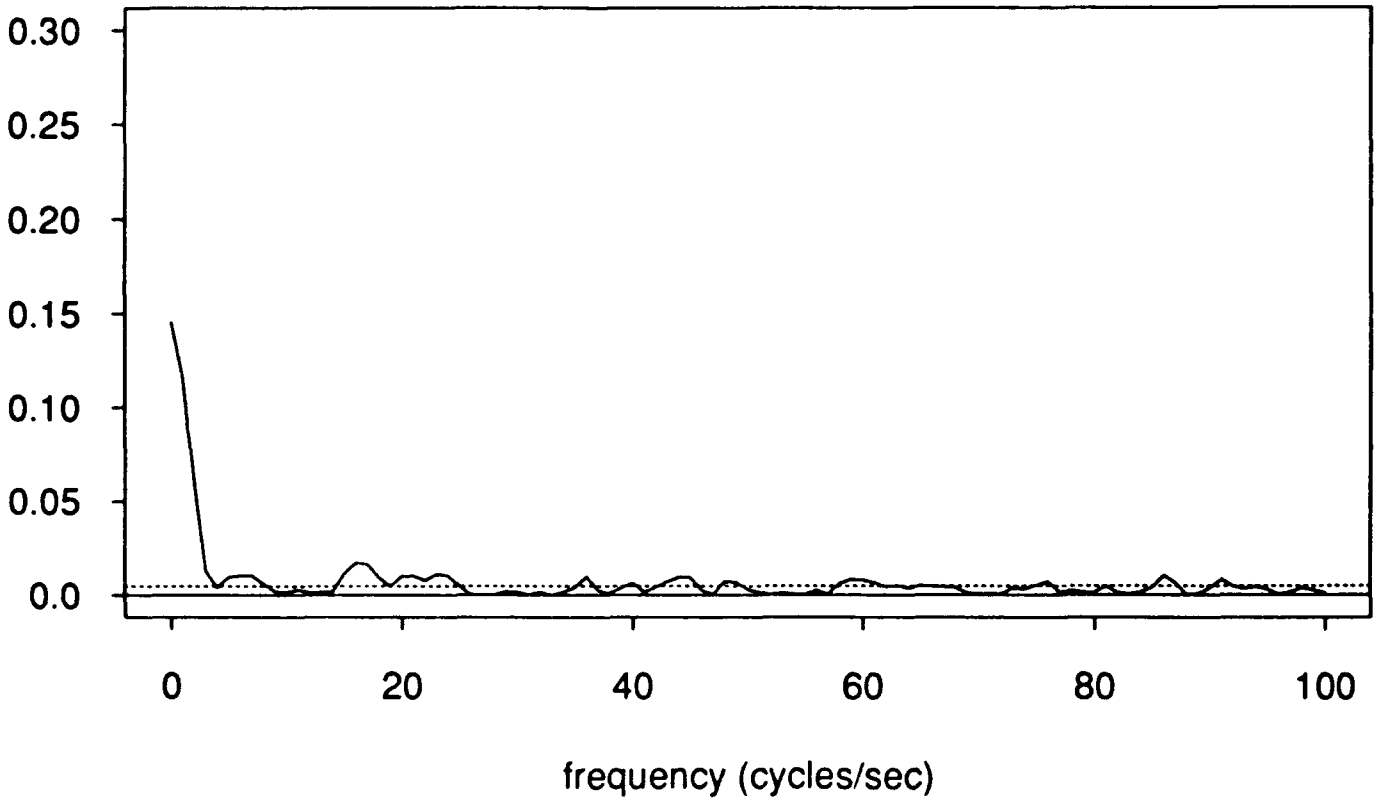
Spontaneous units 3&4, sqrt(crossint)



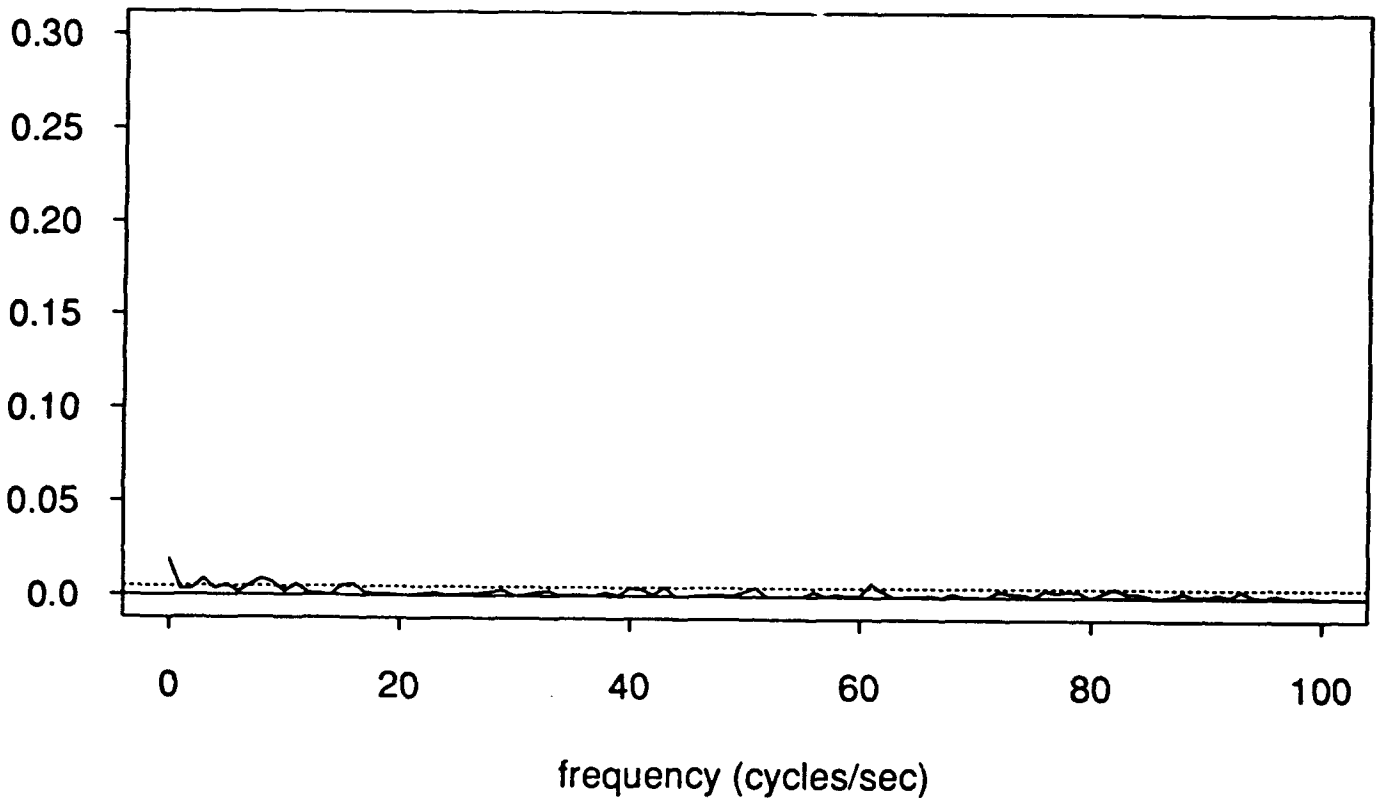
Stimulated units 6&7, coherence



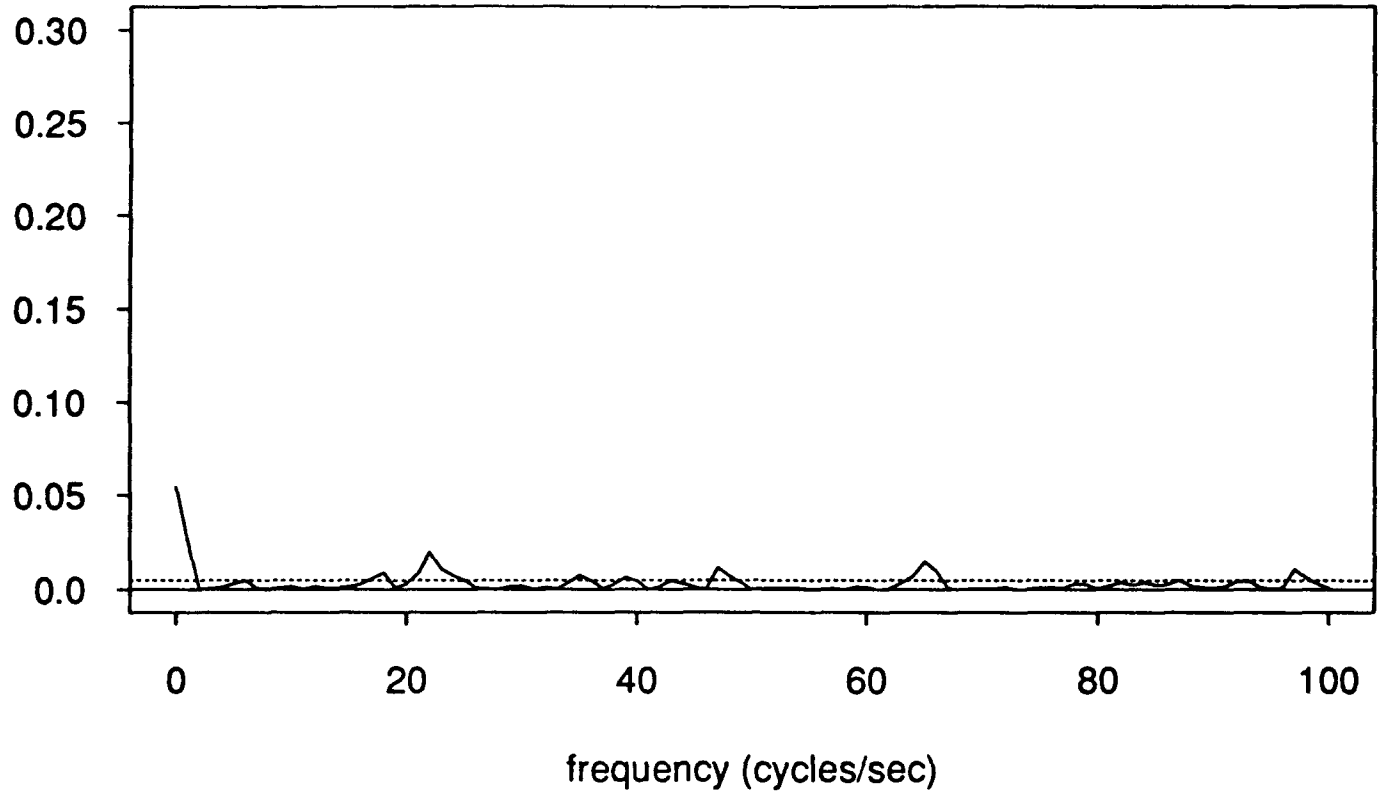
Spontaneous units 3&4, coherence



Units 6&7, partial coherence



Stimulated units 3&4, coherence



Moment-based oscillation properties of mixture models

Bruce Lindsay
Department of Statistics
Pennsylvania State University
and
Kathryn Roeder
*Department of Statistics
Yale University

Abstract

Consider finite mixture models of the form $g(x;Q) = \int f(x;\theta)dQ(\theta)$ where f is a parametric density and Q is a discrete probability measure. An important and difficult statistical problem concerns the determination of the number of support points (usually known as components) of Q from a sample of observations from g. For an important class of exponential family models we have the following result: if P has more than p components, and Q is an appropriately chosen p component approximation of P then $g(x;P) - g(x;Q)$ demonstrates a prescribed sign change behavior, as does the corresponding difference in the distribution functions. These strong structural properties have implications for diagnostic plots for the number of components in a finite mixture.

1 Introduction

Consider a family of univariate probability densities $f(x;\theta)$, with respect to some σ finite measure $d\gamma(x)$, parameterized by $\theta \in \Omega$. Frequently, interest

*The authors were supported by NSF grants DMS-9106895 to Lindsay and DMS-9001421 to Roeder.

lies in mixtures of such densities. The random variable X is said to have a mixture distribution $G(\cdot; Q)$ if X has density

$$g(x; Q) = \int f(x; \theta)\, dQ(\theta), \tag{1}$$

and the mixing distribution Q is a probability measure on Ω. If Q has a finite number of support points $\nu \equiv \nu(Q)$ then we say Q is a finite mixing distribution and we write $Q_\nu = \sum \pi_j \delta(\theta_j)$ with $\theta_1, \ldots, \theta_\nu$ being the support points (often called components) and $\pi_1, \ldots, \pi_\nu$ being the weights.

A problem of longstanding interest in such models is inference on the unknown value of $\nu(Q)$. At the simplest level, this is the problem of determining if ν equals 1, the one component model, or is greater than 1, the multicomponent model. Shaked (1980) presented important results for this problem when the component densities $f(x; \theta)$ are one parameter exponential family. We extend his results in two directions, generalizing to the discrimination between $\nu = p$ versus $\nu > p$, and moving beyond the one parameter exponential family to the normal mixture model in which each component has a different mean, but the same unknown variance.

Here we summarize Shaked's sign crossings results. Suppose we wish to contrast a multicomponent model $g(x; Q)$ with a plausible one component model $f(x; \theta)$. Choose $\theta = \theta^*$ for the one component model so that the observed variable X has the same mean under both densities:

$$\int x g(x; Q)\, d\gamma(x) = \int x f(x; \theta^*)\, d\gamma(x).$$

Our notation for this last equation will be $E[X; Q] = E[X; \theta^*]$. Shaked showed that $g(x; Q) - f(x; \theta^*)$ has exactly two sign changes, in the order

$(+,-,+)$, as x traverses the sample space. That is, $g(x;Q)$ has heavier tails than $f(x;\theta^*)$. Moreover, the difference in distribution functions $G(x;Q) - F(x;\theta^*)$ has exactly one sign change, in the order $(+,-)$.

We extend his results as follows: let P, the nominal true mixing distribution, satisfy $\nu(P) \geq p$; choose Q_p, a candidate p-point probability measure, such that it satisfies

$$E[X^k;P] = E[X^k;Q_p], \quad k = 0, 1, \ldots 2p-1. \tag{2}$$

(In Section 2 we show how to solve for Q_p.) Then, in Theorem 3.2, we show that $g(x;P) - g(x;Q_p)$ has *exactly* $2p$ sign changes in the order $(+,-,\cdots,-,+)$, unless it is identically zero (the case of nonidentifiable P). An exact sign change result for the difference in distribution functions is also given in Section 3. In Section 4, these results are extended to normal densities with unknown variance.

Before proceeding to the mathematical verification of these results, we offer a few brief comments on their potential application. In Figure 1, we plot $\{g(x;P) - g(x;Q_2)\}/\sqrt{g(x;P)}$ for the case when $f(x;\theta)$ is Poisson, P puts mass 1/3 each at (1,3 and 5), and Q_2 is constructed to match moments as specified in (2). We note the clear trimodality of this function, in constrast to the unimodality of the density $g(x;P)$ (Figure 2).

Shaked demonstrated that his sign change results could be used for diagnostic checks to determine if the data were from a mixture of specified exponential family densities rather than a one component model. These ideas were further developed in Lindsay and Roeder (1992). When interest lies in assessing the number of components in a finite mixture, the oscillation

results obtained in this article have clear implications for diagnostics plots. In a companion paper these results are used to develop diagnostic plots for the case of normal mean mixtures with unknown variance (Roeder 1992).

2 Background

2.1 The models under investigation

We will be interested in component densities $f(x;\theta)$ where both x and θ have ranges in the real numbers, say $x \in T \subset R$ and $\theta \in [l,u] \subset \Omega$. Furthermore, $f(\cdot;\cdot)$ satisfies regularity conditions which will be expounded in this subsection. Although the most important application of the results to follow is the one parameter exponential family, the results readily extend to other cases of interest for which we need the following terminology.

A real function of two variables, $K(x,\theta)$, ranging over linearly ordered sets T and Ω is said to be *totally positive* (TP) if certain determinantal inequalities hold (Karlin 1968, p. 11, 15). For instance, the functions $\exp(\theta x)$ and $\mathbf{I}(x \leq \theta)$ are TP. In addition, many density functions occuring in statistical theory are TP. For example, the one parameter exponential family with density function $K(x;\theta) = \exp\{\theta x - \psi(\theta)\}$. Other examples include the noncentral-t and noncentral-χ^2 densities. In fact, all of the densities mentioned above are strictly TP (STP; Karlin 1968, p. 12). For a more extensive list, see Karlin 1968, p. 117). We will say that $f(x;\theta)$ is an STP-model if $f(x;\theta)$ is strictly totally positive in x and θ.

2.2 Background on moments and exponential families

In order to apply our results in a particular model we need to establish two important structural features for the component densities $f(x;\theta)$. Our first requirement is as follows: suppose that P is a mixing distribution with p or more support points. Then we need to be able to construct a p-point distribution Q_p such that the first $2p-1$ moments of $g(x;P)$ and $g(x;Q_p)$ match, satisfying (2). Fortunately, there exists an important class of exponential families (the quadratic variance class) in which Q_p satisfying (2) can be shown to exist. This class includes the normal, gamma, Poisson and binomial distributions. The following is a brief review of techniques found in Lindsay (1989).

In the quadratic variance family of exponential family models (Morris 1983), for each k, there exists a polynomial of degree k, call it $\xi_k(x)$, such that

$$\int \xi_k(x) f(x;\theta) d\gamma(x) = (\mu - \mu_0)^k \tag{3}$$

for mean value parameter μ. The choice of μ_0 is arbitrary so we set it to zero. For example, in the Poisson with mean μ, $E[X] = \mu$, $E[X(X-1)] = \mu^2$, $E[X(X-1)(X-2)] = \mu^3$ and so forth. In addition, a classical moment result indicates that for a given distribution P with no fewer than p-points of support, there exists a unique distribution Q_p with exactly p-points of support such that

$$\int \mu^k dQ_p(\mu) = \int \mu^k dP(\mu), \quad k = 1, \ldots, 2p-1. \tag{4}$$

Thus integrating both sides of (3) with respect to $dQ_p(\mu)$ and $dP(\mu)$, and

using (4) yields

$$E[\xi_k(X); P] = E[\xi_k(X); Q_p], \ k = 1, \ldots, 2p-1. \tag{5}$$

Finally, the map taking $(1, x, \ldots, x^{2p-1}) \rightarrow (\xi_0(x), \xi_1(x), \ldots, \xi_{2p-1}(x))$ is invertible, so (5) implies (2).

More details on solving (5) for Q_p are given in Lindsay (1989). The solutions can be obtained algebraically for $p = 2$. For arbitrary p, the problem involves solving a degree p polynomial for its p real roots.

3 One parameter models

In this section we obtain sign change results for one parameter models. The following notation (Karlin 1968, p. 20) will be used. Let $a(x)$ be defined on I where I is a subset of the real line. The number of sign changes of a in I is defined by

$$S^-(a) = \sup \ S^-[a(x_1), \ldots, a(x_m)] \tag{6}$$

where $S^-(y_1, \ldots, y_m)$ is the number of sign changes of the indicated sequence, zero terms being discarded, and the supremum is extended over all sets

$$x_1 < x_2 < \ldots < x_m \ (x_i \in I); \ m < \infty. \tag{7}$$

We assume throughout that $f(x; \theta)$ is an STP kernel and that P and Q_p satisfy (2). The following notation will be used throughout this section: $g_1 \equiv g(x; P)$, $g_2 \equiv g(x; Q_p)$, $G_1 \equiv G(x; P)$ and $G_2 \equiv G(x; Q_p)$.

Remark In the following result we will give exact sign change results for $g_1 - g_2$ with the proviso "the difference $g_1 - g_2$ is not identically zero". If

such an equality in densities occurs, it is clear that there is an identifiability problem; both P and Q_p are generating the same distribution. The results of Lindsay and Roeder (1992) can be used to determine exactly when this will occur. If the sample space is infinite, it will not occur. If the sample space has N points, then p-point distributions Q_p are identifiable when $p \leq (N-1)/2$, and so $g_1 - g_2$ cannot be identically zero. If both P and Q_p have more than $(N-1)/2$ points, then $g_1 - g_2$ cannot have exactly $2p$ sign changes, since we can have at most $N-1$ sign changes as we traverse the sample space. Thus our result proves that P and Q_p generate the same density. ■

Lemma 3.1 *Provided $g_1 - g_2$ is not identically zero, $S^-(g_1 - g_2) \leq 2p$.*

Proof Define the measure $d\lambda(\theta)$ by

$$d\lambda(\theta) = d(P + Q_p)(\theta).$$

Let

$$p^*(\theta) = \begin{cases} P(\{\theta\})/[P(\{\theta\}) + Q_p(\{\theta\})] & \text{if } \theta \in \{\theta_1, \ldots, \theta_p\} \\ 1 & \text{else,} \end{cases}$$

and

$$q^*(\theta) = \begin{cases} Q_p(\{\theta\})/[P(\{\theta\}) + Q_p(\{\theta\})] & \text{if } \theta \in \{\theta_1, \ldots, \theta_p\} \\ 0 & \text{else.} \end{cases}$$

Then p^* and q^* are versions of the Radon-Nikodym derivatives $dP/d\lambda$ and $dQ_p/d\lambda$, so that $g_1 - g_2 = \int f(x;\theta)[p^*(\theta) - q^*(\theta)]d(P + Q_p)(\theta)$.

We now apply Theorem 3.1 (b) of Karlin (1968), noting that $p^*(\theta) - q^*(\theta)$ equals one except possibly at the support of Q_p, where it can be negative. Hence it undergoes a maximum of $2p$ sign changes. Karlin's result then

implies that integration with respect to the STP kernel $f(x;\theta)$ will result in a function, $g_1 - g_2$, with no more sign changes in x than $p^*(\theta) - q^*(\theta)$ has in θ relative to $d\gamma$. This establishes an upper bound of $2p$ sign changes in $g_1 - g_2$. ■

Theorem 3.2 *Provided $g_1 - g_2$ is not identically zero, $S^-(g_1 - g_2) = 2p$, with sign changes in the order $(+, -, \ldots, -, +)$.*

Proof From Lemma 1, we obtain an upper bound on the number of sign changes of $2p$. Because $\int x^k(g_1 - g_2)(x)d\nu(x) = 0$, for $k = 1, \ldots, 2p-1$, any polynomial $A(x)$ of degree $\leq 2p-1$ satisfies

$$\int A(x)(g_1 - g_2)(x)d\gamma(x) = 0.$$

Suppose $S^-(g_1 - g_2) \leq 2p-1$. Then we can construct a polynomial $A(x)$ that matches $g_1 - g_2$ in sign (i.e., it has single roots exactly at the roots of $g_1 - g_2$). It follows that $A(x)(g_1 - g_2) \geq 0$, and since it has zero integral it must be zero except for a set of γ-measure zero. Hence either $g_1 = g_2$, or $g_1 - g_2$ has $2p$ sign changes. ■

Remark As is clear from the proof for this result, our oscillation results still hold if we replace x^k in (2) with any system of functions $\alpha_k(x)$, such as $x^k e^{-x}$, provided that one can construct a polynomial $A(x) = \sum a_k \alpha_k(x)$ which has any prespecified set of $2p-1$ zeroes. Such an approach could be useful in improving on the robustness of the sample moments in applications by using bounded variables such as $\alpha_k(x) = x^k e^{-x}$. The next theorem, however, uses the special form of x^k. ■

Theorem 3.3 *Provided $G_1 - G_2$ is not identically zero, $S^-(G_1 - G_2) = 2p-1$, with sign changes in the order $(+,-,\ldots,+,-)$. The roots occur between the roots of $g_1 - g_2$.*

Proof An upper bound is obtained on the number of sign changes by appealing to the sign change behavior of $g_1 - g_2$. The function $G_1 - G_2$ is increasing on the intervals $[a, b]$ where $g_1 - g_2 \geq 0$:

$$G_1(b) - G_2(b) - (G_1(a) - G_2(a)) = \int I[a < x \leq b]\,(g_1 - g_2)(x)\,d\gamma(x) \geq 0.$$

From this it follows that $G_1 - G_2$ has at most one crossing in each interval where $g_1 - g_2$ is constant in sign, but has none in the first or last interval. Hence $S^-(G_1 - G_2) \leq 2p - 1$. Integration by parts gives

$$0 = \int x d(G_1 - G_2)(x) = \int [G_2 - G_1](x) dx,$$

and more generally

$$0 = \int x^k d(G_1 - G_2)(x) = \int x^{k-1}[G_2 - G_1](x)\,dx,$$

up to $k = 2p - 1$. Now, follow the proof of Theorem 3.2. If $G_2 - G_1$ had $2p - 2$ or fewer sign changes, a polynomial $A(x)$ of degree $2p - 2$ could be constructed with matching signs. Hence $A(x)[G_2 - G_1](x) \geq 0$, but has zero integral. The result follows. ■

For continuous X, a diagnostic plot based on a nonparametric empirical analog of $G_1 - G_2$ can be constructed directly. Let F_n, the empirical distribution function, be an estimate of the alleged distribution G_1 and let $\hat{G}_2$ be an estimate of G_2 constructed by using the method of moments estimates of the

p-component model. Naturally, F_n and $\hat{G}_2$ have $2p-1$ moments in common. It follows that if $F_n - \hat{G}_2$ has the appropriate sign change behavior, then the data provide some support for using more than p components. On the other hand, if a p-point mixture is the correct model, then the asymptotic properties of $F_n - \hat{G}_2$ can be obtained from empirical process theory.

4 Normal Mean Mixtures with Unspecified Variance

In this section we consider a mixture model of great interest — the normal mean mixture. We use the following notation: let $f(x;\mu,\tau)$ denote the density of a $N(\mu,\tau)$ random variable and let $g(x;Q,\tau) = \int f(x;\mu,\tau)\,dQ(\mu)$ denote a mixture of normals with corresponding distribution function $G(x;Q,\tau)$. If τ were known, then this is just a special case of the previous section; however, in practice, τ will typically be unknown and hence we treat it as a free parameter. In this section we extend our results to this case. We first present an existence theorem, due to Lindsay (1989), which extends the classic moment results presented in Section 2 to normal mixtures.

Theorem 4.1 *If Q is a distribution with more than p-points, then there exists a unique p-point distribution Q_p and variance $\tau_p > \tau$ such that*

$$\int x^k dG(x;Q_p,\tau_p) = \int x^k dG(x;Q,\tau) \quad \textit{for} \quad k = 0,1,\ldots,2p. \tag{8}$$

Proof While this is not explicitly stated in Lindsay (1989), it is a consequence of Lemma 5A and Theorem 5C. In the latter, replace the empirical moments with the moments of X under $G(\cdot;Q,\tau)$. ■

Theorem 4.2 *If (Q_p, τ_p) satisfies (8) for $Q = Q_{p+1}$, a $p+1$-point distribution, then*

$$g(x; Q_{p+1}, \tau) - g(x; Q_p, \tau_p)$$

has exactly $2p+2$ sign changes, occuring in the order $(-, +, \ldots, +, -)$.

Proof Since $\tau_p > \tau$, we can represent the above difference as

$$g(x; Q, \tau) - g(x; Q_p^*, \tau)$$

where Q_p^* is the convolution of Q_p with a normal distribution with mean zero and variance $\tau_p - \tau$. By the same argument as in Lemma 1, this means there are a maximum of $2p+2$ sign changes. The polynomial argument used in the proof of Theorem 3.2 can now be used together with (8) to show that there are at least $2p+1$ sign changes. Moreover, since Q_p^* has more mass in the tails than the discrete Q_{p+1}, the difference $g(x; Q, \tau) - g(x; Q_p^*, \tau)$ will have a negative sign in both tails, and so must have an even number of sign changes, hence $2p+2$. ■

Theorem 4.3 *$G(x; Q, \tau) - G(x; Q_p, \tau_p)$ has exactly $2p+1$ sign changes, in the order $(-, +, \ldots, +)$.*

Proof A similar argument to Theorem 3.3. ■

Graphical techniques, such as the normal scores plot (Harding 1948, Cassie 1954) and the modified percentile plot (Fowlkes 1979) have played an important role in identifying whether data follows a mixture of two normal distributions. The geometric characterizations obtained herein extend the arsenal of potential diagnostic plots for normal mixtures.

5 Discussion

Our results above, in the normal case, indicate that

$$g(x; Q_2, \tau) - g(x; \mu, \sigma^2)$$

has 4 sign changes in the order $(-, +, -, +, -)$ provided μ is the mean of Q_2 and $\sigma^2 = Var(X) = \tau + Var(Q_2)$. For this case a supplementary result is available from Roeder (1992). If we instead examine the ratio $R(x) = g(x; Q_2, \tau)/g(x; \mu, \sigma^2)$, we obtain a function proportional to a bimodal normal density. By combining the two results we can see that $R(x)$ is bimodal and that both modes are greater than 1.

In the normal model, with $\pi_1 = \pi_2 = 1/2$, the density $g(x; Q_2, \tau)$ is bimodal if and only if the two separate supports μ_1 and μ_2 satisfy $|\mu_1 - \mu_2| > 2\tau$ (Robertson and Fryer 1969). Thus the ratio function is much more sensitive to the existence of two support points than is the density itself. This sensitivity continues to exist even for very small support weights π_i.

References

[1] Cassie, R.M. (1954). Some Uses of Probability Paper in the Analysis of Size Frequency Distributions. *Australian Journal of Marine Fisheries and Freshwater Research* 5: 513-522.

[2] Fowlkes, E.B. (1979). Some Methods for Studying the Mixture of Two Normal (Lognormal) Distributions. *Journal of the American Statistical Association* 17: 561-575.

[3] Harding, J.P. (1949). The Use of Probability Paper for the Graphical Analysis of Polymodal Frequency Distributions. *Journal of Marine Bilology Association* 28: 141-153.

[4] Karlin, S. (1968). *Total Positivity, Vol 1.* Stanford University Press, Stanford, California.

[5] Lindsay, B.G. (1989). Moment Matrices: Applications in Mixtures *Annals of Statistics* 17: 722-740.

[6] Lindsay, B.G. and Roeder, K. (1992a). Residual Diagnostics for Mixture Models. *Journal of the American Statistical Association*, in press.

[7] Lindsay, B.G. and Roeder, K. (1992b). Uniqueness of Estimation and Identifiability in Mixture Models. Yale University technical report.

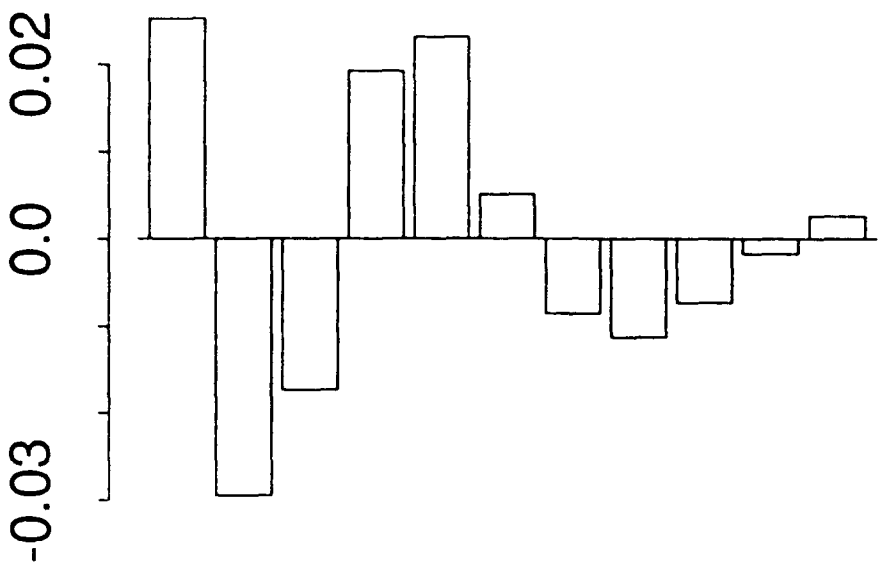
[8] Morris, C.N. (1983). Natural Exponential Families with Quadratic Variance Functions: Statistical Theory. *Statistical Theory* 11: 515-529.

[9] Robertson, C.A. and Fryer J.G. (1969). Some Descriptive Properties of Normal Mixtures. *Skand. Aktur. Tidskr* 52: 137-146.

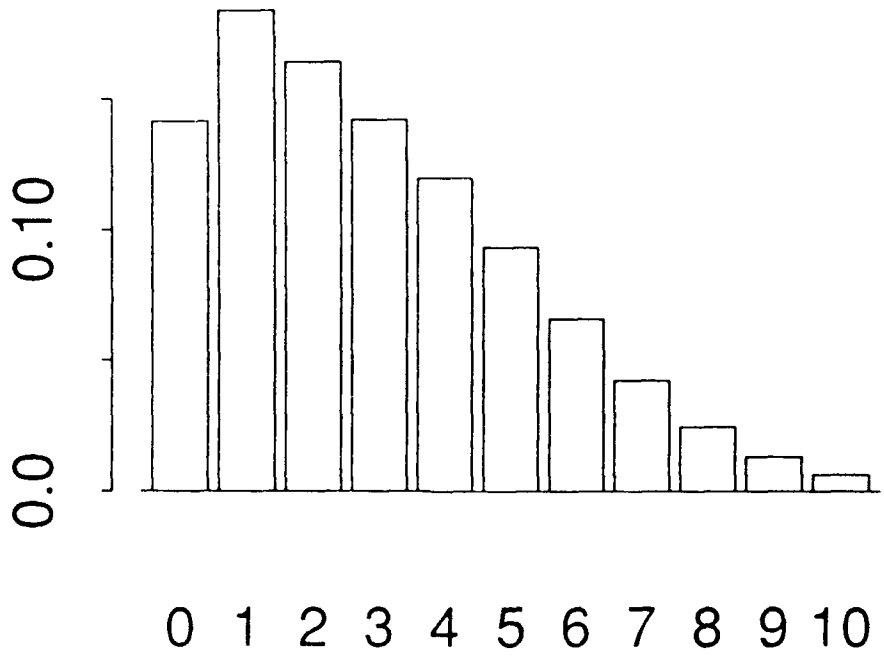
[10] Roeder, K. (1992). A Nonparametric Method for Assessing the Number of Components in a Mixture of Normals. Yale University Technical Report.

[11] Shaked, M. (1980). On Mixtures from Exponential Families. *Journal of the Royal Statistical Society, B.* 42: 192-198.

Difference in Densities



3-Pt Mixed Density



Probability and Moment Calculations for Elliptically Contoured Distributions

by Satish Iyengar *

Department of Mathematics and Statistics

University of Pittsburgh

*Research supported by the Office of Naval Research grant N00014-89-J-1496.

Abstract

The normal distribution has long been the usual model for the analysis of multivariate data. Moment and probability calculations for the multivariate normal are used in applications such as the construction of confidence sets, the assessment of error rates in signal processing, and the construction of optimal quantizers. Recently, the family of elliptically contoured distributions, which includes the normal, has been extensively studied. In this paper, we discuss moment and probability calculations for this broader class, paying particular attention to the approximation of tail probabilities.

1 Introduction

The normal distribution has long been the usual model for the analysis of multivariate data. Moment and probability calculations for the multivariate normal have therefore been well studied for various cases of interest. In statistics, a common application of such quantities is the construction of confidence sets for parameters of the normal distribution. Other examples include the assessment of error rate probabilities in signal processing, the construction of optimal quantizers for a Gaussian process, and the computation of a high order correlation coefficient of the outputs from a zero-memory non-linear device with Gaussian inputs.

The general problem is still intractable, owing to the great difficulty in evaluating high dimensional integrals, but advances in computing technology and recent research has yielded innovative Monte Carlo and numerical integration techniques. These advances have widened the scope of such investigations to include other multivariate distributions. For instance, there are the elliptically contoured distributions and the multivariate Pearson family of distributions, both of which include the multivariate normal. Elliptically contoured distributions, in particular, have been extensively developed: see the collection of papers about them that was recently edited by Anderson and Fang [2].

In this paper, we study the computation of probabilities and moments for certain elliptically contoured distributions, and discuss their applications. There are, of course, many classes of events whose probabilities are of interest, and many functions whose expectations are of interest. Our focus will be on the evaluation of tail probabilities, and on methods for computing product moments, and other non-linear functions of the components of the random vector. In Section 2, we introduce elliptically contoured distributions, and describe their properties. Historically, moment methods have been associated with Pearson's family of distributions. Since some elliptically contoured distributions are also natural multivariate versions of some of Pearson's distributions,

we briefly describe this connection also. In Section 3, we discuss applications of tail probabilities, and describe methods for approximating them accurately. These methods include Monte Carlo with importance sampling, and asymptotic approximations that generalize Mills' ratio for the normal distribution. In Section 4, we turn to moment calculations for elliptically contoured distributions using one of three tools: the characteristic function, a stochastic representation, and a certain partial differential equation satisfied by sufficiently smooth elliptically contoured densities.

2 Elliptically Contoured Distributions and Pearson Families

A p-dimensional vector X has an elliptically contoured distribution if there is a non-negative definite matrix $\Sigma = (\sigma_{ij})$ such that the characteristic function of X is $\hat{f}(t) = e^{it'\mu}\psi(t'\Sigma t)$, where ψ is a real-valued function on $\mathbb{R}_+ = [0,\infty)$. Then X has the stochastic representation

$$X = \mu + \tau\Sigma^{1/2}U_p, \tag{1}$$

where μ is the center of symmetry, the radial part τ is a non-negative random variable, and U_p is uniformly distributed on Ω_p, the surface of the unit sphere in p-dimensions; τ and U_p are independent. The matrix $\Sigma^{1/2}$ is a square root of Σ: for computations, it is convenient to take $\Sigma^{1/2}$ to be the lower triangular matrix from the Cholesky decomposition, or the non-negative definite symmetric square root derived from the spectral representation of Σ. When X has a density f, it is of the form

$$f(x;\mu,\Sigma) = |\Sigma|^{-\frac{1}{2}}\, g(Q), \tag{2}$$

where $Q = Q(x,\mu,\Sigma) = (x-\mu)'\Sigma^{-1}(x-\mu)$, $g : \mathbb{R}_+ \to \mathbb{R}_+$,

$$a_p \int_0^\infty r^{p-1} g(r^2)\, dr = 1, \tag{3}$$

and a_p is the area of Ω_p; the level curves of f are ellipses determined by $\{x : Q = c\}$. In this case, r has the density $h_r(r) = a_p r^{p-1} g(r^2)$. Examples of elliptically contoured distributions include the normal, for which $g(r) = \psi(r) = e^{-r/2}$, and the p-variate t distribution with ν degrees of freedom, for which

$$f_{p,\nu}(x;\mu,\Sigma) = \frac{\Gamma((p+\nu)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)}(1+Q/\nu)^{-(p+\nu)/2}. \tag{4}$$

Another example is due to Iyengar [12] (see also [15]):

$$f_{p,k}(x;\mu,\Sigma,\eta) = \frac{\Gamma(p/2)}{\Gamma(k+p/2)}\,|\eta\pi\Sigma|^{-1/2}\,(Q/\eta)^k \exp(-Q/\eta)\,. \tag{5}$$

where $\eta > 0$ and $k \geq 0$. When $k = 0$, (5) yields the normal distribution. For the bivariate case, Kotz [20] has also studied this family. The uniform distribution on Ω_p is yet another example which will be used for moment calculations below; it does not have a density. For further discussion of elliptically contoured distributions, see Anderson and Fang [2], Das Gupta, et al. [8], and Cambanis, et al. [5].

In one dimension, Pearson's family of distributions is defined by the following differential equation satisfied by their densities (see Cramér [7]):

$$\frac{d\,\log f(x)}{dx} = \frac{x+a}{b_0+b_1x+b_2x^2}. \tag{6}$$

Within this family, the first four moments determine the distribution. Several types of Pearson distributions (depending on a, b_0, b_1, and b_2) have been identified. In addition to the normal, the common types are the beta (Type II), gamma (Type III), and Student's t (Type VII). The elliptically contoured distributions given by (4), and (5) are multivariate versions of Types VII and III, respectively. For example, when $R = I$ and $\mu = 0$, the density for the p-variate t distribution with ν degrees of freedom, satisfies the following differential equation:

$$\nabla \log f_{p,\nu}(x;0,I) = -\frac{(p+\nu)x}{\nu+x'x}\,. \tag{7}$$

However, there is an important difference between (4) and (5). For (5), if $\mu = 0$ and $k > 0$, then the density at the origin is 0, and the modal value, or peak, of the density occurs on the surface of the ellipsoid $\{x : x'\Sigma^{-1}x = k\eta\}$. On the other hand, the density in (4) has its peak at the origin, and it is unimodal. Several results that apply to the normal and (4) do not generalize to (5); see Tong [34] for further details.

3 Tail Probabilities

If X is a random variable with density f and cumulative distribution function F, the tail probability of X refers to

$$\theta = 1 - F(a) = \int_a^\infty f(x)\,dx \tag{8}$$

for large values of a. In many statistical applications, such as hypothesis testing, the tail probability of interest is around 0.05. For such cases, the computation of, say, p-values is usually straightforward. In other applications, especially in engineering, much smaller probabilities are of interest. For instance, in signal processing, the tail probability arises as the error rate of a complex communications system (Scharf [30], Wessel, et al. [35]); and in reliability theory, it arises as the failure rate of a system component (Lawless [22]). Often such systems have redundancies built into them, so that their error or failure rates are very low. A simple model of failure regards X as an overall index of stress, and considers very large values of the failure threshold, a.

In this formulation of the problem, two difficulties arise. First, the usual quadrature rules and Monte Carlo methods for evaluating θ are not sufficiently accurate, so specialized methods are needed for evaluating tail probabilities. We will turn to some of these methods below. Next, the basis for the choice of probabilisitic model (that is, F) is tenuous. This is because for a complex system, the theoretical derivation of F based on individual component characteristics is

intractable; also, data to estimate θ is sparse since the event of interest is rare. While information about the central region (near the mean or median) of F is usually available, the tail behavior is usually unknown, so extrapolation is necessary. One way of addressing this problem is to consider a wide range of plausible models for the tail behavior to derive a range of values for the tail probability. For one example of just such an approach, see Lavine [21], who studied shuttle O-ring data.

Multivariate versions of this problem arise in similar fashion: for instance, a system with two components may fail when each component's stress exceeds its respective threshold, leading to the failure probability $P(X_1 \geq a_1, X_2 \geq a_2)$. A number of new difficulties also arise. First, multiple integration is still a hard problem in general, so with few exceptions multivariate tail probabilities are not well studied. Also, a tail region can take on many shapes, for example, $\{x : x_1 \geq a_1, x_2 \geq a_2\}$, $\{x : \alpha_1 x_1 + \alpha_2 x_2 \geq a\}$, or $\{x : x_1^2 + x_2^2 \geq a^2\}$. Below, we restrict attention to convex regions that are far from the center of the distribution, eliminating the last example from consideration.

There are two main sources of error in assessing tail probabilities. The first is numerical: it is generally hard to evaluate a small quantity with small relative error. For a deterministic method, if $\tilde{\theta}$ is an approximation to θ, the relative error is $(\tilde{\theta} - \theta)/\theta$. For a Monte Carlo method, the coefficient of variation (the ratio of the standard deviation to the mean of an estimator) is a measure of the relative error. If the unbiased estimator $\hat{\theta}_n$ of θ is an average of n independent replicates, its squared coefficient of variation (cv^2) is

$$\mathrm{cv}^2(\hat{\theta}_n) = \frac{\mathrm{var}(\hat{\theta}_n)}{\theta^2} = \frac{1}{n}\left[\frac{E(\hat{\theta}_1^2)}{\theta^2} - 1\right]. \tag{9}$$

Below, we study the use of Monte Carlo with importance sampling to derive estimators for which the cv^2 is small. If B is a tail region, and f is the density, importance sampling uses the

expression

$$\theta = \int_B \frac{f(x)}{g(x)} g(x)\,dx = \int_B l(x) g(x)\,dx, \tag{10}$$

for some "sampling density" g to get an unbiased estimator which is the average of n independent replicates (over the set B) of the likelihood ratio $l(Y)$, where Y has density g. We seek those g for which the cv^2 is bounded as the tail probability tends to zero.

The second source of error is statistical: the uncertainty in the choice of the model F makes the tail probability estimate uncertain, even if there were no numerical error. There are several ways to address this issue. One is to introduce a plausible family of models, and compute a range of tail probabilities for that family. Another is to follow the approach of Johnstone [19], for the Pearson family. He estimates the parameters of the family from available data, and then provides an estimate of a given quantile with its standard error. Yet another approach is Bayesian: first model the uncertainty in F by putting a prior on it, and then use available data to compute the posterior distribution of the tail probability.

We start with the univariate case to motivate the multivariate case below. If X has density f, l'Hôpital's rule says that with suitable regularity, the asymptotic behavior of $P(X > a)/f(a)$ is the same as that of $r(a) = -f(a)/f'(a)$. The regularity conditions are that $f'(t) \neq 0$ for all sufficiently large t, and that the ratio $r(a)$ have a limit as $a \to \infty$; these conditions are met in many cases of interest. Writing

$$\int_a^\infty f(x)dx = r(a)f(a) \int_0^\infty \frac{f(x+a)}{r(a)f(a)} dx, \tag{11}$$

it is clear that (under the same regularity conditions) the last integral in (11) approaches 1 as $a \to \infty$; thus, it is bounded away from 0, and estimating it with good relative accuracy can be done using importance sampling. This heuristic has been extended by Gray and Wang [11], where the generalized jackknife is used for evaluating univariate tail probabilities. The method suggested below may be regarded as a Monte Carlo analog of that procedure.

For the normal distribution, (11) yields

$$\int_a^\infty \phi(x)dx = \frac{\phi(a)}{a}\int_0^\infty \frac{a\phi(x+a)}{\phi(a)}dx = \frac{\phi(a)}{a}\int_0^\infty e^{-x^2/2}ae^{-ax}dx\,, \tag{12}$$

which suggests the estimator

$$\hat{\theta} = \frac{\phi(a)}{a}e^{-T^2/2}, \tag{13}$$

where T has the exponential density, ae^{-at} for $t \geq 0$. Now, let $\Phi(x)$ and $\phi(x)$ denote the univariate standard normal distribution and density functions, respectively, and let

$$M(x) = \frac{\Phi(-x)}{\phi(x)} = \frac{1}{x}\int_0^\infty e^{-t^2/2}xe^{-xt}dt \tag{14}$$

denote Mills' ratio. Since M is a convex, decreasing function (Iyengar [13]), the following inequalities are easy to prove:

$$\frac{x}{1+x^2} < M(x) < \frac{1}{x} \quad \text{for } x > 0. \tag{15}$$

These inequalities, in turn, imply that

$$\mathrm{cv}^2(\hat{\theta}) = \frac{M(a/\sqrt{2})}{M(a)^2 a\sqrt{2}} - 1 \sim \frac{2}{a^2} \tag{16}$$

as $a \to \infty$, so that the cv^2 tends to zero as a increases. This estimator results from the sampling density $g(t) = ae^{-a(t-a)}$ for $t \geq a$. The deterministic analog of this result is that

$$0 < \frac{\phi(a)/a - \Phi(-a)}{\Phi(-a)} = \frac{1}{aM(a)} - 1 < \frac{1}{a^2}, \tag{17}$$

so that the relative error in approximating $\Phi(-a)$ by $\phi(a)/a$ decreases to zero as a increases.

The phenomenon observed in (16) is quite general: for a wide class of problems, the coefficient of variation actually tends to zero, hence the relative accuracy improves as the threshold a increases. In addition, this method is feasible since the calculation of $r(a)$ depends on the differentiation of the density rather than its integration; since the behavior of the tail probability

is already captured by $r(a)f(a)$, the evaluation of the remaining integral by Monte Carlo provides a correction term. In practice, either (11) or one of the following two expressions for θ is also useful:

$$\theta = r(a)f(a)\int_0^\infty \frac{f(a+x/a)}{ar(a)f(a)}dx = r(a)f(a)\int_0^\infty \frac{af(a+ax)}{r(a)f(a)}dx\,. \tag{18}$$

Two other examples illustrate this technique. The first involves the generalized inverse Gaussian distribution, whose density is

$$f(t \mid \alpha,\beta,\lambda) = \frac{(\alpha/\beta)^{\lambda/2}}{2K_\lambda((\alpha\beta)^{1/2})}t^{\lambda-1}\exp\left[-\frac{1}{2}(\alpha t+\beta/t)\right],\ \text{for } t>0, \tag{19}$$

where K_λ is the modified Bessel function of the third kind with index λ. The parameter space is the union of the following three sets: $\{\alpha>0,\beta>0\}$, $\{\alpha=0,\beta>0,\lambda<0\}$, and $\{\alpha>0,\beta=0,\lambda>0\}$. This family includes the gamma, the inverse Gaussian, the hyperbola distribution, and their reciprocals, in the sense that if X has density $f(t \mid \alpha,\beta,\lambda)$, then X^{-1} has density $f(t \mid \beta,\alpha,-\lambda)$. For the case $\alpha>0$, $\beta>0$, this method yields the estimator

$$\hat{\theta} = \frac{2}{\alpha}f(a \mid \alpha,\beta,\lambda)\,e^{\beta/2a}(1+\frac{2T}{a\alpha})^{\lambda-1}\exp\left[-\frac{1}{2}\frac{\alpha\beta}{(a\alpha+2T)}\right], \tag{20}$$

for sufficiently large a, where T has a standard exponential density. The second example is the t distribution with k degrees of freedom, with density $f_k(x)$ proportional to $(1+x^2/k)^{-(k+1)/2}$, for which the estimator is

$$\hat{\theta} = \frac{a}{k}f_k(a)\left[\frac{(k+a^2)Y^2}{k+a^2Y^2}\right]^{(k+1)/2}, \tag{21}$$

where Y has the Pareto density k/y^{k+1} for $y \geq 1$. In both cases, the cv^2 decreases to zero as $a \to \infty$. Detailed proofs of these and related results are given in [17].

We now turn to the multivariate case. In 1962, Slepian [32] proved the following inequality. Let $X \sim N_p(0,\Sigma=(\sigma_{ij}))$ and $Y \sim N_p(0,T=(\tau_{ij}))$ with $\sigma_{ij} \geq \tau_{ij}$ and $\sigma_{ii}=\tau_{ii}$; then for any vector a, $P(X \geq a) \geq P(Y \geq a)$, where $x \geq a$ means that $x_i \geq a_i$ for all i. Slepian derived

this result using Plackett's identity (see Section 4 below) in a study of one-sided boundary crossing problems for Gaussian processes. Since Slepian proved his inequality, his result has been generalized in a number of ways. For instance, the inequality holds for all elliptically contoured distributions: see Das Gupta, et al. [8] and Tong [34] for such results.

When $\sigma_{ij} \geq 0$ for all i and j, the inequality $P_\Sigma(X \geq a) \geq P_I(X \geq a)$ yields a lower bound which can be easily computed for the normal, since then it is a product of univariate normal probabilities. However, this lower bound often gives a poor approximation (see Iyengar [14]), so that Slepian's inequality is more useful for theoretical investigations. Thus, in this section, we describe alternative methods that provide good approximations.

Suppose that X is a p-variate vector which has an elliptically contoured distribution with density $|\Sigma|^{-\frac{1}{2}}\ g(x'\Sigma^{-1}x)$; further, let term "tail region" refer to a closed convex region B that is far from 0 (of course, B should have non-empty interior, else the probability will be zero). If $\Sigma = L'L$ is the Cholesky decomposition of Σ, then $Z = L^{-1}X$ has the density $f(z) = f(z;0,I) = g(z'z)$, and $P(X \in B) = P(Z \in A = L^{-1}B)$. Since A is closed and convex, it contains a unique point, α, that is closest to the origin: $|\alpha| \leq |z|$, for $z \in A$, and A is contained in the half plane $\{z : z'a \geq \alpha'\alpha\}$. Since Z has a spherically symmetric distribution, A can be rotated so that $\alpha = re_1$, where e_1 is the unit vector in the z_1 direction, and $r = |\alpha|$. Note that $r = r(A)$ depends upon the set A; for notational convenience, this dependence will be suppressed. Next, if $\beta = L\alpha$, then β minimizes the Mahalanobis distance, $(x'\Sigma^{-1}x)^{1/2}$, of points in B to the origin; also, B is contained in the half plane $\{x : x'\Sigma^{-1}\beta \geq \beta'\Sigma^{-1}\beta\}$. Of course, the problem of finding β is a quadratic programming problem which can be solved using known techniques. For any set A, matrix D, and vector c, let $DA + c$ denote the set, $\{Dx + c : x \in A\}$.

