

AD-A120 772

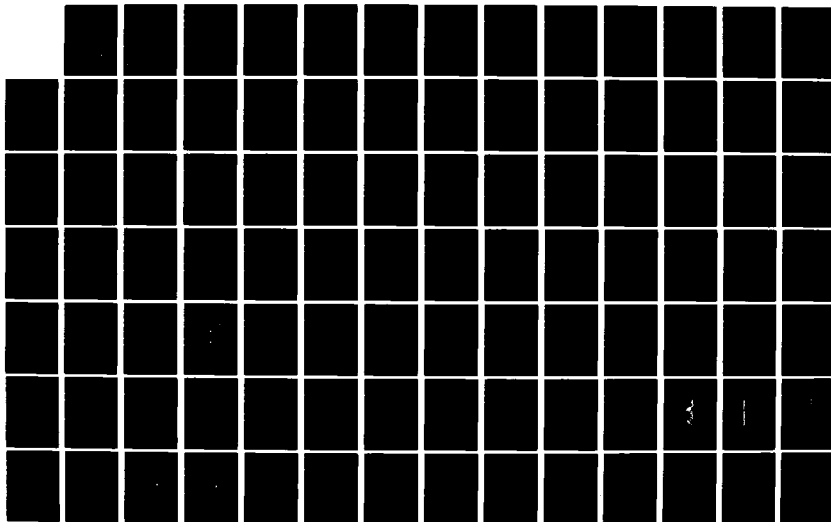
NON-GAUSSIAN AND MULTIVARIATE NOISE MODELS FOR SIGNAL
DETECTION(U) PRINCETON UNIV NJ DEPT OF ELECTRICAL
ENGINEERING AND COMPUTER SCIENCE A B MARTINEZ ET AL.
SEP 82 TR-6 N00014-81-K-0146

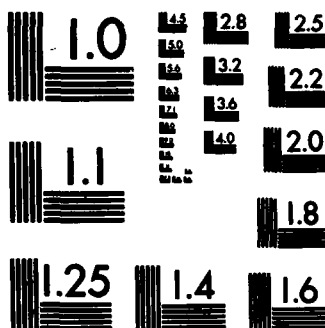
1/2

UNCLASSIFIED

F/G 9/4

NL

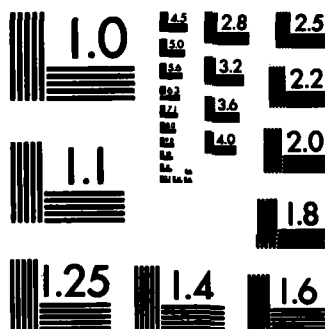