To estimate $\theta = P(Z \in A)$, ordinary Monte Carlo averages n independent replicates of $I(Z \in A)$, where I is an indicator function. This estimator's variance is $(\theta - \theta^2)/n$. An

alternative approach is to use $f(z-\alpha)$ as a sampling function (Wessel, et al. [35] refer to this as improved importance sampling). The expression

$$\theta = \int_A \frac{f(z)}{f(z-\alpha)} f(z-\alpha)\, dz = \int_{A-\alpha} \frac{f(z+\alpha)}{f(z)} f(z)\, dz \tag{22}$$

suggests the unbiased estimator

$$\hat{\theta} = \frac{f(Z+\alpha)}{f(Z)} I(Z \in A-\alpha). \tag{23}$$

If g is a decreasing function — that is, f is unimodal, as is the case with the p-variate normal or t, but not the family given in (5) — then $f(z) \leq f(z-\alpha)$ for $z \in A$, and

$$E(\hat{\theta}^2) = \int_A \frac{f(z)}{f(z-\alpha)} f(z)\, dz \leq \theta, \tag{24}$$

so that $\hat{\theta}$ has a smaller variance (and smaller cv^2) than ordinary Monte Carlo. However, it can be shown that for several cases (the normal and the t), the cv^2 tends to infinity as $\alpha \to \infty$ (see [17]). Thus, we turn to multivariate analogs of the method described in (12) above.

Although a direct generalization of (12) is not available, the analog is to write $A_0 = A - \alpha$, and

$$\theta = \int_A f(z)\, dz = f(\alpha) \int_{A_0} \frac{f(z+\alpha)}{f(\alpha)}\, dz, \tag{25}$$

and to manipulate the ratio $f(z+\alpha)/f(\alpha)$ to derive an estimator that has bounded cv^2 as the region A moves outward to infinity. Just as in the one-dimensional case, there is no generic method that will work for all g; and unlike the one-dimensional case, the shape of A (or equivalently the shape of B and the dependence among the random variables as given by Σ) plays an important role in the choice of sampling function. We now sketch the details for the normal and t distributions.

For the normal with density $\phi_p(z) = \phi_p(z; 0, I)$, (25) becomes

$$\theta = \phi_p(\alpha) \int_{A_0} \frac{\phi_p(z+\alpha)}{\phi_p(\alpha)}\, dz = \frac{(2\pi)^{(p-1)/2} \phi_p(\alpha)}{|\alpha|} \int_{A_0} |\alpha|\, e^{-|\alpha| z_1 - z_1^2} \phi_{p-1}(u)\, du\, dz_1, \tag{26}$$

where $u = (z_2, \ldots, z_p)$. Next, for the t density $f_p(z) = f_{p,\nu}(z; 0, I)$, a slight modification of (12) is needed. Let $A_1 = A/|\alpha|$ to get

$$\theta = f_p(\alpha) \int_A \left(\frac{\nu + |\alpha|^2}{\nu + |z|^2} \right)^{(p+\nu)/2} dz = f_p(\alpha) |\alpha|^p \int_{A_1} \left(\frac{\nu + |\alpha|^2}{\nu + |\alpha|^2 |z|^2} \right)^{(p+\nu)/2} dz. \tag{27}$$

Now using the sampling density which is proportional to $|z|^{-(p+\nu)}$ on A_1, we get

$$\theta = f_p(\alpha) |\alpha|^p \int_{A_1} \left(\frac{(\nu + |\alpha|^2) |z|^2}{\nu + |\alpha|^2 |z|^2} \right)^{(p+\nu)/2} \frac{dz}{|z|^{p+\nu}}. \tag{28}$$

Such expressions provide guidelines on the nature of the sampling function to use for importance sampling. The specific choice depends, as mentioned before, on the nature of A, specifically, on the shape of A near the origin (or A_1 near the point e_1). In particular, let $B = \{x : x_1 \geq b_1, x_2 \geq b_2\}$, where the b_i are positive; without loss of generality, suppose that $b_1 \leq b_2$. When the correlation between X_1 and X_2 is ρ, the point, β, that is closest to the origin (using Mahalanobis distance) is

$$\beta = \begin{cases} (b_1, b_2) & \text{if } \rho < b_1/b_2 \\ (\rho b_2, b_2) & \text{if } \rho \geq b_1/b_2. \end{cases} \tag{29}$$

Transforming to the independent case and rotating so that the nearest point, α, is in the e_1 direction gives

$$\alpha = \begin{cases} ([b' R^{-1} b]^{1/2}, 0) & \text{if } \rho < b_1/b_2 \\ (b_2, 0) & \text{if } \rho \geq b_1/b_2. \end{cases} \tag{30}$$

The region A is given in Figures 1 for $\rho < b_1/b_2$, and 2 for $\rho \geq b_1/b_2$. Since the nature of $A_0 = A - \alpha$ at the origin is determined by the difference $\rho - b_1/b_2$, the ratio b_1/b_2 will be preserved in the calculations above: in effect, the region B will be moved outward towards infinity in the direction of the vector $b = (b_1, b_2)$.

{FIGURES HERE}

We will now provide some of the details for the normal distribution; for a fuller account, see [17]. When the correlation coefficient ρ is not large ($\rho < b_1/b_2$ when $b_1 \leq b_2$), the bivariate sampling function consisted of a product of two exponential densities, and when ρ is large, the sampling function consisted of the product of an exponential and a normal. This is intuitively plausible, since for small ρ, the bivariate normal density is not far from the independent case, while for large ρ, it is not far from the singular case, for which the exponential given in (113) yields accurate estimates. Transforming back to X (with $\rho_{12} = \rho$), the estimators are given by the following. For $\rho < b_1/b_2$,

$$\frac{\phi_2(b;\Sigma)(1-\rho^2)^2}{(b_1-\rho b_2)(b_2-\rho b_1)} e^{-T'R^{-1}T/2}, \tag{31}$$

where $T = (T_1, T_2)$ has independent exponentially distributed components with mean vector $((1-\rho^2)/(b_1-\rho b_2), (1-\rho^2)/(b_1-\rho b_2))$. And for $\rho \geq b_1/b_2$ it is

$$\frac{\phi(b_2)}{b_2} e^{-T^2/2} I[(T,U) \in A_0], \tag{32}$$

where T and U are independent with densities $|\alpha|\ e^{-|\alpha|t}$ and $\phi(u)$, respectively, and $A_0 = A - (b_2, 0)$ is the translate of the set given in Figure 2. For both of these cases, it can be shown that the cv^2 for the estimators given above all tend to zero as $\alpha \to \infty$, that is, as the tail probability diminishes. The proof for the normal case is given in [17]. We omit the proof for the t distribution. Instead, we turn to the key quantity that is used in the proofs, Mills' ratio.

Several definitions of the multivariate normal Mills' ratio are available. The first definition is due to Savage, [29] for the case of orthants:

$$M_1(B;R) = \frac{P(X \in B)}{\phi_p(b;R)}, \tag{33}$$

for $X \sim N_p(0, R)$. Another definition is gotten by first transforming to the spherically symmetric case with Z, A, and α replacing X, B, and β respectively. For $r = |\alpha|$ let

$$M_2(A;I) = \frac{P(Z \in A)}{\phi(r)}. \tag{34}$$

This definition applies to convex regions A, not just orthants. However, the two definitions do not coincide when B is an orthant. For $R \neq I$,

$$M_2(A;I) = \frac{P(Z \in A)}{(2\pi)^{(p-1)/2}\phi_p(\alpha;I)} = \left(\frac{2\pi}{|2\pi R|}\right)^{1/2} \frac{P(X \in B)}{\phi_p(\beta;R)}, \tag{35}$$

so that the two definitions differ in two respects. First, in place of β, it uses the vertex b; for example, when $(b_1, b_2) = (1,2)$ and $\rho = 0.95$, $(\beta_1, \beta_2) = (1.9, 2)$. This is an important difference, because when the correlation is high, importance sampling centered at b can be much worse than that centered even at the origin (see [17]). Second, the new definition has the factor $(2\pi/\mid 2\pi R \mid)^{1/2}$; this is not an important difference, but it does mean that proper comparisons of the two must first adjust for this factor.

For the multivariate normal, the following inequalities for M_2 generalize (15):

$$M_2(A;I) < \frac{1}{r} P[(T,U) \in A_0], \tag{36}$$

and

$$\begin{aligned} M_2(A;I) &> \frac{1}{r}\left[P[(T,U) \in A_0] - \int_{A_0} \frac{t^2}{2} r e^{-rt} \phi(u) du dt\right] \\ &> \frac{1}{r}\left[P[(T,U) \in A_0] - \int_0^\infty \frac{t^2}{2} r e^{-rt} \phi(u) du dt\right] \\ &= \frac{1}{r}\left[P[(T,U) \in A_0] - \frac{1}{r^2}\right], \end{aligned} \tag{37}$$

where (T,U) is as in (32). When $A = L^{-1}B$, where B is a quadrant, explicit expressions for the bounds in (36) and the first line of (37) are available. Such inequalities are not available for M_1. These inequalities are used in [17] to prove that the estimators in (31) and (32) have cv^2 tending to zero as $\alpha \to \infty$.

Mills' ratio for elliptically contoured densities are defined analogously: the numerator is $P(X \in B)$, while the denominator is either $\phi_p(b;R)$ or $\phi_p(\beta;R)$ for M_1 and M_2, respectively. In [9], Fang and Xu give a detailed account of M_1 They show that if X has an elliptically

contoured distribution given by (2), where g is a non-increasing function, then the function $-P(X \in B)$ is a Schur convex function; they use this fact, along with standard majorization results to provide inequalities for M_1. A detailed study of the analog of M_2 for other elliptically contoured distributions has not yet been done.

4 Computation of Moments

In his paper, Brillinger [4] noted that a moment generalizes the notion of a probability, since the latter is the first moment of an indicator function, which is a building block of integrable functions. Here, we use the term moment to denote the expected value, when it exists, of some function of a random vector, that is, $E[g(X)] = E[g(X_1, \ldots, X_p)]$. Conventionally, (product) moments are defined as $E\left[\prod_{i=1}^{p} X_i^{k_i}\right]$, where k_i are non-negative integers. In this section, we discuss three methods for computing moments for elliptically contoured distributions. The first uses the characteristic function when it is available, the second uses the stochastic representation (1) when the moments of τ are available, and the third uses several partial differential equations that are given below. Throughout, let $X = \mu + \tau\Sigma^{1/2}U_p$, as in (1).

The first two methods, which are due to Li [23], are of course equivalent; computational convenience dictates the choice of method. Let the k^{th} moment (when it exists) of the vector X be given by the matrix $\Gamma_k(X)$, where

$$\Gamma_k(X) = (\gamma_{rs}^{(k)}) = \begin{cases} E[X \otimes X' \otimes X \ldots \otimes X'] & \text{if } k \text{ is even} \\ E[X \otimes X' \otimes X \ldots \otimes X' \otimes X] & \text{if } k \text{ is odd,} \end{cases} \tag{38}$$

where $\otimes$ denotes the Kronecker product, which has k terms in (38). This definition reduces to the usual mean vector and covariance matrix when $k = 1$ and 2, respectively; $\Gamma_1(X) = \mu$ whenever the first moment exists. For $k \geq 3$, the following recipe tells us where to find $E\left[\prod_{i=1}^{p} X_i^{k_i}\right]$ (with $\sum_{i=1}^{p} k_i = k$) in $\Gamma_k(X)$: if the terms in the product are strung out thus, $\gamma_{rs}^{(k)} = E(X_{i_1} X_{i_2} \ldots X_{i_k})$,

then

$$r = 1 + \sum_{j=1}^{[(k+1)/2]} (i_{2j-1} - 1)\, p^{[(k+1)/2]-j} \tag{39}$$

and

$$s = 1 + \sum_{j=1}^{[k/2]} (i_{2j-1} - 1)\, p^{[k/2]-j}, \tag{40}$$

where $[a]$ is the greatest integer in a.

Using this notation, the matrices $\Gamma_k(X)$ can be expressed in two ways. First, if the characteristic function is known, repeated differentiation of it gives the following expressions for $k = 2$ and 3:

$$\begin{aligned} \Gamma_2(X) &= \mu\mu' - 2\psi'(0)\Sigma, \\ \Gamma_3(X) &= \mu \otimes \mu' \otimes \mu - 2\psi'(0)[\mu \otimes \Sigma + \Sigma \otimes \mu + \mathrm{vec}(\Sigma)\mu'], \end{aligned} \tag{41}$$

where $\mathrm{vec}(\Sigma) = (\sigma_{11}, \sigma_{21}, \ldots, \sigma_{p1}, \ldots, \sigma_{1p}, \ldots, \sigma_{pp})'$ strings out the columns of Σ into one long vector.

This formulation is useful for the family (5), for the characteristic function is given by

$$\psi_{p,k}(t;\eta) = e^{-\eta t/4} \sum_{m=0}^{k} \binom{k}{m} \frac{\Gamma(p/2)}{\Gamma(m + p/2)} (-\eta t/4)^m, \tag{42}$$

so that $-2\psi'(0) = \eta(2k + p)/2p$. A proof of this result is given in Iyengar and Tong [15]. When the characteristic function is not available, but the moments of τ are available, the representation (for $\mu = 0$ and $\Sigma = I$) $X = \tau U_p$ implies that $\Gamma_k(X) = \tau^k \Gamma_k(U_p)$. Since $\Gamma_k(U_p)$ can be derived from the known properties of the normal distribution, $-2\psi'(0)$ is replaced by $E(\tau^2)/p$ in (41). For instance, for the multivariate t, the characteristic function is intractable, but the density of τ is proportional to

$$r^{p-1}(1 + r^2/\nu)^{-(p+\nu)/2}, \quad r > 0, \tag{43}$$

which yields the finite moments upon integration. Expressions for the fourth moment Γ_4 that involve $\psi''(0)$ or $E(\tau^4)$ are given in [23]; even higher order moments can be computed along

the lines outlined there. Since quadratic forms in elliptically contoured distributions arise in standard testing procedures (see Anderson and Fang [2]), Li also provides expressions for their moments.

In a related study, Xu and Fang [36] define an $n \times p$ matrix has a matrix elliptically contoured density if TX has the same distribution as X for every $n \times n$ orthogonal matrix T. The density then has the form $c_{n,p} f(x'x)$; if $Y = X\Sigma^{1/2}$ for a $p \times p$ covariance matrix Σ, the density of Y is given by

$$c_{n,p} \, |\Sigma|^{-n/2} \, f(\Sigma^{-1/2} y' y \Sigma^{-1/2}). \tag{44}$$

In their paper, Xu and Fang give the expected values of zonal polynomials and other symmetric functions of $W = Y'Y$. The expressions are rather involved, so we omit them.

The third method of computing moments has a longer history. In 1958, Price [27] proved the following result. Let $N_p(\mu, \Sigma)$ denote a p-variate normal with mean μ and covariance matrix $\Sigma = (\sigma_{ij})$. Suppose that $X = (X_1, \ldots, X_p)$ has a $N_p(\mu, \Sigma)$ distribution (written $X \sim N_p(\mu, \Sigma)$), and let $g_1(X_1), \ldots, g_p(X_p)$ be differentiable functions of the components of X, each admitting a Laplace transform; then

$$\frac{\partial}{\partial \sigma_{ij}} E\left[\prod_1^p g_k(X_k)\right] = E\left[\frac{\partial^2}{\partial x_i \partial x_j} \prod_1^p g_k(X_k)\right] \quad \text{for } i \neq j. \tag{45}$$

Conversely, if this identity holds for arbitrary $g_1, \ldots, g_p$ (with both expectations above defined) then X has a multivariate normal distribution. Price and others used this theorem to facilitate studies in signal processing. In particular, suppose that a zero-memory non-linear input-output device with Gaussian input X_i that yields output $g_i(X_i)$. The p^{th}-order correlation coefficient of the outputs is a quantity of interest which requires the computation of the expectation of $\prod_1^p g_k(X_k)$. The differential equation of Price's theorem provides a useful computational tool for such calculations. Consider the following trivial example: if $h(\rho) = E(X_1 X_2)$, where $\rho_{12} = \rho$

is the correlation between the standardized variates X_1 and X_2, then $h'(\rho) = 1$, and $h(\rho) = \rho$ follows.

Although Price's theorem is an elegant result, it has several limitations. In fact, Pawula [25] (see also Papoulis [24]) noted that when $p = 2$, and the right hand side of (45) can be evaluated explicitly, there is a single differential equation to solve. But for larger p, there are $p(p-1)/2$ differential equations to solve simultaneously. Furthermore, Price's result only applied to a product of functions of individual components only. Pawula used a result of Plackett [26] to overcome these limitations. In 1954, Plackett proved the following identity while investigating a reduction formula for multivariate normal probabilities: if the density of a $N_p(\mu, \Sigma)$ variate is $\phi_p(x - \mu, \Sigma)$, then

$$\frac{\partial}{\partial \sigma_{ij}} \phi_p(x - \mu; \Sigma) = \frac{\partial^2}{\partial x_i \partial x_j} \phi_p(x - \mu; \Sigma), \quad \text{for } i \neq j. \tag{46}$$

For the case $i = j$, we have the diffusion equation

$$\frac{\partial}{\partial \sigma_{ii}} \phi_p(x - \mu; \Sigma) = \frac{1}{2} \frac{\partial^2}{\partial x_i^2} \phi_p(x - \mu; \Sigma). \tag{47}$$

Pawula used Plackett's identity to extend Price's theorem thus: if $g(x_1, \ldots, x_p)$ is sufficiently smooth and vanishes rapidly near infinity, then

$$\frac{\partial}{\partial \sigma_{ij}} E\left[g(X_1, \ldots, X_p)\right] = E\left[\frac{\partial^2}{\partial x_i \partial x_j} g(X_1, \ldots, X_p)\right] \quad \text{for } i \neq j. \tag{48}$$

This extension allowed the study of more general functions, such as the "linear rectifier correlator," $g(x_1, x_2) = |x_1 + x_2| - |x_1 - x_2|$.

Pawula then used the following method, also due to Plackett, to reduce the number of differential equations to solve from $p(p-1)/2$ to one. For a given Σ define a line between it and the identity matrix I, $\Sigma_t = (1 - t)I + t\Sigma$ for $0 \leq t \leq 1$. The chain rule then gives

$$\frac{\partial}{\partial t} \phi_p(x - \mu; \Sigma_t) = \sum_{i<j} \sigma_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \phi_p(x - \mu; \Sigma_t), \tag{49}$$

so that

$$\frac{\partial}{\partial t} E_t [g(X_1, \ldots, X_p)] = E_t \left[\sum_{i<j} \sigma_{ij} \frac{\partial^2}{\partial x_i \partial x_j} g(X_1, \ldots, X_p) \right], \tag{50}$$

where E_t denotes the expectation with respect to $N(\mu, \Sigma_t)$. When the right hand side of (50) can be evaluated, a single ordinary differential equation results. By solving it, Pawula showed how to compute the moments of various functions of X, such as products of Hermite polynomials or error functions. In some cases, higher order derivatives with respect to t are needed: they are just iterates of the partial differential operator on the right of (50).

The search for bounds for certain probabilities and expectations has recently led to several generalizations of Plackett's identity to elliptically contoured distributions. The first is a result of Joag-dev, et al. [18] which only requires that g in (2) be differentiable:

$$\frac{\partial}{\partial \sigma_{ij}} f(x; \mu, \Sigma) = -\frac{\partial}{\partial x_j} (\sum_{k=1}^{p} \sigma^{ik} x_k) f(x; \mu, \Sigma), \tag{51}$$

where σ^{ik} is the i, k element of Σ^{-1}. Another is due to Iyengar ([12], see also Iyengar and Tong [15]), who proved the following identity for $f_{p,k}$:

$$\frac{\partial}{\partial \rho_{ij}} f_{p,k}(x; \mu, \Sigma, \eta) = \frac{\eta}{2} \sum_{m=0}^{k} \frac{k!}{m!} \frac{\Gamma(\frac{p}{2} + m)}{\Gamma(\frac{p}{2} + k)} \frac{\partial^2}{\partial x_i \partial x_j} f_{p,m}(x; \mu, \Sigma, \eta). \tag{52}$$

This specializes to Plackett's identity when $k = 0$. Finally, Gordon [10] proved a definitive version of Plackett's identity for elliptically contoured densities (the proof of which he traced back to [8,18]). He showed that the following two statements about functions g and h, each mapping $\mathbb{R}_+$ into itself and vanishing at ∞, are equivalent:

$$h(t) = \frac{1}{2} \int_t^{\infty} g(r)\, dr \tag{53}$$

and

$$\frac{\partial}{\partial \sigma_{ij}} g_\Sigma(x) = \frac{\partial^2}{\partial x_i \partial x_j} h_\Sigma(x), \tag{54}$$

where $g_\Sigma(x) = |\Sigma|^{-1/2}\, g(x'\Sigma^{-1}x)$, and similarly for h. When g is an exponential or an appropriately chosen gamma density the identities of Plackett and Iyengar, (46) and (52), respectively, follow. Next, for the p-variate t with ν degrees of freedom, we have

$$h(t) = \frac{\Gamma((p+\nu)/2)}{(\pi\nu)^{p/2}\Gamma(\nu/2)} \frac{\nu}{(p+\nu-2)}(1+t)^{-(p+\nu-2)/2}. \tag{55}$$

These extensions of Plackett's identity have been used principally for theoretical investigations, in particular, for studying the nature of the dependence among the components of X. A systematic study of their use for the computation of moments of various functions (other than the usual product moments given by Γ_k) has not yet been done. The mathematical basis for Plackett's identity goes back to the 19^{th} century work of Schläfli [31] on hyperspherical simplices, and the later work of the geometer Coxeter [6]. For more on the geometrical aspects of Plackett's identity and related issues, see Abrahamson [1] Iyengar [16] and Ruben [28].

5 Conclusion

In this paper, we have discussed recent developments in probability and moment calculations for elliptically contoured distributions. These developments should allow the use of models other than the multivariate normal for high dimensional data. Clearly, much more work needs to be done. For instance, since Monte Carlo is an increasingly popular method for assessing the performance of various systems, a more systematic study of appropriate sampling functions is needed. Only the beginnings of such a study are given here.

6 Acknowledgment

This research was partly supported by the Office of Naval Research grant N00014-89-J-1496. This paper was written while I was visiting the Statistics Department at Carnegie-Mellon Uni

versity. I thank the department for the use of their facilities.

References

1. Abrahamson, I. (1964) Orthant Probabilities for the Quadrivariate Normal Distribution. *Annals of Mathematical Statistics.* 35, 1685-1703.

2. Anderson, T.W. and Fang, K.T., eds. (1990) *Statistical Inference in Elliptically Contoured and Related Distributions.* Allerton Press. New York.

3. Andrews, D. (1973) A General Method for the Approximation of Tail Probabilities. *Annals of Statistics.* 1, 367-372.

4. Brillinger, D. (1992) Higher Order Moments and Spectra. This volume.

5. Cambanis, S., Huang, S., Simons, G. (1981) On the Theory of Elliptically Contoured Distributions. *J. of Multivariate Analysis.* 11, 368-385.

6. Coxeter, H. (1973) *Regular Polytopes.* Dover. New York.

7. Cramér, H. (1946) *Methods of Mathematical Statistics.* Princeton U. Press. Princeton.

8. Das Gupta, S., Eaton, M., Olkin, I., Perlman, M., Savage, J., Sobel, M. (1972) Inequalities on the Probability Content of Convex Regions for Elliptically Contoured Distributions. *Proceedings of the Sixth Berkeley Symposium on Mathematics, Statistics, and Probability.* 2, 241-264.

9. Fang, K.T., Xu, J.L. (1990) The Mills' Ratio of Multivariate Normal Distributions and Spherical Distributions. In *Statistical Inference in Elliptically Contoured and Related Distributions.* K.T. Fang and T.W. Anderson, eds. Allerton Press. New York, 457-468.

10. Gordon, Y. (1987) Elliptically Contoured Distributions. *Probability Theory and Related Fields.* 76, 429-438.

11. Gray, H. and Wang, S. (1991) A General Method for Approximating Tail Probabilities. *J. of the American Statistical Association.* 86, 159-166.

12. Iyengar, S. (1984) A Geometric Approach to Probability Inequalities. University of Pittsburgh Technical Report.

13. Iyengar, S. (1986) On a Lower Bound for the Multivariate Normal Mills' Ratio. *Annals of Probability.* 14, 1399-1403.

14. Iyengar, S. (1988) Evaluation of Normal Probabilities of Symmetric Regions. *SIAM Jour-*

nal of Scientific and Statistical Computing. 9, 418-424.

15. Iyengar, S., Tong, Y. (1990) Convexity of Elliptically Contoured Distributions with Applications. *Sankhya A.* 51, 13-29.

16. Iyengar, S. (1990) Plackett's Identity, its Generalizations, and their Uses. To appear in the Festschrift for Charles Dunnett.

17. Iyengar, S. (1990) The Approximation of Tail Probabilities. U. of Pittsburgh Technical Report.

18. Joag-Dev, K., Perlman, M., Pitt, L. (1983) Association of Normal Random Variables and Slepian's Inequality. *Annals of Probability.* 11, 451-455.

19. Johnstone, I. (1986) A Program for Estimating Uncertainties in Quantile Estimates Derived from Empirical Pearson Fits. Stanford U. Technical Report.

20. Kotz, S. (1974) Multivariate Distributions at a Cross-Road. *Statistical Distributions in Scientific Work. Vol. 1.* G. P. Patil, S. Kotz, J.K. Ord, eds. 247-270.

21. Lavine, M. (1991) Problems in Extrapolation with Space-Shuttle O-Ring Data. *J. of the American Statistical Association.* 86, 919-921.

22. Lawless, J. (1982) *Statistical Models and Methods for Lifetime Data.* Wiley. New York.

23. Li, G. (1990) Moments of a Random Vector and its Quadratic Forms. In *Statistical Inference in Elliptically Contoured and Related Distributions.* K.T. Fang and T.W. Anderson, eds. Allerton Press. New York, 433-440.

24. Papoulis, A. (1965) *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill. New York.

25. Pawula, R. (1967) A Modified Version of Price's Theorem. *IEEE Transactions on Information Theory.* IT 13, 285-288.

26. Plackett, R. (1954) Reduction Formula for Multivariate Normal Integrals. *Biometrika*, 41, 351-360.

27. Price, R. (1958) A Useful Theorem for Non-linear Devices Having Gaussian Inputs. *IRE Transactions on Information Theory.* IT 4, 69-72.

28. Ruben, H. (1954) On the Moments of Order Statistics in Samples from Normal Populations. *Biometrika.* 41, 200-227.

29. Savage, I.R. (1962) Mills' Ratio for Multivariate Normal Distributions. *J. of Research of*

the National Bureau Standards. B 66, 93-96.

30. Scharf, L. (1991) *Statistical Signal Processing.* Addison-Wesley. New York.

31. Schläfli, L. (1858,1860) On the Multiple Integral $\int^n dxdy\ldots dz$, whose Limits are $p_1 = a_1x + b_1y + \ldots + h_1z > 0, p_2 > 0, \ldots, p_m > 0, x^2 + y^2 + \ldots z^2 < 1$. *Quarterly J. of Pure and Applied Mathematics.* 2, 261-301; 3, 54-68; 3, 97-107.

32. Slepian, D. (1962) The One-Sided Barrier Problem for Gaussian Noise. *Bell System Technical J.* 41, 463-501.

33. Thisted, R. (1988) *Elements of Statistical Computing.* Chapman and Hall. New York.

34. Tong, Y.L. (1990) *The Multivariate Normal Distribution.* Springer. New York.

35. Wessell, A., Hall, E., and Wise, G. (1988) Some Comments on Importance Sampling. *Proceedings of the 22nd Conference on Information Sciences and Systems.*

36. Xu, J.L., Fang, K.T. (1990) The Expected Values of Zonal Polynomials of Elliptically Contoured Distributions. In *Statistical Inference in Elliptically Contoured and Related Distributions.* K.T. Fang and T.W. Anderson, eds. Allerton Press. New York, 469-479.

FIGURE 1

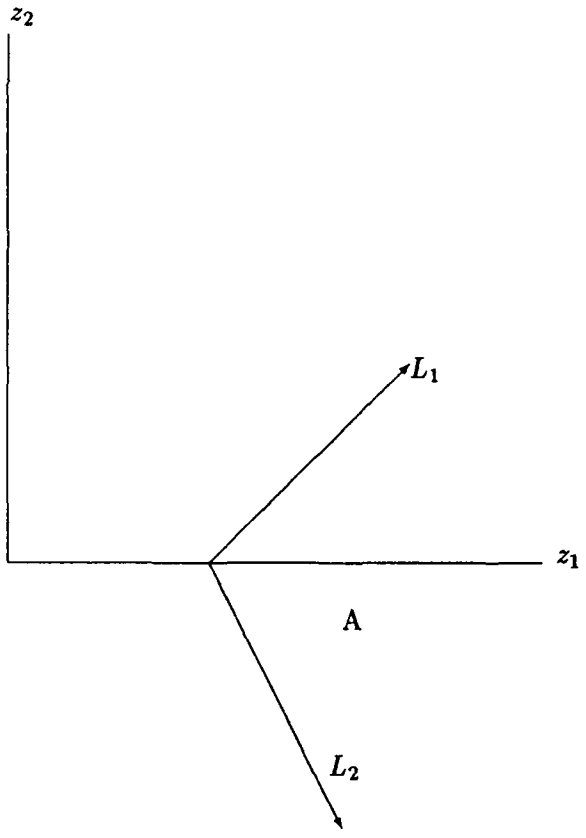
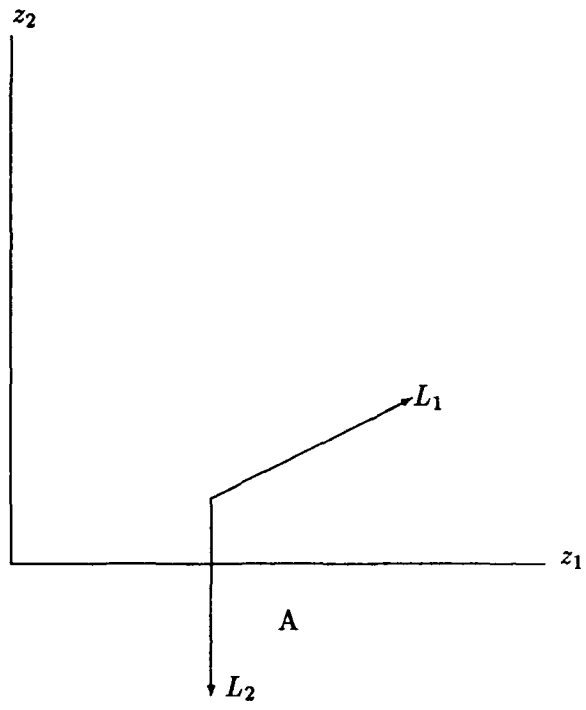


FIGURE 2



Legends for the two Figures

FIGURE 1: $\rho < b_1/b_2$; A is bounded by L_1 and L_2.

$$L_1 : z_2 = \frac{b_1(1-\rho^2)^{1/2}}{(b_2 - \rho b_1)}(z_1 - (b'R^{-1}b)^{1/2}), \text{ for } z_1 \geq (b'R^{-1}b)^{1/2}$$

$$L_2 : z_2 = \frac{-b_2(1-\rho^2)^{1/2}}{(b_1 - \rho b_2)}(z_1 - (b'R^{-1}b)^{1/2}), \text{ for } z_1 \geq (b'R^{-1}b)^{1/2}$$

FIGURE 2: $\rho \geq b_1/b_2$; A is bounded by L_1 and L_2.

$$L_1 : z_2 = \frac{(\rho z_1 - b_1)}{(1-\rho^2)^{1/2}}, \text{ for } z_1 \geq b_2$$

$$L_2 : z_1 = b_2, \text{ for } z_2 \leq \frac{(\rho b_2 - b_1)}{(1-\rho^2)^{1/2}}$$

Model Discrimination using Higher Order Moments*

by
David Guy and Keh-Shin Lii
Department of Statistics
University of California, Riverside
Riverside, CA 92521

Abstract

In this paper we discuss the problem discriminating among various non-linear time series models. While the method we propose is of a general nature we consider a restricted class of models that share an identical AR(1) equivalent correlation function structure ;hence, identical spectral density. Consequently, the possibility of discriminating among them on the basis of second order moments is theoretically, and practically, impossible. The approach being taken is aimed at discriminating among the models on the basis of higher order moments i.e. the higher order cumulant structure. Specifically, we shall focus on the 3^{rd}-order cumulant structures as our initial step beyond the conventional covariance structure.

Key Words : Time series, Linear, Non-linear, Gaussianity, Stationarity, Autoregressive, Exponential Models, PAR(1), ARE(1), EAR(1), TEAR(1), NEAR(1), Robertson's Fixed and Random Models, Correlation and Cumulant Structure.

*This research is supported by ONR contract N0004-85-k-0468 and ONR grant N00014- 92-J-1086.

1 Introduction

Statistical methods based on moment information have been used extensively. In terms of model identification the time series literature has been devoting a considerable attention to the problem of identifying the p and q order under the general linear framwork of ARMA(p,q) modelling. Second order correlation information (e.g. acf and pacf) became a main tool in the process of of selecting p and q. While second order information is of paramount importance in the case where the roots of the AR and MA polynomials remain outside the unit circle, higher order cumulant information becomes crucial in deciding on the locations of the zeros or poles of possibly non-invertible, non-causal and non-Gaussian ARMA models. Of course there are many very useful statistical tools for solving the above mentioned problems which are not based on moments. For example, the use of information based criteria such as AIC, MAIC and BIC in selecting orders of an ARMA model, the use of MLE in locating roots of a mixed phase ARMA process, ect. While these non-moment based methods might be more efficient than moments methods, the moments methods are generally simpler, easier and intuitvely appealing both in theory and computation. It is often the case that one needs the initial point supplied by such a method to start an efficient but complicated non-moment based method.

The introduction of non-linear time series models in recent years (e.g. bilinear, threshhold, random coefficient, ect.) amplified the importance of using higher order cumulant information in discriminating among the various non-linear models. It was shown that different models are capable of producing an identical correlation function of the linear autoregressive type; thus, giving rise to a class of models characterized as 2^{cd}-order equivalent. Consequently, efforts have been diverted to the analysis of the higher order cumulant structure with the hope of exploiting differences among the models at higher order correlation dependency structure. The basic idea underlying the search for information in the higher order cumulant structure in order to distinguish two models may be stated as follows. Within the class of

moment determination, moment sequences of two different stochastic processes cannot be identical. Specifically, given two stationary series $\{X_t\} \neq \{Y_t\}$ there exists $(u_1, u_2, \ldots, u_k)$ such that k^{th}-order moments or cumulants with lags $(u_1, u_2, \ldots, u_k)$ of $\{X_t\}$ and $\{Y_t\}$ are not equal, i.e.

$$C_x(u_1, u_2, \ldots, u_k) \neq C_y(u_1, u_2, \ldots, u_k).$$

In practice, one hopes that the above is true for a small order k, and the difference is large relative to a given sample size. Otherwise, the search for a discriminatory power in the higer order cumulant structure might turn out to be fruitless.

The problem of discrimination among non-linear time series models has been considered by many authors. Lawrence and Lewis [24] considered special 3^{rd}-order structure of the form

$$Cov(R_t^{(p)}, X_{t+k}^2) \quad , \quad Cov([R_t^{(p)}]^2, X_{t+k})$$

where $R_t^{(p)}$ are the linear autoregressive residuals of order p for RCA and PAR models . Within the class of bilinear models Li [26] and Gabr [10] considered quantities of the form

$$Cov(X_t^2, X_{t+k}^2)$$

$$Cov(X_t^2, X_{t+k})$$

respectively. Auestad and Tjøstheim [4] considered the use of non-parametric methods aimed at the conditional mean and variance of various non-linear time series models. Anderson [1] approached this problem differently by observing differences in the sample paths generated by the exponential family. Using a fluctuating type statistic he was able to discriminate among simulated traces for a reasonable number of observations. In his work the moments do not play a role in the proposed discrimination procedure and as such may provide an alternative in situations where moments up to the desired order do not exist. Tsay [37] offers a very general method for selecting a model depending on the type of characteristic one is interested to investigate.

We propose a new approach which relies on the conjecture that the information required for discrimination among the models is available in the higher order moments or equivalently,

in the higher order cumulant structure. Specifically, we shall concentrate our attention on the third order cumulant structure given by

$$C(r,s) = E[(X_t - \mu_x)(X_{t-r} - \mu_x)(X_{t-s} - \mu_x)]. \tag{1.1}$$

The family of exponential time series models will be the framework within which we shall show the parametric equality of the correlation (hence, the spectral density) functions, and the way in which the theoretical higher order cumulant structure points out to the differences among the models. We demonstrate the method for a restricted case where we consider a family of non-linear time series models with known marginal distributions and a common AR(1) equivalent correlation structure. This family consists of marginal exponentially distributed time series models which include :

- (i) Product Autoregressive Model [PAR(1)]
- (ii) Exponential Autoregressive Model [EAR(1)]
- (iii) Transposed Exponential Autoregressive Model [TEAR(1)]
- (iv) Newer Exponential Autoregressive Model [NEAR(1)]
- (v) Robertson's Fixed Model
- (vi) Robertson's Random Model

In addition we shall consider the linear autoregressive model with exponential innovation process which we shall call ARE(1). As opposed to the family mentioned above the ARE(1) does not have a known marginal distribution ;however, its moments can be computed. This model, though, shares the same correlation structure as the non-linear exponential family. The underlying objective is to discriminate among realizations produced by the models we consider. This task is impossible to accomplish since they have identical second order

structure. We should note that the models being considered by no means exhaust all such second order equivalent ones.

The plan for the paper is as follows, we shall start with a brief review of the traditional approach to time series analysis , followed by a presentation of the family of exponential time series models in section 3. In that section we shall state the form taken up by each model, show the type of sample traces they are capable of producing and develop their correlation functions. Then we give a brief review of higher order cumulants in section 4. Subsequently, the results we obtained for the 3^{rd}-order cumulant structure for the seven models under considerations are presented in section 5. General methodology is presented in section 6. The results of the simulation part are the topic of section 7. There we also briefly discuss the way in which the sample traces, correlation functions and 3^{rd}-order cumulants were generated empirically. A brief conclusion is given in section 8.

2 Stationarity, Linearity and Gaussianity

Over the last 50 years statisticians have developed a large body of theory and methods aimed at the analysis of time series data. A comprehensive account of their work culminated in books such as Kendall and Stuart [17], Jenkins and Watts [15], Box and Jenkins [5], Hannan [12], Anderson [2], Brillinger [7], Chatfield [9], Koopmans [18], Priestley [30], Rosenblatt [32], and Brockwell and Davis [8], to name a few. The foundations of classical time series analysis, as described in the above references, were thought to be based on two underlying assumptions, stating that :

1. The time series is *stationary* to an order of at least two. The process is assumed to remain in equilibrium about a constant mean level with the proportion of ordinates not exceeding any given level is about equal over any time interval spanned by the sample. In case the observed series does not exhibit such behavior, it is further assumed that

weak stationarity can be achieved by applying an appropriate transformation e.g. linear filtering.

2. The time series, viewed as a stochastic process, $\{X_t, t \in T\}$, is an output from a *linear filter* whose input is the *white noise* process $\{Z_t\}$; hence, the observed sample realization can be represented as a linear function of past and present values of $\{Z_t\}$ - a one sided representation.

In recent years the validity of these twin assumptions - as reasonable approximations to sample trace realizations - has been questioned as data from a wider variety of sources became available. Coupled with advances in the field of non-linear dynamics (deterministic chaos theories), research in the field of non-stationary, non-linear and non-Gaussian time series methodology have been in progress. Subsequent efforts to bring non-linear time series literature under one unified framework resulted in the publication of books like Priestley [29] and Tong [36]. The reader is also referred to Mohler [28] for a collection of papers on theory, computational methods and applications in the area of non-linear signal processing. Tong [36] discusses properties of the Gaussian stationary linear model (GSLM) which may possibly be violated :

- (a) Time series that exhibit strong asymmetric behavior cannot be expected to confirm to the GSLM. Such models are characterized by symmetric joint cumulative density functions and that rules out asymmetric sample realizations.

- (b) The GSLM does not give rise to clusters of outliers e.g. sudden bursts of large magnitudes at irregular time intervals. Observed time series in socio-economic related phenomena do tend to exhibit groups of outliers.

- (c) Sample traces that demonstrate strong cycles cannot be modeled by the GSLM since the regression functions at lag (k) i.e. $E[X_t|X_{t-k}]$ are all linear due to the assumed joint normality.

- (d) The Gaussian process $\{X_t\}$ is reversible i.e. $(X_1, ..., X_{t_n})'$ has the same distribution as $(X_{t_n}, ..., X_1)'$. Reversibility is violated in the presence of differences in the rate at which a sample path rises to its maxima, and the rate at which it falls away from it. One simple way for investigating departures from reversibility is to plot the sample on a transparency and then turn it over. If the mirror image is similar to the original plot then the series may be assumed reversible - irreversible otherwise.

One could also test formally for Gaussianity and linearity. Following Brillinger [6], who pointed out to the potential of using the bispectral density function as the basis for classifying a process as linear (and possibly Gaussian) or non-linear, Subba Rao and Gabr [35] and Hinnich [13] developed formal tests for linearity and Gaussianity. The tests are based on the constancy of the normalized bispectral density function under the assumption that $\{X_t\}$ have a linear representation. Tong [36]) provides a comprehensive review of tests for linearity and normality. Priestley [29] considers the case where a stationary process does not fit into a linear representation and concludes that "*a fortiori* many types of non-stationary processes would also fall outside the domain of linear models." In summary, observed time series do not necessarily conform to models such as the GSLM. The degree to which a time series realization represents a trace generated by the GSLM, has a direct bearing on the usefulness of estimating an ARMA(p,q) model. For purposes of prediction, forecasting and control one is better off taking advantage of the non-linear (hence, non-Gaussian) structure of the data during the modeling stage. If indeed the GSLM is deemed inappropriate, one has the choice among several families of non-linear models. We shall turn to some of these explicitly in section 3.

3 AR(1) Type Exponential Models (EAR)

The family of models we consider here is that of the *exponential autoregressive* models, which is composed of the EAR(1) and its generalization to the *transposed exponential autoregressive* model TEAR(1), and the *newer exponential autoregrssive* NEAR(1) model. This type of time series models were proposed by Gaver and Lewis [11], Lawrance and Lewis [20, 21], Jacobs and Lewis [14], Lawrance [19] and further developed by Lawrance and Lewis [22, 23, 24]. Also we consider Robertson's Fixed and Random models [31], and the Product Autoregressive PAR(1) model proposed by McKenzie [27] - where all models being restricted to a first order autoregressive structure.

In contrast with other non-linear time series models (e.g. bilinear and threshold), this class of models is an attempt to capture the behavior of, possibly observed, time series processes with explicit marginal exponential distributions. The family of EAR models is advocated as a way of relaxing the assumption of marginal Gaussianity which underlies the Gaussian linear stationary model. The reasons behind the choice of the exponential distribution as the marginal distribution are given in Gaver-Lewis [11] and Lawrance and Lewis [23]. The standard linear first order autoregressive process, AR(1), with exponential input, ARE(1), will be used for comparison purposes in section 5. This model has an identical correlation and spectral density functions as do the models mentioned above :however, its marginal distribution is not known, thus, it is not to be considered as an exponential model but rather as a linear AR(1) model with exponential input. The fact that it is linear enables us to distinguish it from any other non- linear model, with or without an identical correlation structure, based on the theoretical result stating that a process with a linear representation has a flat (constant) normalized bispectral density, for more details see Subba Rao and Gabr [35].

3.1 PAR(1) Model

A natural extension of the linear AR(1) model was proposed by McKenzie [27] and consists of an exponentiation of the linear model such that the additive form is being transformed into a multiplicative form. Here we consider a sepcial case of the gamma family of marginally distributed time series where the output series has an exponetial marginal distribution of unit mean. Specifically,

$$X_t = X_{t-1}^{\alpha} V_t \tag{3.1}$$

where $\alpha \in (0, 1)$ and V_t is given by a mixture of uniform $(0, \pi)$ and exponential mean one random variables independent of each other.

This model differs from the others we consider in two aspects. First, the innovation process does not posess a known parametric density function and its higher order cumulant structure is expressed in terms of the moments of X_t only. Second, we note that (3.1) may be linearized by taking the logs of both sides of the equation. As such it is classified as an *intrinsically* linear model i.e. a non-linear model which can be linearized. It differs from the following models which cannot be linearized due to their switching nature and are to be considered under the class of *intrinsically* non-linear models i.e. a non-linear model which can not be linearized.

3.2 EAR(1) Model

In the following set up we let $\{E_t\}$ be a sequence of i.i.d exponential (λ) random variables with a probability density function given by

$$f_E(\epsilon) = \begin{cases} \lambda e^{-\lambda \epsilon} & \epsilon \geq 0 \ , \ \lambda > 0 \\ 0 & \text{otherwise} \ . \end{cases} \tag{3.2}$$

We define an EAR(1) model as,

$$
\begin{aligned}
X_t &= \rho X_{t-1} + \varepsilon_t && (3.3)\\
&= \begin{cases} \rho X_{t-1} & \text{with prob. } \rho \\ \rho X_{t-1} + E_t & \text{with prob. } (1-\rho) \end{cases} && (3.4)\\
&= \rho X_{t-1} + I_t E_t && (3.5)
\end{aligned}
$$

with $(0 \leq \rho < 1)$ and $\{I_t\}$ being an i.i.d sequence defined by

$$
I_t = \begin{cases} 0 & \text{with prob. } \rho \\ 1 & \text{with prob. } 1-\rho. \end{cases} \qquad (3.6)
$$

Under this formulation $\{X_t\}$ is marginally distributed as an exponential random variable with parameter λ.

Gaver and Lewis [11] point out to several characteristics of the EAR(1) model:

- Setting $\rho = 0$ yields the special case where $\{X_t\}$ is a sequence of i.i.d exponential random variables.
- ε_t is not a continuous random variable. This feature distinguishes (3.5) from the usual linear AR(1) equation with Gaussian or exponential input.
- The representation (3.5) is one of a random linear combination of an i.i.d exponential sequences; thus, can be easily simulated on a computer.

One problem the EAR(1) model has is called 'zero defect' (see Lawrance and Lewis [22]) and relates to the sample paths it generates. Specifically, the model generates paths in which large values are followed by runs of decreasing values, with the runs having geometrically distributed lengths. The large values arise when E_t is included (i.e $I_t = 1$) while the falling values stem from the deterministic part of (3.5) (i.e $I_t = 0$).

3.3 TEAR(1) Model

A natural extension of the EAR(1) model is to interchange the role of X_{t-1} and E_t in (3.5). This does not affect the exponential (λ) marginal distribution of X_t. Upon replacing ρ by $1-\alpha$ we obtain the *transposed exponential autoregressive* TEAR(1) model

$$X_t = I_t X_{t-1} + (1-\alpha)E_t \tag{3.7}$$

$$= \begin{cases} X_{t-1} + (1-\alpha)E_t & \text{with prob. } \alpha \\ (1-\alpha)E_t & \text{with prob. } 1-\alpha \end{cases} \tag{3.8}$$

where

$$I_t = \begin{cases} 0 & \text{with prob. } 1-\alpha \\ 1 & \text{with prob. } \alpha. \end{cases} \tag{3.9}$$

Note that in this case the innovation process is a continuous random variable scaled by a constant $1-\alpha$. The behavior of a simulated path, for a large α, shows geometrically distributed runs of rising values (i.e. $I_t = 1$) followed by sharp declines when the selection $I_t = 0$ is made. The decline due to the exclusion of the previous value X_{t-1}.

The TEAR(1) model is discussed by Lawrance and Lewis [22] as an extension of the EAR(1) model. However, TEAR(1) is also a special case of Arnold's [3] exponential model driven by past innovations. Specifically, define the random variables

$$N_t = 1 \quad \text{if and only if} \quad U_t = 1$$

$$N_t = i \quad \text{if and only if} \quad U_t = 0, U_{t-1} = 0 \ldots U_{t-i+1} = 1$$

where U_t are i.i.d Bernoulli(p) random variables with N_t being distributed identically but not independently as Geometric(α) random variables with domain $1, 2, 3, \ldots$.

The model, expressed in terms of past innovations, is given by

$$X_t = \alpha \sum_{i=1}^{N_t} \varepsilon_{t-i+1} \tag{3.10}$$

where $\varepsilon_t \sim$ iid Exp(λ) and the sum is multiplied by α to obtain strict stationarity. This representation is obtained if one express the TEAR(1) model (3.8) recurssively.

3.4 NEAR(1) Model

The previous two models, EAR(1) and TEAR(1), are special cases of a more flexible model in which $\{X_{t-1}\}$ in (3.8) is scaled by a coefficient β; thus, simulated realizations generated by such model are of interest as it may circumvent the problem of geometrically distributed runs of falling or increasing values which might not be applicable. Specifically, let $\{X_t\}$ denote the time series variables and $\{E_t\}$ be a sequence of an i.i.d unit mean exponential random variables acting as the innovation process. The NEAR(1) model is defined as

$$X_t = \varepsilon_t + \begin{cases} \beta X_{t-1} & \text{with prob. } \alpha \\ 0 & \text{with prob. } 1-\alpha \end{cases} \tag{3.11}$$

$$= \beta I_t X_{t-1} + \varepsilon_t \tag{3.12}$$

where

$$\varepsilon_t = \begin{cases} E_t & \text{with prob. } p \\ bE_t & \text{with prob. } 1-p \end{cases} \tag{3.13}$$

$$I_t = \begin{cases} 0 & \text{with prob. } 1-\alpha \\ 1 & \text{with prob. } \alpha \end{cases} \tag{3.14}$$

with $b = (1-\alpha)\beta$ and $p = \frac{1-\beta}{1-(1-\alpha)\beta}$. The parameters α and β are allowed to take values over the domain defined by $0 \leq \alpha, \beta \leq 1$ with $\alpha, \beta \neq 1$. Setting ($\alpha = 1$, $0 \leq \beta < 1$) in (3.12) yields the EAR(1) model, where fixing ($\beta = 1$, $0 \leq \alpha < 1$) give rise to the TEAR(1) model. Both are extreme cases of a NEAR(1) process. We note that due to the distributional assumption underlying $\{E_t\}$, the innovation process is not allowed to take on negative values i.e. $P[E_t \leq 0] = 0$. It is obvious how the concept of "switching" comes into play in (3.12). The switch from one linear piece to the other is controlled by an external random mechanism with a prespecified parametric probabilistic structure.

3.5 Robertson's Fixed and Random models

Robertson [31] suggested two exponential models which we shall refer to as Robertson's fixed and random models. Our main concern is to show that these models cannot be identified via the correlation or spectral density functions ;hence, one has to explore the higher order cumulant structure.

3.5.1 The Fixed Model

Consider the following switching structure

$$X_t = \begin{cases} X_{t-1} - ln\beta & \text{with prob. } \beta \\ E_t & \text{with prob. } 1-\beta \end{cases} \tag{3.15}$$

where β is a fixed constant, E_t has a truncated exponential distribution given by

$$f_E(\epsilon) = \begin{cases} \frac{1}{1-\beta}e^{-\epsilon} & 0 < \epsilon < -ln\beta \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

with the marginal distribution of X_t being exponential with unit mean. Alternatively, (3.15) may be represented using an indicator random variable i.e.

$$X_t = I_t(X_{t-1} - ln\beta) + (1 - I_t)E_t \tag{3.17}$$

where

$$I_t = \begin{cases} 1 & \text{with prob. } \beta \\ 0 & \text{with prob. } 1-\beta. \end{cases} \tag{3.18}$$

3.5.2 The Random model

One may generalize the fixed model by allowing β to become a random variable which acts as a mixing distribution, with domain restricted to the interval [0,1]. Specifically, let X_t have

the representation

$$X_t = \begin{cases} X_{t-1} - ln\beta_t & \text{with prob. } \beta_t \\ E_t & \text{with prob. } 1 - \beta_t \end{cases} \tag{3.19}$$

or stated in terms of an indicator random variable

$$X_t = I_t(X_{t-1} - ln\beta_t) + (1 - I_t)E_t \tag{3.20}$$

where

$$I_t = \begin{cases} 1 & \text{with prob. } \beta_t \\ 0 & \text{with prob. } 1 - \beta_t. \end{cases} \tag{3.21}$$

The probability density assigned to β_t is a beta density with parameters $(\alpha, 2)$

$$f_{\beta_t}(\beta) = \begin{cases} \alpha(\alpha+1)(1-\beta)\beta^{\alpha-1} & 0 < \beta < 1 \ , \ \alpha > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.22}$$

The distribution of $ln\beta_t$ is obtained using the standard transformation of variables technique. Let $Y = ln\beta_t$ then

$$f_Y(y) = \begin{cases} \alpha(\alpha+1)(1-e^y)e^{\alpha y} & -\infty < y < 0 \ , \ \alpha > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.23}$$

The probability density function for E_t is appropriately modified

$$f_E(\epsilon) = \begin{cases} \frac{1}{1-\beta_t}e^{-\epsilon} & 0 < \epsilon < -ln\beta_t \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Within this framework one notices that the random variables I_t and E_t are not independent as they both involve the mixing distribution β_t. The marginal distribution of X_t, though, remains exponential with unit mean by construction. We remark that all these models are stationary in the wide sense i.e. strictly stationary.

3.6 Summary

We recall that the models under investigation are the following :

- **ARE(1) :**

$$X_t = \phi X_{t-1} + E_t \tag{3.25}$$

- **PAR(1) :**

$$X_t = X_{t-1}^{\alpha} V_t \tag{3.26}$$

- **EAR(1) :**

$$X_t = \begin{cases} \rho X_{t-1} & \text{with prob. } \rho \\ \rho X_{t-1} + E_t & \text{with prob. } 1-\rho \end{cases} \tag{3.27}$$

- **TEAR(1) :** $(\rho = 1 - \alpha)$

$$X_t = \begin{cases} X_{t-1} + (1-\alpha)E_t & \text{with prob. } \alpha \\ (1-\alpha)E_t & \text{with prob. } 1-\alpha \end{cases} \tag{3.28}$$

- **NEAR(1) :**

$$X_t = \begin{cases} \beta X_{t-1} + \varepsilon_t & \text{with prob. } \alpha \\ \varepsilon_t & \text{with prob. } 1-\alpha \end{cases} \tag{3.29}$$

where,

$$\varepsilon_t = \begin{cases} E_t & \text{with prob. } p \\ bE_t & \text{with prob. } 1-p \end{cases}$$

$$p = \frac{1-\beta}{1-b} \qquad b = (1-\alpha)\beta$$

- **Roberston's Fixed Model :**

$$X_t = \begin{cases} X_{t-1} - ln\beta & \text{with prob. } \beta \\ E_t & \text{with prob. } 1-\beta \end{cases} \tag{3.30}$$

where

$$f_E(\epsilon) = \begin{cases} \frac{1}{1-\beta}\epsilon^{-\epsilon} & 0 < \epsilon < -ln\beta \\ 0 & \text{otherwise} \end{cases}$$

- **Roberston's Random Model :**

$$X_t = \begin{cases} X_{t-1} - ln\beta_t & \text{with prob. } \beta_t \\ E_t & \text{with prob. } 1-\beta_t \end{cases} \tag{3.31}$$

where

$$f_E(\epsilon) = \begin{cases} \frac{1}{1-\beta_t}\epsilon^{-\epsilon} & 0 < \epsilon < -ln\beta_t \\ 0 & \text{otherwise} \end{cases}$$

$$f_{\beta_t}(\beta) = \begin{cases} \alpha(\alpha+1)(1-\beta)\beta^{\alpha-1} & 0 < \beta < 1\ ,\ \alpha > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Table 1: Correlation Functions

ARE(1)	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
ϕ^s	α^s	ρ^s	α^s	$(\alpha\beta)^s$	β^s	$(\frac{\alpha}{\alpha+2})^s$

For all models, but Robertson's and PAR(1), the input process $\{E_t\}$ is assumed to be an i.i.d exponential sequence of unit mean and, with the exception of ARE(1), the output $\{X_t\}$ has a marginal exponential distribution with mean one. The correlation functions for the various models are given in table 1.

Figures 1-3 contain simulated traces produced by the various models. Note that we indexed the parameter values of each model such that the correlation functions produce identical results i.e $\rho(s) = (0.1)^s, (0.5)^s, (0.75)^s$.

4 Higher Order Cumulants

Let $\{X_t\}$ be a real valued strictly stationary random process and let $m(t_1, t_2, \ldots, t_k)$ be the k^{th}-order product moment i.e.

$$m(t_1, t_2, \ldots, t_k) = E[X_{t_1} X_{t_2} \ldots X_{t_k}]. \tag{4.1}$$

For a stationary process of order k, we can write (4.1) as

$$m(t_1, t_2, \ldots, t_k) = m(0, t_2 - t_1, t_3 - t_1, \ldots, t_k - t_1). \tag{4.2}$$

Now let the characteristic function (cf) of $\{X_t\}$ be defined by

$$\phi_X(\zeta_1, \zeta_2, \ldots, \zeta_k) = E[e^{i(\zeta_1 X_{t_1} + \zeta_2 X_{t_2} + \ldots + \zeta_k X_{t_k})}] \tag{4.3}$$

then the Taylor series expansion of (4.3) about the origin is given by

$$\phi_X(\underline{\zeta}) = \int \left\{ \sum_{j=1}^{k} \frac{i^j(\zeta_1\zeta_2\ldots\zeta_j)}{j!}(X_{t_1}X_{t_2}\ldots X_{t_j}) + O(|\underline{\zeta}|^k) \right\} dF \tag{4.4}$$

$$= \sum_{j=1}^{k} \frac{i^j(\zeta_1\zeta_2\ldots\zeta_j)}{j!} E[X_{t_1}X_{t_2}\ldots X_{t_j}] + O(|\underline{\zeta}|^k) \tag{4.5}$$

$$= \sum_{j=1}^{k} \frac{i^j(\zeta_1\zeta_2\ldots\zeta_j)}{j!} m(t_1, t_2, \ldots, t_k) + O(|\underline{\zeta}|^k) \tag{4.6}$$

where $|\underline{\zeta}| = \left\{\sum_{i=1}^{k} \zeta_i^2\right\}^{\frac{1}{2}}$ and $F = F_{X_{t_1}, X_{t_2}, \ldots, X_{t_k}}(x_{t_1}, x_{t_2}, \ldots, x_{t_k})$ being the joint cumulative distribution function .

The logarithm of the cf (4.3) is defined as the cumulant generating function (cgf)

$$K_X(\zeta_1, \zeta_2, \ldots, \zeta_k) = log\{E[e^{i(\zeta_1 X_{t_1} + \zeta_2 X_{t_2} + \ldots + \zeta_k X_{t_k})}]\} \tag{4.7}$$

such that $C(t_1, t_2, \ldots, t_k)$, the k^{th}-order joint cumulant of the set of random variables $\{X_{t_1}, X_{t_2}, \ldots, X_{t_k}\}$, is the coefficient of $(\zeta_1, \zeta_2, \ldots, \zeta_k)$ in the Taylor series expansion of (4.7) about the origin. Specifically

$$K_X(\underline{\zeta}) = \sum_{j=1}^{k} \frac{i^j(\zeta_1\zeta_2\ldots\zeta_j)}{j!} C_j(t_1, t_2, \ldots, t_j) + O(|\underline{\zeta}|^k) \tag{4.8}$$

where $C_j(t_1, t_2, \ldots, t_j) = \text{Cumulant}(X_{t_1}, X_{t_2}, \ldots, X_{t_j})$. We note that the cumulant of order greater than two are all zero for a Gaussian process. This feature is used extensively in signal processing to suppress Gaussian noise.

The relationship between moments and cumulants were formalized by Leonov and Shiryaev [25] and are given by

$$m(t_1, \ldots, t_k) = E[X_{t_1}X_{t_2}\ldots X_{t_k}] = \sum_{\nu} C(\nu_1)C(\nu_2)\ldots C(\nu_p) \tag{4.9}$$

where the sum is taken over all partitions $(\nu_1, \ldots, \nu_p)$ which is a partition of $(t_1, \ldots, t_k)$. Relationship (4.9) implies that we can write the moments in terms of the cumulants and if

we invert (4.9) then one can write the cumulants in terms of the corresponding moments; hence, the inversion of (4.9) yields

$$C(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) = \sum_{\nu} (-1)^{p-1}(p-1)! E(\prod_{j \in \nu_1} X_j) \ldots E(\prod_{j \in \nu_p} X_j) \qquad (4.10)$$

and if the process is k^{th}-order stationary then we may write

$$\begin{aligned} C(t_1, t_2, \ldots, t_k) &= C(0, t_2 - t_1, \ldots, t_k - t_1) \\ &= C(\tau_1, \tau_2, \ldots, \tau_{k-1}). \end{aligned}$$

From (4.10) it is seen that the cumulant $C(\tau_1, \tau_2, \ldots, \tau_{k-1})$ is a k^{th}-order polynomial in the moments of no higher than k and conversely, the k^{th}-order moment $m(t_1, t_2, \ldots, t_k)$ is a k^{th}-order polynomial in cumulants of order no higher than k. Consider the specific cases :

$$\begin{aligned} C_1(0) &= E[X_t] = \mu_x \\ C_2(s) &= \mu(s) - \mu_x^2 \\ C_3(s_1, s_2) &= \mu(s_1, s_2) \\ &- \{\mu(s_1) + \mu(s_2) + \mu(s_2 - s_1)\}\mu_x + 2\mu^3 \\ C_4(s_1, s_2, s_3) &= \mu(s_1, s_2, s_3) \\ &- \mu_x\{\mu(s_2 - s_1, s_3 - s_1) + \mu(s_2, s_3) + \mu(s_1, s_2) + \mu(s_1, s_3)\} \\ &+ 2\mu_x^2\{\mu(s_1) + \mu(s_2) + \mu(s_3) + \mu(s_2 - s_1) + \mu(s_3 - s_1) + \mu(s_3 - s_2)\} \\ &- \mu(s_1)\mu(s_3 - s_2) - \mu(s_2)\mu(s_3 - s_1) - \mu(s_3)\mu(s_2 - s_1) - 6\mu_x^4 \end{aligned}$$

where

$$\begin{aligned} \mu(s) &= E[X_t X_{t+s}] \\ \mu(s_1, s_2) &= E[X_t X_{t+s_1} X_{t+s_2}] \\ \mu(s_1, s_2, s_3) &= E[X_t X_{t+s_1} X_{t+s_2} X_{t+s_3}] \end{aligned}$$

Consequently, one may write $C_3(s_1, s_2)$ in the form

$$\begin{aligned} C(r,s) &= E[(X_t - \mu_x)(X_{t+r} - \mu_x)(X_{t+s} - \mu_x)] \\ &= E(X_t X_{t+r} X_{t+s}) - \mu_x[E(X_t X_{t+r}) + E(X_t X_{t+s}) + E(X_{t+r} X_{t+s})] + 2\mu_x^3 \end{aligned} \tag{4.11}$$

and $C_4(s_1, s_2, s_3)$ may be expressed as

$$\begin{aligned} C(r,s,u) &= E[X_t X_{t+r} X_{t+s} X_{t+u}] \\ &- \mu_x(E[X_t X_{t+s-r} X_{t+u-r}] + E[X_t X_{t+s} X_{t+u}] + E[X_t X_{t+r} X_{t+s}] + E[X_t X_{t+r} X_{t+u}]) \\ &+ 2\mu_x^2(E[X_t X_{t+r}] + E[X_t X_{t+s}] + E[X_t X_{t+u}] \\ &+ E[X_t X_{t+s-r}] + E[X_t X_{t+u-r}] + E[X_t X_{t+u-s}]) \\ &- E[X_t X_{t+r}]E[X_t X_{t+u-s}] - E[X_t X_{t+s}]E[X_t X_{t+u-r}] - E[X_t X_{t+u}]E[X_t X_{t+s-r}] - 6\mu_x^4. \end{aligned} \tag{4.12}$$

For a detailed account of the relations between moments and cumulants the reader is advised to consult Kendall, Stuart and Ord [16]. Cumulants and their relationship to spectral analysis are discussed by Sesay [34] and Rosenblatt [33]. Sesay [34] discusses the various uses of cumulants and cumulants spectra, specifically

- Cumulant spectra is used in tests aimed at discriminating between linear and non-linear non-Gaussian processes (see Subba Rao and Gabr [35]).
- The asymptotic-distributions in some non-linear theory may be obtained using cumulants.
- Time reversibility may be determined by verifying $C(-s_1, \ldots, -s_{k-1}) = C(s_1, \ldots, s_{k-1})$ or equivalently the imaginary part of the k^{th}-order spectrum is equal to zero.
- Cross-cumulants, and cross-cumulant spectra, can be used in the estimation of the parameters of a non-linear difference equation through the use of transfer functions that arise in the Volterra expansion (see Priestley [30]).

5 The 3^{rd}-Order Cumulant Structure

In the following we shall present the 3^{rd}-order cumulant structure for each of the models discussed in section 3. For each of the models a closed form solution to the 3^{rd}-order cumulant structure is given. These solutions are based on closed form expressions obtained for the expectation terms which define the 3^{rd}-order cumulant structure. For all - but ARE(1) model - the output process is marginally distributed as an exponential process with unit mean. The results presented in this section are based on the marginal moments given by

$$\mu_x = 1 \quad \mu_{x,2} = 2 \quad \mu_{x,3} = 6$$

The input process is taken as an i.i.d exponential process with unit mean; hence, with identical moments as stated above. Robertson's and PAR(1) models form a separate class, in this respect, since the innovation process is defined by a sequence of i.i.d truncated exponential random variables and a mixture of exponetial and uniform random variables, respectively. The introduction of a mixing distribution in Robertson's random model further complicates the structure of the innovation process. Tables 2, 3 and 4 list the 3^{rd}-order cumulant structure for these models. We recall that the models under investigation are given by (3.25)-(3.31).

The following expressions are used in tabels 2, 3, 4, 5 and 6

$$\begin{aligned}
\mu_x &= \frac{1}{1-\phi} \\
\mu_{x,2} &= \frac{2}{(1-\phi^2)(1-\phi)} \\
\mu_{x,3} &= \frac{6}{(1-\phi^3)(1-\phi^2)(1-\phi)} \\
\gamma(\tau) &= \frac{1-\phi^\tau}{1-\phi} \\
\gamma_2(\tau) &= \frac{1-\phi^{2\tau}}{1-\phi^2} \\
\gamma(2\tau) &= \frac{1-\phi^{2\tau}}{1-\phi}
\end{aligned}$$

$$
\begin{aligned}
\mu_{x,\alpha} &= E[X^{\alpha}] = \Gamma(1+\alpha)\ ,\ \alpha \in (0,1) \\
\mu_{\varepsilon} &= p + (1-p)b \\
\mu_{\varepsilon,2} &= 2[p + (1-p)b^2] \\
\gamma_{\beta}(\tau) &= \frac{1-\beta^{\tau}}{1-\beta} \\
\gamma_{\alpha,\beta}(\tau) &= \frac{1-(\alpha\beta)^{\tau}}{1-\alpha\beta} \\
\gamma_{\lambda}(\tau) &= \frac{1-\lambda^{\tau}}{1-\lambda} \\
\lambda &= \alpha\beta^2 \\
a &= E[ln\beta_t I_t] - \frac{1}{1+\alpha} + E[E_t I_t] \\
b_t &= E[ln\beta_t I_t^2] = E[ln\beta_t I_t] \\
c &= E[(ln\beta_t)^2 I_t^2] + \frac{2}{(1+\alpha)^2} - E[E_t^2 I_t] \\
d &= E[X_t X_{t-1}] = \varrho\mu_{x,2} - a\mu_x \\
\gamma(\tau) &= \frac{1-\varrho^{\tau}}{1-\varrho} \\
\varrho &= \frac{\alpha}{\alpha+2} \\
E[ln\beta I_t] &= E[ln\beta I_t^2] = \alpha(\alpha+1)\left\{\frac{1}{(\alpha+2)^2} - \frac{1}{(\alpha+1)^2}\right\} \\
E[(ln\beta)^2 I_t] &= E[(ln\beta)^2 I_t^2] = \alpha(\alpha+1)\left\{\frac{2}{(\alpha+1)^3} - \frac{2}{(\alpha+2)^3}\right\} \\
E[E_t I_t] &= E[E_t I_t^2] = \alpha(\alpha+1)\left\{\frac{1}{\alpha+1} - \frac{1}{\alpha+2} - \frac{1}{(\alpha+2)^2}\right\} \\
E[E_t^2 I_t] &= E[E_t^2 I_t^2] = 2\alpha(\alpha+1)\left\{\frac{1}{\alpha+1} - \frac{1}{\alpha+2} - \frac{1}{(\alpha+2)^2} - \frac{1}{(\alpha+2)^3}\right\}
\end{aligned}
$$

Given the information summarized in these tables one may standardize the rate of decay of the correlation function such that the correlation functions are identical for these models for a given parameter value. Our goal is to investigate how would the 3^{rd}-order cumulant structure behave subject to a standardized correlation function. It is our conjecture that one might be able to discriminate among signal paths produced by the various models on the basis of higher order moments. It is obvious that the correlation functions can not be

Table 2: 3^{rd}-Order Cumulant Structure : The Linear Model

	ARE(1)
C(0,0)	$\mu_{x,3} - 3\mu_{x,2}\mu_x + 2\mu_x^3$
$C(0,\tau)$	$\phi^\tau[\mu_{x,3} - 2\mu_{x,2}\mu_x]$ $+[\gamma(\tau) - \mu_x][\mu_{x,2} - 2\mu_x^2]$
$C(\tau,\tau)$	$\phi^{2\tau}\mu_{x,3} + 2\mu_{x,2}[\gamma(2\tau) - \gamma(\tau)]$ $+2\mu_x[\gamma_2(\tau) - \mu_x\gamma(\tau)]$ $+2\mu_x\phi[\frac{1}{1-\phi}\{\gamma_2(\tau) - \phi^{\tau-1}\gamma(\tau)\}]$ $-\mu_{x,2}\mu_x[1 + 2\phi^\tau] + 2\mu_x^3$
$C(1,1+\tau)$	$\phi^{\tau+2}\mu_{x,3} + [\phi\mu_{x,2} + \mu_x][2\phi^\tau + \gamma(\tau)]$ $-\mu_x[\mu_{x,2}\{\phi^\tau + \phi^{\tau+1} + \phi\} + \mu_x\{\gamma(\tau) + \gamma(\tau+1) + 1\}]$ $+2\mu_x^3$
$C(h,h+\tau)$	$\phi^{\tau+2h}\mu_{x,3}$ $+2\phi^\tau[\mu_{x,2}\{\gamma(2h) - \gamma(h)\} + \mu_x\{\gamma_2(h) + \frac{\phi}{1-\phi}\{\gamma_2(h) - \phi^{h-1}\gamma(h)\}\}]$ $+\mu_{x,2}\phi^h\gamma(\tau) + \mu_x\gamma(h)\gamma(\tau)$ $-\mu_x[\mu_{x,2}\{\phi^\tau + \phi^{\tau+h} + \phi^h\} + \mu_x\{\gamma(\tau) + \gamma(\tau+h) + \gamma(h)\}] + 2\mu_x^3$

Table 3: 3^{rd}-Order Cumulant Structure : The Intrinsically Linear Model

	PAR(1)
C(0,0)	2
$C(0,\tau)$	$\frac{1}{\mu_{x,\alpha^\tau}}[\mu_{x,\alpha^\tau+2} - 2\mu_{x,\alpha^\tau+1}]$
$C(\tau,\tau)$	$\frac{2\mu_{x,2\alpha^\tau+1}}{\mu_{x,2\alpha^\tau}} - 2\frac{\mu_{x,\alpha^\tau+1}}{\mu_{x,\alpha^\tau}}$
$C(1,1+\tau)$	$\frac{\mu_{x,\alpha(\alpha^\tau+1)+1}\mu_{x,\alpha^\tau+1}}{\mu_{x,\alpha^\tau}\mu_{x,\alpha(\alpha^\tau+1)}} - \left\{\frac{\mu_{x,\alpha^\tau+1}}{\mu_{x,\alpha^\tau}} + \frac{\mu_{x,\alpha^\tau+1+1}}{\mu_{x,\alpha^\tau+1}} + \frac{\mu_{x,\alpha+1}}{\mu_{x,\alpha}}\right\} + 2$
$C(h,h+\tau)$	$\frac{\mu_{x,\alpha^h(\alpha^\tau+1)+1}\mu_{x,\alpha^\tau+1}}{\mu_{x,\alpha^\tau}\mu_{x,\alpha^h(\alpha^\tau+1)}} - \left\{\frac{\mu_{x,\alpha^\tau+1}}{\mu_{x,\alpha^\tau}} + \frac{\mu_{x,\alpha^\tau+h+1}}{\mu_{x,\alpha^\tau+h}} + \frac{\mu_{x,\alpha^h+1}}{\mu_{x,\alpha^h}}\right\} + 2$

used as a tool for discrimination purposes and consequently nor can the spectral densities.

To illustrate the shape of the 3^{rd}-order cumulant structure, see figure 4, we set $\phi, \rho, \alpha, \beta, \alpha\beta, \frac{\alpha}{\alpha+2} = 0.5$. First, we observe that certain ratios in tables 2,3,4 and 5 yield a clear characterization of the cumulant surfaces. Consider the ratios, presented in table 6, for the models with a simple close form i.e. EAR(1), TEAR(1) and Robertson's fixed model.

While such simple expressions are not available for the remaining models it is possible to investigate the behavior of these ratios numerically. Two of the above ratios turn out to be

Table 4: 3^{rd}-Order Cumulant Structure : The Intrinsically Non-linear Models

	EAR(1)	**TEAR(1)**	**NEAR(1)**
$C(0,0)$	2	2	2
$C(0,\tau)$	$2\rho^{\tau}$	$2\alpha^{\tau}$	$2(\alpha\beta)^{\tau}$
$C(\tau,\tau)$	$2\rho^{2\tau}$	$2\alpha^{\tau}[1+\tau(1-\alpha)]$	$6\lambda^{\tau}$ $+4(\alpha\beta)^{\tau}[\mu_{\epsilon}\gamma_{\beta}(\tau)-1]$ $+2\mu_{\epsilon}[\mu_{\epsilon}\alpha\beta\frac{1}{1-\alpha\beta}\{\gamma_{\lambda}(\tau)$ $-(\alpha\beta)^{\tau-1}\gamma_{\beta}(\tau)\}-\gamma_{\alpha\beta}(\tau)]$ $+\mu_{\epsilon,2}\gamma_{\lambda}(\tau)$
$C(1,1+\tau)$	$2\rho^{\tau+2}$	$2\alpha^{\tau+1}(2-\alpha)$	$2(\alpha\beta)^{\tau+1}[3\beta+2\mu_{\epsilon}]$ $+(\alpha\beta)^{\tau}\mu_{\epsilon,2}$ $+\gamma_{\alpha\beta}(\tau)\mu_{\epsilon}[\mu_{\epsilon}+2\alpha\beta]$ $-[2\{(\alpha\beta)^{\tau}(1+\alpha\beta)+\alpha\beta\}$ $+\mu_{\epsilon}\{\gamma_{\alpha\beta}(\tau)$ $+\gamma_{\alpha\beta}(\tau+1)+1\}]+2$
$C(h,h+\tau)$	$2\rho^{\tau+2h}$	$2\alpha^{\tau+h}[1+h(1-\alpha)]$	$6(\alpha\beta)^{\tau}\lambda^{h}$ $+4(\alpha\beta)^{\tau+h}\mu_{\epsilon}\gamma_{\beta}(h)$ $+2(\alpha\beta)^{h}\mu_{\epsilon}\gamma_{\alpha\beta}(\tau)$ $+2(\alpha\beta)^{\tau+1}\mu_{\epsilon}^{2}\frac{1}{1-\alpha\beta}\{\gamma_{\lambda}(h)$ $-(\alpha\beta)^{h-1}\gamma_{\beta}(h)\}$ $+(\alpha\beta)^{\tau}\mu_{\epsilon,2}\gamma_{\lambda}(h)$ $+\mu_{\epsilon}^{2}\gamma_{\alpha\beta}(h)\gamma_{\alpha\beta}(\tau)$ $-[2\{(\alpha\beta)^{\tau}(1+(\alpha\beta)^{h})+(\alpha\beta)^{h}\}$ $+\mu_{\epsilon}\{\gamma_{\alpha\beta}(\tau)+\gamma_{\alpha\beta}(\tau+h)$ $+\gamma_{\alpha\beta}(h)\}]+2$

more informative for the purpose of discriminating among the models : $\frac{C(\tau,\tau)}{C(0,\tau)}$ and $\frac{C(1,1+\tau)}{C(\tau,\tau)}$. In the simulation context, however, since the cumulant surfaces decay rapidly towards 0, the computation of these ratios become difficult as we attempt to divide by very small values . These numerical considerations unstabilize the use of the ratios as a tool for discriminating among the models. The computed ratios (as functions of the lag τ), indexed by a set of parameter values such that the correlation function of each model exhibits an identical behavior (e.g. $\rho(s)=(0.5)^{s}$) are also given, figures 5-6, so to demonstrate the shapes of the expressions given in the first and fourth rows of table 6.

Given the plots of the ratios and the cumulant surfaces for the six models we may classify them into three categories. EAR(1) forms its own class. Robertson's models and TEAR(1)

Table 5: 3^{rd}-Order Cumulant Structure : The intrinsically Non-linear Models (cont.)

	Robertson's Fixed Model	**Robertson's Random Model**
$C(0,0)$	2	2
$C(0,\tau)$	$2\beta^{\tau}$	$2\varrho^{\tau}$
$C(\tau,\tau)$	$2\beta^{\tau}(1-\tau ln\beta)$	$2\varrho^{\tau}[1-2\tau b_r\varrho^{-1}]$ $+(\alpha+2)ab_r[\gamma(\tau-1)-(\tau-1)\varrho^{\tau-1}]$ $+\gamma(\tau)[2a+c]$
$C(1,1+\tau)$	$2\beta^{\tau+1}(1-ln\beta)$	$4\varrho^{\tau+1}$ $-\varrho^{\tau}[2(2b_r+1)-c]$ $+a[\gamma(\tau)+\gamma(\tau+1)+1]$ $-ad\gamma(\tau)-2(\varrho-1)$
$C(h,h+\tau)$	$2\beta^{\tau+h}(1-hln\beta)$	$4\varrho^{\tau+h}$ $-2\varrho^{\tau+h-1}[2hb_r+a]$ $-2\varrho^{h}[a\gamma(\tau-1)+1]$ $+\varrho^{\tau}[(\alpha+2)ab_r\{\gamma(h-1)-(h-1)\varrho^{h-1}\}-2]$ $+\gamma(h)[a^2\gamma(\tau)+c\varrho^{\tau}]$ $+a[\gamma(\tau)+\gamma(\tau+h)+\gamma(h)]+2$

form a separate group. NEAR(1) and PAR(1) form an additional class. Note that the cumulant surface produced by NEAR(1) is a combination of EAR(1) and TEAR(1) and that it looks very much like the surface produced by PAR(1). However, the two models seem to differ in their behavior when one observe the plots of the theoretical ratios. Closer look at the vertical axis for NEAR(1) and PAR(1) in figures 5 and 6 shows that the ranges are similar and much smaller than the ranges of the vertical axis for the other models.

6 Methodology

In the following we propose a discrimination procedure that may be applied to the models under investigation (3.25)-(3.31) or to any set of competing models.

Let

$$M = \{ \text{ a finite set of finit parameter models } \}.$$

Table 6: Ratios of the 3^{rd}-Order Cumulant Structure

	EAR(1)	**TEAR(1)**	**Robertson's Fixed Model**
$\frac{C(\tau,\tau)}{C(0,\tau)}$	ρ^{τ}	$1+(1-\alpha)\tau$	$1-\tau ln\beta$
$\frac{C(1,1+\tau)}{C(0,\tau)}$	ρ^{2}	$\alpha(2-\alpha)$	$\beta(1-ln\beta)$
$\frac{C(h,h+\tau)}{C(0,\tau)}$	ρ^{2h}	$\alpha^{h}[1+h(1-\alpha)]$	$\beta^{h}(1-hln\beta)$
$\frac{C(1,1+\tau)}{C(\tau,\tau)}$	$\rho^{2-\tau}$	$\frac{\alpha(2-\alpha)}{1+\tau(1-\alpha)}$	$\frac{\beta(1-ln\beta)}{1-\tau ln\beta}$
$\frac{C(h,h+\tau)}{C(\tau,\tau)}$	$\rho^{2h-\tau}$	$\frac{\alpha^{h}[1+h(1-\alpha)]}{1+\tau(1-\alpha)}$	$\frac{\beta^{h}(1-hln\beta)}{1-\tau ln\beta}$
$\frac{C(h,h+\tau)}{C(1,1+\tau)}$	$\rho^{2(h-1)}$	$\frac{\alpha^{h-1}[1+h(1-\alpha)]}{2-\alpha}$	$\frac{\beta^{h-1}(1-hln\beta)}{1-ln\beta}$

Our objective is to identify the most compatible model $m \in M$ with $\{X_t\}_{t=1}^{n}$. Specifically, given $\{X_t\}_{t=1}^{n}$, find a model $m \in M$ such that $m \sim \{X_t\}_{t=1}^{n}$.

Procedure :

1. Compute $\hat{C}_x(u_1,\ldots,u_k)$, $k = 0,1,2,\ldots,$ $u_i \in I$ integer. We call it the empirical k^{th}-order cumulant structure based on the data $\{X_t\}_{t=1}^{n}$.

2. For each $m \in M$

   (a) Estimate , using $\{X_t\}_{t=1}^{n}$, the parameter θ_m (possibly a vector) for model m.

   (b) Compute $C_{\theta_m}(u_1,\ldots,u_k)$ for model m empirically or using the theoretical cumulant structure. We shall call it **Method 1** if the computation of the cumulant structure is done using the known theoretical cumulant structure. We shall call it **Method 2** if the computation of the cumulant structure is done empirically based on $\{X_t\}_{t=1}^{n}$.

3. Given the above quantities we seek to minimize, for a norm $\| \ \|$

$$\text{Min}_{m\in M} \parallel \hat{C}_x(u_1,\ldots,u_k) - C_{\theta_m}(u_1,\ldots,u_k) \parallel . \tag{6.1}$$

Alternatively,

$$\text{Min}_{m\in M} \parallel \hat{f}_x(\lambda_1,\ldots,\lambda_k) - f_{\theta_m}(\lambda_1,\ldots,\lambda_k) \parallel \tag{6.2}$$

where $f_{\theta_m}(\lambda_1, \ldots, \lambda_k)$ is the k^{th}-order spectrum (i.e. polyspectrum). The general distance measure may be specified as e.g.

$$\| g \| = \sum_{\{u_i\} \in S} | g |^2 .$$

There are several issues that need to be considered under the proposed procedure. First, various properties of the model such as stationarity, ergodicity, moment conditions, moment calculations, parameter estimation and simulation aspects of sample traces must be investigated. Second, statistical properties of the formal test statistics based on (6.1) or (6.2) have to be studied. In order to do so the sampling properties of the proposed procedure must be investigated. In the following section we consider the simulation aspects of (6.1) and present some simulation results for both methods 1 and 2.

7 Simulation Results

In order to verify the possibility of discrimating among the various models on the basis of their respective 3^{rd}-order cumulant surfaces, it is necessary to obtain reasonable agreements among the theoretical and simulated cumulants. In the following we discuss issues related to the simulation aspects of the sample traces, correlation functions, 3^{rd}-order cumulant surfaces and ratios.

7.1 Simulating Sample Traces

The simulation aspects of the NEAR(1) model and its special cases, EAR(1) and TEAR(1), were considered by Lawrance and Lewis [20]. The algorithm they give is being used in our simulation to generate sample realizations for the NEAR(1) family. The subcases, EAR(1) and TEAR(1), are simulated by setting ($\alpha = 0.99$, $0 \leq \beta < 1$) and ($\beta = 0.99$, $0 \leq \alpha < 1$) respectively, in the same program that generates the simulated paths for NEAR(1) model.

We follow Lawrance and Lewis in setting the degenerate parameters to 0.99 so to avoid complications in the simulation of the traces.

Robertson's fixed and random models are generated by two different programs. One which allows a selection of a branch with a fixed probability and one which allows the selection of a branch with a random probability generated according to a beta random variable with parameters $(\alpha, 2)$. The input signal is a truncated exponential; hence, needs to be simulated accordingly. Since no IMSL subroutine is available we generate a realization from a truncated exponential random variable using the cumulative distribution function technique. Realizations from the AR(1) model are easily simulated and no further explanations are required. McKenzie [27] discusses the simulation of PAR(1) models. The innovation process V_t is generated according to

$$V = E^{1-\alpha} b(U)$$

where U is distributed as a uniform $(0, \pi)$ sequence of random variables which is independent of E - a sequence of exponential mean one random variables. The function b is defined by

$$b(\phi) = sin\phi(sin\alpha\phi)^{-\alpha}(sin(1-\alpha)\phi)^{-(1-\alpha)}.$$

Thus, $\{V_t\}$ is generated as a mixture of uniform and exponential sequences of independent random variables.

All the simulated paths are generated by FORTRAN programs that call IMSL subroutines which are used to simulate continous uniform, beta and exponential realizations.

7.2 Simulating Higher Order Moments

One FORTRAN program is employed in simulating the correlation functions, 3^{rd}-order cumulant surfaces and certain slices of these surfaces . Smoothing considerations lead us to simulate each model 30 times where the length of each simulated trace is 1010 data points.

Table 7: Distance Measure (6.1) - $\rho(s) = (0.25)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	0.26	0.21	0.17	0.06	0.58	0.40
EAR(1)	0.018	0.015	0.68	0.08	1.40	1.14
TEAR(1)	0.78	0.70	0.018	0.37	0.08	0.03
NEAR(1)	0.12	0.09	0.31	0.008	0.84	0.63
Robertson's Fixed	1.80	1.65	0.22	1.06	0.02	0.07
Robertson's Random	0.73	0.66	0.02	0.34	0.12	0.05

The program computes two expectation terms : $E[X_t X_{t+r}]$, over the range of lags 0 to 9, and $E[X_t X_{t+r} X_{t+r+s}]$, over the range of lags, -9 to -9. Then the smoothed empirical correlation function and the smoothed 3^{rd}-order cumulant surface are computed using their definitions. In the computations of the expectation terms we use :

$$E[X_t X_{t+r}] = \frac{1}{1010} \sum_{t=1}^{1001} X_t X_{t+r}$$

$$E[X_t X_{t+r} X_{t+r+s}] = \frac{1}{1010} \sum_{t=1}^{1001} X_t X_{t+r} X_{t+r+s}$$

In order to determine how accuratly the simulated cumulant surfaces match their theoretical couterparts we plot the empirical correlation functions, the empirical $C(\tau, \tau)$ slice and the complete simulated surfaces in figures 7-9. This is done for various parameter values and shown for those that correspond to $\rho(s) = (0.5)^s$.

7.3 Discrimination Procedure : Method 1

The results of the simulation study are summarized in tables 7-12. Tables 7-9 are examples of typical values obtained by a single run of the simulation. Tables 10-12 provide the proportions of correct model identification out of 30 repetitions. Note that in table 7 the diagonal line contains the minimum values of rows 2-5. This is precisely how we would expect the procedure to perform for any parameter value indexing a standardized correlation function. However, errors occure at the first and last rows where the method fails to select the correct

Table 8: Distance Measure (6.1) : $\rho(s) = (0.5)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	0.67	0.41	0.74	0.02	1.44	1.43
EAR(1)	0.07	0.01	2.24	0.30	3.34	3.42
TEAR(1)	3.59	2.97	0.085	1.39	0.087	0.10
NEAR(1)	0.64	0.37	0.93	0.005	1.67	1.69
Robertson's Fixed	4.92	4.27	0.32	2.36	0.06	0.13
Robertson's Random	2.36	1.91	0.07	0.76	0.26	0.32

Table 9: Distance Measure (6.1) : $\rho(s) = (0.75)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	1.91	1.12	2.28	0.03	3.07	3.02
EAR(1)	1.82	0.02	6.53	0.85	7.79	7.69
TEAR(1)	4.14	3.85	0.46	1.25	0.79	0.93
NEAR(1)	1.83	0.63	3.26	0.09	4.19	4.15
Robertson's Fixed	4.82	4.59	0.31	1.63	0.57	0.70
Robertson's Random	9.43	10.38	0.51	5.32	0.24	0.36

model. The PAR(1) model is being identified as a NEAR(1) model and Robertson's Random model is being identified as a TEAR(1) model. The theoretical plots of the 3^{rd}-order cumulant structure support this confusion as they show that these models produce very similar surfaces that are hard to distinguish. In table 8 we note that the procedure fails again to select PAR(1) and Roberson's random models. Errors occur at the first and last two rows of table 9 where the procedure fails to distinguish PAR(1), the fix and random models. The incorrect selection that appears in the above tables is consistent with our previous remark regarding the grouping of the models into three categories. Robertson's models and TEAR(1) were identified as sharing a very similar 3^{rd}-order cumulant structure and so were PAR(1) and TEAR(1). Thus, one would expect to have difficulties in discriminating among models that belong to the same family. The pattern established in the previous tables is consistent in the 30 repetitions we consider in tables 10-12. PAR(1) is consistently confused with NEAR(1), and TEAR(1) and Robertson's models stand out as a separate class. The random model is by large the hardest to identify and typically is mistaken for TEAR(1) model. Although the procedure is successful in identifying TEAR(1) and the fixed model it

Table 10: Proportions of Correct Identification : $\rho(s) = (0.25)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
EAR(1)	0.03	0.97	0.0	0.0	0.0	0.0
TEAR(1)	0.0	0.0	1.0	0.0	0.0	0.0
NEAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
Robertson's Fixed	0.0	0.0	0.0	0.0	0.73	0.27
Robertson's Random	0.0	0.0	0.7	0.0	0.03	0.27

Table 11: Proportions of Correct Identification : $\rho(s) = (0.5)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
EAR(1)	0.0	1.0	0.0	0.0	0.0	0.0
TEAR(1)	0.0	0.0	0.7	0.03	0.27	0.0
NEAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
Robertson's Fixed	0.0	0.0	0.17	0.0	0.83	0.0
Robertson's Random	0.0	0.0	0.63	0.0	0.37	0.0

Table 12: Proportions of Correct Identification : $\rho(s) = (0.75)^s$

	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Robertson's Fixed	Robertson's Random
PAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
EAR(1)	0.0	1.0	0.0	0.0	0.0	0.0
TEAR(1)	0.0	0.0	0.67	0.07	0.26	0.0
NEAR(1)	0.0	0.0	0.0	1.0	0.0	0.0
Robertson's Fixed	0.0	0.0	0.47	0.0	0.53	0.0
Robertson's Random	0.0	0.0	0.53	0.0	0.47	0.0

Table 13: Proportions of Correct Identification : $\rho(s) = (0.25)^s$

	TEAR(1)	**Robertson's Fixed**	**Robertson's Random**
TEAR(1)	0.57	0.03	0.4
Robertson's Fixed	0.13	0.87	0.0
Robertson's Random	0.17	0.1	0.73

Table 14: Proportions of Correct Identification : $\rho(s) = (0.5)^s$

	TEAR(1)	**Robertson's Fixed**	**Robertson's Random**
TEAR(1)	0.7	0.17	0.13
Robertson's Fixed	0.27	0.46	0.27
Robertson's Random	0.3	0.5	0.2

is the confusion in selecting the random model that makes it difficult to judge the aduquacy of TEAR(1) or the fixed model. However, since the three models share very similar traces and 3^{rd}-order cumulant structure one may choose to accept each of the three as compatible with any of that group.

To remedy this problem we may apply the proposed discrimination procedure to the 4^{rd}-order cumulant structure for these three models. One may argue that since the models share an identical 2^{rd}-order moment structure and a similar 3^{rd}-order cumulant structure (but too similar so their differences can not be captured by (6.1)), then it might be possible to reveal their true identity through the use of the 4^{rd}-order cumulant structure. Tables 13-15 contain the results of the simulation study applied to the 4^{rd}-order cumulant structure of TEAR(1) and Robertson's models. The choice among the models is not clear cut as the proportions

Table 15: Proportions of Correct Identification : $\rho(s) = (0.75)^s$

	TEAR(1)	**Robertson's Fixed**	**Robertson's Random**
TEAR(1)	0.63	0.3	0.07
Robertson's Fixed	0.13	0.57	0.0
Robertson's Random	0.17	0.47	0.06

Table 16: Proportions of Correct Identification : $\rho(s) = (0.25)^s$

	ARE(1)	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Rob's Fixed	Rob's Random
ARE(1)	1.0	0.0	0.0	0.0	0.0	0.0	0.0
PAR(1)	0.0	0.73	0.0	0.0	0.27	0.0	0.0
EAR(1)	0.0	0.0	0.93	0.0	0.07	0.0	0.0
TEAR(1)	0.0	0.0	0.0	0.63	0.0	0.03	0.33
NEAR(1)	0.0	0.3	0.03	0.0	0.67	0.0	0.0
Rob's Fixed	0.0	0.0	0.0	0.07	0.0	0.67	0.26
Rob's Random	0.0	0.0	0.0	0.3	0.0	0.23	0.47

of correct identification are not large enough to enable a reasonable degree of discrimination power among the three competing models. This result was expected to hold given the theoretical expressions as expressed through the plots for the theoretical 4^{rd}-order cumulant structure, figure 10. In these plots the models are shown to produce similar behavior at various frames of C(r,s,u); thus, there is no reason to expect a high degree of discrimination power among the models on the basis of the proposed procedure and the 4^{rd}-order cumulant structure.

7.4 Discrimination Procedure : Method 2

In tables 16-18 we provide the results of our simulation study according to (6.1) based on the empirical cumulant structure only. Note that we added ARE(1) for comparison purposes. Since the marginal moments of ARE(1) are different from the remaining models we standarize its mean to equal one so the mean of the exponential innovation process becomes $1 - \phi$. The higher order moments are not standarized to equal those of the exponential models. The results in tables 16-18 are by large consistent with the results obtained under the previous method. The main difference appears to be in the improved separation between PAR(1) and NEAR(1) under the second method while under the first method, which involved the theoretical cumulant structure, PAR(1) is consistantly mistaken for NEAR(1). We use method 2 with the 1^{th}-order empirical cumulant structure for TEAR(1) and Robertson's models. The results are summarized in tables 19-21. Figure 11 contains the plots of

Table 17: Proportions of Correct Identification : $\rho(s) = (0.5)^s$

	ARE(1)	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Rob's Fixed	Rob's Random
ARE(1)	1.0	0.0	0.0	0.0	0.0	0.0	0.0
PAR(1)	0.0	0.87	0.0	0.0	0.13	0.0	0.0
EAR(1)	0.0	0.0	1.0	0.0	0.0	0.0	0.0
TEAR(1)	0.0	0.0	0.0	0.43	0.0	0.27	0.3
NEAR(1)	0.0	0.47	0.0	0.0	0.53	0.0	0.0
Rob's Fixed	0.0	0.0	0.0	0.07	0.0	0.63	0.3
Rob's Random	0.0	0.0	0.0	0.3	0.0	0.4	0.3

Table 18: Proportions of Correct Identification : $\rho(s) = (0.75)^s$

	ARE(1)	PAR(1)	EAR(1)	TEAR(1)	NEAR(1)	Rob's Fixed	Rob's Random
ARE(1)	1.0	0.0	0.0	0.0	0.0	0.0	0.0
PAR(1)	0.0	0.57	0.0	0.0	0.43	0.0	0.0
EAR(1)	0.0	0.0	1.0	0.0	0.0	0.0	0.0
TEAR(1)	0.0	0.0	0.0	0.43	0.03	0.2	0.34
NEAR(1)	0.0	0.53	0.0	0.0	0.47	0.0	0.0
Rob's Fixed	0.0	0.0	0.0	0.23	0.0	0.43	0.34
Rob's Random	0.0	0.0	0.0	0.3	0.0	0.3	0.4

the simulated 4^{th}-order cumulant structure for the three models. The results confirm our previous comment regarding the difficulties encounterd by the discrimination procedure in distinguishing among these three models.

8 Conclusions

The problem of discrimination among non-linear time series models is considered in this paper through the family of exponential models. In this specific case we are able to develop the theoretical 3^{rd}-order cumulant structure and confirm it by simulation. The procedure we

Table 19: Proportions of Correct Identification : $\rho(s) = (0.25)^s$

	TEAR(1)	Robertson's Fixed	Robertson's Random
TEAR(1)	0.40	0.27	0.33
Robertson's Fixed	0.13	0.54	0.33
Robertson's Random	0.40	0.33	0.27

Table 20: Proportions of Correct Identification : $\rho(s) = (0.5)^s$

	TEAR(1)	**Robertson's Fixed**	**Robertson's Random**
TEAR(1)	0.27	0.20	0.53
Robertson's Fixed	0.23	0.40	0.37
Robertson's Random	0.27	0.3	0.43

Table 21: Proportions of Correct Identification : $\rho(s) = (0.75)^s$

	TEAR(1)	**Robertson's Fixed**	**Robertson's Random**
TEAR(1)	0.30	0.30	0.40
Robertson's Fixed	0.27	0.33	0.40
Robertson's Random	0.27	0.30	0.43

propose is not restricted to the class of AR(1) type models or the class of models for which analytical results for the 3^{rd}-order cumulant structure are available. It is a general procedure with the potential for a wide range of non-linear models. It is based in the understanding that different models cannot have an identical moment sequence ; hence, the discrimination among them would become possible at some stage in the higher order cumulant structure. In our specific case we are able to obtain a significant improvement in our discriminatory power just by going one step above the traditional second order moment analysis i.e. the correlation function. While second order moments play a dominating role in linear model discrimination they are very limited in the non-linear case. When the 2^{nd}-order analysis fails to provide enough information we propose to apply higher order moment analysis for the purpose of model discrimination.

References

[1] D. N. Anderson. *Some Time Series Models with Non-Additive Structure.* PhD thesis, University of California, Riverside, 1990.

[2] T. W. Anderson. *The Statistical Analysis of Time Series.* Wiley, New York, 1971.

[3] B. C. Arnold. A logistic process constructed using geometric minimization. *Statistics and Probability Letters*, 7:253–257, 1989.

[4] B. Auestad and D. Tjøstheim. Identification of nonlinear time series: First-order characterization and order determination. *Biometrika*, 77(4):669–87, 1990.

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

[6] D. R. Brillinger. An introduction to polyspectra. *Ann. Math. Statist.*, 36:1351–1374, 1965.

[7] D. R. Brillinger. *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York, 1975.

[8] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, Berlin, 1987.

[9] C. Chatfield. *The Analysis of Time Series: Theory and Practice.* Chapman and Hall, London, 1975.

[10] M. M. Gabr. On the third-order moment structure and bispectral analysis of some bilinear time series. *J. Tim. Sr. An.*, 9:11–20, 1988.

[11] D. P. Gaver and P. A. W. Lewis. First-order autoregressive gamma sequences and point processes. *Adv. Appl. Prob.*, 12:727–745, 1980.

[12] E. J. Hannan. *Multiple Time Series.* Wiley, New York, 1970.

[13] M. Hinnich. Testing for gaussianity and linearity of a stationary time series. *J. Time Series Anal.*, 3:169–176, 1982.

[14] P. A. Jacobs and P. A. W. Lewis. A mixed autoregressive-moving average exponential sequence and point process (earma 1,1). *Adv. Appl. Prob.*, 9:87–104, 1977.

[15] G. M. Jenkins and D. G. Watts. *Spectral Analysis and its Applications.* Holden-Day, San Francisco, 1968.

[16] A. Kendall, M. G. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics, Vol. 1.* Oxford University Press, New York, 1987.

[17] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics, 3 Vols.* Griffin, London, 1966.

[18] L. H. Koopmans. *The Spectral Analysis of Time Series.* Academic Press, New York, 1974.

[19] A. J. Lawrance. The mixed exponential solution to the first-order autoregressive model. *J. Appl. Prob.*, 12:522–546, 1980.

[20] A. J. Lawrance and P. A. W. Lewis. Simulation of some autoregressive markovian sequences of positive random variables. In R. J. Shannon H. J. Highland, M. G. Spiegel, editor, *1979 Winter Simulation Conference*, pages 301–307, New York, 1979. IEEE.

[21] A. J. Lawrance and P. A. W. Lewis. The exponential autoregressive-moving average earma(p,q) process. *J. R. Statist. Soc. B*, 42:150–161, 1980.

[22] A. J. Lawrance and P. A. W. Lewis. A new autoregressive time series model in exponential variables (near(1)). *Adv. Appl. Prob.*, 13:826–845, 1981.

[23] A. J. Lawrance and P. A. W. Lewis. Modeling and residual analysis of non-linear autoregressive time series in exponential variables (with discussion. *J. Roy. Statist. Soc. Ser. B*, 47(2):165 202, 1985.

[24] A. J. Lawrance and P. A. W. Lewis. Higher-order residual analysis for nonlinear time series with autoregressive correlation structure. *Inter. Stat. Rev.*, 55(1):21–35, 1987.

[25] V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theor. Prob. Appl.*, 4:319–329, 1959.

[26] W. K. Li. On the autocorrelation structure and identification of some bilinear time series. *J. Tim. Sr. An.*, 5(3):173–181, 1984.

[27] E. McKenzie. Product autoregression : a time-series characterization of the gamma distribution. *J. Appl. Prob.*, 19(2):463–468, 1982.

[28] R. R. Mohler, editor. *Nonlinear Time Series and Signal Processing.* Springer-Verlag, Berlin, 1988.

[29] M. B. Priestley. *Non-Linear and Non-Stationary Time Series Analysis.* Academic Press, London, 1989.

[30] M. B. Pristley. *Spectral Analysis and Time Series, 2 Vols.* Academic Press, London and New York, 1981.

[31] C. A. Robertson. An exponential process via a switching structure. Department of Statistics, University of California, Riverside, 1989.

[32] Murry Rosenblat. *Stationary Sequences and Random Fields.* Birkhaüser, Boston, 1986.

[33] M. Rosenblatt. *Cumulants and cumulant spectra.* North-Holland, Amsterdam, 1981. In Handbook of Statistics, Vol. 3. Edited by P. R. Krishnaiah.

[34] S. A. O. Sesay. *Frequency-domain methods of estimation and higher order moment analysis for the bilinear model BL(p,0,p,1).* PhD thesis, University of Manchester Institute of Science and Technology, 1985.

[35] T. Subba Rao and M. M. Gabr. *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Springer-Verlag, Berlin, 1984.

[36] H. Tong. *Non-linear Time Series: A dynamical System Approach.* Clarendon Press, Oxford, 1990.

[37] R. S. Tsay. Model checking via parametric bootstraps in time series analysis. *Appl. Statist.*, 41(1):1–15, 1992.

Figure 1

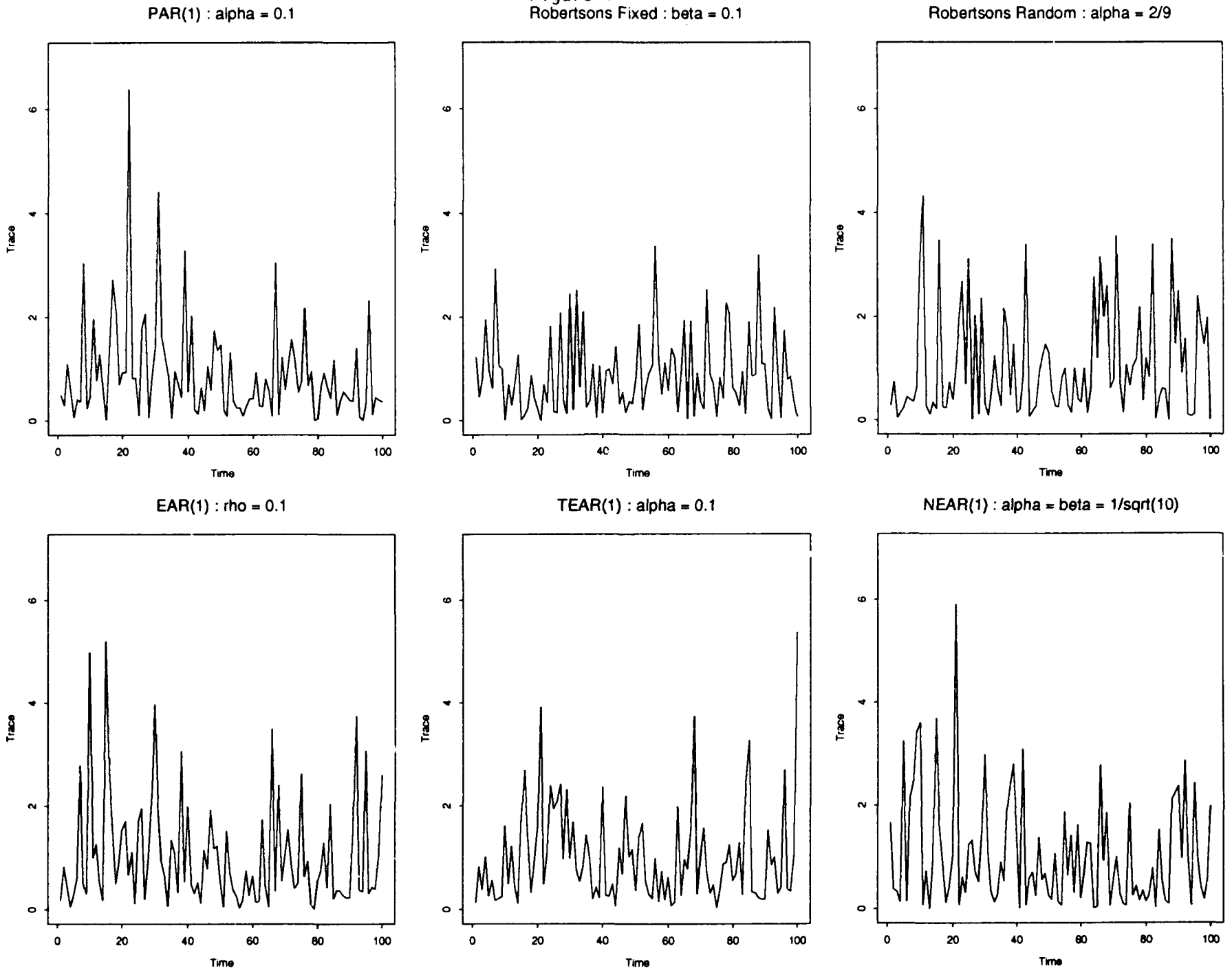

PAR(1) : alpha = 0.1
Robertsons Fixed : beta = 0.1
Robertsons Random : alpha = 2/9
EAR(1) : rho = 0.1
TEAR(1) : alpha = 0.1
NEAR(1) : alpha = beta = 1/sqrt(10)
Trace
Time
0
2
4
6
0
20
40
60
80
100

Figure 2

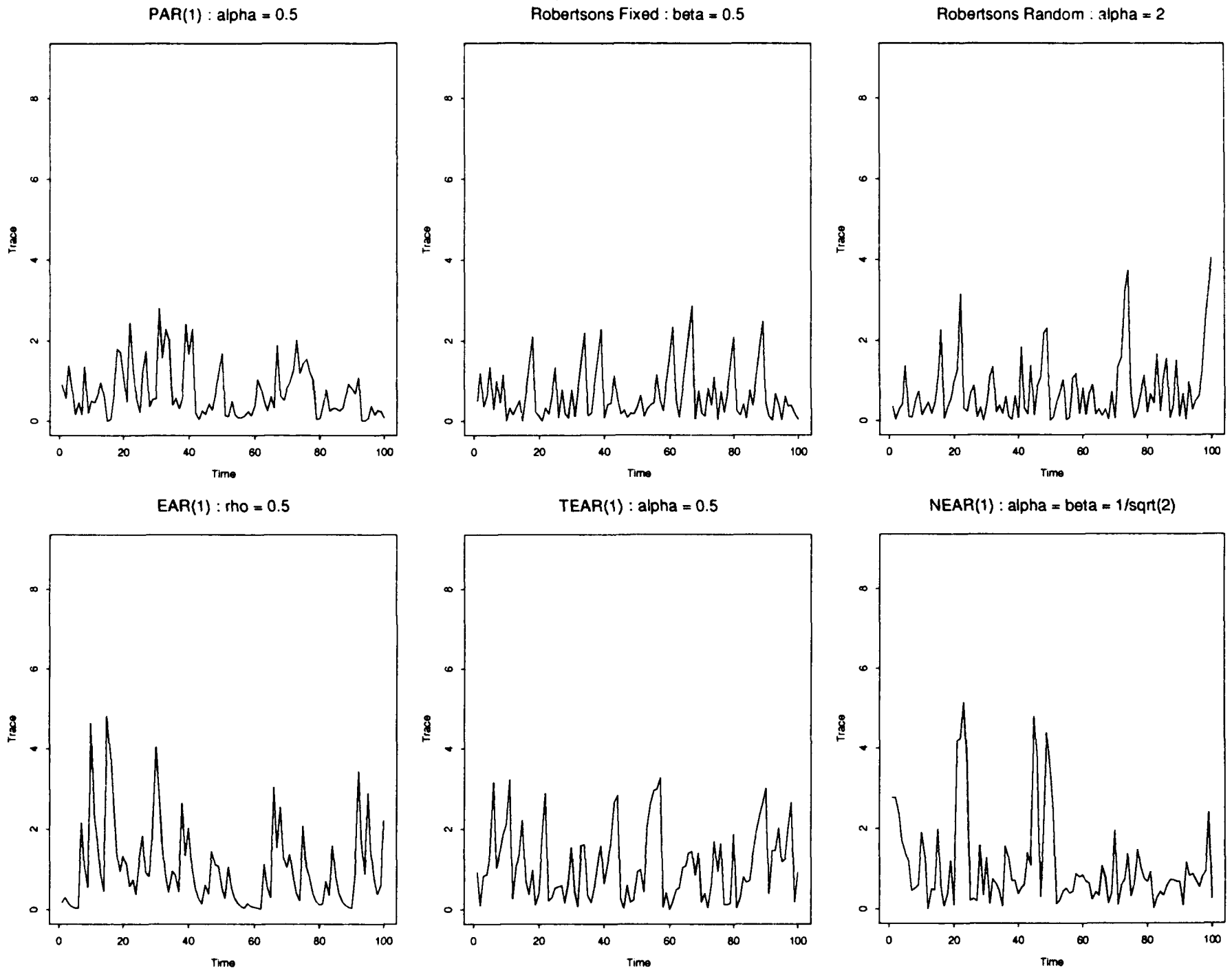
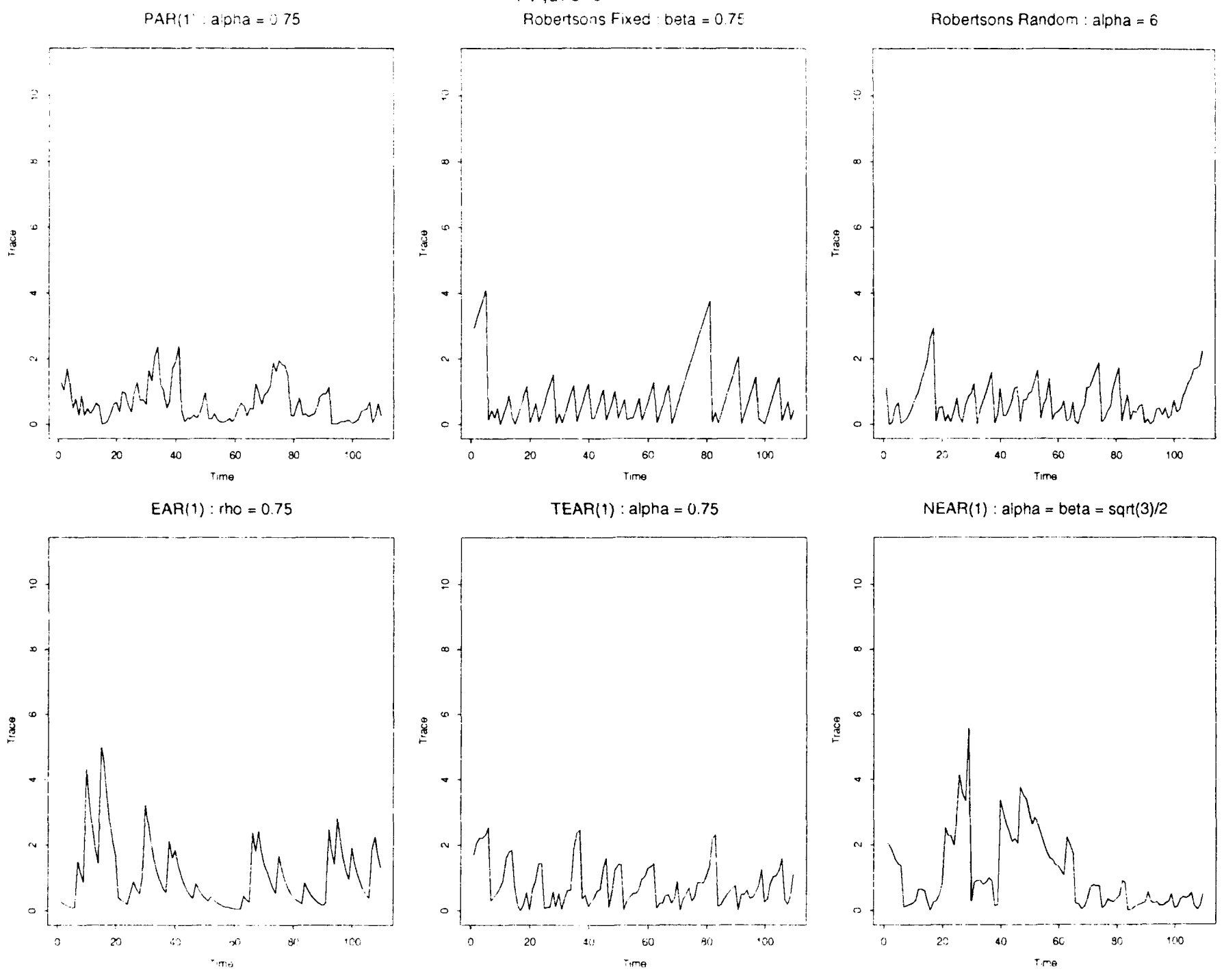

PAR(1) : alpha = 0.5
Robertsons Fixed : beta = 0.5
Robertsons Random : alpha = 2
EAR(1) : rho = 0.5
TEAR(1) : alpha = 0.5
NEAR(1) : alpha = beta = 1/sqrt(2)
Trace
Time
0
2
4
6
8
0
20
40
60
80
100

Figure 3


PAR(1) : alpha = 0.75
Robertsons Fixed : beta = 0.75
Robertsons Random : alpha = 6
EAR(1) : rho = 0.75
TEAR(1) : alpha = 0.75
NEAR(1) : alpha = beta = sqrt(3)/2
Trace
Time
0
2
4
6
8
10
0
20
40
60
80
100

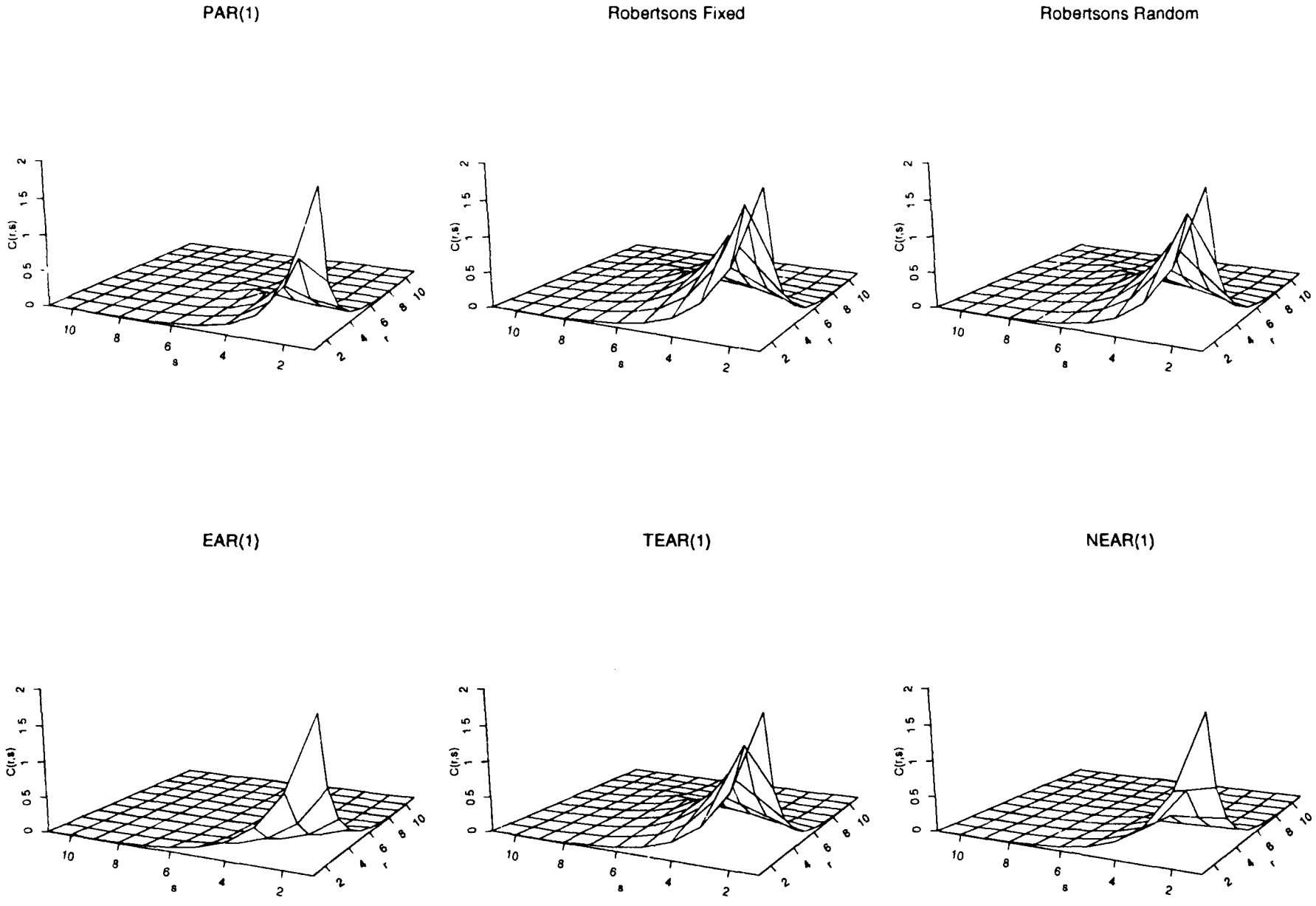


Figure 4 - C(r,s) : rho(s) = (0.50)**s

Figure 5

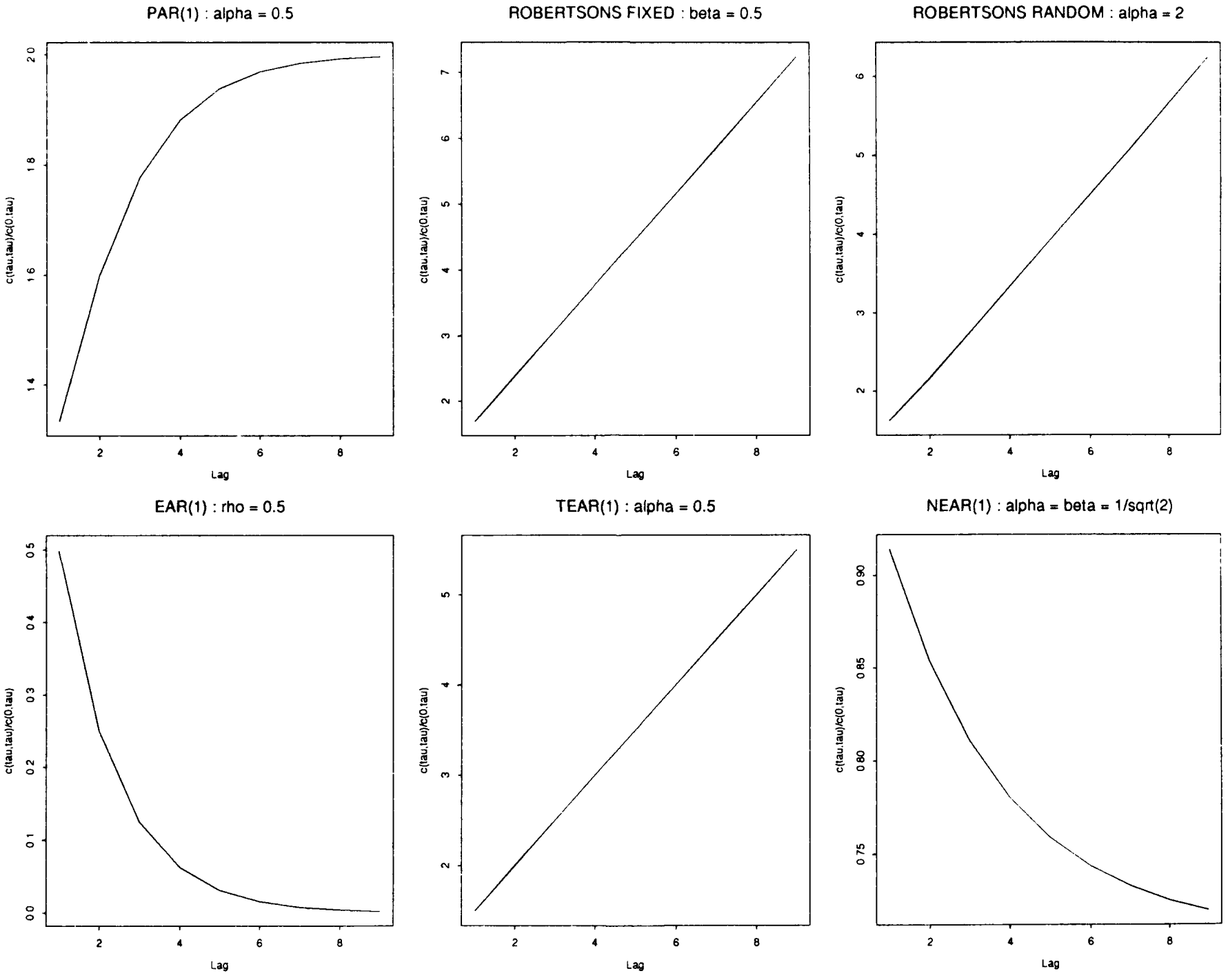

PAR(1) : alpha = 0.5
ROBERTSONS FIXED : beta = 0.5
ROBERTSONS RANDOM : alpha = 2
EAR(1) : rho = 0.5
TEAR(1) : alpha = 0.5
NEAR(1) : alpha = beta = 1/sqrt(2)
c(tau,tau)/c(0,tau)
Lag

Figure 6

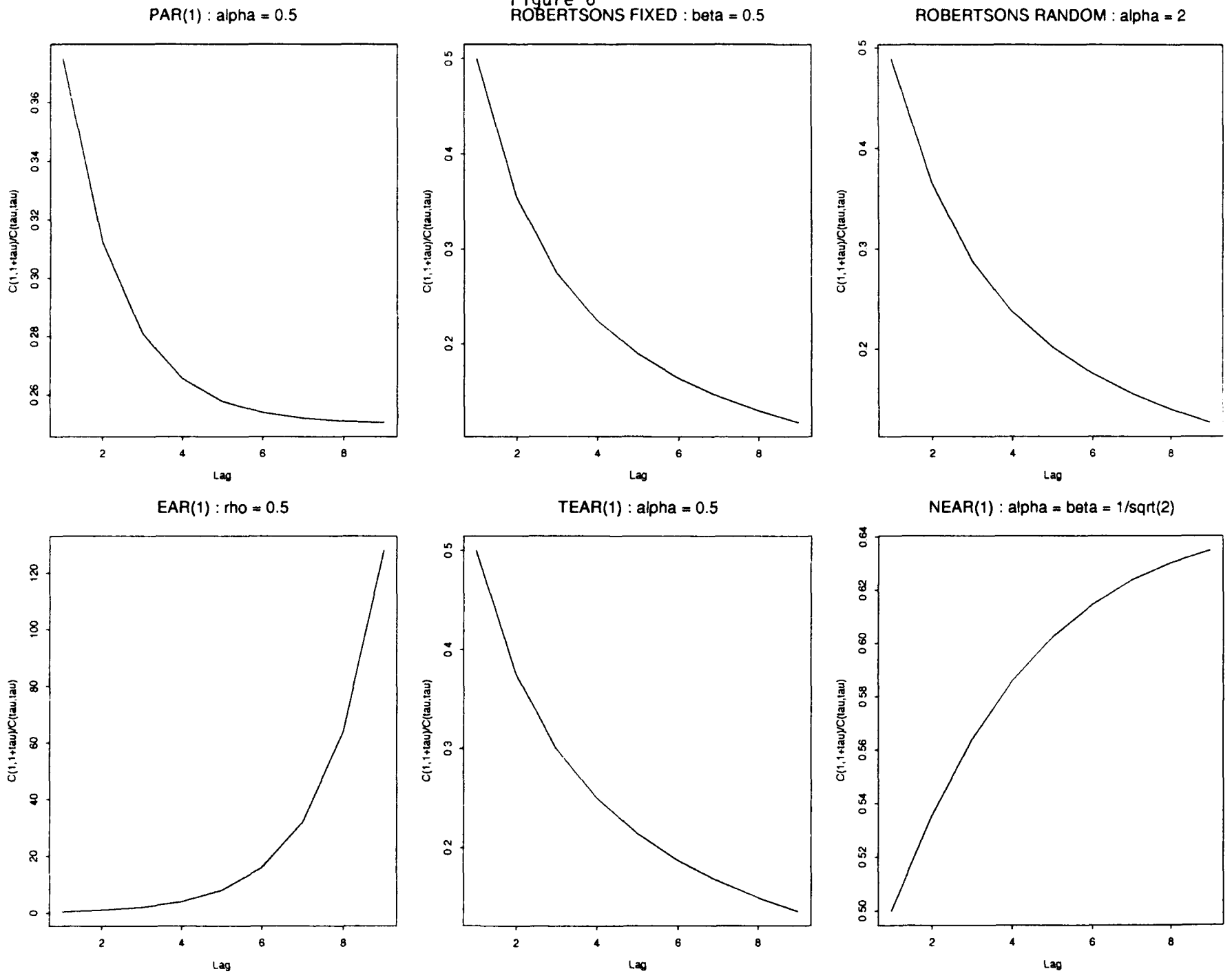
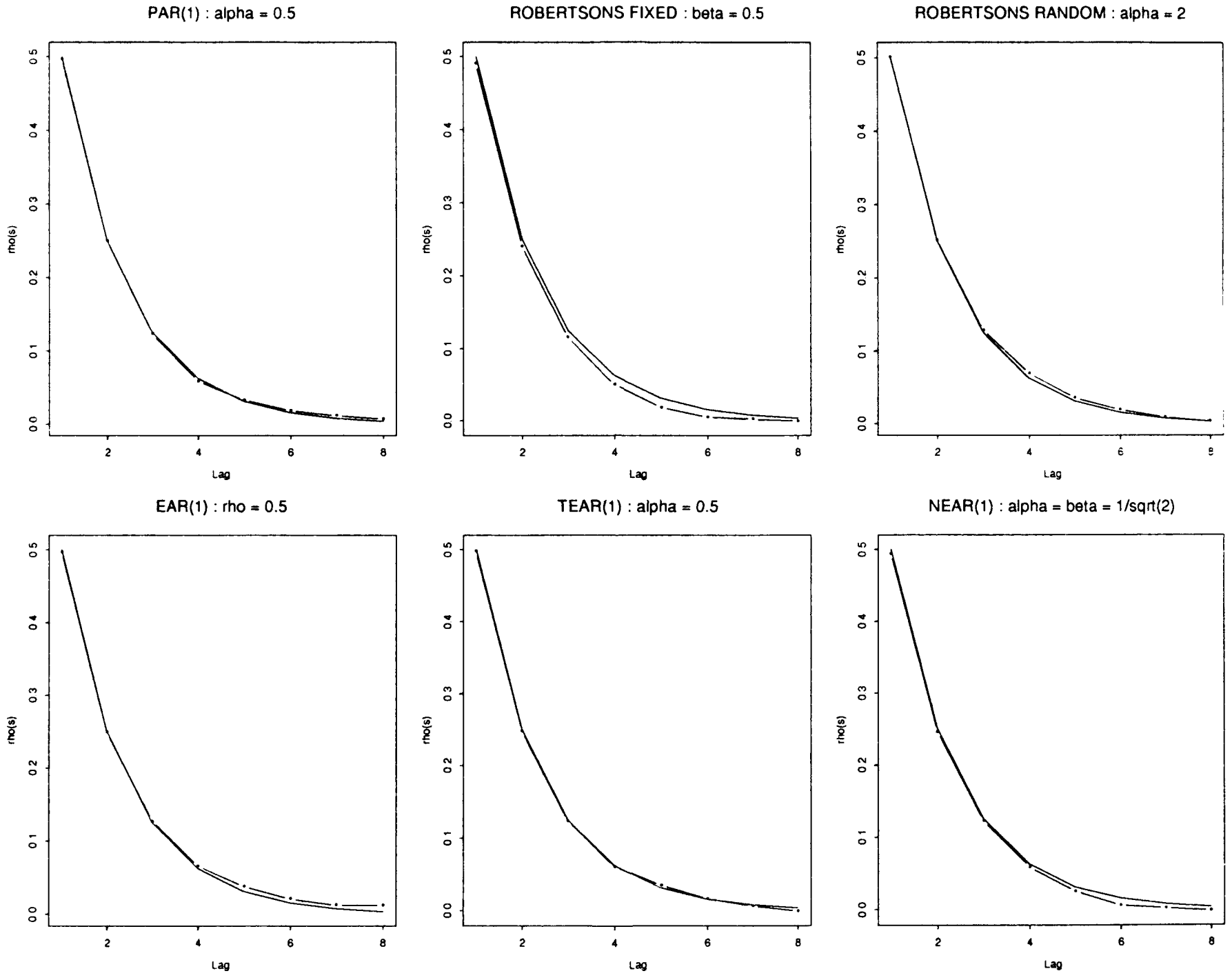

PAR(1) : alpha = 0.5
ROBERTSONS FIXED : beta = 0.5
ROBERTSONS RANDOM : alpha = 2
EAR(1) : rho = 0.5
TEAR(1) : alpha = 0.5
NEAR(1) : alpha = beta = 1/sqrt(2)
C(1,1+tau)/C(tau,tau)
Lag

Figure 7


PAR(1) : alpha = 0.5
ROBERTSONS FIXED : beta = 0.5
ROBERTSONS RANDOM : alpha = 2
EAR(1) : rho = 0.5
TEAR(1) : alpha = 0.5
NEAR(1) : alpha = beta = 1/sqrt(2)
rho(s)
Lag
0.0
0.1
0.2
0.3
0.4
0.5
2
4
6
8

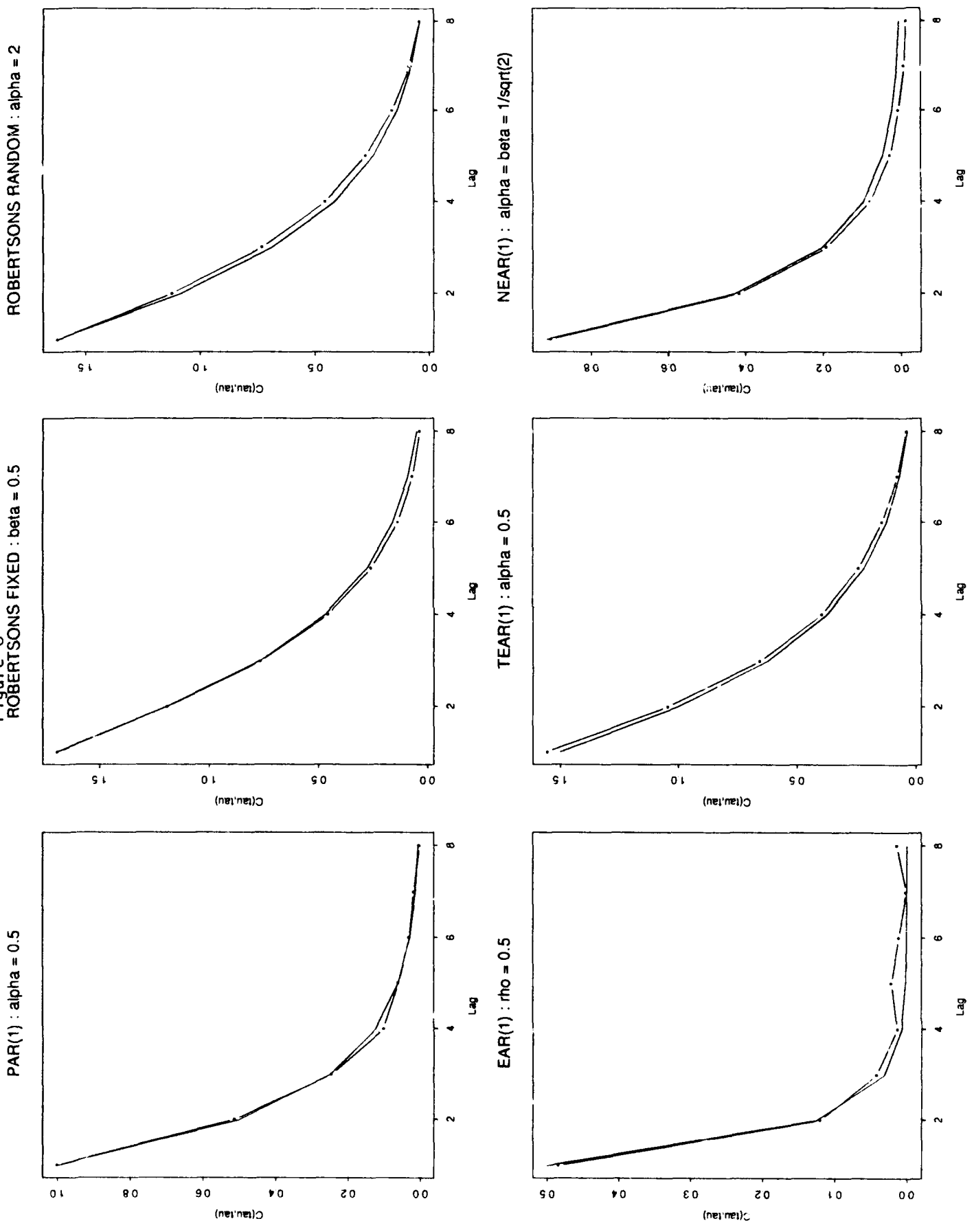
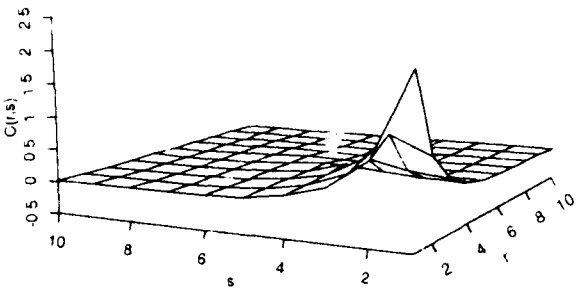
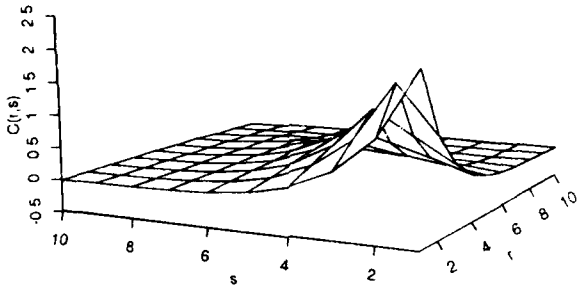

ROBERTSONS RANDOM : alpha = 2
NEAR(1) : alpha = beta = 1/sqrt(2)
ROBERTSONS FIXED : beta = 0.5
TEAR(1) : alpha = 0.5
PAR(1) : alpha = 0.5
EAR(1) : rho = 0.5
C(tau,tau)
Lag

Figure 8

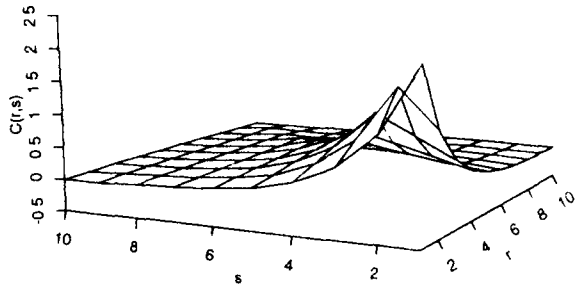
PAR(1)



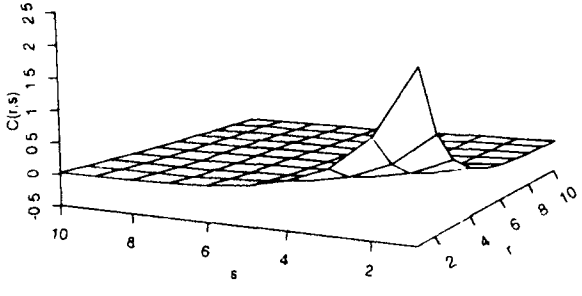
Robertsons Fixed



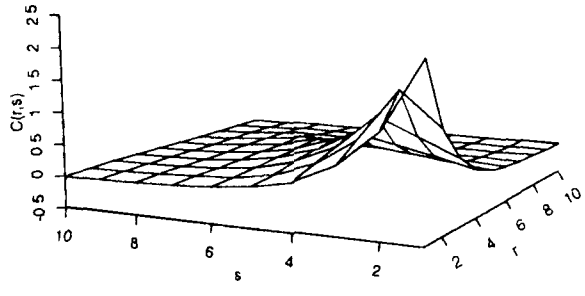
Robertsons Random



EAR(1)



TEAR(1)



NEAR(1)

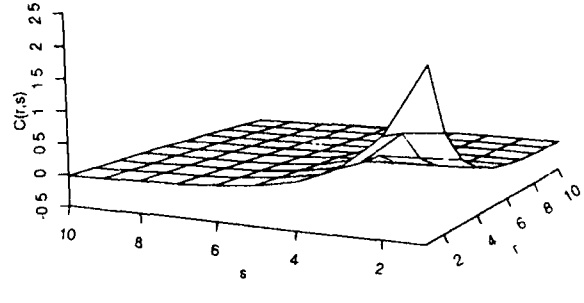
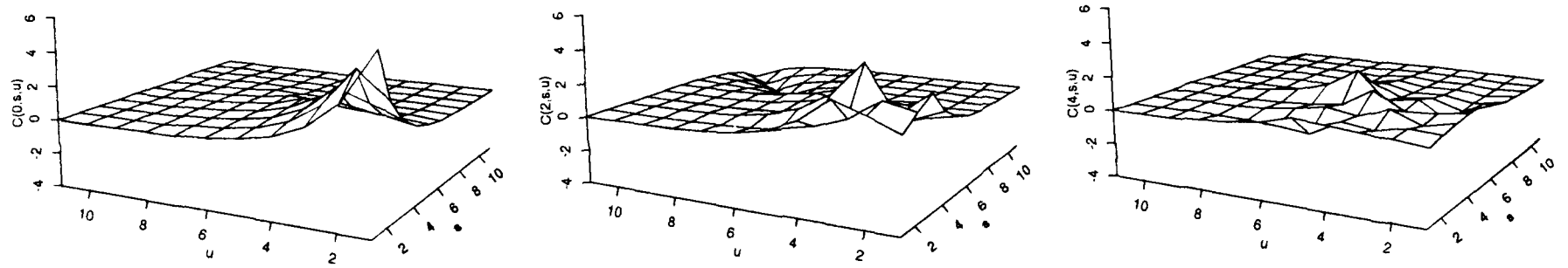
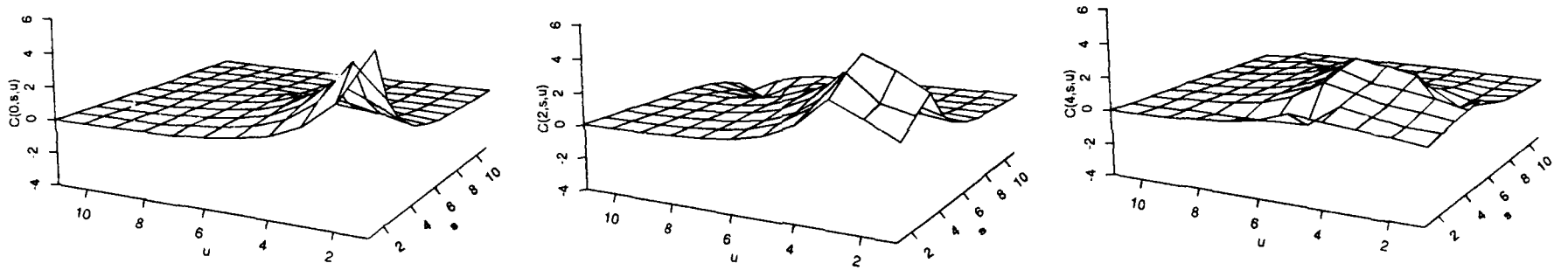


Figure 9 - Simulated C(r,s) : rho(s) = (0.50)**s



ROBERTSONS FIXED MODEL



ROBERTSONS RANDOM MODEL

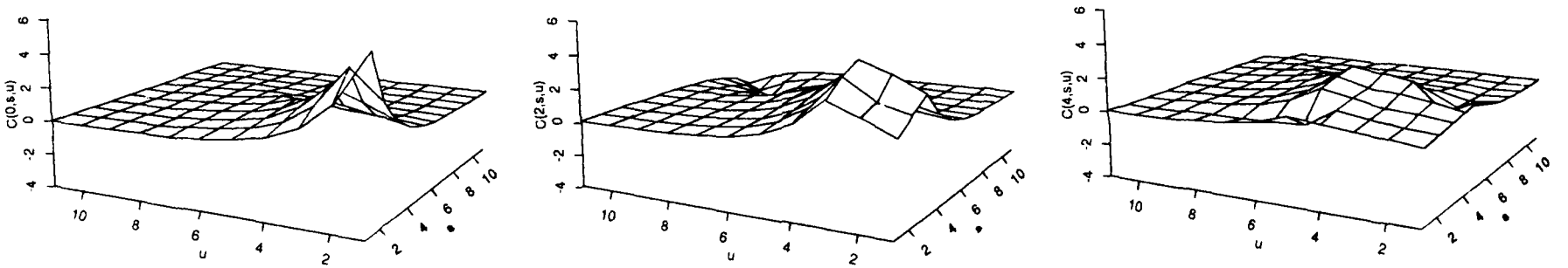
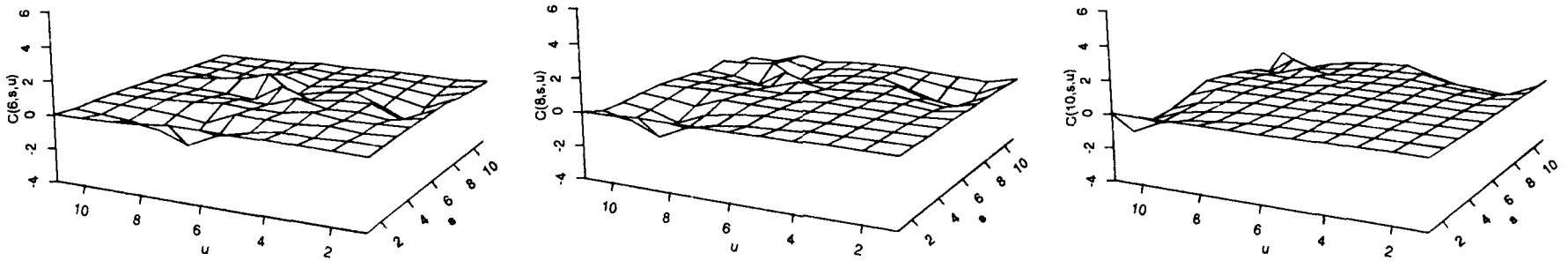
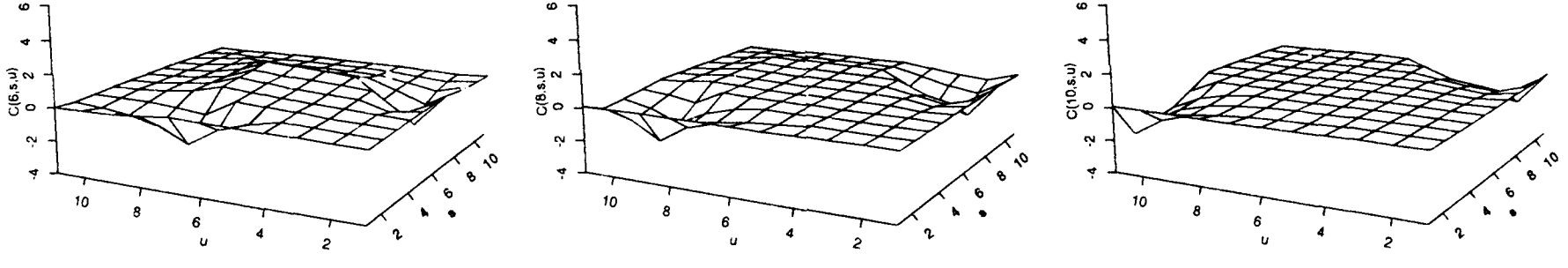


Figure 10 - C(r,s,u) : rho(s) = (0.50)**s

TEAR(1)



ROBERTSONS FIXED MODEL



ROBERTSONS RANDOM MODEL

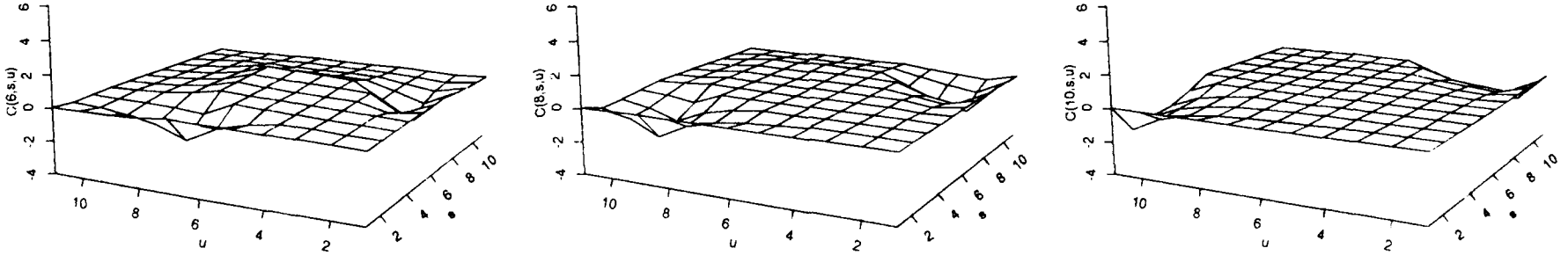


Figure 10 (cont.)

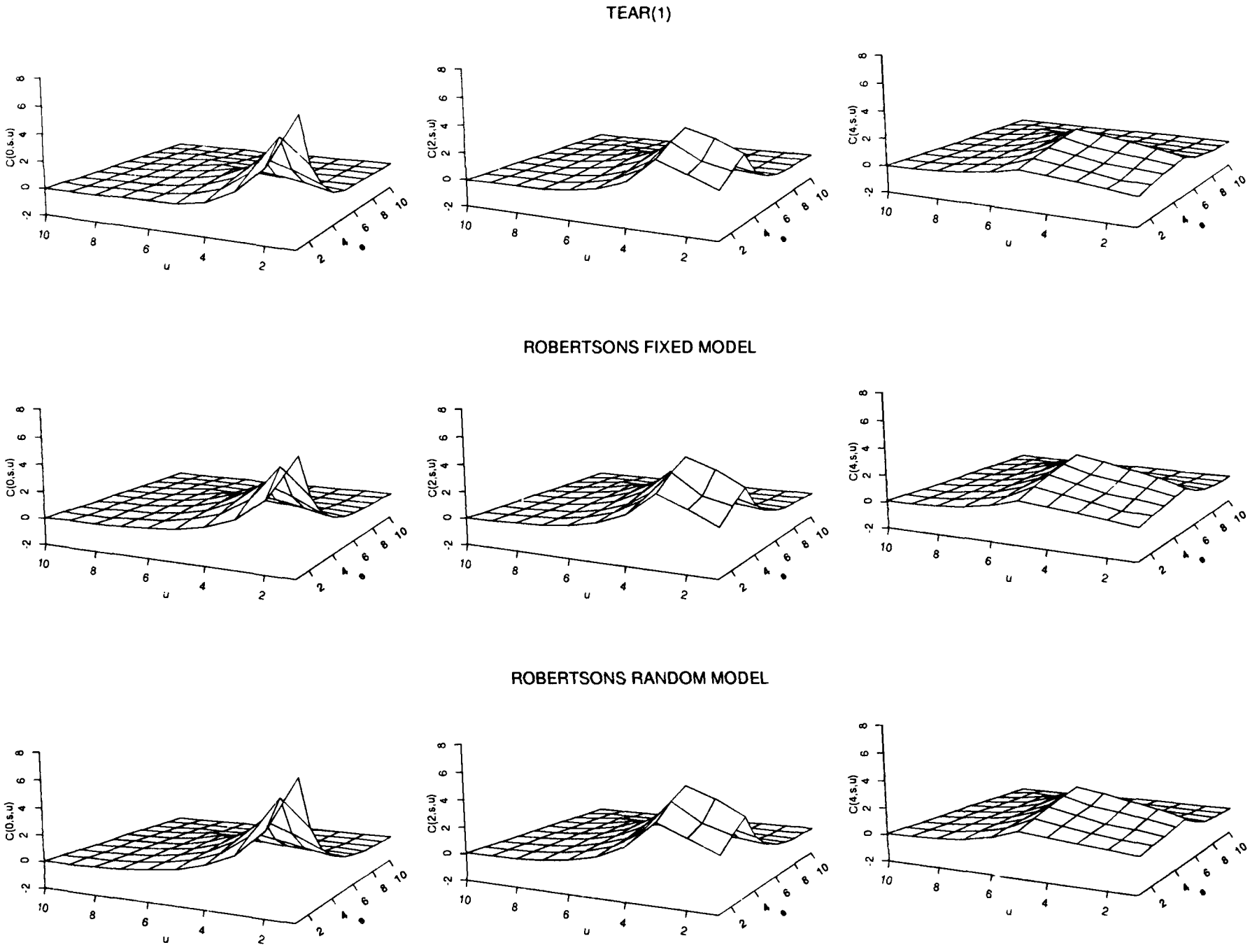
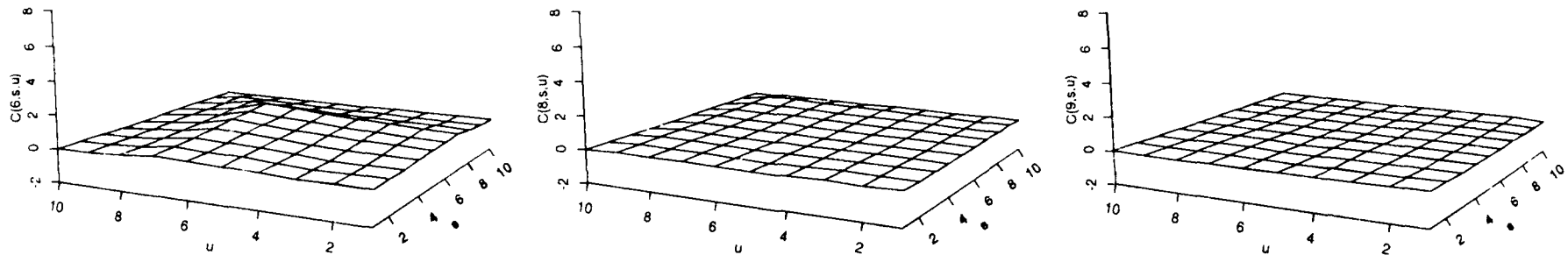
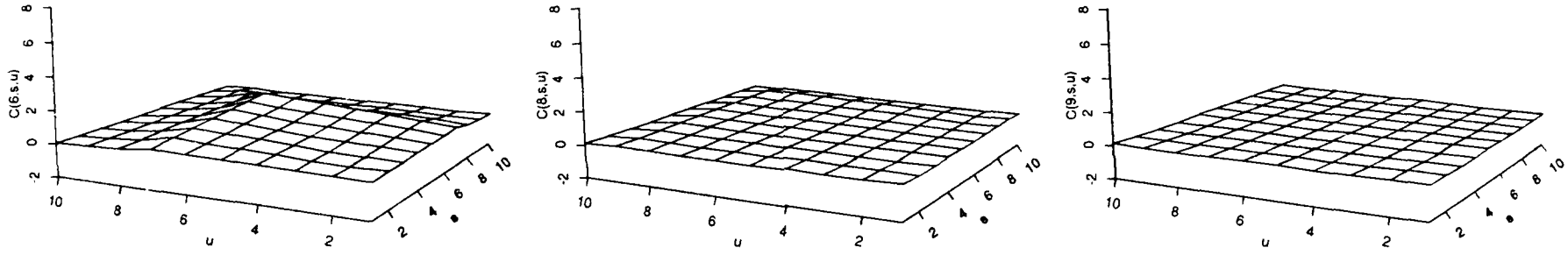


Figure 11 - Simulated C(r,s,u) : rho(s) = (0.50)**s

TEAR(1)



ROBERTSONS FIXED MODEL



ROBERTSONS RANDOM MODEL

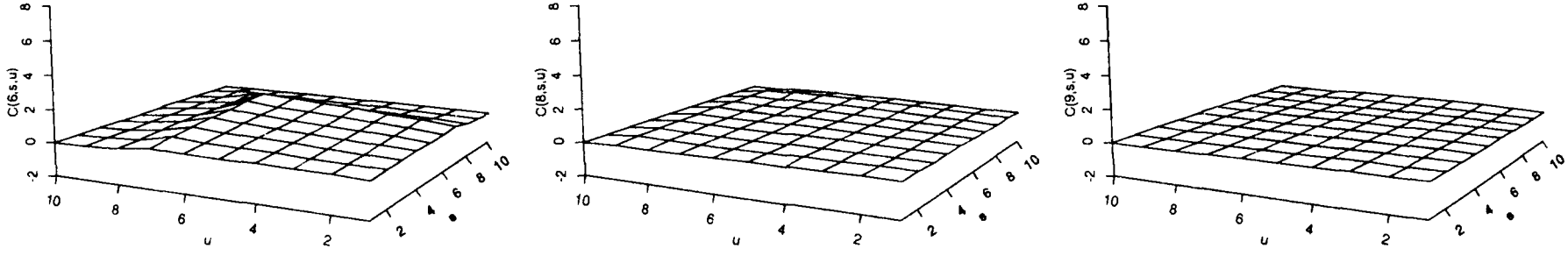


Figure 11 (cont.)

MOMENTS IN STATISTICS: APPROXIMATIONS TO DENSITIES AND GOODNESS-OF-FIT

Michael A. Stephens,
Simon Fraser University, Burnaby, B. C., Canada V5A 1S6

Summary

In this article we discuss ways in which moments are used (a) to approximate distributions, and (b) to test fit to a given distribution.

1 Approximating distributions using moments

Solomon and Stephens (1977) give a number of examples of statistics X for which the first few, or even all, the moments or cumulants may be found, but whose density $f(x)$ and distribution $F(x)$, assumed continuous, are intractable. A good example is the statistic S whose distribution is the weighted sum of independent chi-square variables, each with one degree of freedom, written

$$S = \sum_{i=1}^{k} \lambda_i (u_i)^2 \tag{1}$$

where u_i are i. i. d. $N(0,1)$, and λ_i are known weights. Many quantities in statistics have distributions (often asymptotic distributions) like S; for example, the Pearson X^2 statistic, used in testing fit to a distribution when the distribution tested contains unknown parameters which are estimated by maximising the usual likelihood, rather than the multinomial likelihood, has this distribution with some $\lambda_i \neq 1$. Other goodness-of-fit statistics, of Cramer–von Mises type, based on the empirical distribution function (EDF), also have such asymptotic distributions (see, for example, many examples in Stephens, 1986a).

One of the first examples of S to be tabulated, for $k = 2$, involved errors in target hitting during World War 2: tables for S were produced with some labour by Grad and Solomon (1955) using analytic methods. These have been extended by various authors to higher values of k, but the analysis after $k = 5$ or 6 rapidly

becomes very difficult. Thus in general it is difficult to find exact percentage points of S, but the cumulants κ_r, $r = 1, 2, \ldots$, are very easily obtained:

$$\kappa_r = \sum_{i=1}^{k} \lambda_i^r 2^{r-1}(r-1)! \tag{2}$$

2 Moments and cumulants

In this section we list definitions. The r-th moment about the origin of a random variable X, or equivalently of its distribution $f(x)$, will be called μ_r'; the r-th moment about the mean will be μ_r. The moment generating function $M_X(t)$ of X is defined by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx; \tag{3}$$

when expanded as a Taylor series,

$$M_X(t) = 1 + \mu t + \frac{\mu_2' t^2}{2!} + \frac{\mu_3' t^3}{3!} + \cdots + \frac{\mu_r' t^r}{r!} + \cdots \tag{4}$$

where $\mu = \mu_1'$ is the mean of X.

Cumulants κ_r are defined through the cumulant generating function $C_X(t) = \log M_X(t)$, where "log" refers to natural logarithm. Then

$$C_X(t) = \kappa_1 t + \frac{\kappa_2 t^2}{2!} + \frac{\kappa_3 t^3}{3!} + \cdots + \frac{\kappa_r t^r}{r!} + \cdots \tag{5}$$

Thus in principle we must find $M_X(t)$ before finding $C_X(t)$.

The following relationships exist between low-order moments and cumulants: $\kappa_1 = \mu_1' = \mu$; $\kappa_2 = \mu_2 = \sigma^2$; $\kappa_3 = \mu_3$; $\kappa_4 = \mu_4 - 3\mu_2^2$. Further relationships may be found in Kendall and Stuart (1977, vol 1).

Suppose $Z = X_1 + X_2 + X_3 + \ldots + X_k$ where X_i are independent random variables. Then a property of moment generating functions is

$$M_Z(t) = M_{X_1}(t)\, M_{X_2}(t)\, M_{X_3}(t) \ldots M_{X_k}(t),$$

so that

$$C_Z(t) = C_{X_1}(t) + C_{X_2}(t) + \cdots + C_{X_k}(t), \tag{6}$$

and it quickly follows, using obvious notation, that

$$\kappa_r(Z) = \kappa_r(X_1) + \kappa_r(X_2) + \cdots + \kappa_r X_k. \tag{7}$$

This additive property makes it very easy to find cumulants of sums of independent random variables, and hence, for example, the cumulants of S.

Two important $M_X(t)$ are those of the $N(\mu, \sigma^2)$ distribution, $M_X(t) = \exp(\mu t + \sigma^2 t^2/2)$, and the χ^2_p distribution, $M_X(t) = 1/(1-2t)^{p/2}$. Finally, it is easily shown that $\mu_r(aX+b) = a^r \mu_r(X)$, for $r \geq 2$, where a and b are any real constants, and $\kappa_r(aX+b) = a^r \kappa_r(X), r \geq 2$.

As an example, consider S. If X has a χ^2_1 distribution, the MGF of X is $1/(1-2t)^{1/2}$; thus $C_X(t) = -\frac{1}{2}\log(1-2t)$, and expansion gives $C_X(t) = t + 2t^2 + \frac{8t^3}{3!} + \frac{48t^4}{4!} + \cdots$. Thus the r-th cumulant of X is $\kappa_r = 2^{r-1}(r-1)!$, that of $\lambda_i X$ is $\lambda_i^r \kappa_r$, and by the additive property (7), the r-th cumulant of S is given by the expression (2).

3 Mathematical approximations

The approximations in this section are called "mathematical" because they are based on mathematical analysis, with known properties of accuracy and convergence, in contrast to those to be considered later.

Suppose $n(t)$ is the standard normal density

$$n(t) = e^{-t^2/2}/\sqrt{2\pi} \tag{8}$$

and let $f(x)$ be the (continuous) density of X. Then it is (nearly always) possible to expand $f(x)$ as

$$f(x) = n(x)\left\{1 + \frac{1}{2}(\mu_2 - 1)H_2(x) + \frac{1}{6}\mu_3 H_3(x) + \frac{1}{24}(\mu_4 - 6\mu_2 + 3)H_4(x) + \ldots\right\} \tag{9}$$

called a Gram-Charlier series. The $H_r(x)$ are Hermite polynomials. Lists of Hermite polynomials, and also conditions for convergence, etc., are given in Kendall and Stuart (1977, vol. 1).

The basic technique involved in deriving (9) rests on the fact that Hermite polynomials are orthogonal with respect to the kernel $n(x)$; thus

$$\int_{-\infty}^{\infty} H_i(x)\, H_j(x)\, n(x)\, dx = \begin{cases} 0, i \neq j \\ j!, i = j. \end{cases} \tag{10}$$

Then if $f(x) = \sum_i c_i n(x) H_i(x)$, multiplication by $H_j(x)$ on both sides, and integration, gives

$$c_j = \int_{-\infty}^{\infty} f(x)\, H_j(x)\, dx/j!$$

. When worked out, $c_2 = (\mu_2 - 1)/2, c_3 = \mu_3/6$, etc.

If an *infinite* set of moments is available, as for S, the density can be approximated very accurately using a Gram-Charlier series of sufficient length, but there are many statistics in practical applications for which it is difficult even to get the first four moments — see Solomon and Stephens (1977) for examples. There are two other important drawbacks:

1. A k-term fit might, at any one value of x, be worse than a $(k-1)$-term fit.

2. Gram-Charlier series with finite numbers of moments can give a negative density $f(x)$, particularly in the tails.

3.1 Percentage points approximation

A Gram-Charlier-type expansion can also be found for $F(x)$, the distribution function of X; this can be inverted to give a percentage point for a given cumulative area α. Thus suppose $F(x_\alpha) = \alpha$; we want an approximation to x_α. A **Cornish-Fisher expansion** gives $x - \xi$ as a series in Hermite polynomials in x, or (more practically useful) in ξ, where ξ is the percentile corresponding to α for the normal distribution, that is, ξ is the solution of

$$\int_{-\infty}^{\xi} n(x)\,dx = \alpha. \tag{11}$$

Again, problems can arise with the convergence to the desired x_α. For more details on mathematical expansions of Gram-Charlier or Cornish-Fisher type, see Kendall and Stuart (1977, vol. 1).

4 Pearson curves and other systems

We now turn to a method of approximation which can be thought of as "laying one curve upon another" — the approximating curve has parameters which can be varied to make a good fit. The parameters are usually chosen by matching moments or cumulants. Percentage points of the approximating curve, which are tabulated or otherwise easily found, are then used as approximations to the desired points.

A family of approximating curves is the Pearson system, where the (continuous) density $f(x)$ is approximated by $f^*(x)$, given by

$$\frac{1}{f^*(x)}\frac{df^*(x)}{dx} = \frac{a+x}{b_0 + b_1 x + b_2 x^2}. \tag{12}$$

According to the values of the constants a, b_0, b_1, b_2, integration of the right-hand side will take many forms, giving great flexibility to the system of densities $f^*(x)$. With considerable algebra (see Elderton and Johnson, 1969, for details), the constants may be put in terms of the moments:

$$\text{Suppose } A = 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2; \text{ then} \tag{13}$$

$$a = \frac{\mu_3(\mu_4 + 3\mu_2^2)}{A}, \tag{14}$$

$$b_0 = \frac{-\mu_2(4\mu_2\mu_4 - 3\mu_3^2)}{A}, \tag{15}$$

$$b_1 = -a; \tag{16}$$

$$b_2 = \frac{-(2\mu_2\mu_4 - 3\mu_3^2 - 12\mu_2^2)}{A}. \tag{17}$$

Thus knowledge of the first four moments or cumulants of X will fix the constants above: a further constant C enters on integrating, but is fixed by the fact that the total integral of $f^*(x)$ must be 1.

4.1 Percentage points

When the constants are known, the density $f^*(x)$ may be integrated and percentage points solved for numerically. Over the years, this was done, at first very laboriously, for a small range of possibilities, but a quite extensive tabulation was made, using electronic computers, in the late '60s. These tables are in *Biometrika Tables for Statisticians*, vol. II. The form of the tables is as follows. The percentage points for X, the *standardised* X-variable given by $X = (x - \mu)/\sigma$, are plotted in a two-way table indexed by the skewness and kurtosis parameters β_1 and β_2. These are defined by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}; \tag{18}$$

they have been defined to be scale-free, and $\sqrt{\beta_1}$ takes the sign of μ_3. β_1 measures skewness: a large (positive) $\sqrt{\beta_1}$ means the curve is skewed towards positive values (long tail is to the right) and *vice versa* for negative $\sqrt{\beta_1}$. A large β_2 (always positive) means the density has heavy tails. Of course, all symmetric distributions have $\beta_1 = 0$; a benchmark to measure kurtosis is the normal distribution for which $\beta_2 = 3$. Since $\kappa_4 = \mu_4 - 3\mu_2^2$, the parameter $\gamma_2 = \beta_2 - 3 = \kappa_4/\kappa_2^2$ can also be regarded as measuring kurtosis, with value $\gamma_2 = 0$ for the normal distribution.

Suppose, for a given S, we have $\sqrt{\beta_1} = 0.8$ and $\beta_2 = 4.6$. To use *Biometrika Tables*, one enters the appropriate $\sqrt{\beta_1}$ table, $\sqrt{\beta_1} = 0.8$, and travels down the left-hand column until the β_2 value, 4.6, is reached. Along the row are 17 tabulated percentage points for X, from $\alpha = 0.00$ to $\alpha = 1.00$. Interpolation must be used for $\sqrt{\beta_1}, \beta_2$ values not explicitly given.

4.2 *Un peu d'histoire*

At this point, perhaps, it might be permitted to enliven the account with what the *Guide Michelin* calls *un peu d'histoire*. At the time *Biometrika Tables* Vol. II were being prepared, I was fortunate enough to know Professor E. S. Pearson, then retired but still very active, especially as Editor of *Biometrika*. He had collaborated with workers in the U. S. to get the tables (Johnson, Nixon,

Amos and Pearson, 1963) and had carefully compiled the full set by hand. He had introduced me to Pearson curves, which, to put it mildly, did not figure prominently in statistical training of the day, and had shown me how effective they could be. He gave me a copy of the tables to use. I undertook to write a Fortran program on the IBM 650, to interpolate and find points, given the first four moments. All 20 tables were then typed onto punched cards; in the end, I got it down to approximately 45 minutes per table. This is not such a dramatic piece of history as *Michelin* usually provides (assignations and assassinations often play a prominent role), but a diminishing generation of modern readers will still empathise with the fears of losing the boxes of cards, getting them wet in the snows of Montréal, etc., not to mention the awful discovery of a wrongly-typed number!

Since then, programs have been written to integrate the density equation for $f^*(x)$ numerically and to solve for x_α for given α, or to provide the tail area for given x; one of these, kindly given to me by Amos and Daniel (1971), has been added to my program; this greatly increases the range of β_1 and β_2 for which Pearson curve approximations can be found. However, points are still output from both the Amos and Daniel part of the program and by the *Biometrika Tables* part, ostensibly as a check where available, but truthfully as a sentimental tribute to E. S. P.

Later on, Charles Davis and I (Davis and Stephens, 1983) added to the program to enable a fit to be made using knowledge of an end point (for example, that the left-hand endpoint of S is zero) and *three* moments. This is especially valuable for the type of statistic for which each successive moment requires exponentially increasing hard work — for example, the distribution of areas, or perimeters, of polygons formed by randomly dropping lines on a plane — see Solomon and Stephens (1977). The Pearson-curve fitting program is available from the author.

Further developments have included algorithms to facilitate use of Pearson curves — see, for example, Bowman and Shenton (1979a, 1979b).

4.3 Accuracy of Pearson curve fits

(a) Pearson curve densities are unimodal, or possibly J- or U-shaped, but never multimodal. They are also never negative.

(b) Percentage points or tail areas found from Pearson curve fitting have been found, for unimodal long-tailed distributions, to be very accurate in the long tail, at least for tail areas bigger then 0.005, or the 0.5% point. Pearson and Tukey (1965) discuss this issue; Solomon and Stephens (1977) give comparisons. (In making comparisons, one must of course compare the Pearson curve fit with the correct x_α, or the correct area for given x, for a distribution which is ***not itself*** a member of the Pearson family.)

(c) Davis (1975) has made extensive comparisons with Gram-Charlier fits using only four moments. Pearson curve fits are better than Gram-Charlier fits everywhere except for distributions very close to the normal, as measured by the β_1, β_2 values.

4.4 Other systems

Johnson (1949) has proposed another family (divided into three parts) of curves defined by four moments: for example, the S_U curves are those given by the relation

$$\xi = \gamma + \delta \sinh^{-1} X \tag{19}$$

where $X = (x - \mu)/\sigma$, and γ, δ are to be chosen to make the distribution of ξ as close as possible to $N(0,1)$. A discussion, and tables to facilitate the calculation of γ and δ, are in *Biometrika Tables for Statisticians* Vol. II. Other authors have also proposed families of distributions, but they have not come into such common use for the purpose of approximating percentage points.

5 Use of higher moments

We now turn to the first of two interesting questions — can higher moments be used to improve the accuracy of Pearson curve fits in the long tail of the distribution? The long tail will be supposed to lie to the right, as for the distribution of S; then, since higher values of x will contribute more to the higher moments than smaller values, we might suppose that fits using higher moments will improve accuracy. Unfortunately it is not easy to establish the four constants in terms of higher moments — of course, only four of these would be needed to fix the constants. A recursion formula exists to generate higher moments, for $r = 2, 3, \ldots$:

$$r b_0 \mu'_{r-1} + \{(r+1)b_1 + a\}\mu'_r + \{(r+2)b_2 + 1\}\mu'_{r+1} = 0 \tag{20}$$

In this recursion, the constants a, b_0, b_1 and b_2 occur, and this means that one cannot reverse the recursion and generate , say, μ and σ^2 from μ_3, μ_4, μ_5 and μ_6.

Nevertheless, one can generate the fifth and sixth moments of the Pearson curve with the same first four moments of, say, S, and compare them with the *true* fifth and sixth moments of S. The first two moments are then slightly changed, and the procedure successively repeated, until the third, fourth, fifth and sixth moments of each curve match. This will mean that the mean and variance of the Pearson curve will not be exactly the same as those for S, although they will be close, and this will probably make a worse fit in the lower tail; but for higher x the fit could improve. I have made some comparisons using this procedure, but, as one might expect, there appears to be no systematic improvement. In discussion, when this paper was first presented, the suggestion

was made to use Least Squares to make "closest" fits, in order to compare the six moments. More work is needed to compare Pearson curve fits along these various lines, but it is not likely that the improvement will be sure, or will extend to points far into the tails. In the end it must be remembered that one curve is simply being laid on top of another, with only four parameters to vary, and there is no mathematical analysis that will *guarantee* accuracy.

Other methods for developing accuracy in the extreme tails include numerical inversion of the Characteristic Function (essentially the MGF with it replacing t, where $i = \sqrt{-1}$), or saddlepoint approximations. A method due to Imhof (1961) uses numerical inversion for distributions such as S, but the computer time needed increases rapidly as the distance into the tails increases (to give small tail areas). Field (1992) has recently examined saddle-point approximations for S. These would seem to give more promise of tail-end accuracy in the long run.

6 Use of sample moments

The second interesting question is: how accurate are Pearson curve fits when sample moments are used to make the fit? In the earliest days, this was the use to which Pearson curves were applied — to find a smooth density to describe a set of data, such as lengths of beans, or width of skulls. Kendall and Stuart (1977, Vol. 1) gives details of such a fit. In general, the Pearson curves will give very good fits to a unimodal set of data, or even to J-shaped or U-shaped sets, but it is important to assess the accuracy of extrapolation from the sample to the supposed population from which it came. More precisely, we ask how close the sample fit estimate of, say, the upper-tail 5% point is to the true population 5% point, and, further, whether or not the Pearson-curve point is better than the estimated point derived from choosing the appropriate order statistic — in a sample of 1000, the 951st value in ascending order, or in a sample of size 10000, the 9501st value. Some investigation of these questions has been undertaken in two quite different ways, by Johnstone (1988) and by myself (Stephens, 1991).

The accuracy of the Pearson curve point will depend on:

1. the sample size n,
2. the α-level (tail area) of the point required,
3. the true skewness and kurtosis of the density approximated,
4. higher moments.

Johnstone gives a small study, for samples from populations with the following range of parameters:

β_1	0.0	0.0	1.0	1.0	2.0
β_2	3.3	4.0	5.25	6.0	7.5

Johnstone gives plots of the estimated coefficient of variation, CV, of the Pearson curve x_α against $-\log\alpha$, where the base of logarithms is 10. Thus the CV of the estimated $x_{0.01}$ is plotted against 2, that of the estimated $x_{0.001}$ is plotted against 3, etc . The coefficient of variation is estimated using a Taylor series approximation. As one might expect, the CV goes up markedly as α gets smaller (so $-\log\alpha$ gets larger on the x-axis), and the steepness of the rise is greater for the more skew distributions .

In Stephens (1991), Monte Carlo samples were taken from populations for which exact percentage points could be found, and the exact points were compared with those obtained from (a) Pearson curve fits using the moments of each sample, and (b) the order statistic estimate from each sample. The order statistic estimate will be asymptotically unbiased, while one can say nothing exact about the point obtained by laying one curve on another; recall that sample moments, especially the third and fourth, are extremely variable, even for quite large samples. The results showed, as expected, that the Pearson curve points were more biased. However, somewhat surprisingly, they had smaller mean square error. Therefore, it might well be preferable to use the Pearson curve points, although, again, more investigations should be made especially if the points required are far into the tail.

7 Goodness of fit using moments

In this second part of the paper, we discuss how moments are used in *Goodness-of-Fit*, that is, to test whether a random sample comes from a given (continuous) distribution. The distribution will often have unknown parameters, which must be estimated from the given sample.

7.1 Tests based on skewness and kurtosis

Suppose the r-th sample moment m_r about the mean is defined by

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r. \tag{21}$$

The sample skewness and sample kurtosis are then defined by

$$b_1 = \frac{m_3^2}{m_2^3}, b_2 = \frac{m_4}{m_2^2}. \tag{22}$$

These statistics are not unbiased estimates of β_1 and β_2, but they are consistent, that is, the bias diminishes with increasing sample size. The sample skewness and kurtosis are time-honoured statistics for testing normality, having been used in a rather *ad hoc* manner for most of this century; b_1 is compared with zero,

and b_2 with 3, the value of β_2 for the normal distribution. However, distribution theory of b_1 and b_2 is difficult, and it is only since computers have been available that extensive and reliable tables of significance points have existed for these statistics. Further, b_1 and b_2 can be combined to give one overall statistic (d'Agostino and Pearson, 1973, 1974; d'Agostino, 1986). For other distributions Bowman and Shenton (1986) have also given tables for these statistics. Studies have shown that skewness and kurtosis, especially combined, provide good omnibus tests for normality, although less is known for other distributions. For the important discrete distribution, the Poisson, all cumulants are equal to the mean, denoted by the parameter λ; a time-honoured test for the Poisson is based on the ratio of sample variance to sample mean, which of course should be about one. Again, this simple statistic appears to compete well with others in terms of power.

7.2 A formal technique based on moments

Perhaps because of the variability of sample moments, which makes calculation of significance points difficult for statistics based on these moments when calculated from samples of reasonable size, it took some time to formalize a technique based on moments. Gurland and Dahiya (1970) and Dahiya and Gurland (1972) have however devised a general procedure. The essential steps are as follows:

1. A vector ζ of length s, say, must be found, whose components ζ_i are functions of the theoretical moments, and such that each component ζ_i is linear in the parameters. (This might involve re-parametrising the distribution from its usual form).

2. The estimate h of ζ is obtained by replacing theoretical moments by sample moments.

3. The test statistic is then based on the difference $h - \zeta$.

Suppose that Σ is the covariance matrix of h, θ is the q-vector of unknown parameters, and W is the $s \times q$ matrix such that $\zeta = W\theta$. Then define

$$\hat{Q}_t = n(h - W\hat{\theta})'\hat{\Sigma}^{-1}(h - W\hat{\theta}),$$

where $\hat{\theta} = (W'\hat{\Sigma}^{-1}W)^{-1}W'\hat{\Sigma}^{-1}h$. The statistic $\hat{\theta}$ is the regression estimate of θ obtained by generalized least squares, and $\hat{\Sigma}$ is Σ with the estimate $\hat{\theta}$ used wherever θ appears.

Gurland and Dahiya (1970, 1972) showed that, asymptotically, the test statistic $\hat{Q}_t$ has the χ^2 distribution with $t = s - q$ degrees of freedom. Currie and Stephens (1986, 1990) have studied the procedure, and show several properties of $\hat{Q}_t$. Among these are the fact that the test statistic $\hat{Q}_t$ can be broken into t components, each with asymptotic distribution χ^2_1, and each testing different

features of the distribution. Each component is a function of moments or cumulants. For example, consider the test for normality, that is, for the distribution $N(\mu, \sigma^2)$. Gurland and Dahiya (1970) took $\zeta' = \{\mu, \log\mu_2, \mu_3, \log(\mu_4/3)\}$, so that $h' = \{\bar{x}, \log m_2, m_3, \log(m_4/3)\}$. The matrix W is $W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 2 \end{bmatrix}$, and $\theta = \begin{bmatrix} \mu \\ \log\sigma^2 \end{bmatrix}$. The test statistic $\hat{Q}_2$ becomes $\hat{c}_1 + \hat{c}_2$, where the two components are $\hat{c}_1 = nm_3^2/6m_2^3$ and $\hat{c}_2 = (3n/8)\{\log(m_4/3m_2^2)\}$. Thus the method leads to $nb_1/6$ and $(3n/8)\log(b_2/3)$ as test statistics, equivalent to the "old-fashioned" b_1 and b_2.

However, it should be noted that the components are not unique; they depend on how ζ is formed. Currie and Stephens (1986, 1990) discuss these questions in some detail.

8 Components of other goodness-of-fit statistics

Other goodness-of-fit statistics also have components which are functions of moments. The oldest of these was proposed by Neyman (1937), in connection with a test for uniformity.

A test for a fully specified continuous distribution (that is, all parameters known) can always be converted to a test for uniformity by means of the Probability Integral Transformation, and a test for the exponential distribution can also be so converted, even when the scale and origin parameters are not known, so that Neyman's test has wider applicability than it might at first appear. (For details of these transformations, see Stephens, 1986a, 1986b).

Neyman's test is as follows: suppose the test is that Z has a uniform distribution between 0 and 1, written $U(0,1)$. On the alternative, let the logarithm of the density of Z be expanded as a series of Legendre polynomials:

$$\log(f(z)) = A(c)\{1 + c_1L_1(z) + c_2L_2(z) + c_3L_3(z) + \cdots\}, \qquad (23)$$

where the c_i are coefficients, components of the vector c, $L_i(z)$ is the i-th Legendre polynomial, and $A(c)$ is a normalising constant.

A test for uniformity is then a test that all $c_i = 0$. The estimates of c_i are

$$\hat{c}_i = \sum_{j=1}^{n} L_i(z_j) \qquad (24)$$

where $z_1, z_2, \ldots, z_n$ is the given sample.

The first few Legendre polynomials are best expressed in terms of $y = z - 0.5$. Then

$$L_1(z) = 2\sqrt{3}y, \quad (25)$$

$$L_2(z) = \sqrt{5}(6y^2 - 0.5), \quad (26)$$

$$L_3(z) = \sqrt{7}(20y^3 - 3y), \quad (27)$$

so that the estimate $\hat{c}_1$ becomes a function of the first moment about the known mean 0.5, the second estimate $\hat{c}_2$ becomes a function of the second moment, $\hat{c}_3$ a function of both the third and the first moments, etc.

Neyman shows that the suitably normalised $\hat{c}_i$ have asymptotic $N(0,1)$ distributions, and his overall test statistic is the sum of the squares of these normalised estimates. Thus the overall statistic has an asymptotic χ^2 distribution, just as for the Dahiya-Gurland statistic, and the individual terms, based on moments, are the components of the overall test statistic.

9 EDF statistics

Another important family of goodness-of-fit statistics is that derived from the Empirical Distribution Function (EDF) of the z-sample. This family includes the well-known Kolmogorov-Smirnov statistic and the Cramer-von Mises family of statistics (for details and tests for many distributions based on these, see Stephens, 1986a).

One of the most important of the Cramer-von Mises class is A^2, introduced by Anderson and Darling (1954). The definition of A^2 is based on an integral involving the difference between the EDF and the tested distribution $F(x)$ (with parameters estimated if necessary). The working formula is

$$A^2 = -n - \frac{1}{n}\sum_i (2i-1)\left[\log z_{(i)} + \log(1 - z_{(n+1-i)})\right], \quad (28)$$

where $z_i = F(x_i)$, and $z_{(i)}$ are the order statistics.

As an omnibus test statistic, A^2 has been shown to perform well in many test situations.

Anderson and Darling showed that the asymptotic distribution of A^2 is, like S of Section 1, a sum of weighted χ^2 variables. The individual terms in the sum can again be regarded as components of the entire statistic, and Stephens (1974) has investigated these components in some detail. A remarkable result is that they too are based on Legendre polynomials, so that they are effectively the same as the Neyman components, based on moments of the z-sample. There has been some investigation of components of these and other statistics, as individual test statistics for the distribution under test; references are given by Stephens(1986a). As for the Gurland-Dahiya components, they can

be expected to be sensitive to different departures from the tested distribution. The complete test statistics of Neyman and of Anderson-Darling combine the same components, but with different weightings.

References

Amos, D.E. and Daniel, S.L. (1971) Tables of percentage points of standardized Pearson distributions. Research Report SC-RR-71 0348, Sandia Laboratories, Albuquerque, New Mexico.

Anderson, T.W. and Darling, D.A. (1954) A test of goodness of fit. *J. Amer. Statist. Assoc.* **49** 765–769.

Bowman, K. O. and Shenton, L.R. (1979a) Approximate percentage points for Pearson distributions. *Biometrika* **66** 145–151.

Bowman, K. O. and Shenton, L.R. (1979b) Further approximate Pearson percentage points and Cornish-Fisher. *Comm. Stat. B — Simula. Computa.* **8** 231–234.

Bowman, K. O. and Shenton, L. R. (1986) Moment $(\sqrt{b_1}, b_2)$ techniques. In d'Agostino, R. B. and Stephens, M. A., eds., 1986.

Currie, I. D. and Stephens, M. A. (1986a) Gurland-Dahiya statistics for the exponential distribution. Technical report, Dept. of Mathematics and Statistics, Simon Fraser University.

Currie, I. D. and Stephens, M. A. (1986b) Gurland-Dahiya statistics for the inverse Gaussian and gamma distributions. Technical report, Dept. of Mathematics and Statistics, Simon Fraser University.

Currie, I. D. and Stephens, M. A. (1990) Tests of fit using sample moments. *Scand. J. Statist.* **17** 147–156.

d'Agostino, R. B. (1986) Tests for the normal distribution. Chapter 9 in *Goodness-of-Fit Techniques*, (d'Agostino, R. B. and Stephens, M. A., eds.) Marcel Dekker: New York.

d'Agostino, R. B. and Pearson E. S. (1973) Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika* **60** 613–622.

d'Agostino, R. B. and Pearson E. S. (1974) Correction and amendment. Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika* **61** 647.

Dahiya, R. C. and Gurland, J. (1972) Goodness-of-fit tests for the gamma and exponential distributions. *Technometrics* **14** 791–801.

Davis, C. S. (1975) Unpublished MSc. thesis, Dept. of Mathematics, McMaster Univ., Hamilton, Ontario.

Davis, C. S. and Stephens, M. A. (1983) Approximate percentage points using Pearson curves. Algorithm AS 192, *Appl. Stat.* **32** 322-327.

Elderton, W. P. and Johnson, N. I (1969) *Systems of frequency curves.* Cambridge University Press: London.

Field, C. (1992) Tail areas of sums of weighted chi-squares. To appear.

Grad, A. and Solomon, H. (1955) Distribution of quadratic forms and some approximations. *Ann. Math. Statist.* **26** 464–477.

Gurland, J. and Dahiya, R. C. (1970) A test of fit for continuous distributions based on generalised minimum chi-squared. *Statistical Papers in honor of G. W. Snedecor* 115–127 (ed. T. A. Bancroft), Iowa State Univ. Press.

Imhof, J. P. (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* **48** 417–426.

Johnson, N. L. (1949) Systems of frequency curves generated by methods of translation. *Biometrika* **36** 149–176.

Johnson, N. L., Nixon, E., Amos, D. E. and Pearson, E. S. (1963) Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2, expressed in standard measure. *Biometrika* **50** 459–498.

Johnstone, I. (1986) A program for estimating uncertainties in quantile estimates derived from empirical Pearson fits. Technical report no. 381, Dept. of Statistics, Stanford Univ.

Kendall and Stuart (1977) *The Advanced Theory of Statistics*, Vol. 1, Charles Griffin: London.

Neyman, J. (1937) "Smooth" tests for goodness-of-fit. *Skand. Aktuarietidskr.* **20** 149–199.

Pearson, E. S. and Tukey, J. W. (1965) Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52** 533 546.

Solomon, H. and Stephens, M. A. (1977) Distribution of a sum of weighted chi-square variables. *J. Amer. Statist. Assoc.* **72** 881–885.

Stephens, M. A. (1974) Components of goodness-of-fit statistics. *Ann. Inst. Henri Poincaré* B. **10** 37–54.

Stephens, M. A. (1986a) Tests based on EDF statistics. Chapter 4 in *Goodness-of-Fit Techniques*, (d'Agostino, R. B. and Stephens, M. A., eds.) Marcel Dekker: New York.

Stephens, M. A. (1986b) Tests for the exponential distribution. Chapter 11 in *Goodness-of-Fit Techniques*, (d'Agostino, R. B. and Stephens, M. A., eds.) Marcel Dekker: New York.

Stephens, M. A. (1991) Computer problems in goodness-of-fit. *The Frontiers of Statistical Computation, Simulation and Modeling* (Dudewicz, J., Nelson, P., Ozturk, A., eds.) American Sciences Press: Syracuse.

Recent Applications of Higher-Order Statistics to Speech and Array Processing

Mithat C. Doğan and Jerry M. Mendel
Signal and Image Processing Institute
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089-2564.

April 6, 1992

Abstract

Higher-order statistics (HOS) are now very widely used. Two areas where they begin receiving considerable attention are array and speech processing. This paper describes some recent applications of HOS in both areas by the authors [19]-[20].

In our speech processing application, we demonstrate a way to better discriminate between voiced and unvoiced speech. This is accomplished by observing the behavior of a cumulant-based adaptive filter, and makes use of the fact that unvoiced speech is Gaussian, whereas voiced speech is definitely non-Gaussian. We have also shown a way to utilize the prediction residual from the adaptive filter to estimate the pitch period for voiced speech.

Array processing encompasses a multitude of problems, including beamforming and direction-of-arrival (DOA) estimation. We have developed fourth-order cumulant-based blind optimum beamforming algorithms that outperform existing methods. The term *blind* indicates that our methods do not require a priori knowledge of array geometry and DOA, nor they are affected by multipath propagation and presence of smart jammers. Extensive simulations support our theoretical claims on the optimality of our beamforming procedure.

1 Introduction

Our work on speech processing describes a method that consists of an adaptive predictor, a voicing decision (V/UV), and a pitch period estimator. The focus of this study is on robust detection of speech state and estimation of pitch period. This is accomplished by observing the behavior of an adaptive predictor which processes the speech signal. Higher-order- statistical analysis is proposed for discrimination of speech states. Comparing the energy of the original speech signal with that of the prediction-error residual yields the decision method. Both covariance and cumulant-based

prediction methods are investigated and the latter is shown to be a more robust way of making (V/UV) decision. Pitch estimation is accomplished by using correlation-based approaches that operate on the energy estimate of the cumulant-based prediction residual rather than the original speech signal. Pitch estimation by our method yields better performance than currently existing batch procedures.

Array processing work, as described in this paper, addresses the problem of blind optimum beamforming for a non-Gaussian desired signal in the presence of interference. Sensor response, location uncertainty and use of sample statistics can severely degrade the performance of optimum beamformers. In this paper, we propose blind estimation of the source steering vector in the presence of multiple, directional, correlated or coherent Gaussian interferers via higher-order-statistics. In this way, we employ the statistical characteristics of the desired signal to make the necessary discrimination, without any a-priori knowledge of array manifold and direction-of-arrival information about the desired signal. We then improve our method to utilize the data in a more efficient manner. In any application, only sample statistics are available, so we propose a robust beamforming approach that employs the steering vector estimate obtained by cumulant-based signal processing. We further propose a method that employs both covariance and cumulant information to combat finite sample effects. We analyze the effects of multipath propagation on the reception of the desired signal. We show that even in the presence of coherence, cumulant-based beamformer still behaves as *the* optimum beamformer that maximizes the Signal to Interference plus Noise Ratio (SINR). Finally, we propose an adaptive version of our algorithm. Simulations demonstrate the excellent performance of our approach in a wide variety of situations.

2 Cumulant-Based Adaptive Analysis of Speech Signals

Voiced/Unvoiced (V/UV) decision is an important problem in speech processing. Almost all speech coding, recognition and speaker identification systems require this information for an effective processing of speech data. In addition, low-delay speech processing systems require this decision be provided in real-time. In [2] some commonly employed features are described, and a subset of them are used to train an artificial neural network to perform V/UV decision.

In frame-based analysis of speech signals, feature extraction is performed on the current block of data, and a decision is given at the end of the period. For this reason, frame-based methods are incapable of tracking rapid changes in signal characteristics. Transitions of the state of speech within a frame period affect the decisions resulting from a frame-based analyzer. In general, this mixed state of speech within a period can not be identified and incorrect decisions will be made. This will degrade the performance of the overall speech processing system. In addition, frame-based analysis introduces delay, which may not be tolerable in low-delay systems.

Severe non-stationarity observed in speech signals and low-delay requirements of the contemporary speech processing systems motivate the use of adaptive algorithms for feature extraction in place of their batch counterparts. In general, adaptive processing techniques are designed to minimize some least-squares error criterion. Their use is motivated by the assumption that the processes are Gaussian and the performance analysis is tractable with this assumption [3]; however, this approach ignores the non-Gaussian nature of the underlying signal.

Adaptive prediction of the incoming signal and continuous monitoring of prediction error power

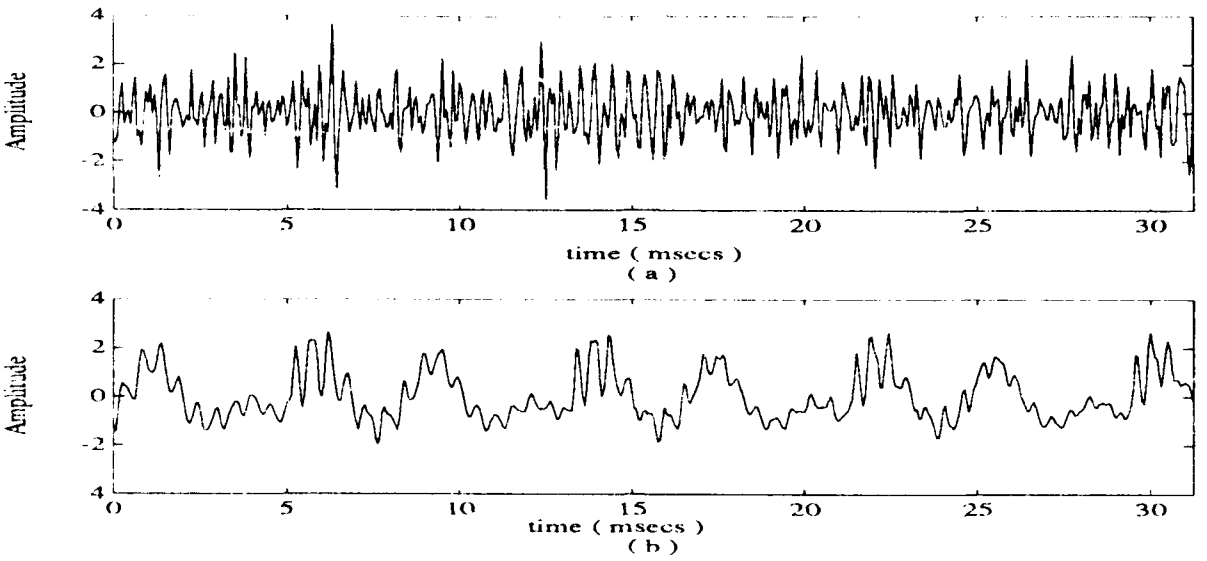


Figure 1: Typical speech signals: (a) Unvoiced speech, (b) Voiced speech.

makes detecting changes in the spectral characteristics of the process possible. We may consider such a change as an *event*. After an event, an adaptive unit will require a period to adjust itself for the new configuration. During this learning period, prediction error power will temporarily increase. This observation was used in [35], to detect abrupt changes in the autoregressive (AR) parameters of a linear process. If a lattice form is used rather than a finite impulse response (FIR) filter, reflection coefficients will be available for monitoring purposes. In addition, adaptive lattice filters exhibit better learning characteristics than their FIR counterparts. This may improve the ability to localize the event when prediction error power is monitored.

In this study, we shall investigate the application of adaptive prediction methods to detect V/UV transitions in speech signals; hence, events of interest will be V/UV or UV/V transitions. Our approach will take the speech production model into account and utilize higher than second-order statistics of speech signals.

2.1 Speech Production Model

The state of speech signal belongs to three categories: voiced, unvoiced and silence. Silent periods can be detected easily by monitoring zero crossing rate and energy of the received signals [53]. For this reason, we shall concentrate on voiced/unvoiced classification of speech.

Unvoiced sounds are generated by forming a constriction at some point in the vocal tract and forcing air through the constriction at a high velocity to produce turbulence. This creates a broad spectrum noise source to excite the vocal tract. The energy concentration is shifted to the high-frequency end of the spectrum for unvoiced sounds, but the spectrum is relatively flat when compared with that of voiced speech. Due to large number of random effects involved in the production of unvoiced speech, Gaussian noise is a valid candidate as the excitation source. This assumption is validated by Wells [73]. In his work, the bispectrum is used to make V/UV decision. It has been found that bispectrum of English fricatives tend to zero, but for vowels the situation is

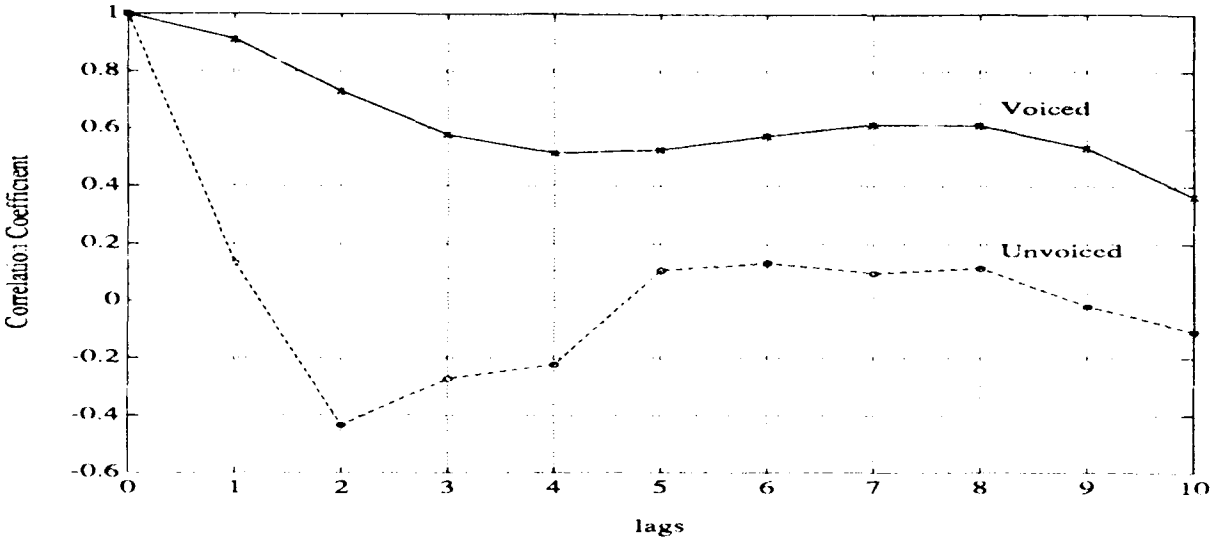


Figure 2: Adjacent sample correlation of speech signals.

just the opposite. A typical unvoiced segment of speech is shown in Fig. 1a.

Voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they vibrate in a relaxation oscillation, thereby producing quasi-periodic pulses of air which excite the vocal tract. This excitation is clearly non-Gaussian. The energy concentration is in the low-frequency side of the spectrum in the form of a fundamental component and its harmonics. In addition, voiced sounds have more energy than unvoiced sounds. A typical voiced speech segment is shown in Fig. 1b.

For voiced sounds, the vocal tract can be modelled as an all-pole linear system. The same model also holds for unvoiced sounds but the AR order is less. Correlation between adjacent samples is high for voiced sounds. On the other hand, unvoiced speech resembles white noise since its spectrum is relatively flat, yielding small correlation between adjacent samples. Correlation sequences for voiced and unvoiced cases are illustrated in Fig. 2.

The differences in the excitation and correlation properties for these two cases can be used to discriminate between them; however, with second-order statistics we can only use the correlation properties but can not utilize the information about the excitation model. This motivates the use of higher-order cumulants of speech signals.

2.2 Our Approach

In the previous section, we mentioned the distinctions between voiced and unvoiced sounds: correlation among adjacent samples and excitation models. In this section, we shall investigate methods that fully utilize this information.

Linear prediction (LP) methods are employed to accomplish our goal; however, we shall not use batch-type methods for reasons outlined previously. Linear prediction can be based on second-or higher-order statistics, however the former is usually employed. Linear prediction is essentially identifying the inverse of a linear system driven by white noise; hence, it can be considered as a

system identification problem. The system under consideration can be approximated by an AR model, so an FIR prediction filter will whiten the spectrum of the incoming signal. We shall investigate the differences between cumulant-and covariance-based adaptive prediction methods in this section.

2.2.1 Second-order statistics based adaptive filtering

Correlation-based adaptive prediction filters tend to minimize the prediction error power at the output of the filter. Since correlation among adjacent samples is high for voiced signals, we can remove a large proportion of energy from the original speech signal using prediction. On the other hand, in the case of unvoiced sounds, LP will not be that successful due to small correlation among samples. Therefore, a comparison of the input signal power with the power in the prediction residual will reveal the state of the speech signal.

Lattice prediction filters enable monitoring the variation of prediction error power with model order due to their specific structure. Autoregressive model-order-selection can be performed by selecting the tap which results in minimum prediction-error power. This leads to another discrimination between voiced and unvoiced sounds , since this order will be relatively lower for the unvoiced case.

2.2.2 Fourth-order statistics based adaptive filtering

In this section, we shall investigate the behavior of a fourth-order cumulant- based adaptive filter. An adaptive algorithm for estimating the parameters of nonstationary AR processes, excited by non-Gaussian signals is proposed in [65], and some modifications are suggested in [22]. We used the method of [65], which is in the software package $Hi-Spec^{TM}$ (trademark of United Signals and Systems, Inc.) [33]. The ideas for the covariance-based filter directly apply to this case with one important exception: the cumulant-based adaptive filter provides the solution to the cumulant-based normal equations, and this solution is *not* the one that minimizes the prediction- error power; however, one may argue that if the speech production system can be identified accurately, then the prediction error should be close to the minimum possible value.

With higher-order statistics, we have the diversity of using the excitation information: for voiced sounds , the excitation is non-Gaussian; hence, the speech production mechanism can be identified by cumulant-based AR equations. On the other hand, for unvoiced sounds the excitation is Gaussian, *making the identification problem ill-posed*[1] *The cumulant-based adaptive filter will not be able to identify the system and, since there is no associated output-power minimization criterion, prediction-error power may arbitrarily increase.* In this case, a cumulant-based filter may even amplify the speech signal making the power reduction by prediction comparison more clear than when using a covariance-based method.

To validate our ideas about covariance and cumulant-based adaptive prediction of speech signals, we performed some experiments using data from the TIMIT speech recognition database. The results verify our claims and are provided in the next section.

[1] A cumulant-based filter provides the solution of cumulant-based normal equations in an adaptive fashion; however, this set of equations becomes trivial when the input to be analyzed is a Gaussian linear process, because higher than second-order cumulants of Gaussian processes are zero.

2.3 Experiments

We start our experiments by investigating the prediction performance of correlation-and cumulant-based linear predictors in voiced speech case. An indication of performance is the energy of prediction-error residual at the output of the filter. For this purpose, we selected a voiced speech segment from the TIMIT database and performed adaptive filtering based on both correlation and cumulants. We expected that the correlation-based filter would yield better performance, since it is designed to minimize prediction-error power. The original speech signal is scaled so that estimate of its variance is unity. The results of this experiment are shown in Fig. 3. Energy values reported in this figure represent the estimate of the variance of the signal averaged over the data window. Interestingly enough, the cumulant-based filter performed better than its covariance counterpart, although the latter is designed to minimize the power of the prediction residual. We repeated this experiment with other speech segments and in all of the cases, cumulant-based filter outperformed covariance-based filter.

In voiced speech, a conventional system identification approach for estimating the AR parameters, using a least-squares fit procedure, suffers due to the nature of the excitation sequence. It is known that, for voiced speech, the source is definitely non-Gaussian ; it is quasi-periodic in nature with spiky excitations. The impulsive nature of the excitation in voiced speech is exploited in [40], by making a Bernoulli-Gaussian assumption to develop a multipulse coding scheme. In [39] , a robust linear prediction algorithm is proposed which takes into account the non-Gaussian nature of source excitation for voiced speech by assuming the excitation is from a mixture distribution, such that a large portion of the excitation sequence is from a normal distribution with small variance while a small portion comes from an unknown distribution of higher variance. Such a distribution is called *heavy-tailed Gaussian*. Based on the above mixture model, a linear prediction algorithm is devised which employs robust statistical procedures (developed in [34]) that operate in a batch mode. Although satisfactory performance is observed, the method can not track the transitions in the input data. This points out a very important fact : conventional linear prediction can be unsatisfactory due to incorrect modelling of the excitation. Of course, this carries over to the adaptive domain, i.e., a correlation-based adaptive algorithm may not be able to yield the best possible fit in the presence of outliers in the data. On the other hand, a non-Gaussian excitation is required by higher-order-statistics-based identification algorithms. A cumulant-based adaptive filter is able to reduce the power in the signal by effective prediction. although it is not based on a criterion for minimizing the power of prediction residual. Power reduction may be even more than that provided by a covariance-based filter due to the just described outlier problem.

To analyze the behavior of adaptive predictors in voiced and unvoiced speech states, we selected a 250 msec period of speech segment in which there are two transitions: voiced (0-75 msec), unvoiced (75-190 msec) and again voiced (190-250 msec). This signal is shown in Fig. 4.

We used an order ten predictor for adaptive filtering of the speech waveform. Figure 5 shows the prediction-error from a covariance-based filter. Observe that an adaptive filter based on a power minimization criterion will turn off during the unvoiced period: hence, this segment passes undistorted through the filter. The reason for this (as explained previously) is the small adjacent-sample correlation for unvoiced sounds which makes the process unpredictable. To minimize the output power,the filter turns off; however, during voiced segments deconvolution is successful. We observe a quasi-periodic pulse train for the prediction residual, which is in accordance with the excitation model for voiced speech production.

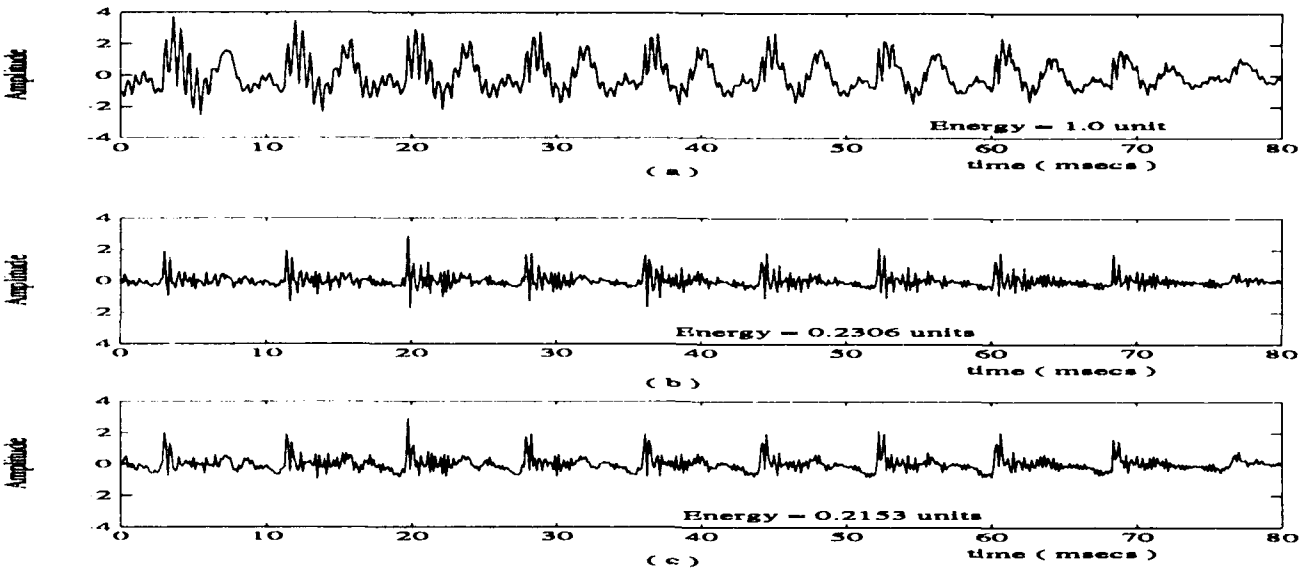


Figure 3: Energy comparisons. (a) Original speech signal; (b) prediction residual from covariance-based filter; and (c) prediction residual from cumulant-based filter.

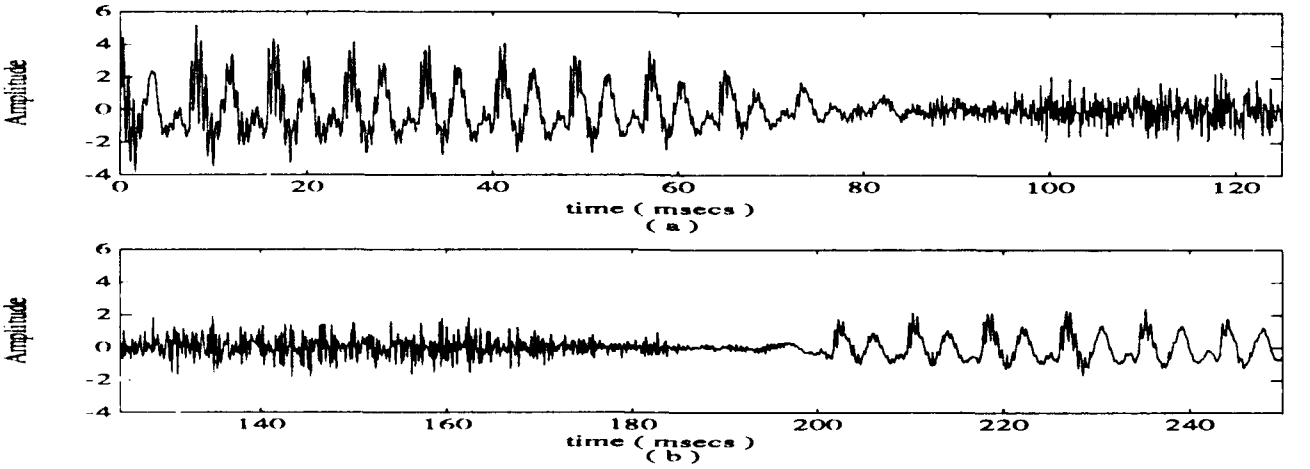


Figure 4: Speech signal to be used in the experiment: (a) first 125 msecs, (b) last 125 msecs.

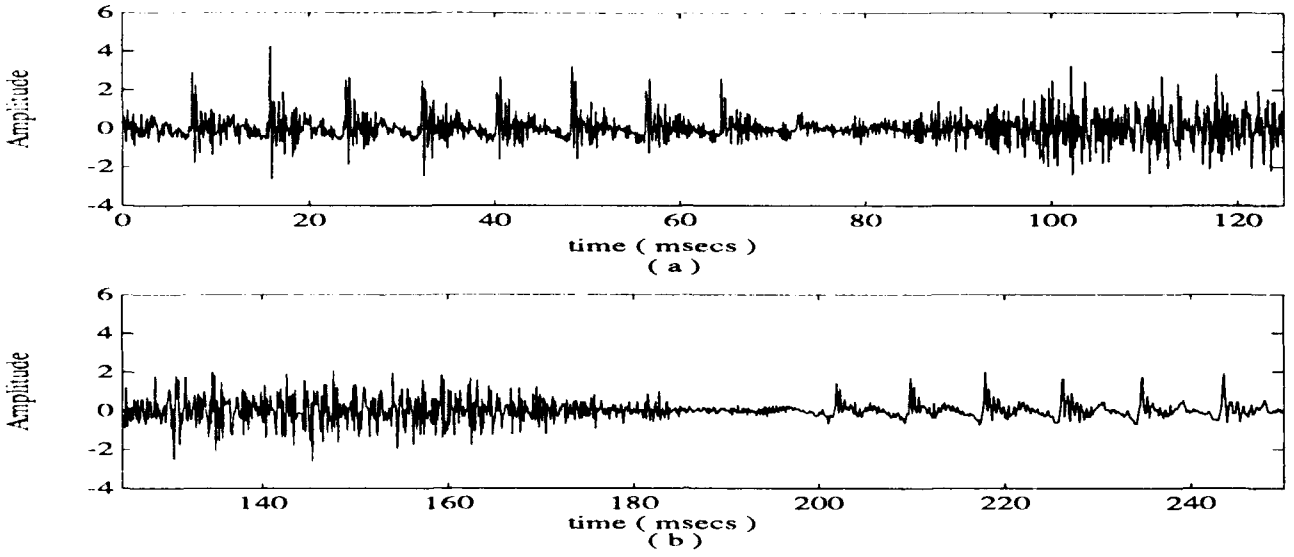


Figure 5: Prediction residual from covariance-based adaptive filter: (a) first 125 msecs, (b) last 125 msecs.

Figure 6 depicts the cumulant-based filter residual. During voiced periods, successful deconvolution is possible since the excitation is non-Gaussian, and again a quasi-periodic pulse train is observed at the output of the filter. Now, however, the filter amplifies the speech signal during the unvoiced segment. As explained before, during this mode of operation, the system identification task is ill-posed, and, since this filter has no power minimization criterion, the power of the prediction residual becomes higher than the unvoiced speech signal.

To make better comparisons concerning the energy of the original speech and prediction residuals, obtained via the two different filters, we illustrate the energy estimates in Fig. 7. Energy is estimated by first squaring the signal and then performing low-pass filtering using a 15 point Hamming window. Fig. 7 shows that, by comparing the prediction-residual power and the original-signal power, it is possible to make reliable V/UV decisions. With the cumulant-based method, even better results are obtained, because it amplifies the input data during unvoiced periods.

The observations from this experiment, validate our earlier statements; however, using a predictor may bring additional advantages as well. One important by-product is pitch period estimation. Pitch period is the time difference between the quasi-periodic excitation pulses during voiced speech. After the V/UV detection step, better pitch estimation is possible by operating on the energy estimate of prediction-residual rather than on the original speech signal. From Fig. 7, we observe that the peaks in the energy estimate sequence are spaced by a pitch period during voiced periods and they are sharper than the ones in the original speech signal due to combined filtering and squaring operations. Consequently, we may apply the correlation-based approach described in [18] to the energy estimate sequence, for a reliable, simple but robust calculation of pitch period. In [18], pitch estimation is accomplished as follows: low-pass filtered speech signal is quantized to three levels; -1,0,1 and the correlation sequence of this quantized signal is obtained. Covariance calculation is simple with the quantized sequence, since it can be performed only by addition. Finally, a peak-picking method estimates the pitch period. Peak-search is performed on the possible range of values

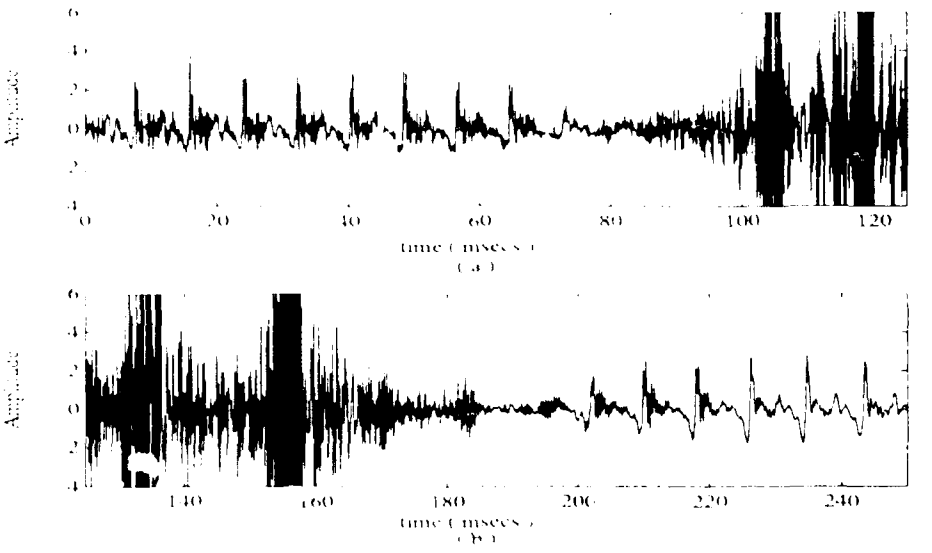


Figure 6. Prediction residual from cumulant based adaptive filter: (a) first 125 msecs, (b) last 125 msecs.

that pitch period can take, which is called the admissible pitch range. We applied this method to the energy estimate of prediction residual from the cumulant based predictor that processes the speech segment in Fig. 4. The original signal and pitch estimates are given in Fig. 8. The decisions and estimates agree with the signal characteristics. Results from the correlation-based filter are also accurate for this speech segment; however, the accuracy of the correlation-based method depends more on the threshold employed in comparing the power of prediction residual to that of the input, than in the cumulant based counterpart since the latter amplifies unvoiced speech. Therefore, we can observe degradation in the correlation based case since it is sensitive to the value of the threshold.

The second voiced speech segment in Fig. 4 is an example of the situation when harmonics are stronger than the fundamental frequency component. In general, correlation-based approaches operating directly on the speech signal fail when this event is present. To demonstrate this, we implemented the method described in [37]. In [37] pitch estimation is accomplished by calculating the correlation sequence of the low pass filtered speech signal, and employing a peak-picking algorithm on the correlation sequence. Peak searching is done on the admissible pitch range. For reliability purposes, the algorithm also investigates the possibility of pitch errors, by checking for peaks at one half, one third, one fourth, one fifth, and one sixth of the first estimate of the pitch period, if they are in the admissible pitch range. If a peak at these locations is larger in amplitude than half of that of the current estimate, the pitch estimate is changed to the location of this peak. In our experiment, the pitch detector of [37] locates the major peak at lag 68; however, its decision rule identifies another peak around lag 34 which is in the admissible pitch range. Since the amplitude of the peak at lag 34 is larger than half of that of the major peak, the final pitch estimate is chosen to be half of the correct value, which is a gross error.

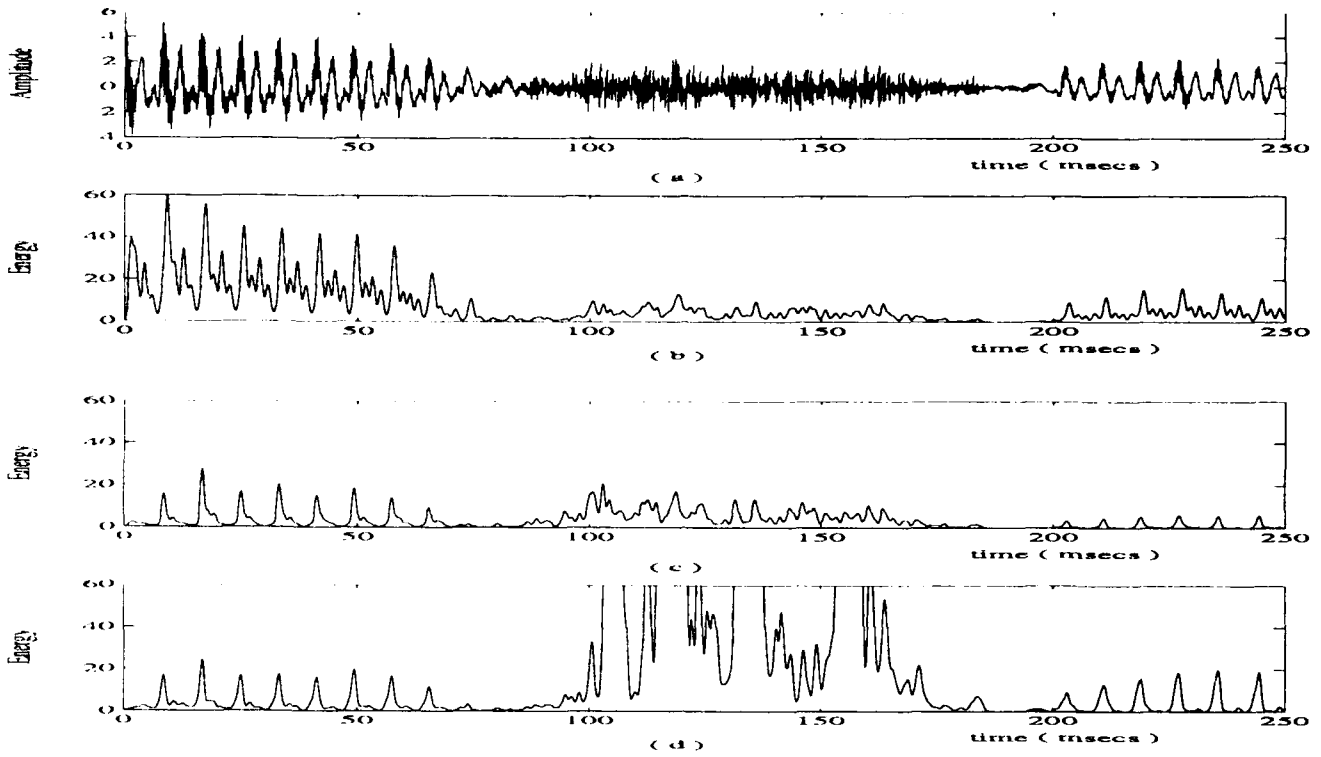


Figure 7: Energy estimates. (a) Original speech signal; (b) energy estimate of original speech signal; (c) energy estimate of prediction-error residual from covariance-based filter, (d) energy estimate of prediction-error residual from cumulant-based filter.

2.4 Conclusions

In this work, we showed that it is possible to track transitions in the state of speech using adaptive linear prediction. Both covariance and cumulant-based methods are investigated, and greater contrast between V/UV cases is demonstrated by the latter method because cumulants can use the difference in the excitation model of the two speech states.

Pitch-period estimation is also possible by linear prediction. Rather than operating on the original signal, we prefer to employ the prediction-error residual available from an adaptive filter. Cumulant-based approach operating on the power estimate of the residual process is shown to be a practical way of pitch estimation.

We investigated the prediction performance of adaptive predictors based on correlation and cumulants and found that cumulant-based prediction can outperform correlation-based prediction, although the latter is designed to minimize the power of the prediction residual. We conjectured that outliers in the excitation model of voiced speech result in this phenomena. Better prediction performance obtained via cumulants is worth investigating analytically; however, this is not tractable with real or synthesized speech since there are many parameters involved. Simpler cases, such as a single sinusoid in Gaussian noise can be analyzed to evaluate the performance of cumulant and covariance-based adaptive-line-enhancers.

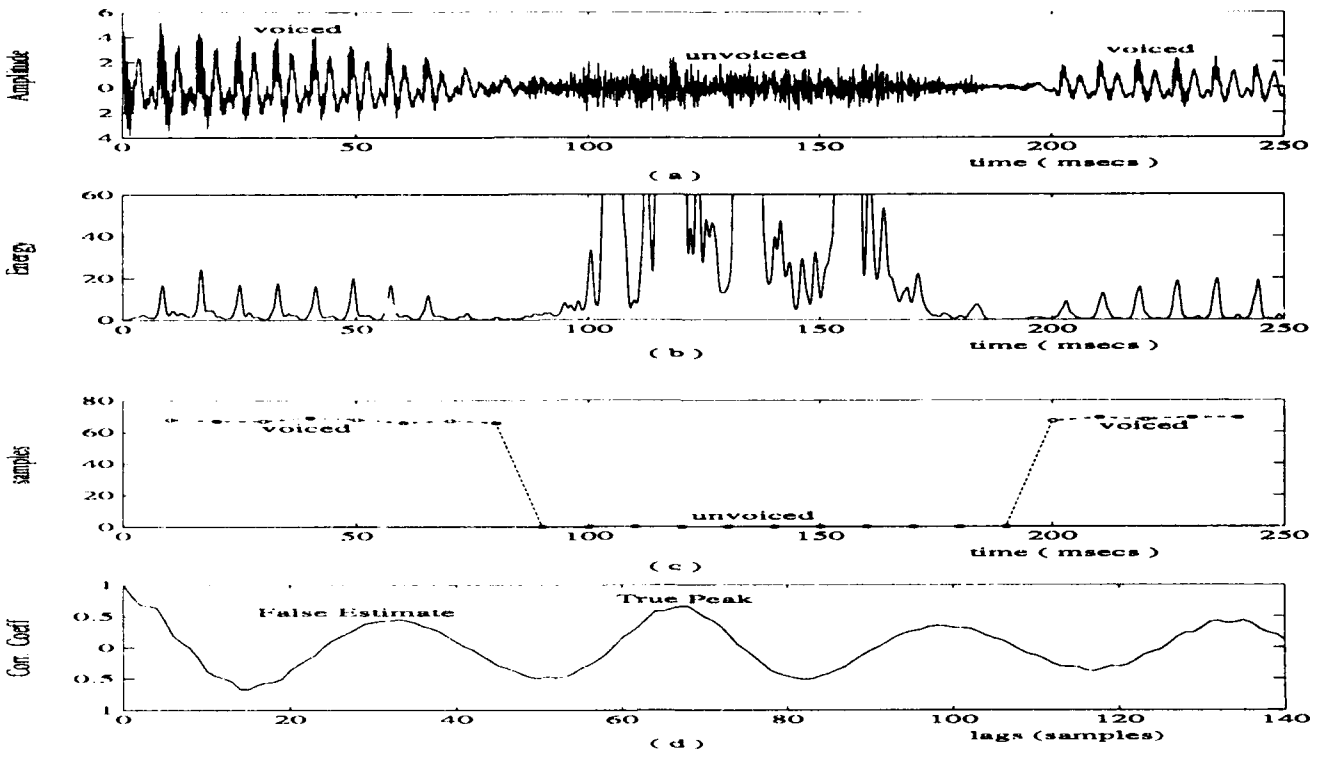


Figure 8: Pitch-period estimation experiment. (a) Original speech signal; (b) energy estimate of prediction-error residual from cumulant-based filter; (c) pitch contour obtained by processing energy-estimate sequence using the method in [18]; and (d) autocorrelation sequence of the second voiced speech segment processed by the method in [37], leading to a gross error.

3 Cumulant-Based Blind Optimum Beamforming

Array processing techniques play an important role in enhancement of signals in the presence of interference. A number of books, and an extensive literature [13,30-32,42,44,50,64,68] have already been published. Capon's minimum-variance distortionless response (MVDR) beamformer [8] has been a starting point for both signal enhancement and high-resolution direction-of-arrival (DOA) estimation.

In recent years, there has been an increasing interest in high-resolution array processing techniques based on eigendecomposition of the covariance matrix of received signals [4,17,26-27,36,38,56,60-61,69,71]. To recover the signal of interest in the presence of interfering signals, the so-called COPY function [58] is used. In this procedure, DOA's for all signals are first estimated, and then the minimum-variance processor that reconstructs the desired signal and minimizes the contribution of all interference sources is implemented. All of the previously referenced methods rely on complete knowledge of responses and locations of array elements and/or DOA information of the desired signal.

If the array manifold is unknown, or there are uncertainities, it is then necessary to calibrate

the array [55,72] ; however, this is not a practical thing to do, since calibration must be done quite frequently, and, each time, array-manifold information must be stored. In addition, calibration sources may be required. Even small errors in the calibration procedure may considerably degrade the performance. Sensitivity analyses of high-resolution methods and MVDR beamforming have been presented in [11,12,14,16,24-25,29,70,76].

In this study, we shall employ higher-order statistics of received signals to estimate the steering vector of the non-Gaussian desired signal in the presence of directional Gaussian interferers with unknown covariance structure. We assume no knowledge of array manifold, and DOA information about the desired signal. Desired signal may be voiced speech, sonar signal, radar return or a communication signal. In our work, we specialize to the communications scenario, which requires the use of fourth-order cumulants. Following a mathematical formulation of the problem in Section 3.1, we describe blind estimation and optimum beamforming procedures in Section 3.2.

Any estimation procedure is subject to errors, as is our cumulant-based source steering vector estimation method. In theory, cumulants are blind to Gaussian noise; however, their estimates are corrupted by such noise. In order to obtain satisfactory results, longer data lengths are necessary in cumulant-based signal processing. To alleviate the effects of estimation error in the beamforming step, we propose a more efficient estimation procedure that fully utilizes the data acquired by the array. We further suggest a method of combining cumulant and covariance information to yield better estimates. Then we employ a robust beamforming method based on artificial noise injection to combat mismatch in the source steering vector. We consider the estimation error as a mismatch and successfully apply this robust approach to our problem. These methods are presented in Section 3.3.

In a communications environment, multipath propagation almost always take place. In this case, all eigendecomposition-based techniques and MVDR fail. Only in some specific array configurations is it possible to decorrelate incoming signals and then estimate their DOA's. We analyze the behavior of our cumulant-based approach in Section 3.4. We show that our proposed approach behaves as *the* optimum beamformer that maximizes the Signal to Interference plus Noise Ratio (SINR).

For real-time operation (a necessary requirement in communications applications) we propose an adaptive implementation of the cumulant-based beamformer in Section 3.5. We then present simulation experiments to indicate the performance of our approach in Section 3.6. Finally, we draw our conclusions in Section 3.7.

3.1 Problem Formulation

We formulate our problem in a narrowband fashion. In array processing, a problem is classified as narrowband if the signal bandwidth is small compared to the reciprocal of the time required for the signal wavefront to propagate across the array. For a discussion on bandwidth, see [60,63].

In our formulation, lower and upper case italic letters are used to represent scalars, lower case bold font letters are used for vectors, and, upper case bold font letters are used for matrices.

3.1.1 Signal Model

Consider an array of M elements, with arbitrary sensor response characteristics and locations. Assume there are J Gaussian interference signals $\{ i_j(t),\ j = 1,2,\ldots,J \}$, and a non-Gaussian

desired signal $d(t)$, centered at frequency ω_o. We assume sources are far away from the array so that a planar wavefront approximation is possible. The additive noise present is assumed to be Gaussian with unknown covariance. With these assumptions, the received signal at the kth sensor can be expressed, as

$$r_k(t) = a_k(\theta_d)\, d(t) + \sum_{j=1}^{J} a_k(\theta_{i_j})\, i_j(t) + n_k(t) \tag{1}$$

where,

- θ_x : the direction-of-arrival of the wavefront corresponding to emitter x.
- $a_k(\theta_x)$: response of the kth sensor to xth signal wavefront, including the phase factor associated with the travel time of the signal wavefront with respect to a reference point; without loss of generality, this point can be taken as the first sensor location.
- $d(t)$: the desired non-Gaussian signal as received at sensor 1, with variance σ_d^2.
- $i_j(t)$: the jth interferer waveform as received at sensor 1; interference signals are assumed to be independent of the desired signal, and they are Gaussian processes.
- $n_k(t)$: the additive noise at the kth sensor.

Equation (1) can be rewritten in matrix notation, as

$$\begin{bmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_M(t) \end{bmatrix} = \begin{bmatrix} \mathbf{a}(\theta_d), \mathbf{a}(\theta_{i_1}), \cdots, \mathbf{a}(\theta_{i_J}) \end{bmatrix} \begin{bmatrix} d(t) \\ i_1(t) \\ \vdots \\ i_J(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_M(t) \end{bmatrix} \tag{2}$$

where $\mathbf{a}(\theta_x)$ represents the Mx1 steering vector for the wavefront from emitter x, which can be expressed as

$$\mathbf{a}(\theta_x) = \left[a_1(\theta_x), a_2(\theta_x), \ldots, a_M(\theta_x) \right]^T \tag{3}$$

We define the *array manifold* as the collection of steering vectors over all DOA's of interest. Alternative expressions for the received signal vector are,

$$\mathbf{r}(t) = \mathbf{A}\,\mathbf{z}(t) + \mathbf{n}(t) = \mathbf{a}(\theta_d)\, d(t) + \mathbf{A_I}\,\mathbf{i}(t) + \mathbf{n}(t) \tag{4}$$

In this last expression, we partitioned the Mx$(J+1)$ steering matrix $\mathbf{A}$ as,

$$\mathbf{A} = \left[\mathbf{a}(\theta_d), \mathbf{A_I} \right] \tag{5}$$

where the MxJ matrix $\mathbf{A_I}$ is the steering matrix for interference sources.

In this paper, we address the problem of optimum beamforming with an array of sensors whose responses and locations are completely unknown; hence, although we may have a priori knowledge about the direction-of-arrival of desired signal, we can not perform beamforming due to the lack of knowledge of array manifold. In [23], this problem is addressed; however, [23]'s algorithm is limited to a single interference signal. We investigate the possibility of a more general solution; namely, signal recovery in the presence of multiple interferers whose correlation structure is unknown. Before presenting our approach, which employs higher-order statistics, we demonstrate the limitations of covariance based array processing for this problem.

3.1.2 Covariance-Based Approaches

Currently used high-resolution methods of DOA estimation and minimum-variance distortionless response beamforming (MVDR) employ the covariance matrix of signals received by the array. The wavefront covariance matrix, **S**, is defined as the covariance of the source signals as received at the reference point, i.e., at sensor 1:

$$\mathbf{S} = \mathcal{E}\{\mathbf{z}(t)\,\mathbf{z}^H(t)\} \tag{6}$$

where $(\cdot)^H$ denotes complex conjugate transpose. Using the received signal model in (4), we can express the MxM covariance matrix **R** of array measurements in the following two ways:

$$\mathbf{R} = \mathcal{E}\{\mathbf{r}(t)\,\mathbf{r}^H(t)\} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{R}_n = \sigma_d^2\,\mathbf{a}(\theta_d)\,\mathbf{a}^H(\theta_d) + \mathbf{R}_u \tag{7}$$

where $\mathbf{R}_n$ is the noise covariance matrix,

$$\mathbf{R}_n = \mathcal{E}\{\mathbf{n}(t)\,\mathbf{n}^H(t)\} \tag{8}$$

and, $\mathbf{R}_u$ is the covariance matrix of the undesired signals, i.e.,

$$\mathbf{R}_u = \mathcal{E}\{[\mathbf{A_I}\,\mathbf{i}(t) + \mathbf{n}(t)]\,[\mathbf{A_I}\,\mathbf{i}(t) + \mathbf{n}(t)]^H\} \tag{9}$$

In general, the noise covariance matrix, $\mathbf{R}_n$, is unknown. With some restrictions on array orientation and noise covariance structure, some approaches for high resolution DOA estimation are proposed in [47,52] that do not require this information; however, these techniques have their limitations due to involved assumptions. Even with complete knowledge of noise covariance structure, source localization is still impossible without the knowledge of array manifold. In [56], ESPRIT algorithm is devised to overcome this problem; however, ESPRIT requires transitionally equivalent subarrays with known displacement vectors, which may also be impractical due to all the constraints on array orientation. In [21], an eigendecomposition-based beamforming approach is proposed which assumes the identifiability of the signal subspace and availability of the steering vector information for the signal of interest. Good results were obtained under these assumptions; however, this method can not handle coherent interference and spatially colored noise.

In [9-10,57], blind estimation of steering vectors for independent emitters is discussed with the following conclusion:

> Blind estimation of source steering vectors is not possible with only second-order statistics, but employing higher-than-second-order cumulants, it is possible to estimate source steering vectors up to a scale factor.

MVDR beamforming is an alternate approach for signal recovery. This approach however, requires knowledge of the steering vector for the desired source up to a scale factor and uses the covariance matrix **R** of received signals for processing. The output of the MVDR beamformer $y(t)$ can be expressed as [8]

$$y(t) = \mathbf{w}^H\,\mathbf{r}(t) = [\,\beta_1\,\mathbf{R}^{-1}\,\mathbf{a}(\theta_d)\,]^H\,\mathbf{r}(t) \tag{10}$$

where the constant β_1 is present to maintain a specified response for the desired signal and **w** denotes the weight vector of the processor.

From the above expression, it is clear that MVDR beamforming requires knowledge of $\mathbf{a}(\theta_d)$. Without knowledge of array manifold, it is not possible to determine $\mathbf{a}(\theta_d)$ even in the case of known

θ_1. Therefore, MVDR beamforming can not be directly applied to our problem. In addition, the MVDR beamformer is quite sensitive to errors in assumed sensor locations and characteristics [11-12,14,29,70,76].

In many applications, multipath propagation takes place resulting in coherent sources. Coherence presents a serious problem to DOA methods; it leads to a singular source covariance matrix **S**, for which it is not possible to estimate source locations except in some specific array configurations [18-19,61-62,66,74,75]. In the MVDR case, source coherency does not represent a problem as long as there is no source correlated with the desired signal; however, this situation is rarely met in practice. In general, the desired signal is subject to multipath propagation, and performance of MVDR approach degrades severely [54,78]. An optimum beamforming procedure has been suggested in [6] to overcome the coherence problem by using a linear array of elements with identical directional characteristics.

We are therefore looking for a method that can overcome all these problems. In the next section, we present an approach that accomplishes this by combining cumulant-based blind estimation and MVDR beamforming.

3.2 Cumulant-Based Optimum Beamforming

In the previous section, we discussed the problem of optimum beamforming and concluded that it is not possible to recover a desired signal in the presence of multiple interferers, unknown sensor noise covariance, multipath propagation and without any information about array manifold. In this section, we propose a method to overcome these problems. We propose a two-step procedure: higher-order-statistics for blind estimation of the source steering vector, followed by MVDR beamforming based on second-order statistics of received signals and steering vector estimate provided by the first step.

3.2.1 Estimation of desired signal steering vector

In this section, we employ cumulants of received signals, to estimate the steering vector of the desired signal up to a constant factor. Third-order cumulants are blind to signals with symmetric probability density function. On the other hand, most signals in communication environments have symmetric density functions, which motivates the use of fourth-order cumulants[2]. First, we define the *fourth-order zero-lag cumulant* operator of complex processes $\{x_1(t), x_2(t), x_3(t), x_4(t)\}$, as

$$cum\{x_1(t), x_2(t), x_3(t), x_4(t)\} \stackrel{\Delta}{=} E\{x_1(t)x_2(t)x_3(t)x_4(t)\} - E\{x_1(t)x_2(t)\}E\{x_3(t)x_4(t)\}$$
$$- E\{x_1(t)x_3(t)\}E\{x_2(t)x_4(t)\} - E\{x_1(t)x_4(t)\}E\{x_2(t)x_3(t)\} \quad (11)$$

Next, consider the vector $\mathbf{c} = [c_1, c_2, \ldots, c_M]^T$, defined as

$$c_l \stackrel{\Delta}{=} cum\{x_1(t), x_1^H(t), x_1^H(t), x_l(t)\} \quad l = 1, 2, \ldots, M. \quad (12)$$

As suggested in [13], there are various ways of defining fourth-order statistics of complex random processes. We follow the approach presented in [51] in (12). Since interference signals are independent of the desired signal and they are Gaussian with zero fourth-order cumulants, we can express

[2] An estimation procedure based on third order statistics is presented in[19].

c_l as

$$c_l = cum\{a_1(\theta_d)s_d(t), a_1^H(\theta_d)s_l^H(t), a_1^H(\theta_d)s_d^H(t), a_l(\theta_d)s_d(t)\} \tag{13}$$

Using properties of cumulants, we obtain

$$c_l = |a_1(\theta_d)|^2 \, a_1^H(\theta_d) \, \gamma_{d,4} \, a_l(\theta_d) \tag{14}$$

where $\gamma_{d,4}$ denotes the *zeroth* lag of the *fourth*-order cumulant of the desired signal. Defining $\beta_2 = |a_1(\theta_d)|^2 \, a_1^H(\theta_d) \, \gamma_{d,4}$ we have the following expression for the Mx1 vector $\mathbf{c}$:

$$\mathbf{c} = \beta_2 \, \mathbf{a}(\theta_d) \tag{15}$$

Observe that the vector $\mathbf{c}$ is a *replica of the steering vector of the desired signal up to a scale factor.* We show in the next section how this information can be used to recover the desired signal.

3.2.2 Interference Rejection

With the knowledge of the steering vector of the desired signal, interference rejection is possible using the following minimum-variance distortionless response formulation: find the weight vector $\mathbf{w}$ that minimizes the power, $\mathbf{w}^H \, \mathbf{R} \, \mathbf{w}$, at the output of the beamformer subject to the constraint $\mathbf{w}^H \, \mathbf{c} = 1$, where $\mathbf{c}$ is obtained via the cumulant-based estimation procedure described in Subsection 3.2.1. The solution to this optimization problem is well-known [8], and can be expressed as

$$\mathbf{w}_{cum} = \beta_3 \, \mathbf{R}^{-1} \, \mathbf{c} \tag{16}$$

where the constant $\beta_3 = (\mathbf{c}^H \, \mathbf{R}^{-1} \, \mathbf{c})^{-1}$ is present in order to maintain the linear constraint.

Due to the constraint $\mathbf{w}^H \, \mathbf{c} = 1$, the power minimization procedure does not cancel the desired signal, but rejects all interference components and sensor noise in the best possible manner. Note that this is accomplished without knowledge of covariance structure of interference signals, sensor noise or array manifold. In the sequel, we refer to the processor in (16) as CUM_1. The proof that this cumulant-based beamformer is identical to the maximum SINR processor is provided in Section 3.4, where the general multipath case is treated.

3.3 Robust Beamforming

In this section, we first propose an approach that utilizes the received data in the estimation of the source steering vector in a more efficient manner. We then suggest a method that uses both cumulants and covariance information under some scenarios. Finally, we employ a robust method to combat the effects of estimation errors.

3.3.1 Efficient Utilization of Array Data

In the previous section, we presented a method of blind estimation of the desired source steering vector from the received data; however, the proposed approach is rather inefficient in the sense that only the first sensor is taken as reference. For example, if the connection from this element to the processor is broken, then the estimation objective can not be accomplished. Similarly, due to poor receiving circuitry following this array element, the reference signal may be very noisy, degrading the quality of the estimate. We can overcome these difficulties by using multiple reference elements.

Define the matrix $\mathbf{C}$ with the (k,l)th element,

$$C_{k,l} \triangleq cum\{r_k(t), r_k^H(t), r_k^H(t), r_l(t)\} \text{ where } k,l = 1,\ldots,M. \quad (17)$$

With true statistics, the cross-cumulant matrix $\mathbf{C}$ will have rank 1, since all its columns are scaled replicas of the desired source steering vector; however, with sample statistics this condition never holds. The left singular vector of $\mathbf{C}$ with the largest singular value can be used as the estimate of the desired source steering vector removing the effects of noise. In this way, we utilize array data more efficiently[3]. The beamformer that employs the steering vector estimate obtained in the way described above is referred to as the CUM_2 beamformer in the sequel.

In addition, the Total Least Squares algorithm, that takes the errors in both the received data covariance matrix estimate and the steering vector estimate into account, is a better choice for computing the optimum weight vector, as suggested in [78], but it is computationally expensive. If extra computations are feasible, we suggest the use of the Constrained Total Least Squares algorithm [1], for even better numerical results.

3.3.2 Covariance-Cumulant (C^2) Approach

In some array processing applications, sensor noise covariance structure has a definite structure enabling a whitening operation on the received data. The principal eigenvectors of the covariance matrix of this processed data reveal the subspace spanned by the steering vectors of directional signals illuminating the array [58]. Hence, the steering vector estimate obtained by the cumulant-based approach can be improved by projecting this estimate on the subspace spanned by the principal eigenvectors of the covariance matrix. This improved estimate can then be used in the beamforming procedure of Section 3.2.2. The motivation behind this approach is that covariance estimates exhibit less variance than cumulant estimates, but in the covariance domain we can not identify the source steering vector if there are multiple sources. This procedure yields an estimate of the steering vector from covariance-matrix information by employing the cumulant-based estimate as side information. A mathematical description of this approach is presented below:

1. From the received da a, estimate the covariance matrix $\mathbf{R}$ and the desired signal steering vector $\mathbf{c}$ by the cumulant-based procedure.

2. Perform an eigendecomposition of the sample covariance matrix, to reveal the signal and noise subspaces: the eigenvectors of $\mathbf{R}$ with the repeated minimum eigenvalue span the noise subspace [58], while the rest span the signal subspace.

3. Assume the signal subspace is $(J+1)$ dimensional. Then, the basis vectors for the signal subspace, obtained from the eigendecomposition procedure, can be sorted in an $Mx(J+1)$ matrix $\mathbf{E}_s$ with the column space identical to the signal subspace.

4. Project the cumulant-based steering vector estimate $\mathbf{c}$, on the signal subspace to obtain an improved estimate $\mathbf{c}_{imp}$, as

$$\mathbf{c}_{imp} = \mathbf{E}_s\mathbf{E}_s^H\mathbf{c}$$

5. Compute the weights for the beamformer, as

$$\mathbf{w}_{imp} = \mathbf{R}^{-1}\mathbf{c}_{imp}$$

[3] A method that utilizes the array data even more efficiently is presented in[19].

3.3.3 Robustness Constraint

Any estimation procedure is inevitably subject to errors. MVDR beamforming is extremely sensitive to mismatch [11-12,14,16,29,70,76], especially in high SNR conditions and in arrays with large number of elements. A variety of constraints have been summarized in [68] assuming perfect knowledge of element characteristics and locations; however, in our case these methods are not applicable since there is no available information about the array manifold to design effective constraints.

Errors in the steering vector estimate result in signal cancellation. This mismatch condition, arising from non-perfect estimation, can be viewed as the problem of optimum beamforming with an array of sensors at slightly perturbed locations. In [15], a method that constrains the white noise gain of the processor is proposed for the solution of the latter problem. In this section, we use the same approach to alleviate the effects of estimation errors in cumulant-based optimum beamforming.

In order to understand the mismatch problem and find a way to alleviate its effects, we need to analyze the problem analytically. Consider the power response of a beamformer with a weight vector $\mathbf{w}$, as a function of DOA θ, defined as

$$P(\theta) \triangleq |\mathbf{w}^H \mathbf{a}(\theta)|^2 \tag{18}$$

with $\mathbf{a}(\theta)$ denoting the steering vector for an arrival from θ. The derivative, $\partial P(\theta)/\partial\theta$, can be expressed, as

$$\frac{\partial P(\theta)}{\partial \theta} = 2Re\{\, \mathbf{w}^H \mathbf{a}(\theta) \, [\, \sum_{l=1}^{M} w_l \frac{\partial}{\partial \theta} \, a_l^H(\theta) \,\,]\} \tag{19}$$

Now consider the following scenario: we have an MVDR processor *looking* at θ_o, which is the expected DOA for the desired signal. Instead, the source illuminates the array from θ_d which is very close but not equal to θ_o. In this case, the beamformer treats the desired signal as interference and nulls it; however, due to the distortionless response constraint for θ_o, and since the angles are very close, the derivative $\partial P(\theta)/\partial\theta$ must be large in magnitude for θ between θ_d and θ_o. From the derivative expression (19), it is clear that this is possible only if the norm of the weight vector increases, since the inner product, $\mathbf{w}^H\mathbf{a}(\theta)$, and, the derivatives, $\{\frac{\partial}{\partial\theta}\, a_l^H(\theta)\}_{l=1}^M$ are bounded. In this situation, the constraint is maintained by increasing the angle between the weight vector and the look-direction steering vector. This phenomena was exploited in [77], for tuning the beamformer to acquire a weak desired signal in the presence of strong interference.

Note that the white-noise amplification factor for any processor with a weight vector $\mathbf{w}$ is $\mathbf{w}^H\mathbf{w}$; hence, the nulling phenomena can be prevented if the white noise level at the processor is sufficiently high so that output power minimization criterion limits the increase in the norm of $\mathbf{w}$. This can be achieved by perturbing the covariance matrix estimate of array measurements by a scaled identity matrix as,

$$\mathbf{R}_p = \mathbf{R} + \epsilon \mathbf{I} \tag{20}$$

where ϵ is a non-negative parameter which adjusts the strength of perturbation. Alternatively, it is possible to coin a term *virtual* SNR, SNR_v, defined as

$$\mathrm{SNR}_v = \mathrm{SNR} - 10\log_{10}\left(\frac{\epsilon + \sigma_n^2}{\sigma_n^2}\right) \tag{21}$$

We then determine the weight vector as,

$$\mathbf{w} = \mathbf{R}_p^{-1}\,\mathbf{a}(\theta_o) \tag{22}$$

A recent method presented in [15] performs this procedure in an adaptive fashion by a simple scaling of the weight vector. In our case, we do not have source DOA information, but we do have an estimate of the steering vector. It is therefore possible to use this estimate in place of $\mathbf{a}(\theta_o)$ in (22) to formulate the cumulant-based processor with limited signal nulling property.

3.4 Multipath Phenomena

Eigendecomposition-based high-resolution methods [4,17,26-27,36,38,56,60-61,69,71] have proven to be effective means of obtaining bearing estimates of far-field narrowband sources from noisy measurements. The performance of these algorithms is severely degraded when coherence is present. Several methods have been proposed to solve the coherent signals problem with restrictions on array geometry [48-49,61-62,66,74,75]; however, with lack of knowledge of array manifold it is not possible to solve the coherence problem. MVDR beamforming also fails to perform optimally, when interference signals are correlated with the desired signal [54,78]. In some scenarios, even the conventional beamformer outperforms the MVDR approach due to signal cancellation in the MVDR beamformer.

In Section 3.2, we showed that the cumulant-based beamformer is not affected by the presence of coherence among interfering Gaussian signals as long as they are not correlated with the desired signal. The same is not possible for high-resolution DOA estimation methods; but, the MVDR beamformer may perform equally well if the desired signal steering vector is known and a satisfactory estimate of $\mathbf{R}$ is available. In this section, we show that the cumulant-based approach is not affected by the presence of multipath propagation of the desired signal. In addition, we show that *the cumulant-based processor turns out to be the maximal-ratio-combiner* [5] *that maximizes the SINR.*

With the presence of multipath propagation or smart jamming, our signal model in (1) changes to

$$r_k(t) = d(t) \sum_{l=1}^{L} a_k(\theta_{d_l})\,\eta_l + \sum_{j=1}^{J} a_k(\theta_{i_j})\,i_j(t) + n_k(t) \tag{23}$$

or in vector form

$$\mathbf{r}(t) = \left[\, \mathbf{a}(\theta_{d_1}),\ \mathbf{a}(\theta_{d_2}),\ \cdots,\ \mathbf{a}(\theta_{d_L}) \,\right] \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_L \end{bmatrix} d(t) \;+\; \mathbf{A_I}\,\mathbf{i}(t) \;+\; \mathbf{n}(t) \tag{24}$$

where the set of scalars $\{\, \eta_1, \eta_2, \ldots, \eta_L \,\}$ constitute the multipath coefficients for an L-ray scenario. The set of vectors, $\{\, \mathbf{a}(\theta_{d_1}),\ \mathbf{a}(\theta_{d_2}),\ \ldots,\ \mathbf{a}(\theta_{d_L}) \,\}$ are the corresponding steering vectors of the L-ray model. Letting

$$\mathbf{b} \;\triangleq\; \left[\, \mathbf{a}(\theta_{d_1}),\ \mathbf{a}(\theta_{d_2}),\ \cdots,\ \mathbf{a}(\theta_{d_L}) \,\right] \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_L \end{bmatrix} \;=\; \mathbf{A_D}\,\eta \tag{25}$$

we can reduce the signal model for multipath phenomena to the single-ray propagation model of Section 3.1.1,

$$\mathbf{r}(t) = \mathbf{b}\, d(t) + \mathbf{A_I}\, \mathbf{i}(t) + \mathbf{n}(t) \tag{26}$$

because we can view the vector $\mathbf{b}$ as a *generalized* steering vector for a single desired signal although it may not be a vector in the array manifold. Therefore, following our work in Section 3.2, cumulant-based blind estimation procedure will yield

$$\mathbf{c} = \beta_4\, \mathbf{b} \tag{27}$$

where $\beta_4 = |b_1|^2\, b_1^H\, \gamma_{d,4}$, in which b_1 is the first component of $\mathbf{b}$. Incorporating (27) into the constrained power minimization procedure, we obtain the following weight vector,

$$\mathbf{w}_{cum} = \beta_5\, \mathbf{R}^{-1}\, \mathbf{c} = \beta_4\beta_5\, \mathbf{R}^{-1}\, \mathbf{b} \tag{28}$$

where $\beta_5 = (\, \mathbf{c}^H\, \mathbf{R}^{-1}\, \mathbf{c}\,)^{-1}$.

Next, we find an alternate expression for $\mathbf{w}_{cum}$. Recall that the optimization problem which results in $\mathbf{w}_{cum}$ is: minimize $\mathbf{w}^H\mathbf{R}\mathbf{w}$ subject to $\mathbf{w}^H\mathbf{c} = 1$, or by (27), $\mathbf{w}^H\mathbf{b} = 1/\beta_4$. We can express the output power in the following way by using (9) and (26),

$$\mathbf{w}^H\mathbf{R}\mathbf{w} = \sigma_d^2\, |\, \mathbf{w}^H\, \mathbf{b}\, |^2 + \mathbf{w}^H\mathbf{R}_u\mathbf{w} \tag{29}$$

but, due to the constraint $\mathbf{w}^H\mathbf{b} = 1/\beta_4$, the first term in the above expression is a constant. Therefore, the original optimization problem can be translated into : minimize $\mathbf{w}^H\mathbf{R}_u\mathbf{w}$, subject to $\mathbf{w}^H\mathbf{c} = 1$ or equivalently, $\mathbf{w}^H\mathbf{b} = 1/\beta_4$. The solution to this problem is

$$\mathbf{w}_{cum} = \beta_6\, \mathbf{R}_u^{-1}\, \mathbf{c} \tag{30}$$

where $\beta_6 = (\mathbf{c}^H\, \mathbf{R}^{-1}\, \mathbf{c})^{-1}$. Of course, this solution can also be expressed in terms of $\mathbf{b}$, as

$$\mathbf{w}_{cum} = \beta_7\, \mathbf{R}_u^{-1}\, \mathbf{b} \tag{31}$$

where $\beta_7 = \beta_4\beta_6$.

Note that although (30) and (31) are alternate expressions for $\mathbf{w}_{cum}$, they are not the way to actually compute $\mathbf{w}_{cum}$, since $\mathbf{R}_u$ is not available in general.

Next, we determine the weight vector that yields the maximum SINR. SINR can be expressed as a function of the weight vector of the beamformer, as

$$\mathrm{SINR}(\mathbf{w}) = \sigma_d^2\, \frac{\mathbf{w}^H\, \mathbf{b}\, \mathbf{b}^H\, \mathbf{w}}{\mathbf{w}^H\, \mathbf{R}_u\, \mathbf{w}} \tag{32}$$

Defining, $\mathbf{v} = \mathbf{R}_u^{1/2}\, \mathbf{w}$ so that $\mathbf{w} = \mathbf{R}_u^{-1/2}\, \mathbf{v}$, we can reexpress (32), as

$$\mathrm{SINR}(\mathbf{w}) = \mathrm{SINR}(\, \mathbf{R}_u^{-1/2}\, \mathbf{v}) = \sigma_d^2 \frac{|\, \mathbf{v}^H\, \mathbf{R}_u^{1/2}\, \mathbf{b}\, |^2}{\mathbf{v}^H\, \mathbf{v}} \tag{33}$$

Applying the Schwarz inequality [50] to (33), we find that

$$\mathrm{SINR}(\mathbf{w}) = \mathrm{SINR}(\, \mathbf{R}_u^{-1/2}\, \mathbf{v}) \leq \sigma_d^2 \|\, \mathbf{R}_u^{-1/2}\, \mathbf{b}\, \|^2 = \sigma_d^2\, \mathbf{b}^H\mathbf{R}_u^{-1}\mathbf{b} \tag{34}$$

where equality holds if and only if

$$\mathbf{v} = \beta_8 \, \mathbf{R}_u^{-1/2} \, \mathbf{b} \tag{35}$$

in which β_8 is a non-zero constant. Consequently, the optimum weight vector $\mathbf{w}_{\text{SINR}}$, which yields the maximum SINR, can be determined from $\mathbf{w} = \mathbf{R}_u^{-1/2}\mathbf{v}$ and (35), as

$$\mathbf{w}_{\text{SINR}} = \beta_8 \, \mathbf{R}_u^{-1} \, \mathbf{b} \tag{36}$$

Based on this derivation, some comments are in order. It is clear, by comparing (31) and (36), that the cumulant-based beamformer does indeed yield the maximum possible SINR, since $\mathbf{w}_{cum}$ is just a scaled version of $\mathbf{w}_{\text{SINR}}$. This observation proves that the cumulant-based beamformer is optimal. In addition, $\mathbf{w}_{cum}$ can be computed from the received data, whereas $\mathbf{w}_{\text{SINR}}$, as implemented in (36), requires knowledge of $\mathbf{R}_u$, which can not be determined from the received data in the presence of the desired signal. Finally, note that robust approaches presented in Section 3.3 are directly applicable in the presence of multipath.

3.5 Adaptive Processing

In real-world applications, adaptive beamforming is an important requirement, especially when the desired signal source is in relative motion with respect to the array. In this section, we address this problem by providing an "estimate and plug" type of adaptive algorithm for the CUM_1 method.

The beamforming procedure (16) requires the inverse of the sample covariance matrix to compute the weights. We can estimate the covariance matrix recursively, as

$$\hat{\mathbf{R}}_t = (1 - \alpha_1)\hat{\mathbf{R}}_{t-1} + \alpha_1 \mathbf{r}(t)\mathbf{r}^H(t) \tag{37}$$

Since we need to propagate the inverse of $\hat{\mathbf{R}}_t$, we use the Sherman-Morrison formula [46], to obtain

$$\hat{\mathbf{R}}_t^{-1} = \frac{1}{1-\alpha_1}\left[\hat{\mathbf{R}}_{t-1}^{-1} - \alpha_1 \frac{\hat{\mathbf{R}}_{t-1}^{-1}\mathbf{r}(t)\mathbf{r}^H(t)\hat{\mathbf{R}}_{t-1}^{-1}}{1 - \alpha_1[1 - \mathbf{r}^H(t)\hat{\mathbf{R}}_{t-1}^{-1}\mathbf{r}(t)]}\right] \quad t = 1, 2, \ldots \tag{38}$$

with $\hat{\mathbf{R}}_0^{-1} = \gamma \mathbf{I}$ where γ is a large positive number and α_1 controls the learning rate for second-order statistics.

To compute the weight vector, we also need the cumulant-based estimate of the source steering vector $\mathbf{c}$. We can estimate it recursively as

$$\hat{c}_l(t) = (1 - \alpha_2)\hat{c}_l(t-1) + \alpha_2[\,|r_1(t)|^2 r_1^H(t) r_l(t) - 2p(t)q(t) - v^H(t)x(t)] \tag{39}$$

with the auxilary processes defined as

$$p(t) = (1 - \alpha_3)p(t-1) + \alpha_3 |r_1(t)|^2$$

$$q(t) = (1 - \alpha_3)q(t-1) + \alpha_3 r_1^H(t) r_l(t)$$

$$v(t) = (1 - \alpha_3)v(t-1) + \alpha_3 r_1^2(t)$$

$$x(t) = (1 - \alpha_3)x(t-1) + \alpha_3 r_1(t) r_l(t)$$

The auxiliary processes are required in order to implement the cross-correlation terms in (11). The initial values for the auxilary processes can be set to zero. Different learning rates are provided

to emphasize the fact that higher-order statistics require longer periods to acquire the required information.

We can perform adaptive beamforming by computing the weight vector at each time as

$$\mathbf{w}(t) = \hat{\mathbf{R}}_t^{-1}\hat{\mathbf{c}}(t) \tag{40}$$

and obtain the array output, as

$$y(t) = \mathbf{w}^H(t)\mathbf{r}(t). \tag{41}$$

Adaptive versions of CUM_2 and C^2 methods will appear in a later publication.

3.6 Simulations

In this section we present various experiments to illustrate the performance of cumulant-based beamforming. In all of the experiments we employed a uniformly spaced linear array, rather than an arbitrary geometry. This is done for two reasons: Covariance-based techniques are mainly designed for this type of array structure, e.g., the spatial smoothing algorithm [48-49,61-62,66,74,75], so that it will be possible to compare both previous and future work with our current results. In addition, allowing a sufficient number of multipath rays, it is possible to represent any arbitrary steering vector by the linear array, since the steering vectors of the uniformly spaced isotropic linear array exhibit Vandermonde structure, resulting in linearly independent vectors for different DOA's. In all batch type of experiments, the record length is 1000 snapshots and the array has 10 isotropic elements with uniform half-wavelength spacing.

3.6.1 Experiment 1: Desired Signal in White-Noise

In this experiment, we employ the linear array described above for optimum reception of a BPSK signal, which is expected to arrive from broadside in the presence of temporally and spatially white, equal power, circularly symmetric sensor noise; however, the desired source illuminates the array from 5° broadside.

Our first MVDR beamformer, $MVDR_1$, looks to broadside, i.e., a mismatch condition. Our second MVDR beamformer, $MVDR_2$, uses exact knowledge of DOA of the desired signal. We also employ the cumulant-based beamformer of Section 3.2, CUM_1, and the improved cumulant-based beamformer CUM_2 of Section 3.3.1. We investigate the performance of these processors for the following two elemental SNR levels: 20 dB for a strong signal and 0 dB for a weak signal. Note that the white-noise gain of any processor is limited to 10 dB by the number of sensors [15].

The beampattern responses (18), and white-noise gains of these beamformers are presented in Fig. 9 for SNR=20 dB. All responses are normalized to have a maximum value of 0 dB. For comparison purposes, the optimum beamformer response, calculated by using true statistics in (16), is presented as the dashed curves. Observe that due to the mismatch condition, $MVDR_1$ nulls the desired signal. More interestingly, the $MVDR_2$ processor that utilizes the true DOA information does not improve the SNR, due to the mismatch arising from the use of a sample-data covariance matrix. The cumulant-based processors, CUM_1 and CUM_2, yield excellent performance without any knowledge of source DOA. *It is very important to observe that the performance of cumulant-based processors are better than that of the MVDR with exactly known look-direction.*

We performed 100 Monte-Carlo runs to investigate the performance in a better way. The results are given in Table 1.

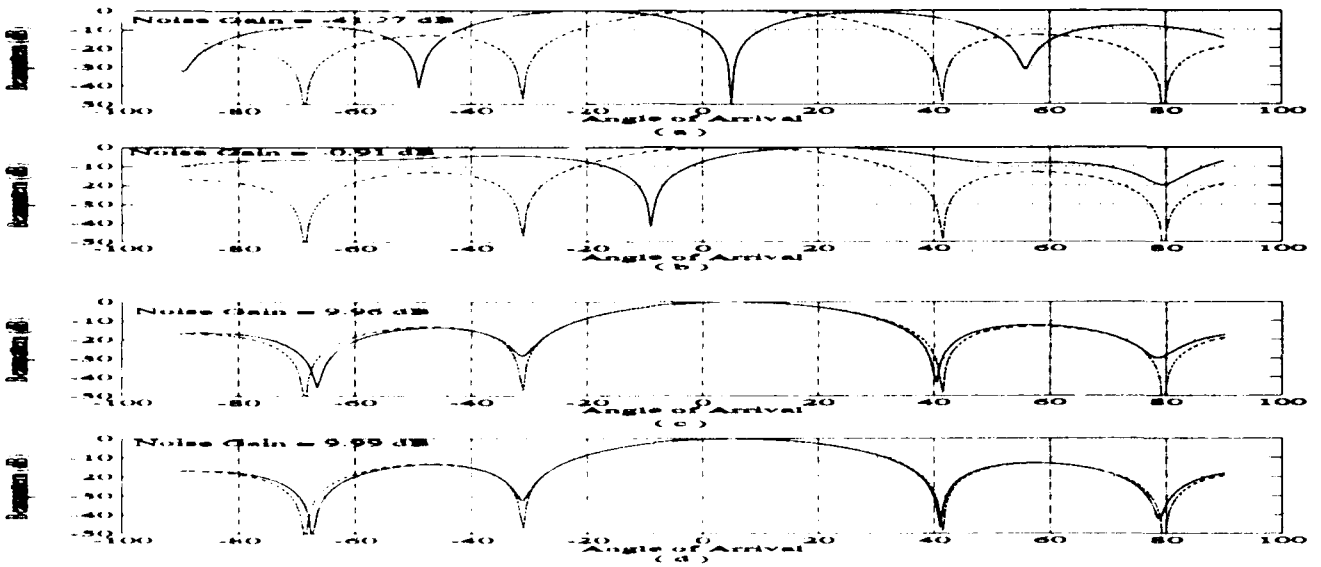


Figure 9: Beampatterns and white-noise gains of processors in a single realization for SNR = 20 dB : (a) $MVDR_1$, (b) $MVDR_2$, (c) CUM_1, (d) CUM_2. The optimum pattern is illustrated in dashed lines for comparison purposes.

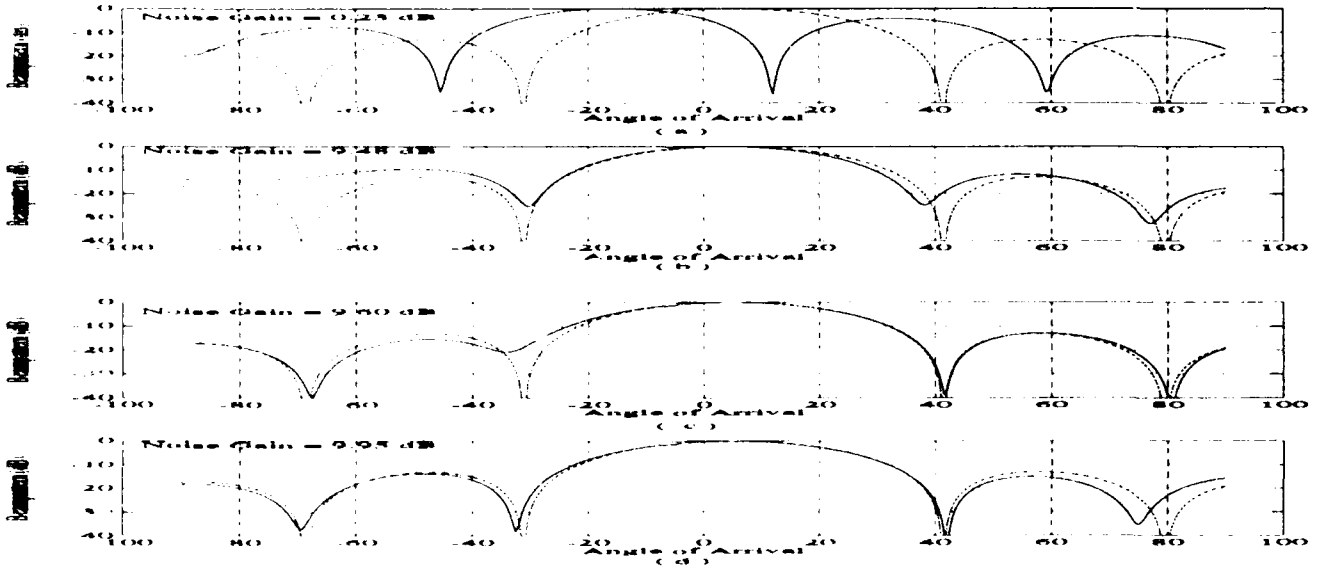


Figure 10: Beampatterns and white-noise gains of processors in a single realization for SNR = 0 dB : (a) $MVDR_1$, (b) $MVDR_2$, (c) CUM_1, (d) CUM_2. The optimum pattern is illustrated in dashed lines for comparison purposes.

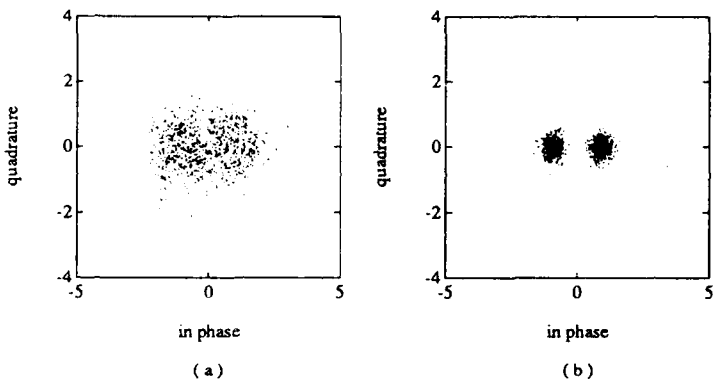


Figure 11: Power of cumulant-based beamforming: (a) received signal at the reference element at SNR = 0 dB, (b) output of CUM_2 processor.

From these results, it is clear that cumulant-based processors are superior and the extra computation involved in CUM_2 reduces the variations. Note, also, that variations in the MVDR processors are significantly larger than those of the cumulant-based counterparts. This agrees with the previous remarks about the sensitivity of MVDR processing to experimental conditions in a high-SNR environment.

Table 1: Results from 100 Monte-Carlo Runs for Experiment 1

Processor	White-Noise Gain (dB)			
	SNR=20dB		SNR=0dB	
	Mean	Std.	Mean	Std.
$MVDR_1$	-38.130	1.579	0.413	0.281
$MVDR_2$	0.179	1.360	9.583	0.131
CUM_1	9.954	0.015	9.058	0.359
CUM_2	9.990	0.003	9.959	0.014

We performed the same experiment for 0 dB SNR condition. Figure 10 illustrates the beampattern responses and white-noise gains of the processors. Monte-Carlo results are also given in Table 1. In this low-SNR condition, MVDR results are expected to improve since the mismatch conditions for the desired signal will be masked by the presence of white noise of comparable power, as explained in Section 3.3. $MVDR_1$ processor does not offer a significant gain due to the persistent mismatch condition, but $MVDR_2$ yields a near-optimum result, since presence of higher-level noise masks the mismatch due to the use of a sample-covariance matrix. The performance of CUM_1

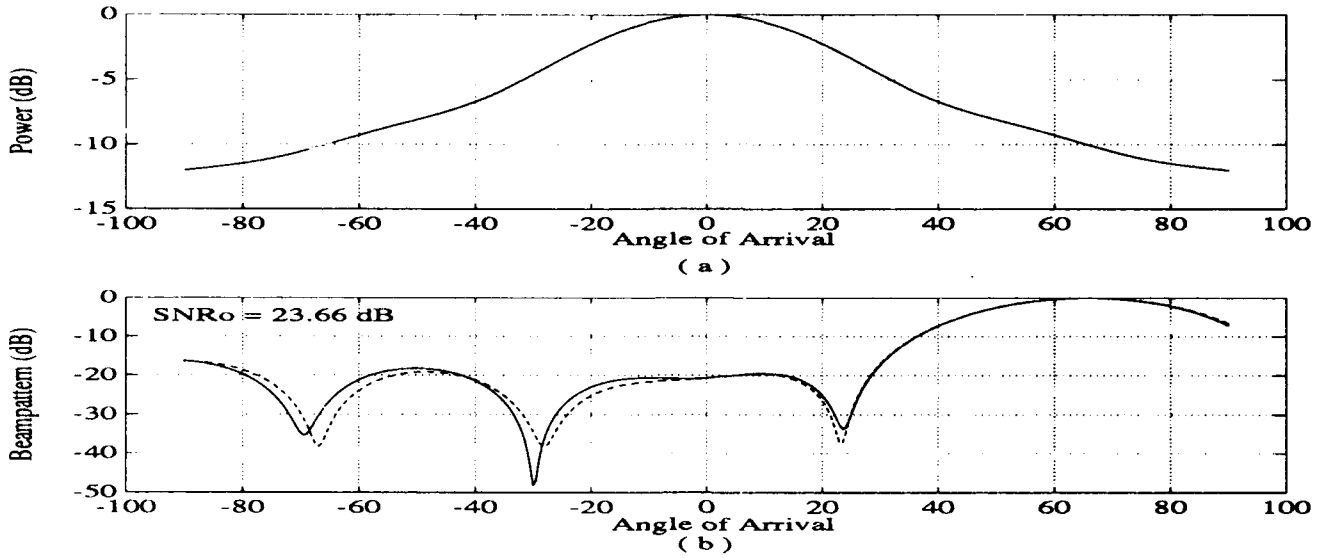


Figure 12: Beamforming in the presence of spatially colored noise: (a) Spatial Power Spectral Density of noise, (b) Beampattern of CUM_2 processor. The optimum pattern is illustrated in dashed lines for comparison purposes.

processor is slightly below than that of $MVDR_2$ and exhibits more variations. This is due to the inefficient use of the array data, since a high-level of noise corrupts the cumulant estimates and with CUM_1 there are no precautions to combat these errors. As expected, CUM_2 overcomes this problem by using SVD. Results in Table 1 indicate that CUM_2 achieves the best performance with minimum variations.

Finally, to demonstrate the power of cumulant-based beamforming, we illustrate the received signal and the output of CUM_2 processor for SNR=0 dB case in Fig. 11. It is clear that CUM_2 is capable of sufficient noise rejection for performing correct decisions.

3.6.2 Experiment 2: Spatially Colored Noise and Multipath Propagation

In this experiment, we investigate the performance of the proposed approach in the presence of spatially colored noise. We employ the linear array of the previous experiment. We assume that the noise field is created by a set of point sources distributed symmetrically about the broadside of the linear array. As suggested in [67], this source structure is typical when the noise field is spherically or cylindrically isotropic. In this case, the noise covariance matrix is symmetric-Toeplitz. In our experiment, we use the following structure for the covariance matrix of undesired components,

$$\mathbf{R}_u(i,j) = 0.8^{\,|i-j|} \tag{42}$$

The spatial power spectrum of undesired components is illustrated in Fig. 12a. It is clear that most of the noise leaks into the system from broadside. The desired signal illuminates the array from broadside, with an SNR of 10 dB. To illustrate the optimum combining property of our approach, we implanted an exact replica of the desired signal illuminating the array from 60^o, where

Table 2: Results from 100 Monte-Carlo Runs for Experiment 2

Processor	SNR_o (dB)	
	Mean	Std
CUM_1	23.641	0.017
CUM_2	23.645	0.015

noise power is relatively less when compared to that from broadside. The beampattern of CUM_2 processor is given in Fig.12b. For comparison purposes, we present the response of the optimum beamformer based on exact statistical information, as a dashed curve. The maximum-possible SNR at the output is 23.689 dB for this scenario. It is clear that the response of CUM_2 is almost identical to that of the optimum beamformer: both processors emphasize the signal illuminating the array from 60^o, since the noise contribution is less in this region. We performed 100 Monte-Carlo runs for this scenario, and the results are presented in Table 2. It is clear that both cumulant-based processors perform equally well. The reason for this phenomenon is the presence of the multipath from 60^o through a low-noise background that virtually increases the effective SNR, which, in turn, alleviates the effects of estimation errors. Note that the peak of the beampattern is slightly shifted from 60^o, in order to receive less interference. Similar behavior is observed in covariance-based direction-of-arrival estimation in the presence of colored noise resulting in biased estimates of parameters.

3.6.3 Experiment 3: Effects of Robustness Constraint

In this experiment, we illustrate the effects of the robustness constraint of Section 3.3.3, on a CUM_1 processor in the presence of white noise. We employ the same array as in the previous experiments. We employ CUM_1, since this processor uses the data inefficiently, and requires a robust approach. In our experiment, we consider the situation with SNR=0 dB. Figure 13 illustrates the beampatterns of CUM_1 processor for several SNR_v values. It is clear from the results that, as the perturbation increases, the patterns match better since the mismatch due to estimation errors in the steering vector estimate are masked by the presence of virtual increased level of noise. This method should be used sparingly in the presence of jammers, because virtually increasing the noise level results in diverting the capability of the array from nulling the directional interference.

3.6.4 Experiment 4: Multiple Interferers

In this experiment, we consider the problem of beamforming in a multipath environment in the presence of multiple jammers. We employ the same array as in the previous experiments. The signal of interest originates from a BPSK communication source, and it is expected from broadside; however, due to multipath effects, multiple delayed and shifted replicas are received. There are two jammers, and one is subject to multipath as well. Table 3, summarizes the signal structure.

Note that there are 10 wavefronts illuminating the array and it is not possible to estimate their DOA's with any existing high-resolution method; hence, signal-COPY algorithms [58] can not be used even with perfect knowledge of the array manifold.

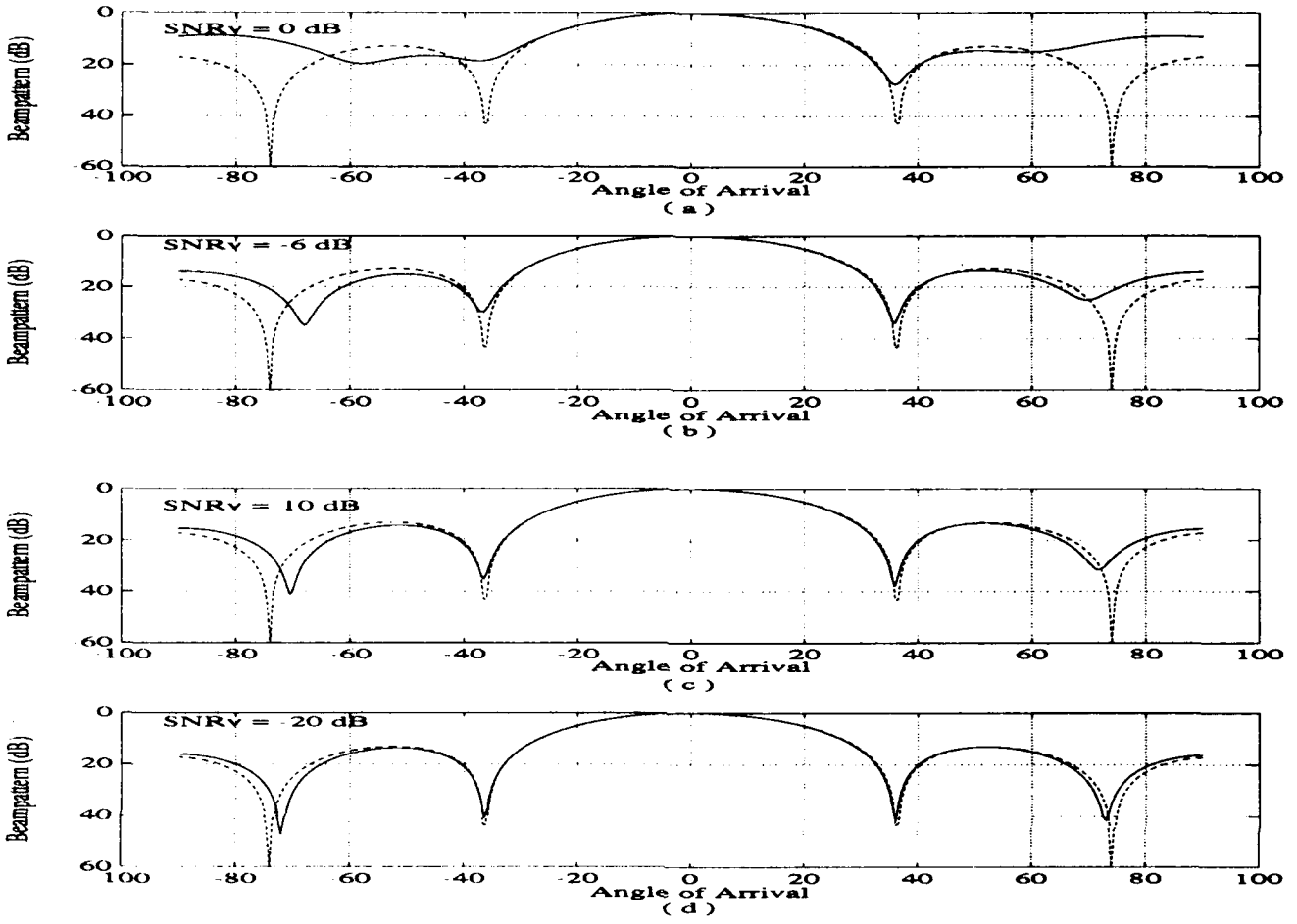


Figure 13: Beampattern of CUM_1 processor for varying virtual SNR: (a) 0 dB, (b) -6 dB, (c) -10 dB, (d) -20 dB. The optimum pattern is illustrated in dashed lines for comparison purposes.

Due to presence of coherent wavefronts, second-order statistics are not spatially stationary along the array; hence, it is not meaningful to define SINR at an array element. Instead, we compute the SINR at the output of the optimal processor by employing true statistics. The maximum possible $SINR_o$ is found from (34) to be 12.677 dB. From Table 4, we observe that CUM_2 performs very well under these severe conditions. Performance of CUM_1 is effected by strong interferers since this processor does not utilize all of the available information. Finally, we observe that MVDR with correct look-direction cancels the desired signal due to coherence. Note that CUM_2 exhibits less variations than other processors.

To gain more insight into the operation of the processors, we illustrate the beampatterns for MVDR and CUM_2 in Fig. 14. We focus on the region where the wavefronts are received by the array. It is observed that the MVDR processor does not null the jammer from -1^o, since it maintains the look-direction constraint for 0^o and tries to minimize the output power by destructively combining the coherent wavefronts. On the other hand, CUM_2 is blind to Gaussian interferers, and, as in

Table 3: Signal structure for Experiment 4

Source	Power (dB)	Multipath Coeff.	DOA
BPSK	10	(0.0,-0.5)	-10^o
		(0.9895,-0.0311)	-2^o
		(1.0,0.0)	0^o
		(-0.6472,-0.4702)	6^o
		(-0.8,0.0)	8^o
		(0.1414,0.1414)	11^o
		(0.0462,0.0191)	18^o
$JAMMER_1$	10	(1.0,0.0)	26^o
		(0.5657,0.5657)	32^o
$JAMMER_2$	10	(1.0,0.0)	-1^o
NOISE	0	—	—

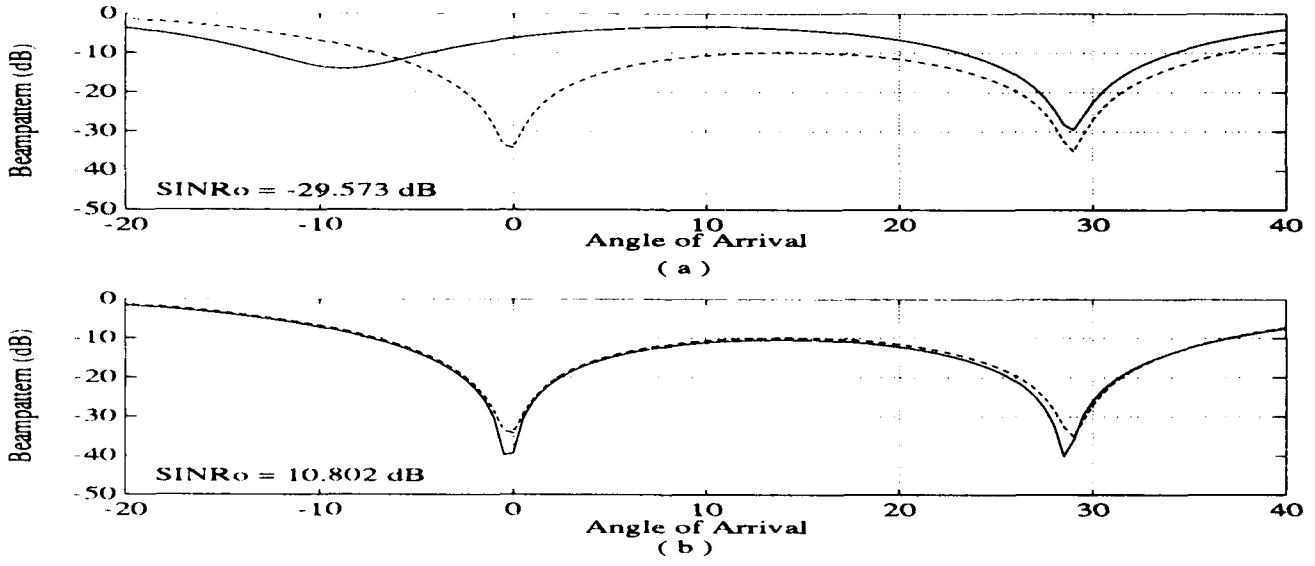


Figure 14: Beampatterns and array gains of processors: (a) MVDR with correct look direction. (b) CUM_2. The optimum pattern is illustrated in dashed lines for comparison purposes.

Table 4: Results from 100 Monte-Carlo Runs for Experiment 4

Processor	$SINR_o$ (dB)	
	Mean	Std
MVDR	-28.424	4.405
CUM_1	4.110	2.118
CUM_2	10.290	0.746
C^2	11.879	0.627

Experiment 2, it estimates the *generalized* steering vector of the desired signal and combines the wavefronts to enhance SINR at the output. CUM_2 puts a null on the jammer from -1^o, destructively combines the wavefronts from the first jammer by weight-phasing rather than null-steering, and reinforces the wavefronts from the desired source.

Finally, we implement the C^2 beamformer suggested in Section 3.3.2: we first estimate the steering vector as done for CUM_2, but then further project it into the subspace spanned by the principal eigenvectors of the sample covariance matrix. We use the resultant vector as the estimate of the desired signal steering vector, and construct an MVDR beamformer based on it. The performance of the resultant processor is demonstrated in Table 4.

We observe that by combining cumulants with covariance information, we obtain the best results.

3.6.5 Experiment 5: Adaptive Processing

In this section, we demonstrate the results from the adaptive version of CUM_1 approach as described in Section 3.5. We employ the 10 element uniform linear array of previous experiments. The initial pattern of the beamformer is designed to be isotropic, by letting $\mathbf{c}(0) = [1, 0, \ldots, 0]^T$. Desired signal illuminates the array from broadside with SNR=10 dB. A jammer with power equal to that of the desired source is present at 30^o. Note that there is no nonstationarity involved in this experiment; our aim is to demonstrate the evolution of the beamforming process and indicate the data lengths required for cumulant and covariance estimation. Tracking properties will be included in our future work, including comparisons with adaptive versions of CUM_2 and C^2 processors.

Figure 15 illustrates the beampattern of the adaptive CUM_1 processor as time evolves. After 100 snapshots, the beampattern is still close to isotropic. At 300 snapshots, covariance matrix estimate is improved, indicating the presence of desired signal from broadside. At this time point, the cumulant-based steering vector estimate has not matured, so it can not prevent the desired signal from being cancelled. After 500 snapshots, cumulant estimates get better, and there is a tendency to cancel the interference rather than the desired signal. Finally, after 700 snapshots the processor removes the interference by null steering.

3.6.6 Experiment 6: Effects of Data Length

In this section, we employ the linear array of Experiment 1, with the same noise conditions, and vary the data length to observe the behavior of the beamformers CUM_1, CUM_2, $MVDR_1$ and

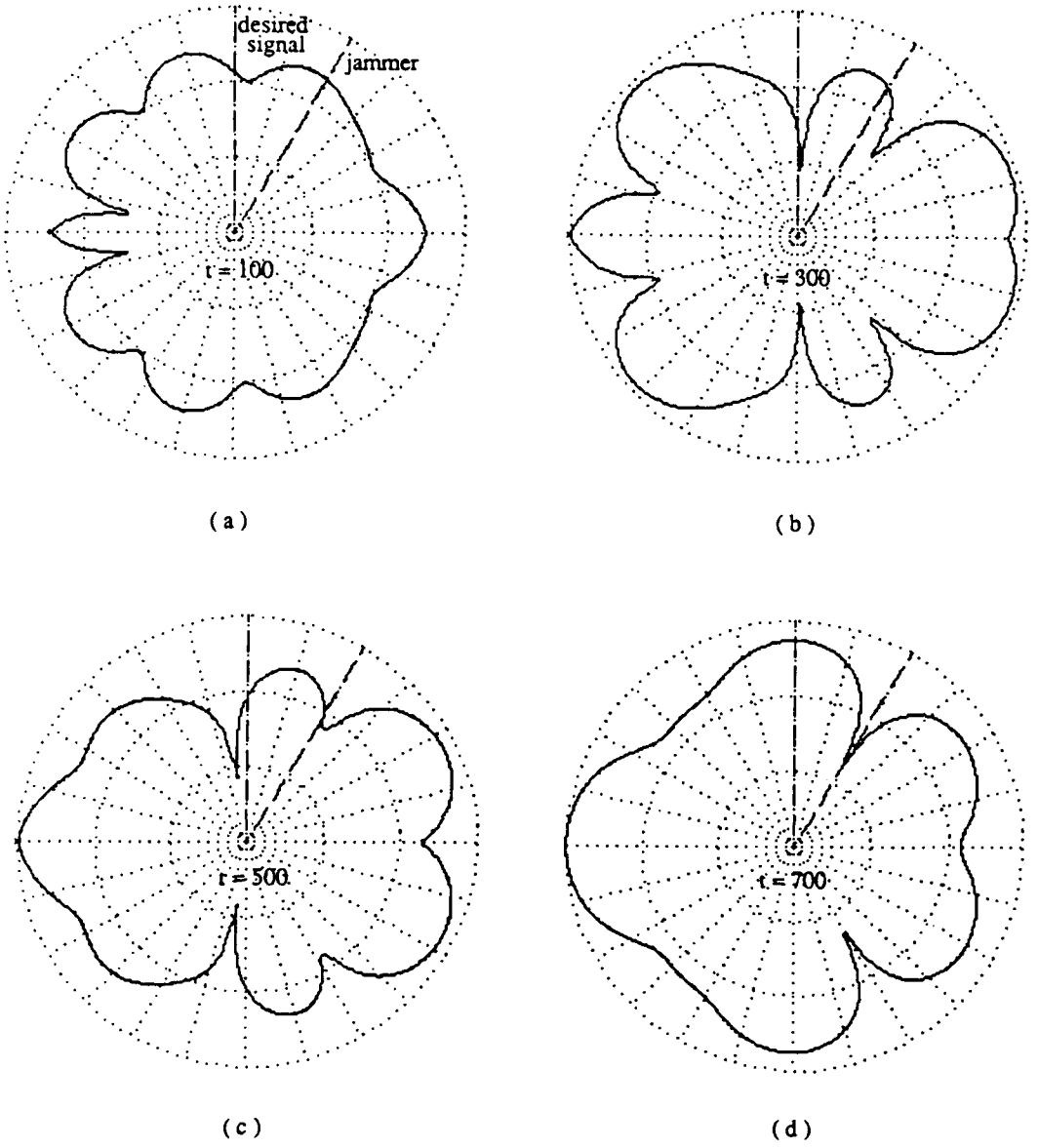


Figure 15: Beampattern of the adaptive CUM_1 processor as a function of time: desired signal is from broadside and the jammer is from 30° as indicated. (a) t=100, (b) t=300, (c) t=500, (d) t=700.

$MVDR_2$. Figure 16 demonstrates the variation of white-noise gain of the processors with data length, for 0dB and 20dB SNR levels. Each point on the plots is obtained by averaging the results from 50 Monte-Carlo simulations.

From Fig. 16a it is clear that CUM_2 outperforms all the processors, including $MVDR_2$ which utilizes the correct look direction for all data lengths. Furthermore, small sample properties of CUM_2 are quite impressive, motivating further research for developing its adaptive version. Low SNR masks the mismatch in $MVDR_2$ due to the use of sample covariance matrix; hence, as can be seen from Fig. 16a, CUM_1 is inferior to $MVDR_2$.

Figures 16b and 16c, indicate the effect of higher SNR on performance. CUM_1 and CUM_2 perform almost identical for all data lengths. Their gain is larger than 9 dB even for less than 50 snapshots. $MVDR_2$ can not recover in this experiment since the mismatch results in severe signal cancellation. We do not include the response of $MVDR_1$, because its performance drifts around -35 dB.

These results indicate that our approach has very promising small sample behavior that deserves more research. This will be a topic of another paper.

3.7 Conclusions

We have presented optimum beamforming algorithms for non-Gaussian signals, which are based on fourth-order cumulants of the data received by the array. Our proposed methods do not make any assumption about the sensor locations and characteristics, i.e., they are blind beamforming methods. Cumulant-based estimation is employed to identify the steering vector of the signal of interest and MVDR beamforming using this estimate is used to remove Gaussian interference components. We have suggested several approaches to combat effects of estimation errors. We have also implemented a recursive version of the method to enable real-time beamforming. Simulation experiments demonstrate the performance of our approaches in a wide variety of situations. It is important to emphasize that the proposed methods outperform an MVDR beamformer with an exactly known look-direction.

In our future work, we shall address the problem of optimum beamforming in the presence of multiple non-Gaussian interferers and design of adaptive algorithms with better convergence properties.

4 Final Comments

In this paper, we summarized our recent research results on the applications of cumulants in speech and array processing. The results are very promising, and encourage further study in these areas.

We acknowledge that especially in speech processing, cumulant applications are still in a very premature state. Array processing, however, captured more attention, particularly after the excellent work in [9]. On the other hand, array processing has many practical problems, such as unknown sensor gain/phase factors, array shape calibration, and DOA estimation for coherent sources in colored noise. It is our aim to develop cumulant-based solutions to those practical problems that still lack reasonable solutions when only second-order statistics are employed.

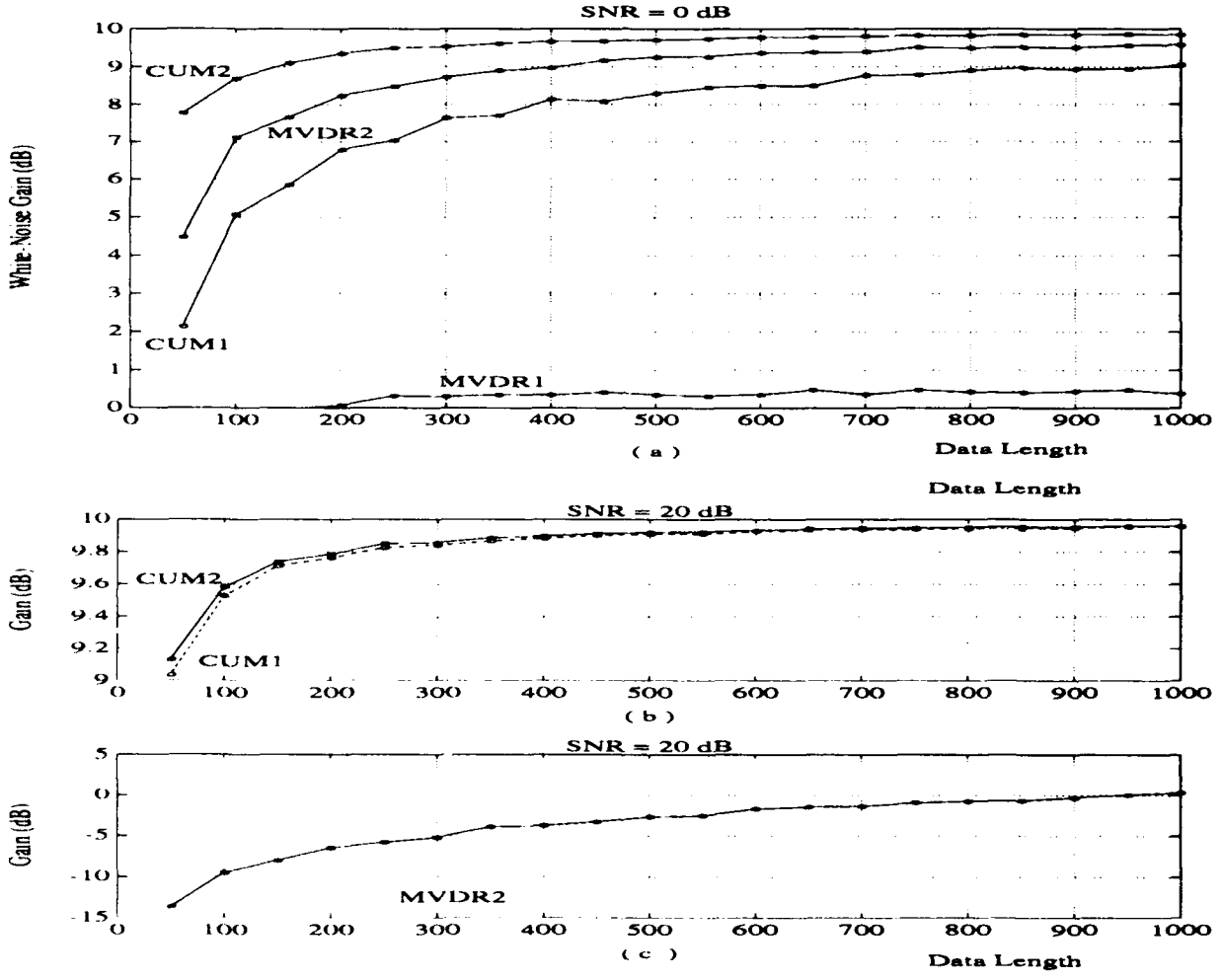


Figure 16: Performance of processors with varying data length: (a) SNR=0 dB, (b) SNR=20 dB, CUM_1 and CUM_2, (c) SNR=20 dB, $MVDR_2$.

References

[1] T.J. Abatzoglou, J.M. Mendel and G.A. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *IEEE Trans. Signal Processing*, vol.39, no.5, pp.1070-1087, May 1991.

[2] A. Bendiksen and K. Steiglitz, "Neural Nets for Voiced/Unvoiced Speech Classification," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.521-524, 1990.

[3] A. Benveniste, M. Metivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, 1990.

[4] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-31, no.5, pp.1235-1247, October 1983.

[5] D.G. Brennan, "On the maximum signal-to-noise ratio realizable from several noisy signals," *Proc. IRE*, vol.43, pg.1350, 1955.

[6] Y. Bresler, V.U. Reddy and T. Kailath, "Optimum beamforming for coherent signal and interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-36, no.6, pp.833-843, June 1988.

[7] D.R. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of kth-order spectra," in *Spectral Analysis of Time Series*, B. Harris, ed., New-York: John Wiley & Sons, pp.189-232, 1967.

[8] J. Capon, "High-resolution frequency-wavenumber spectral analysis," *Proc. of IEEE*, vol.57, no.8, pp.1408-1418, August 1969.

[9] J.F. Cardoso, "Blind identification of independent components with higher-order statistics," *Proc. Vail Workshop on Higher-Order Spectral Analysis*, pp.157-162, June 1989.

[10] P. Comon, "Seperation of stochastic processes," *Proc. Vail Workshop on Higher-Order Spectral Analysis*, pp.174-179, June 1989.

[11] R.T. Compton, "Pointing accuracy and dynamic range in a steered beam adaptive array," *IEEE Trans. Aerosp. Electron. Syst.*, vol.AES-16, pp.280-287, May 1980.

[12] R.T. Compton, "The effect of random steering error vectors in the Applebaum adaptive array," *IEEE Trans. Aerosp. Electron. Syst.*, vol.AES-18, pp.392-400, September 1982.

[13] R.T. Compton, *Adaptive Antennas: Concepts and Performance*, Prentice-Hall, New-Jersey, 1988.

[14] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. of Acoust. Soc. Amer.*, vol.54, no.3, pp.771-785, 1973.

[15] H. Cox, H.M. Zeskind and M.M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-35, no.10, pp.1365-1376, October 1987.

[16] H. Cox, H.M. Zeskind and M.M. Owen, "Effects of amplitude and phase errors on linear predictive array processors," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-36, no.1, January 1988.

[17] S. DeGraaf and D. Johnson, " Capability of array processing algorithms to estimate source bearings," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-33, no.6, pp.1368–1379, December 1985.

[18] Dubnowski, R.W. Schafer and L.R. Rabiner , " Real-time digital hardware pitch detector," *IEEE Trans. Acoust. , Speech , Signal Process.* , vol. ASSP-24, no.1, pp.2–8, February 1976.

[19] M.C. Doğan and J.M. Mendel, "Cumulant-based blind optimum beamforming," *USC-SIPI Report* # 195, Los Angeles, California, January 1992.

[20] M.C. Doğan and J.M. Mendel, "Cumulant-based adaptive analysis of speech signals," *USC-SIPI Report* # 196, Los Angeles, California, January 1992.

[21] D. Feldman and L.J. Griffiths, "A constraint projection approach for robust adaptive beamforming," *in Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, pp.1381–1384, May 1991.

[22] J.R. Fonollosa, J. Vidal and E. Masgrau , " Adaptive system identification based on higher order statistics," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.3437–3440, 1991.

[23] B. Friedlander, "A signal subspace method for adaptive interference cancellation," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-36, no.12, pp.1835–1845, December 1988.

[24] B. Friedlander and B. Porat, " Performance analysis of a null-steering algorithm based on direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-37, no.4, pp.461 466, April 1989.

[25] B. Friedlander, " A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-38, no.10, pp.1740–1751, October 1990.

[26] W. Gabriel, " Spectral analysis and adaptive array superresolution techniques," *Proc. IEEE*, vol.68, no.6, pp.654 666, June 1980.

[27] W. Gabriel, "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. Antennas and Propagation*, vol.AP-34, no.4, pp.291–300, March 1986.

[28] G.B. Giannakis and M. Tsatsanis, "HOS or SOS for parametric modelling," *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Processing*, vol.5, pp.3097 3100, May 1991.

[29] L.C. Godara, "Error analysis of optimal antenna array processors," *IEEE Trans. Aerosp. Electron. Syst.*, vol.AES-22, July 1986.

[30] S. Haykin, ed., *Array Processing: Applications to Radar*, Dowden, Hutchinson & Ross, Inc., 1980.

[31] S. Haykin, ed., *Array Signal Processing*, Prentice-Hall, New-Jersey, 1984.

[32] S. Haykin, ed., *Advances in Spectrum Estimation and Array Processing*. Prentice-Hall, New-Jersey, 1991.

[33] $Hi-Spec^{TM}$: Software package for signal processing with higher-order-spectra, United Signals and Systems, Culver City, California, 1991.

[34] P. Huber, *Robust Statistical Procedures* , CBMS - NSF Regional Conference Series in Applied Mathematics, 1977.

[35] A. Johanson, G. Ahlbom and L.H. Zetterberg, "Event detection using recursively updated lattice filters,"*in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.632–635, 1985.

[36] D. Johnson and S. DeGraaf, " Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-30, no.4, pp.638 647, August 1982.

[37] D.A. Krubsack and R.J. Niederjohn, " An autocorrelation pitch detector and voicing decision with confidence measures developed for noise corrupted speech," *IEEE Trans. Acoust., Speech, Signal Process.* , vol. ASSP-39, no.2, pp. 319-329, February 1991.

[38] R. Kumaresan and D. Tufts, " Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerospace and Electronic Systems*, vol.AES-19, no.1, pp.134–139, January 1983.

[39] C. Lee , " Robust linear prediction of speech," *IEEE Trans. Acoust. , Speech , Signal Processing* , vol. ASSP-36 , no.5, pp.642–650, May 1988.

[40] K.Y. Lee, B.G. Lee, I. Song and S. Ann, " On Bernoulli-Gaussian modelling of speech excitation source," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.217–220, 1990.

[41] K.S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Statist.*, vol.10, pp.1195-1208, 1982.

[42] J. Marr, " A selected bibliography on adaptive antenna arrays," *IEEE Trans. Aerosp. Electron. Syst.*, AES-22, no.6, pp.781- 798, November 1986.

[43] J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *in Proc. IEEE*, vol.79, no.3, pp.278–305, March 1991.

[44] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*. John-Wiley & Sons, Inc., 1980.

[45] C.L. Nikias and M.R. Raghuveer, "Bispectrum estimation: a digital signal processing framework," *Proc. IEEE*, vol.75, no.7, pp.869–891, July 1987.

[46] J.M. Ortega, *Matrix Theory: a second course*. Plenum Press, New-York, 1987.

[47] A. Paulraj and T. Kailath, "Eigenstructure methods for direction of arrival estimation in the presence of unknown noise fields," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-34, no.1, pp.13 20, February 1986.

[48] S. Pei, C.C. Yeh and S.C. Chiu, "Modified spatial smoothing for coherent jammer suppression without signal cancellation," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-36, no.3, pp.412-414, March 1988.

[49] U. Pillai and B. Kwon, " Forward/backward spatial smoothing schemes for coherent signal identification," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-37, no.1, pp.8-15, January 1989.

[50] S.U. Pillai, *Array Signal Processing*. Springer-Verlag, New-York, 1989.

[51] B. Porat and B Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-39, no.9, pp.2016-2024, September 1991.

[52] S. Prasad, R.T. Williams, A.K. Mahalanabis and L.H. Sibul, "A transform-based covariance differencing approach for some classes of parameter estimation problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-36, no.5, pp.631-641, May 1988.

[53] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[54] V. Reddy, A. Paulraj and T. Kailath, "Performance analysis of the optimum beamformer in the presence of correlated sources and its behavior under spatial smoothing," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-35, no.7, pp.927-936, July 1987.

[55] Y. Rockah and P.M. Schultheiss, "Array shape calibration using sources in unknown locations-part I: far-field sources," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-35, no.3, pp.286-299, March 1987.

[56] R. Roy and T. Kailath, "ESPRIT Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-37, no.7, pp.984-995, July 1989.

[57] P. Ruiz and J.L. Lacoume, "Extraction of independent sources from correlated sources: a solution based on cumulants," *Proc. Vail Workshop on Higher-Order Spectral Analysis*, pp.146-151, June 1989.

[58] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol.AP-34, no.3, pp.276-280, March 1986.

[59] R.O. Schmidt and R.E. Franks, "Multiple source DF signal processing: an experimental system," *IEEE Trans. Antennas and Propagation*, vol.AP-34, no.3, pp.281-290, March 1986.

[60] R.A. Scholtz, "How do you define bandwidth," *Proceedings of the International Telemetering Conference*, Los Angeles, California, pp.281-288, October 1972.

[61] T. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-33, no.3, pp.527-536, June 1985.

[62] T. Shan, M. Wax and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing* , vol.ASSP-33, no.4, pp.806-811, August 1985.

[63] D. Slepian, "On bandwidth," *Proc. of IEEE*, vol.64, pp. 292-300, 1976.

[64] Special Issue on Adaptive Antenna Systems, *IEEE Antennas Propagat.*, vol.AP-34, March 1986.

[65] A. Swami and J.M. Mendel, " Adaptive system identification using cumulants," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.2248-2251, 1988.

[66] Y. Su, T. Shan and B. Widrow, " Parallel spatial processing: a cure for signal cancellation in adaptive arrays," *IEEE Trans. Antennas and Propagation*, vol.AP-34, no.3, pp.347-355, March 1986.

[67] R.J. Talham, "Noise correlation functions for unisotropic noise fields," *J. Acoust. Soc. Amer.*, vol.69, pp.213-215, January 1981.

[68] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, pp.4-24, April 1988.

[69] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-39, no.5, pp.1110-1121, May 1991.

[70] A.M. Vural, "Effects of perturbations on the performance of optimum/adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol.AES-15, pp.76-87, January 1979.

[71] A.J. Weiss, A.S. Willsky and B.C. Levy, "Eigenstructure approach for array processing with unknown intensity coefficients," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-36, no.10, pp.1613-1617, October 1988.

[72] A.J. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations-a maximum-likelihood approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-37, no.12, pp.1958-1966, December 1989.

[73] B. Wells , " Voiced/Unvoiced decision based on the bispectrum," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1589-1592, 1985.

[74] B. Widrow, K.M. Duvall, R.P. Gooch and W.C. Newman, "Signal cancellation phenomena in adaptive antennas: causes and cures," *IEEE Trans. Antennas and Propagation*, vol.AP-30, no.3, pp.469-478, May 1982.

[75] R. Williams, S. Prasad, A.K. Mahalanabis and L.H. Sibul, "An improved spatial smoothing technique for bearing estimation in a multipath environment," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-36, no.4, pp.425-432, April 1988.

[76] C.L. Zahm, "Effects of errors in the direction of incidence on the performance of an adaptive array," *Proc. IEEE*, vol.60, pp.1008-1009, August 1972.

[77] C.L. Zahm, "Application of adaptive arrays to suppress strong jammers in the presence of weak signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-9, pp.260-271, 1973.

[78] M. Zoltowski, "On the performance analysis of the MVDR beamformer in the presence of correlated interference," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-36, no.6, pp.945-947, June 1988.

Moments and Wavelets in Signal Estimation

Edward J. Wegman[1]
Center for Computational Statistics
George Mason University

Hung T. Le[2]
International Business Machines

[1]This research was supported by the Office of Naval Research under Grant N00014-92-J-1303, the Army Research Office under Contract DAAL03-91-G-0039, and the National Science Foundation under Grant DMS9002237. This paper was presented as an invited talk at the ONR-sponsored workshop on Moments and Signal Processing held in Monterey, CA on March 30 and 31, 1992. Wegman used to be an important Navy employee as Division Director of the Mathematical Sciences Division at ONR, but now he is only a quasi-important as Theory and Methods Editor of *JASA*. After December 31, 1993, he will resume being unimportant.

[2]Dr. Le is an Adjunct Assistant Professor of Operations Research and Applied Statistics at George Mason University. This work is performed in part while he was on educational leave of absence from IBM.

Moments and Wavelets in Signal Estimation

Abstract: The problem of generalized nonparametric function estimation has received considerable attention over the last two decades. Most of the approaches have assumed smoothness of the function to be estimated generally in the form of continuity of higher order derivatives and/or bounded variation and have used convolution kernels or splines as the estimation devices. Generally focus has been on density estimation or nonparametric regression. The spline and kernel-based methods may be inappropriate if either smoothness assumptions are violated or if additional side conditions are present. Wegman (1984) introduced a general framework for optimal nonparametric function estimation which applies to a much wider class of problems than simply density estimation or nonparametric regression. In this framework, a class of admissible estimators is regarded as a compact, convex subset of a Banach function space and a convex objective functional is to be optimized over this set. Recent work on wavelets suggests a powerful method for constructing orthonormal bases to span the set of admissible estimators. Moreover, older work on frames has re-emerged to some level of prominence because of the work on wavelets. The optimal estimates can be computed as weighted linear combinations of the orthonormal bases. The weight coefficients are computed as moments of the basis functions. We illustrate these methods with some numerical examples.

Moments and Wavelets in Signal Estimation

1. Introduction.

The method of moments is a time-honored traditional technique in statistical inference while wavelet analysis has recently burst upon the mathematical scene to capture the enthusiasm and imagination of many applied mathematicians and engineers both because of their important applications in signal and image processing and other engineering applications and also because of the inherent elegance of the techniques. In this paper we bring these tools together to illustrate their application to transient signal processing. Wavelets are described in detail in a number of locations. Much of the fundamental work was done by Daubechies and is reported in Daubechies, Grossmann and Meyer (1986) and Daubechies (1988). Heil and Walnut (1989) provide a survey from a mathematical perspective while Rioul and Vetterli (1991) provide a survey from a more engineering perspective. The new book by Chui (1992) is an excellent integrated treatment which I believe is more mathematically sophisticated than the author supposes. In spite of its title as an introduction, it requires somewhat more mathematical depth and maturity and is best regarded as more of a monograph.

This present paper describes the basic wavelet theory in the context of the general statistical problem of nonparametric function estimation. It will be show that traditional moment based techniques have an interesting and useful connection to modern nonparametric functional inference for signal processing via wavelets. Wegman (1984) describes a basic framework for optimal nonparametric function estimation. This framework captures the optimal estimation of a wide variety of practical function estimation problems in a common theoretical construct. Wegman (1984), however, only discusses the existence of such optimal estimators. In the present paper, we are interested in combining this optimality framework with more general wavelet algorithms as computational devices for general optimal nonparametric function estimation. A new application of optimal nonparametric function estimation is found in Le and Wegman (1991). A second application will be discussed in this paper.

In section 2, we discuss the optimal nonparametric function estimation framework. In section 3, we turn to a discussion of the general function analytic framework which leads to bases and frames. Section 4 introduces the notion of a wavelet basis and demonstrates the connection with Fourier series and Parseval's

Theorem. In section 5 we turn to transient signal estimation, develop an optimization criterion and illustrate the computation of a transient signal estimator.

2. Optimal Nonparametric Function Estimation.

Consider a general function, f(x), to be estimated based on some sampled data, say $x_1, x_2,\ldots,x_n$. This is, in fact, the most elementary estimation problem in statistical inference. Often the function, f, in question is the probability distribution function or the probability density function and most frequently the approach taken is to place the function within a parametric family indexed by some parameter, say θ. Rather than estimate f directly, the parameter θ is estimated with f_θ then being estimated by $\hat{f}_\theta = f_{\hat{\theta}}$. Under a variety of circumstances, it is much more desirable to take a nonparametric approach so as to avoid problems associated with misspecification of parametric family. This is particularly the case when data is relatively plentiful and the information captured by the parametric model is not needed for statistical efficiency.

Probability density estimation and nonparametric, nonlinear regression are probably the two most widely studied nonparametric function estimation problems. However, other problems of interest which immediately come to mind are spectral density estimation, transfer function estimation, impulse response function estimation, all in the time series setting, and failure rate function estimation and survival function estimation in the reliability/biometry setting. While it may be the case that we simply may want an unconstrained estimate of the function, it is more often the case that we wish to impose one or more constraints, for example, positivity, smoothness, isotonicity, convexity, transience and fixed discontinuities to name a few appropriate constraints. By far, the most common assumption is smoothness and frequently the estimation is via a kernel or convolution smoother. We would like to formulate an optimal nonparametric framework.

We formulate the optimization problem as follows. Let $\mathcal{H}$ be a Hilbert space of functions over $\mathbf{R}$, the real numbers (or $\mathbf{C}$, the complex numbers). For purposes of the present paper, we assume $\mathbf{R}$ rather than $\mathbf{C}$ unless otherwise specified. The techniques we outline here are not limited to a discussion of $L_2(\mathbf{R})$ although quite often we do take $\mathcal{H}$ to be L_2. In this case, we take

$$< f, g > = \int f(x)\, g(x)\, d\mu(x),$$

where μ is Lebesgue measure. We emphasize that this is not absolutely required. As

usual $\|f\| = \sqrt{<f, f>}$. A functional $\mathcal{L}:\mathcal{H}\to\mathbf{R}$ is ***linear*** if

$$\mathcal{L}(\alpha f + \beta g) = \alpha\mathcal{L}(f) + \beta\mathcal{L}(g), \text{ for every } f, g \in \mathcal{H} \text{ and } \alpha, \beta \in \mathbf{R}.$$

$\mathcal{L}$ is ***convex*** on $S \subseteq \mathcal{H}$ if

$$\mathcal{L}(tf + (1-t)g) \le t\mathcal{L}(f) + (1-t)\mathcal{L}(g), \text{ for every } f, g \in S \text{ with } 0 \le t \le 1.$$

$\mathcal{L}$ is ***concave*** if the inequality is reversed. $\mathcal{L}$ is ***strictly convex (concave)*** on S if the inequality is strict. $\mathcal{L}$ is ***uniformly convex*** on S if

$$t\mathcal{L}(f) + (1-t)\mathcal{L}(g) - \mathcal{L}(tf + (1-t)g) \ge ct(1-t)\|f-g\|^2$$

for every $f, g \in S$ and $0 \le t \le 1$.

We wish to use $\mathcal{L}$ as the general objective functional in our optimization framework. For example, if we are concerned with likelihood, we may consider the log likelihood,

$$\mathcal{L}(f) = \sum_{i=1}^{n} \log f(x_i), \; x_i \text{ are a random sample from f.}$$

If we have censored samples we may wish to consider

$$\mathcal{L}(g) = \sum_{i=1}^{n} \delta_i \log g(x_i) + \sum_{i=1}^{n} (1-\delta_i) \log \bar{G}(x_i),$$

x_i again a random sample, δ_i a censoring random variable, $\bar{G} = 1 - G$, and $G(x) = \int_{-\infty}^{x} g(u)\,du$. This is the censored log likelihood. Another example is the penalized least squares. In this case

$$\mathcal{L}(g) = \sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int_a^b (Lg(u))^2\,du.$$

Here L is a differential operator and the solution of this optimization problem over appropriate spaces is called a penalized smoothing L-spline. If $L = D^2$, then the solution is the familiar cubic spline.

The basic idea is to construct $S \subseteq \mathcal{H}$ where S is the collection of functions, g,

which satisfy our desired constraints such as smoothness or isotonicity. We wish to optimize $\mathcal{L}(g)$ over S. The optimized estimator will be an element of S and hence will inherit whatever properties we choose for S. The estimator will optimize $\mathcal{L}(g)$ and hence will be chosen according to whatever optimization criterion appeals to the investigator. In this sense we can construct designer estimators, i.e. estimators that are designed by the investigator to suit the specifics of the problem at hand.

Of course, in a wide variety of rather disparate contexts, many of these estimators are already known. However, they may be proven to exist in a general framework according to the following theorem.

Theorem 2.1:

Consider the following optimization problem:

$$\text{Minimize (maximize) } \mathcal{L}(f) \text{ subject to } f \in S \subseteq \mathcal{H}.$$

Then

a. If $\mathcal{H}$ is finite dimensional, $\mathcal{L}$ is continuous and convex (concave) and S is closed and bounded, then there exists at least one solution.

b. If $\mathcal{H}$ is infinite dimensional, $\mathcal{L}$ is continuous and convex (concave) and S is closed, bounded and convex, then there exists at least one solution.

c. If $\mathcal{L}$ in a. or b. is strictly convex (concave), the solution is unique.

d. If $\mathcal{H}$ is infinite dimensional, $\mathcal{L}$ is continuous and uniformly convex (concave) and S is closed and convex, then there exists a unique solution.

Proof: A full proof is given in Wegman (1984). For completeness, we outline the basic elements here. a. For the finite dimensional case, S closed and bounded implies that S is compact. Choose $f_n \in S$ such that $\mathcal{L}(f_n)$ converges to $\inf\{\mathcal{L}(f): f \in S\}$. Because of compactness, there is a convergent subsequence f_{n_k} having a limit, say f_*. By continuity of $\mathcal{L}$

$$\mathcal{L}(f_*) = \lim_{k\to\infty} \mathcal{L}(f_{n_k}) = \inf\{\mathcal{L}(f): f \in S\}.$$

f_* is the required optimizer. For part b., we have the same basic idea except that S closed, bounded and convex implies that S is weakly compact. We use the weak continuity of $\mathcal{L}$. Uniqueness follows by supposing both f_* and f_{**} are both minimizers. Then

$$\mathcal{L}(tf_* + (1-t)f_{**}) < t\mathcal{L}(f_*) + (1-t)\mathcal{L}(f_{**}) = \inf\{\mathcal{L}(f): f \in S\}.$$

This implies that neither f_* nor f_{**} is a minimizer which is a contradiction. □

This theorem gives us unified framework for the construction of optimal nonparametric function estimators. It does not, however, give us a definitive method for construction of nonparametric function estimators. We give a constructive framework in the next several sections. In closing this section we refer the reader to Wegman (1984) for the complete proof of Theorem 2.1 and many more examples of the use of this result.

3. Bases and Subspaces.

In this section, we discuss the basic theory of spanning bases and their application to function estimation. Consider $f, g \in \mathcal{H}$. f is said to be ***orthogonal*** to g written $f \perp g$ if $< f, g > = 0$. An element f is ***normal*** if $\| f \| = 1$. A family of elements, say $\{e_\lambda : \lambda \in \Lambda\}$ is ***orthonormal*** if each element is normal and if for any pair e_1, e_2 in the family, $e_1 \perp e_2$. A family $\{e_\lambda : \lambda \in \Lambda\}$ is ***complete*** in $S \subseteq \mathcal{H}$ if the only element in S which is orthogonal to every e_λ, $\lambda \in \Lambda$ is 0. A ***basis*** or ***base*** of S is a complete orthonormal family in S. A Hilbert space has a countable basis if and only if it is separable, i.e. if and only if it has a countable dense subset. Ordinary L_p spaces are separable. We are now in a position to state the basic result characterizing bases of Hilbert spaces or subspaces. We write ***span***$(\{e_\lambda\})$ to be the minimal subspace containing $\{e_\lambda\}$. This is the space generated by the elements $\{e_\lambda\}$.

Theorem 3.1:

Let $\mathcal{H}$ be a separable Hilbert space. If $\{e_k\}_{k=1}^{\infty}$ is an orthonormal family in $\mathcal{H}$, then the following are equivalent.

a. $\{e_k\}_{k=1}^{\infty}$ is a basis for $\mathcal{H}$.

b. If $f \in \mathcal{H}$ and $f \perp e_k$ for every k, then $f = 0$.

c. If $f \in \mathcal{H}$, then $f = \sum_{k=1}^{\infty} < f, e_k > e_k$. (orthogonal series expansion)

d. If $f, g \in \mathcal{H}$, then $< f, g > = \sum_{k=1}^{\infty} < f, e_k > < g, e_k >$.

e. If $f \in \mathcal{H}$, $\| f \|^2 = \sum_{k=1}^{\infty} | < f, e_k > |^2$. (Parseval's Theorem)

Proof:

a => b: Trivial by definition.

b => c: We claim $\mathcal{H} = \text{span}(\{e_k\})$. If not there is $f \neq 0$, $f \in \mathcal{H}$ such that $f \notin \text{span}(\{e_k\})$. This implies that $f \perp e_k$ for every k. But $f \perp e_k$ for every k and $f \neq 0$ is a contradiction to the $\{e_k\}$ being a basis. Let $\mathcal{H}_k = \text{span}(e_k)$. Then $\mathcal{H} = \text{span}(\bigcup_{k=1}^{\infty} \mathcal{H}_k) =$

$\sum_k \mathcal{H}_k$. This implies that for $f \in \mathcal{H}$,

$$f = \sum_{k=1}^{\infty} c_k e_k. \tag{3.1}$$

Substituting (3.1) in the expression for the inner product yields

$$< f, e_j > = < \sum_k c_k e_k, e_j > = \sum_{k=1}^{\infty} c_k < e_k, e_j > .$$

By the orthonormal property, $< e_k, e_j > = 1$, if $k = j$ and $= 0$, otherwise. It follows that $< f, e_j > = c_j$. Thus

$$f = \sum_{k=1}^{\infty} < f, e_k > e_k. \tag{3.2}$$

c => d: $< f, g > = < f, \sum_{k=1}^{\infty} < g, e_k > e_k > = \sum_{k=1}^{\infty} < g, e_k > < f, e_k > .$

d => e: Let $f = g$ in part d.

e => a: If $f \in \mathcal{H}$ and $f \perp e_k$ for every k implies $< f, e_k > = 0$ for every k. This in turn implies that $\| f \| = 0$. Thus $f = 0$. This finally implies $\{e_k\}_k$ is a basis. □

Thus given any basis $\{e_k\}_k$, we can exactly write $f = \sum_{k=1}^{\infty} c_k e_k$ and we can estimate f by $\sum_{k=1}^{N} \hat{c}_k e_k$. Thus a computational algorithm for the optimal nonparametric function estimator can be based on this result from Theorem 3.1.c. However, this does not yet take into account the "design" set, S. In order to more carefully study the structure of S we consider the following result. In the following discussion let $S \subseteq \mathcal{H}$. Then define $S^{\perp} = \{f \in \mathcal{H}: f \perp S\}$.

Theorem 3.2:

If $S \subseteq \mathcal{H}$ is a subset of $\mathcal{H}$, then

a. $S^{\perp}$ is a subspace of $\mathcal{H}$ and $S \cap S^{\perp} \subseteq \{0\}$

b. $S \subseteq S^{\perp\perp} = \text{span}(S)$

c. S is a subspace if and only if $S = S^{\perp\perp}$.

Proof: $S^{\perp}$ is a linear manifold. To see this if $f_1, f_2 \in S^{\perp}$, then for every $g \in S$, $< a_1 f_1 + a_2 f_2, g > = a_1 < f_1, g > + a_2 < f_2, g > = a_1 \cdot 0 + a_2 \cdot 0 = 0$. Thus $a_1 f_1 + a_2 f_2 \in S^{\perp}$. This implies $S^{\perp}$ is a linear manifold which is sufficient to show that $S^{\perp}$ is a subspace

provided we can show $S^{\perp}$ is closed. To see this if $f \in$ closure $(S^{\perp})$, then there exists $\{f_n\} \subseteq S^{\perp}$ such that $f = \lim_n f_n$ and for every $g \in S$, $< f_n, g > = 0$. But $< f, g > = \lim_{n\to\infty} < f_n, g > = \lim_{n\to\infty} 0 = 0$. This implies $f \perp S$ which in turn implies $f \in S^{\perp}$. Part b follows from part a by replacing S by $S^{\perp}$. Part c is straightforward application of the two previous parts. □

Suppose now that we have a basis for $\mathcal{H}$, call it $\{e_k\}_{k=1}^{\infty}$. This basis obviously also spans subset S of $\mathcal{H}$ and hence any of our "designer" functions in S can be written in terms of the basis, $\{e_k\}_{k=1}^{\infty}$. The unnecessary basis elements will simply have coefficients of 0. In a sense, however, this basis is too rich and in a noisy estimation setting superfluous basis elements will only contribute to estimating noise. As part of our "designer" set, S, philosophy, we would like to have a minimal basis set for S. Theorem 3.2 gives us a test for this condition. Consider a basis $\{e_k\}_{k=1}^{\infty}$ for $\mathcal{H}$. Form B_S which is to be a basis for S. We define B_S by the following routine. If there is a $g \in S$ such that $< g, e_k > \neq 0$, then let $e_k \in B_S$. If on the other hand there is a $g \in S^{\perp}$ such that $< g, e_k > \neq 0$, then let $e_k \in B_{S^{\perp}}$. Unfortunately, it may not be that $B_S \cap B_{S^{\perp}} = \emptyset$. But this algorithm yields $\{e_k\} = B_S \cup B_{S^{\perp}}$. Moreover $S \subseteq \text{span}(B_S)$. Thus we may be able to eliminate unnecessary basis elements. We may also be able to re-normalize the basis elements using a Gram-Schmidt orthogonalization procedure to make $B_S \perp B_{S^{\perp}}$. Usually if we know the properties of the set, S, we desire and the nature of the basis set $\{e_k\}$, it will be straightforward to construct a test function, g, with which to construct the basis set, B_S. If S is a subspace, then $S = \text{span}(B_S)$. In any case we can carry out our estimation by

$$\hat{f} = \sum_{e_k \in B_s} \hat{c}_k e_k. \tag{3.3}$$

In a completely noiseless setting (3.1) is really an equality in norm, i.e. $\| f - \sum_k c_k e_k \| = 0$. If $\mathcal{H}$ is $L_2(\mu)$, with μ Lebesgue measure, then (3.1) is really

$$f = \sum_k c_k e_k, \text{ almost everywhere } \mu \text{ with } c_k = < f, e_k >. \tag{3.4}$$

This choice of c_k is a minimum norm choice. However, in a noisy setting, i.e. where we do not know f exactly, we cannot compute c_k directly. However, we may be able to estimate c_k by standard inference techniques.

Example 3.1. Norm Estimate. The minimum norm estimate of c_k is the choice which minimizes $\| f - \sum_k c_k e_k \|$, i.e. $c_k = < f, e_k >$. In the L_2 context,

$$< f, e_k > = \int_R f(x)\, e_k(x)\, d\mu(x).$$

If f is a probability density function, then $< f, e_k > = E[e_k]$ which can simply be estimated by $n^{-1} \sum_{j=1}^{n} e_k(x_j)$, where x_j, $j = 1,\ldots,n$ is the sample of observations. We note that the major approach to estimating the weighting coefficients is via a traditional method of moments.

Example 3.2. General Form of Estimate. In the general context with optimization functional $\mathcal{L}$ we have

$$\mathcal{L}(f) = \mathcal{L}\left(\sum_{e_k \in B_e} c_k\, e_k \right) \stackrel{d}{=} \mathcal{L}(\{c_k\}). \tag{3.5}$$

Since (3.5) is a function of a countable number of variables, $\{c_k\}$, we can find the normal equations and with the appropriate choice of basis, find a solution. For this we will typically assume $\mathcal{L}$ is twice differentiable with respect to all c_k. A wide variety of bases have been studied. These include Laguerre polynomials, Hermite polynomials and other orthonormal systems. Perhaps the most well-known orthonormal system is the system of fundamental sinusoids which span $L_2(0, 2\pi)$. One might reasonable guess that wavelets form another orthogonal system. We discuss the connection in the next section.

4. Fourier Analysis and Wavelets.

4.1 Bases for $L_2(0, 2\pi)$.

Let us consider the set of square-integrable functions on $(0, 2\pi)$ which we denote by $L_2(0, 2\pi)$. $L_2(0, 2\pi)$ is a Hilbert space and a traditional choice of an orthonormal basis for this space has been $e_k(x) = e^{ikx}$, the complex sinusoids. Thus any f in $L_2(0,2\pi)$ has the Fourier representation by Theorem 3.1.c

$$f(x) = \sum_{k=-\infty}^{\infty} c_k\, e^{ikx}$$

where the constants c_k are the Fourier coefficients defined by

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-ikx}dx.$$

This pair of equations represent the discrete Fourier transform and the inverse Fourier transform and is the foundation of harmonic analysis. An interesting feature of this complex sinusoids as a base for $L_2(0, 2\pi)$ is that $e_k(x) = e^{ikx}$ can be generated from the superpositions of dilations of a single function, $e(x) = e^{ix}$. By this we mean that

$$e_k(x) = e(kx),\ k = \cdots,\ -1,\ 0,\ 1,\ \cdots$$

These are ***integral dilations*** in the sense that $k \in J$, the integers. The concept of dilations of a fixed generating function is central to the formation of wavelet bases as we shall see shortly.

A well known consequence of Theorem 3.1.e for the complex sinusoid basis is the Parseval Theorem. For this base, we have

Theorem 4.1: (Parseval's Theorem):

$$\|f\|^2 = \int_0^{2\pi} |f(x)|^2\, dx = \sum_{k=-\infty}^{\infty} |c_k|^2 \tag{4.1}$$

Equation (4.1) is known as Parseval's Theorem in harmonic analysis and states that the square norm in the frequency domain is equal to the square norm in the time domain.

While the space $L_2(0, 2\pi)$ is an extremely useful one, for general problems in nonparametric function estimation we are much more interested in $L_2(\mathbf{R})$. We can think of $L_2(0, 2\pi)$ as with functions on the finite support $(0, 2\pi)$ or as periodic functions on $\mathbf{R}$. In the latter case it is clear that the infinitely periodic functions of $L_2(0, 2\pi)$ and the square integrable functions of $L_2(\mathbf{R})$ are very different. In the latter case the function, $f(x) \in L_2(\mathbf{R})$, must converge to 0 as $x \to \pm\infty$. The generating function $e(x) = e^{ix}$ clearly does not have that behavior and is inappropriate as a basis generating function for $L_2(\mathbf{R})$. What is needed is a generating function, $e(x)$, which also has the property that $e(x) \to 0$ as $x \to \pm\infty$. Thus we want to generate a basis from a function which will decay to 0 relatively rapidly, i.e. we want little waves or ***wavelets.***

4.2 Wavelet Bases.

Let us begin by considering a generating function ψ which we will think of as our ***mother wavelet*** or basic wavelet. The idea is that, just as with the sinusoids, we wish to consider a superposition of dilations of the basic waveform ψ. For technical convergence reasons which we shall explain later we wish to consider dyadic dilations rather than simply integral translations. Thus for the first pass, we are inclined to consider $\psi_j(x) = 2^{j/2}\psi(2^{j/2}x)$. Unfortunately, because of the decay of ψ to 0 as $x \to \pm\infty$, the elements $\{\psi_j\}$ are not sufficient to be a basis for $L_2(\mathbf{R})$. We accommodate this by adding translates to get the doubly indexed functions $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$. We choose ψ such that

$$\int_{\mathbf{R}} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega \text{ exists.}$$

Here $\hat{\psi}$ is the Fourier transform of ψ. Under certain choices of ψ, $\psi_{j,k}$ forms a doubly indexed orthonormal basis for L_2 (actually also for Sobolev spaces of higher order as well). As we shall see in the next section, a wavelet basis due to the dilation-translation nature of its basis elements admits an interpretation of a simultaneous time-frequency decomposition of f. Moreover using wavelets, fewer basis elements are required for fitting sharp changes or discontinuities. This implies faster convergence in "non-smooth" situations by the introduction of "localized" basis elements.

Example 3.1 Continued: Notice that

$$c_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{\infty} 2^{j/2}\psi\left(2^j x - k\right) f(x)\, dx.$$

In the density estimation case

$$c_{j,k} = E\left(2^{j/2}\,\psi\left(2^j x - k\right)\right).$$

Thus a natural estimator is

$$\hat{c}_{j,k} = \frac{2^{j/2}}{n} \sum_{i=1}^{n} \psi\left(2^j x_i - k\right),$$

where x_i, $i = 1,\ldots,n$ is the set of observations. Again we are simply using a method of moments estimator.

Notice that we can construct a Parseval's Theorem for Wavelets.

Theorem 4.2: (Parseval's Theorem for Wavelets)

$$(4.2) \quad \|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2 \, dx = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |c_{j,k}|^2 = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |c_{j,k}|^2.$$

At this stage we are left with the problem of constructing an appropriate mother wavelet, ψ, suitable for constructing the basis. To do this we turn to the device of multiresolution analysis.

4.3 Multiresolution Analysis.

To understand multiresolution analysis let us first consider the construction of space $W_j = \text{span}\{\psi_{j,k}: k \in J\}$. That is we fix the dilation and consider the space generated by all possible translates. We may write $L_2(\mathbf{R})$ as a direct sum of the W_j, $L_2(\mathbf{R}) = \sum_{j \in J} W_j$ so that any function $f \in L_2(\mathbf{R})$ may be written as

$$f(x) = \cdots + d_{-1}(x) + d_0(x) + d_1(x) + \cdots$$

where $d_j \in W_j$. If ψ is an orthogonal wavelet, then $W_j \perp W_k$, $k \neq j$. We shall assume the unknown ψ to be an orthogonal wavelet in what follows. Notice that as j increases, the basic wavelet form $\psi(2^j x - k)$ contracts representing higher "frequencies." For each j we may consider the direct sum V_j given by:

$$V_j = \cdots + W_{j-2} + W_{j-1} = \sum_{m=-\infty}^{j-1} W_m.$$

The V_j are closed subspaces and represent spaces of functions with all "frequencies" at or below a given level of resolution. The set of spaces $\{V_j\}$ has the following properties:

1) They are nested in the sense that $V_j \subseteq V_{j+1}$, $j \in J$.
2) Closure $\left(\cup_{j \in J} V_j\right) = L_2(\mathbf{R})$.
3) $\cap_{j \in J} V_j = \{0\}$.
4) $V_{j+1} = V_j + W_j$.
5) $f(x) \in V_j$ if and only if $f(2x) \in V_{j+1}$, $j \in J$.

1), 4) and 5) follow directly from the definition of V_j. 2) is a straightforward consequence of the fact that $\cup_{j \in J} W_j = L_2(\mathbf{R})$. 3) follows because of the orthogonality property.

Any $f \in L_2(\mathbf{R})$ can be projected into V_j. As we have seen with j increasing the

the "frequency" of the wavelet increases which can be interpreted as higher resolution. Thus the projection, $P_j f$, of f into V_j is an increasingly higher resolution approximation to f as $j \to \infty$. Conversely, as $j \to -\infty$, $P_j f$ is an increasingly blurred (smoothed) approximation to f. We shall take V_0 as the ***reference subspace***. Suppose now that we can find a function ϕ and that we can define $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ such that

$$V_0 = \text{span}\{\phi_{0,k}: k \in J\}.$$

Then by property 5), $V_j = \text{span}\{\phi_{j,k}: k \in J\}$. While we began our discussion with the notion of wavelets and have seen some of the consequences, we could have actually begun a discussion with the function ϕ.

Definition. A function ϕ generates a ***multiresolution analysis*** if it generates a nested sequence of spaces having properties 1), 2), 3) and 5) such that $\{\phi_{0,k}, k \in J\}$ forms a basis for V_0. If so, then ϕ is called the ***scaling function***.

For the final discussion of this section, let us consider a multiresolution analysis in which $\{V_j\}$ are generated by a scaling function $\phi \in L_2(\mathbf{R})$ and $\{W_j\}$ are generated by a mother wavelet function $\psi \in L_2(\mathbf{R})$. Any function $f \in L_2(\mathbf{R})$ can be approximated as closely as desired by f_m for some sufficiently large $m \in J$. Notice $f_m = f_{m-1} + d_{m-1}$ where $f_{m-1} \in V_{m-1}$ and $d_{m-1} \in W_{m-1}$. This process can be recursively applied say l times until we have $f \cong f_m = d_{m-1} + d_{m-2} + \cdots + d_{m-l} + f_{m-l}$. Notice that f_{m-l} is a highly smoothed version of the function. Indeed, this suggests that a statistical procedure might be to form a highly smoothed (even overly smoothed) approximation to a function to be estimated. The sequence d_{m-l} through d_m form the higher resolution wavelet approximations. Many of the wavelet coefficients $c_{m-i,k}$ used for constructing d_{m-i}, $i = 1,\ldots, l$ are likely to be 0 and hence can contribute to a very parsimonious representation of the function f. Indeed, a wavelet decomposition is a natural suggestion for a technology for high definition television (HDTV). If f_{m-l} represents the lower resolution conventional NTSC TV signal, then to reconstruct a high resolution image all that is needed is the difference signal which could be parsimoniously represented by the wavelet coefficients $c_{m-i,k}$, $i = 1,\ldots,l$ and $k \in J$, most of which would be 0.

Most importantly, however, is the observation that the scaling function $\phi \in V_0$

and the mother wavelet $\psi \in W_0$ implies that both are in V_1. Since V_1 is generated by $\phi_{1,k}(x) = 2^{1/2}\phi(2x-k)$, there are sequences $\{g(k)\}$ and $\{h(k)\}$ such that

$$\phi(x) = \sum_{k \in J} g(k)\phi(2x-k) \text{ and } \psi(x) = \sum_{k \in J} h(k)\phi(2x-k). \tag{4.3}$$

This remarkable result gives us a construction for the mother wavelet in terms of the scaling function. These equations are called the ***two-scale difference equations***. We can give a time series interpretation to these equations. Lets consider an original discrete time function, $f(n)$, to which we apply the filter

$$y(n) = \sum_{k \in J} g(k)f(2n-k).$$

First of all we note that there is a scale change due to subsampling by two, i.e. a shift by two in $f(n)$ results in a shift of one in $y(n)$. The scale of y is only half that of f. Otherwise this is a low pass filter with impulse response function g. Let us consider iterating this equation so that

$$y^{(j)}(n) = \sum_{k \in J} g(k)y^{(j-1)}(2n-k). \tag{4.4}$$

Notice that if this procedure converges, it converges to a fixed point which will be ϕ. This iterative procedure with repeated down sampling by two is suggestive of a method for constructing wavelets. If g is a finite impulse response (FIR) filter of length l, the construction of a complementary high-pass filter is accomplished with a FIR filter, h, whose impulse response is given by $h(l-1-n) = (-1)^n g(n)$. This scheme is called ***sub-band coding*** in the electrical engineering literature. The low-pass band is given by

$$y_0(n) = \sum_{k \in J} g(k)f(2n-k) \tag{4.5}$$

while the high-pass band is given by

$$y_1(n) = \sum_{k \in J} h(k)f(2n-k). \tag{4.6}$$

The filter impulses as defined form an orthonormal set so that the f may be reconstructed by

$$f(n) = \sum_{k \in J} [y_0(k)g(2k-n) + y_1(k)h(2k-n)]. \tag{4.7}$$

The sub-band coding scheme may be repeatedly applied to form the nested sequence pf V_j. The nested sequence of $\{V_j\}$ is then essentially obtained by recursively downsampling and filtering a function with a low-pass filter whose impulse response function is $g(\cdot)$.

4.4 Construction of Scaling Functions and Mother Wavelets.

We have already hinted that the scaling function may be constructed as the fixed point of the down-sampled, low-passed filter equation (4.4). This can be formalized by considering what statisticians would call the generating function of $g(n)$ and what electrical engineers call the z-transform of $g(\cdot)$.

$$G(z) = \frac{1}{2} \sum_{j \in J} g(j)\, z^j. \tag{4.8}$$

Notice if $z = e^{-i\omega/2}$, then (4.8) is essentially the Fourier transform of the impulse response function $g(\cdot)$. In this case, the first equation in (4.3) may be written as

$$\hat{\phi}(\omega) = G(z)\hat{\phi}\left(\frac{\omega}{2}\right), \text{ with } z = e^{-i\omega/2}. \tag{4.9}$$

This, of course, follows because the Fourier transform of a convolution is the corresponding product of the Fourier transforms. This recursive equation may be iterated to obtain

$$\hat{\phi}(\omega) = \prod_{k=1}^{\infty} G\left(e^{-i\omega/2^k}\right) \hat{\phi}(0). \tag{4.10}$$

We may take $\hat{\phi}$ to be continuous and $\hat{\phi}(0) = 1$. Based on (4.10) we may recover $\phi(\cdot)$ and based on this result, the equation $h(l-1-n) = (-1)^n\, g(n)$ and the second equation of (4.3) we may recover the mother wavelet, $\psi(\cdot)$. Thus Daubechies' original construction shows that wavelets with compact support can be based on finite impulse response filters which was originally motivated by multiresolution analysis. Theorem 4.3 below summarizes the general form of Daubechies' result.

Theorem 4.3: (Daubechies' Wavelet Construction):

Let $g(n)$ be a sequence such that

a. $\sum_{n \in J} |g(n)| \, |n|^{\epsilon} < \infty$ for some $\epsilon > 0$,

b. $\sum_{n \in J} g(n-2j)\, g(n-2k) = \delta_{jk}$,

c. $\sum_{n \in J} g(n) = 1$.

Suppose that $\hat{g}(\omega) = G(e^{-i\omega/2}) = 2^{-1/2} \sum_{n \in J} g(n)\, e^{-in\omega/2}$ can be written as

$$\hat{g}(\omega) = \left[\tfrac{1}{2}\,(1+\epsilon^{-i\omega/2})^{N}\right] \cdot \left[\sum_{n \in J} f(n)\, e^{-in\omega/2}\right]$$

where

d. $\sum_{n \in J} |f(n)| \, |n|^{\epsilon} < \infty$ for some $\epsilon > 0$

e. $\sup_{\omega \in \mathbf{R}} | \sum_{n} f(n)\, e^{-in\omega/2} | < 2^{N-1}$.

Define

$$h(n) = (-1)^n\, g(-n+1),$$

$$\hat{\phi}(\omega) = \prod_{k=1}^{\infty} G\left(e^{-i\omega/2^k}\right),$$

$$\psi(x) = \sum_{k \in J} h(k)\phi(2x-k).$$

Then the orthonormal wavelet basis is ψ_{jk} determined by the mother wavelet ψ. Moreover, if $g(n) = 0$ for $|n| > n_0$, then the wavelets so determined have compact support. □

We state this result without proof which may be found n Daubechies (1988). We note that Daubechies also shows that the mother wavelet, ψ, cannot be an even function and also have a compact support. The exception to this is the trivial constant function which gives rise to the so-called Haar basis. Daubechies illustrates this computation with the example of g given by $g(0) = (1+\sqrt{3})/8$, $g(1) = (3+\sqrt{3})/8$, $g(2) = (3-\sqrt{3})/8$ and, finally, $g(3) = (1-\sqrt{3})/8$. This wavelet is illustrated in Figure 4.1.

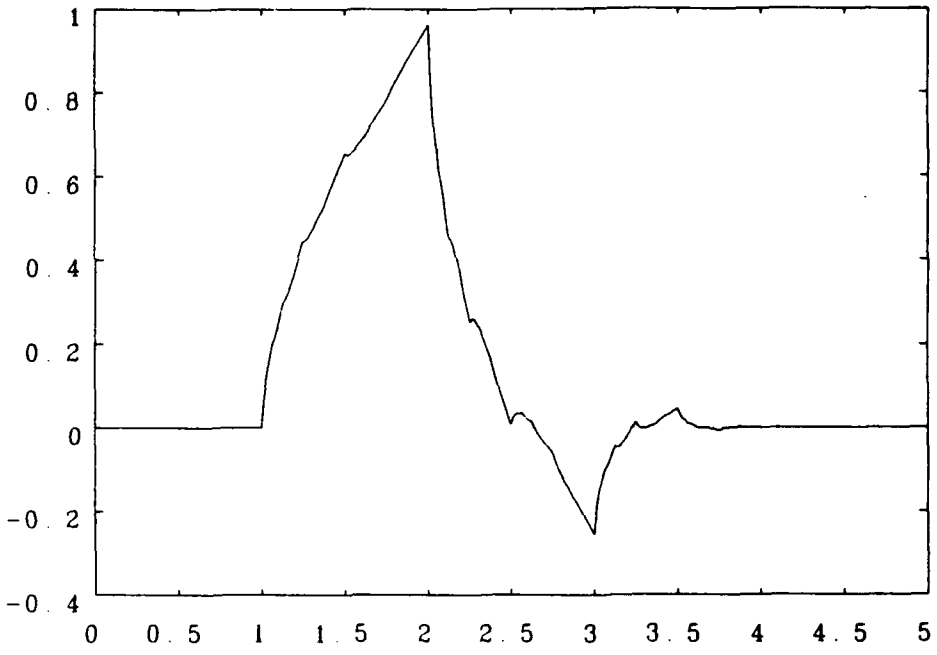


Figure 4.1a. Daubechies' Scaling Function using 4-term FIR filter.

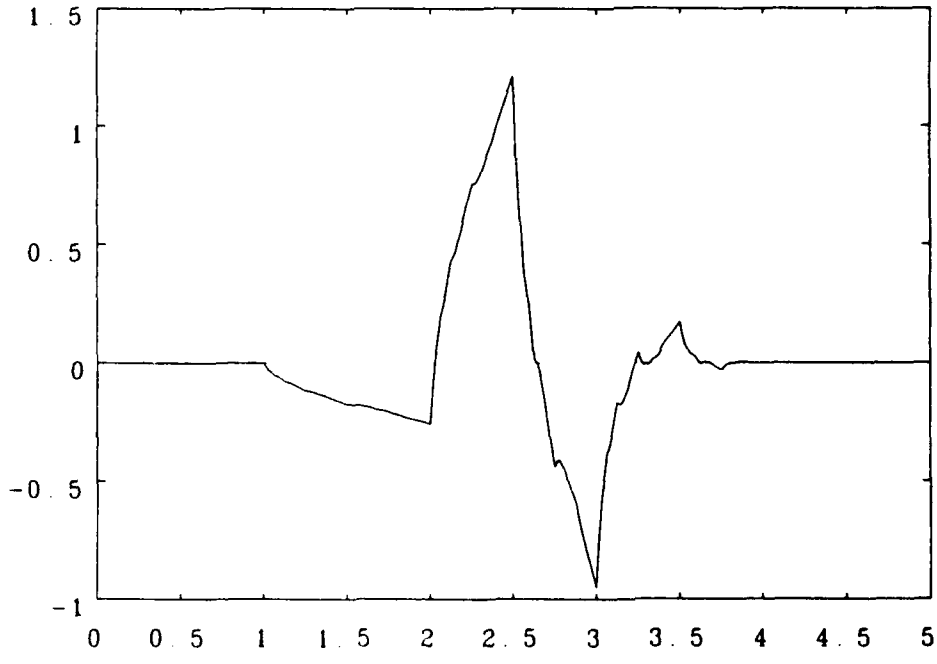


Figure 4.1b. Daubechies' Mother Wavelet using 4-term FIR filter.

5. Transient Signal Function Estimation.

Now with the basic construction of wavelets in hand, we can turn to the transient signal processing application. Wavelets have as one of their prime applications transient signal processing. In particular, since the most effective wavelets are those with compact support, they are a natural basis for transient signal estimation. However, if we are to exploit them in the context of optimal nonparametric function estimation, we must construct an optimality criterion for transient signals. The discussion below outlines an approach to transient signal estimation set in the context of optimal nonparametric function estimation. A fuller treatment can be found in Le and Wegman (1992). We first consider signals. It is well-known that there is no non-zero function in $L_2(\mathbf{R})$ which is both band-limited and time-limited. This being the case, we will assume the signal to be hard band-limited, i.e. with no energy outside a fixed interval, say $[-\nu, \nu]$, but soft time-limited, i.e. with minimal energy in the tails. This particular example demonstrates an elegant application of moments to signal processing.

5.1 Measuring of Out-of-Band Energy

Let $L_2(\mathbf{R})$ be the set of square-integrable, real-valued functions and let $h(t) \in L_2(\mathbf{R})$. Denote by $\hat{f}(\omega)$ the Fourier transform of $f(t)$ such that $\hat{f} \in L_2(\mathbf{R})$. We assume $\hat{f}$ is frequency band-limited so that $\hat{f}(\omega) = 0$, for $|\omega| > \nu$. We propose approximating the class of band-limited time-transient functions by considering functions whose energy time spread is confined to some small level s_0. As a measure of the energy time-spread, we will use analogies to concepts from probability theory to define various moments of $|f(t)|^2$, which plays the role of the energy distribution function. Assuming that

$$\int_{-\infty}^{\infty} |t|^j |f(t)|^2 \, dt < \infty, \; j = 1, 2, \ldots, k,$$

the k^{th} moment of the energy distribution will now be defined as follows

$$M^k = \int_{-\infty}^{\infty} t^k \, |f(t)|^2 \, dt.$$

For $k = 2$, we have the 2nd moment of the energy distribution function as a measure of the energy time spread, given as

$$M^2 = \int_{-\infty}^{\infty} t^2 \, |f(t)|^2 \, dt.$$

Remark: The factor t^k serves as a weight on the energy function which is used to control the degree of spreading in $|f(t)|$. A larger k value implies that more weight is applied at the tail-end of the energy distribution function and, therefore, the process of minimizing M^k requires that more energy be centrally concentrated.

5.2 Optimal Estimation of Band-Limited Processes

For $-\nu$ and ν real numbers, and m and p integers, where $-\infty \leq -\nu < \nu \leq \infty$, and $m \geq 0$ and $p \geq 1$, the Sobolev space $\mathcal{W}^{m,p}[-\nu, \nu]$ of complex-valued functions $\hat{f}$ on $[-\nu, \nu]$ is given by:

$$\mathcal{W}^{m,p}[-\nu, \nu] = \Big\{\hat{f}(\omega): \hat{f}^{(k)}(\omega) \text{ , } k = 0, 1, \ldots, m-1, \text{ are absolutely continuous}$$

$$\text{and, } \int_{-\nu}^{\nu} |\hat{f}^{(m)}(\omega)|^p \, d\omega < \infty\Big\}.$$

We consider observing an actual process, r(t), and we let $\hat{r}(\omega)$ be the Fourier transform of the observed process, r(t). The Fourier transform of the observed process, r(t), will then be modeled as $\hat{r}(\omega) = \hat{g}(\omega) + \xi(\omega)$ where, $\xi(\omega)$ is the spectrum of a stationary noise process, $\hat{g}(\omega) \in \mathcal{W}^{m,2}[-\nu, \nu]$. The fact that $\hat{f}$ belongs to the class $\mathcal{W}^{m,2}[-\nu,\nu]$ of band-limited signals implies that the support of $|f(t)|^2$ is not bounded. The objective is, then, to find a function $\hat{f}(\omega) \in \mathcal{W}^{m,2}[-\nu,\nu]$ which best fits the Fourier transform $\hat{r}(\omega)$ of the observed process r(t) with minimum time-energy spread; specifically we would like to minimize the following functional with $k < m$

$$(5.1) \qquad \min_{\hat{f} \in \mathcal{W}^{m,2}[-\nu,\nu]} \Big[\sum_{j=1}^{n} (\hat{f}(\omega_j) - \hat{r}(\omega_j))^2\Big] \text{ subject to } \int t^{2k} |f(t)|^2 \, dt \leq s_0,$$

where f(t) is the inverse Fourier transform corresponding to $\hat{f}(\omega)$ in $\mathcal{W}^{m,2}[-\nu, \nu]$.

5.3 Moment Connection via Parseval's Theorem

A rather elegant extension of Parseval's Theorem can be constructed under appropriate regularity conditions. The Parseval's Theorem for continuous Fourier transform pairs is

$$\int_{-\nu}^{\nu} |\hat{f}(\omega)|^2 \, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(t)|^2 \, dt.$$

But we know

$$\hat{f}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)\, e^{-it\omega} \, dt.$$

Take k^{th} derivatives with respect to ω

$$\frac{\partial^k \hat{f}(\omega)}{\partial \omega^k} = \frac{1}{2\pi}\int_{-\infty}^{\infty} (-it)^k f(t)\, e^{-it\omega} \, dt$$

so that

$$\hat{f}^{(k)}(\omega) = \frac{\partial^k \hat{f}(\omega)}{\partial \omega^k} \quad \text{is the Fourier transform of } (-it)^k f(t).$$

We can apply Parseval's Theorem to this Fourier transform pair to obtain

Theorem 5.1:

$$\int_{-\nu}^{\nu} |\hat{f}^{(k)}(\omega)|^2 \, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} t^{2k} |f(t)|^2 \, dt. \; \square$$

Thus, our optimization problem (5.1) can now be reformulated as

$$(5.2) \qquad \min_{\hat{f} \in \mathcal{W}^{m,2}[-\nu,\nu]} \Big[\sum_{j=1}^{n} (\hat{f}(\omega_j) - \hat{r}(\omega_j))^2 \Big] \text{ subject to } \int_{-\nu}^{\nu} |\hat{f}^{(k)}(\omega)|^2 \, d\omega \le s_0^*.$$

Using standard Lagrange multiplier techniques, this in turn may be reformulated as

$$(5.3) \qquad \min_{\hat{f} \in \mathcal{W}^{m,2}[-\nu,\, \nu]} \Big[\sum_{j=1}^{n} (\hat{f}(\omega_j) - \hat{r}(\omega_j))^2 + \lambda \int_{-\nu}^{\nu} |\hat{f}^{(k)}(\omega)|^2 \, d\omega \Big].$$

Indeed expression (5.3) is the form of optimization problem which results in a solution which is a generalized polynomial spline of degree $2k-1$. This result may be substantially generalized by the theorem given below which is developed in Le and Wegman (1992).

Theorem 5.2: Let $\hat{g}(\omega)$ be a band-limited spectral process with transient inverse Fourier transform and $\hat{r}(\omega)$ be the observed spectral process defined over some finite band $-\nu \le \omega \le \nu$. We model this spectral process as

$$\hat{r}(\omega) = \hat{g}(\omega) + \xi(\omega)$$

where $\xi(\omega)$ is some stationary white noise process. Let Λ be the time spread measure, defined as follows:

$$\Lambda(\hat{f}) = a_0\Lambda_j(\hat{f}) + a_1\Lambda_k(\hat{f})$$

where,

$$\Lambda_k(\hat{f}) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} t^{2k}\,|f(t)|^2\,dt,$$

and where a_0 and a_1 are the appropriately chosen weights. Here f is the inverse Fourier transform of $\hat{f}$ belonging to $L_2(\mathbf{R})$. Then, the optimal band-limited representation in the Sobolev space $\mathcal{W}^{m,2}[-\nu,\nu]$ is $\hat{f}_\lambda(\omega)$ where $\hat{f}_\lambda(\omega)$ is the solution to the problem:

$$\underset{\hat{f}\,\in\,\mathcal{W}^{m,2}[-\nu,\nu]}{\text{minimize}} \sum_{j=1}^{n} [\hat{f}(\omega_j) - \hat{r}(\omega_j)]^2 + \lambda\Lambda(\hat{f})\ .$$

$\hat{f}_\lambda$ is a generalized L-spline, and λ is known as the smoothing parameter. □

For a general discussion of L-splines, see Wegman and Wright (1983). Notice that if $\Lambda(\hat{f}) = \Lambda_k(\hat{f})$ for some large k, then we are constructing a band-limited transient signal estimator with little energy in the tail of the signal estimate, f_λ, where f_λ is the inverse Fourier transform of $\hat{f}_\lambda$. If $k = 2$, then

$$\Lambda_2(\hat{f}) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} t^4\,|f(t)|^2\,dt = \int_{-\nu}^{\nu} |\hat{f}^{(2)}(\omega)|^2\,d\omega$$

and our solution is the well-known cubic spline. However, much more interesting and physically meaningful solutions may be found. If $\Lambda(\hat{f}) = a_0\Lambda_0(\hat{f}) + a_1\Lambda_k(\hat{f})$, then for k odd

$$\Lambda(\hat{f}) = \frac{1}{2\pi}\left\{a_0^2\int_{-\infty}^{+\infty} |f(t)|^2\,dt + a_1^2\int_{-\infty}^{+\infty} t^{2k}\,|f(t)|^2\,dt\right\}.$$

Thus, we may also want to impose a total energy restriction on the estimated signal space. This imposed restriction may, for example, have resulted from a requirement to minimize channel bandwidth utilization from data transmission systems. Such modification, thus, yields the following optimization problem for k odd

$$\min_{\hat{f} \in W^{m,2}(-\nu,\nu)} \left[\sum_{i=1}^{n} (\hat{f}(\omega_j) - \hat{r}(\omega_j))^2 + \lambda_1 \int_{-\nu}^{\nu} | \hat{f}(\omega) |^2 \, d\omega + \lambda_2 \int_{-\nu}^{\nu} | \hat{f}^{(k)}(\omega) |^2 \, d\omega \right].$$

Hence, by our theorem the optimal solution is again an L-spline.

5.4 Computing Band-limited Transient Estimators and Example

The rather elegant result that our band-limited transient estimators are generalized L-splines makes the numerical computation of the estimators rather more routine since algorithms already exist for computing L-splines. The fact that we can impose total energy limits as well as tail-energy limits is an unexpected bonus. Our interpretation of Theorem 5.2 is as follows. We recommend doing an initial spectral estimation to establish the bandwidth, $-\nu \leq \omega \leq \nu$, over which we want to estimate $\hat{g}(\omega)$ (or more precisely the signal, g(t), its inverse Fourier transform). This initial spectral estimate will also allow us to select the sampling frequencies, ω_j. We recommend selecting these ω_j as the frequencies with the largest spectral mass. Notice that we may regard a transient signal, g(t), as the product of a signal of infinite support with an indicator function of a closed interval. It is well-known that Fourier transform of an indicator function is the so-called Dirichlet kernel which has a large central lobe and decreasing side lobes. By choosing sampling frequencies ω_j at the location of the central and side lobes, our technique allows us to to recover the indicator to an excellent approximation. Thus not only do we estimate the transient signal because of the penalty term for out-of-band energy, but because of the choice of sampling frequencies as well. Figure 5.1 graphically illustrates the results of our technique.

References

Chui, C. K. (1992), ***An Introduction to Wavelets***, Academic Press: Boston

Daubechies, I. (1988), "Orthonormal bases of compactly supported wavelets," ***Comm. on Pure and Appl. Math.***, 41, 909-996

Daubechies, I., Grossmann, A. and Meyer, Y. (1986), "Painless nonorthogonal expansions," ***J. Math. Phys.***, 27, 1271-1283

Heil, C. and Walnut, D. (1989), "Continuous and discrete wavelet transforms," ***SIAM Review***, 31, 628-666

Le, H. T. and Wegman, E. J. (1991), "Generalized function estimation of underwater transient signals," ***J. Acoust. Soc. America***, 89, 274-279

Le, H. T. and Wegman, E. J. (1992), "A spectral representation for the class of band-limited functions," to appear ***Signal Processing***

Rioul, O. and Vetterli, M. (1991), "Wavelets and signal processing," ***IEEE Sign. Proc. Mag.***, 8, 14-38

Wegman, E. J. and Wright, I. W. (1983), "Splines in statistics," ***J. Am. Statist. Assoc.***,78, 351-365

Wegman, E. J. (1984), "Optimal nonparametric function estimation," ***J. Statist. Planning and Infer.***, 9, 375-387

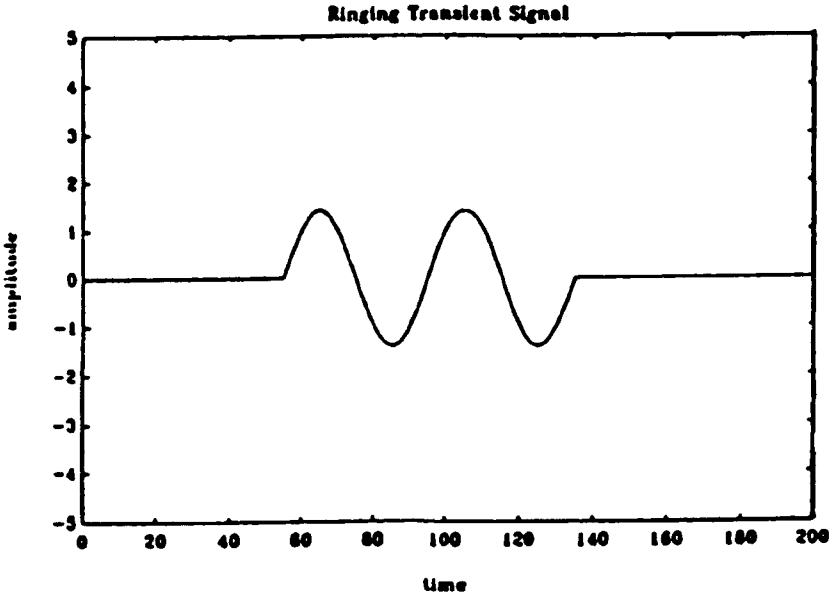


Figure 5.1a. A two-cycle transient signal.

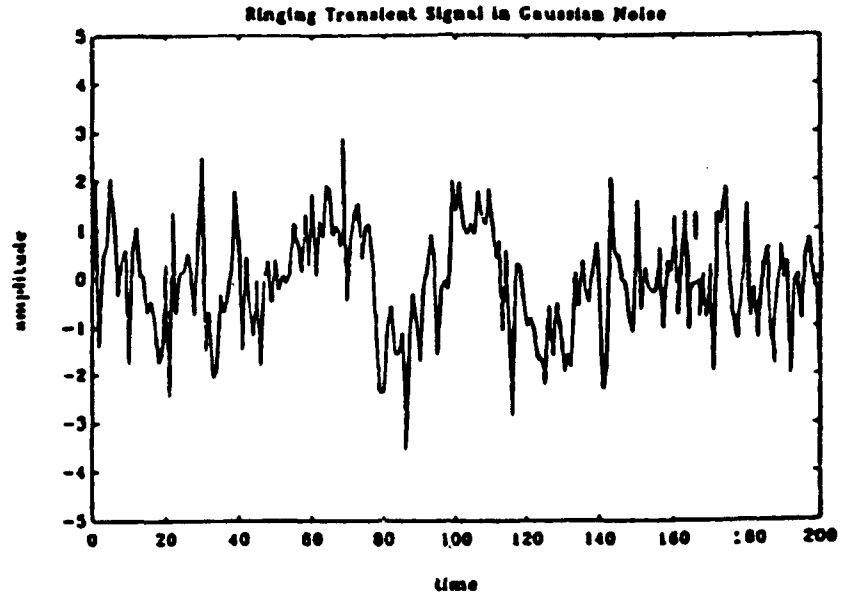


Figure 5.1b. Same two cycle signal buried in Gaussian noise.

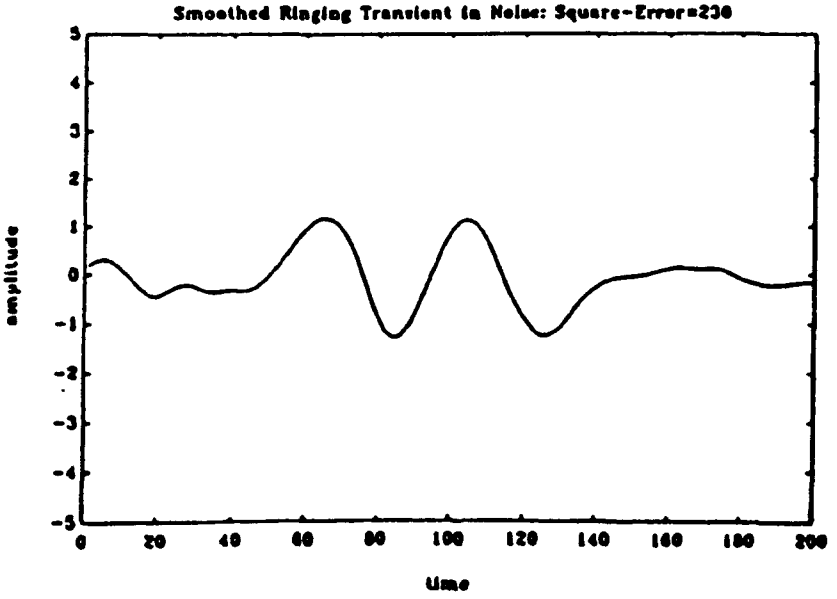


Figure 5.1c. Recovery of two-cycle signal waveform
by optimal band-limited techniques.

THE NAVAL POSTGRADUATE SCHOOL
AND
STANFORD UNIVERSITY

- CONFERENCE -

MOMENTS AND SIGNAL PROCESSING
30-31 MARCH 1992

- AGENDA -

Monday, 30 March 1992
(All sessions will be held in RM 101A, Spanagel Hall, Naval Postgraduate School.)

0830-0900: Registration, Spanagel 101A

0900-0915: Opening Remarks: P. Purdue, NPS
H. Solomon, Stanford University

0915-1045: **Session 1** Chair: P. Purdue, NPS
Speaker: D. Brillinger, U.C. Berkeley
Title: *Moments, Cumulants, and Applications to Random Processes*

1045-1115: Coffee Break

1115-1245: **Session 2** Chair: H. Solomon, Stanford
Speaker: M. Stephens, Simon Fraser University
Title: *Applications of Moments in Statistics and Probability*

1245-1400: Lunch Break

1400-1530: **Session 3** Chair: C. Therrien, NPS
Speaker: J. Mendel, University of Southern California
Title: *Recent Applications of Higher Order Statistics to Speech Processing and Array Processing*

1530-1600: Coffee Break

1600-1730: **Session 4** Chair: P. Purdue, NPS
Speaker: S. Iyengar, University of Pittsburgh
Title: *Probability Calculations for Multivariate Pearson Families*

1800-2000: "Social Hour," La Novia Terrace, Herrmann Hall

Tuesday, 31 March 1992
(All sessions will be held in RM 101A, Spanagel Hall, Naval Postgraduate School.)

0900-1030: **Session 5** Chair: H. Solomon, Stanford
Speaker: M. Nikias, University of Southern California
Title: *Blind Deconvolution using Higher-order Statistics*

1030-1100: Coffee Break

1100-1230: **Session 6** Chair: P. Jacobs, NPS
Speaker: E. Wegman, George Mason University
Title: *A Spectral Representation for the Class Band-limited Functions*

1230-1400: Lunch Break

1400-1530: **Session 7** Chair: L. Whitaker, NPS
Speaker: B. Lindsay, Penn State University
Title: *Moment-based Oscillation Properties of Mixture Models*

1530-1600: Coffee Break

1600-1730: **Session 8** Chair: P. Lewis, NPS
Speaker: K-S Lii, U.S. Riverside
Title: *Nonlinear Discrimination within the Higher Order Cumulant Structure*

1730-1800: **Final Remarks:** P. Purdue
H. Solomon

DISTRIBUTION LIST

1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304	2
2.	Library, Code 55 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Research Office, Code 08 Naval Postgraduate School Monterey, CA 93943-5000	1
4.	Peter Purdue, Code OR/Pd Chairman and Professor Operations Research Department Naval Postgraduate School Monterey, CA 93943-5000	350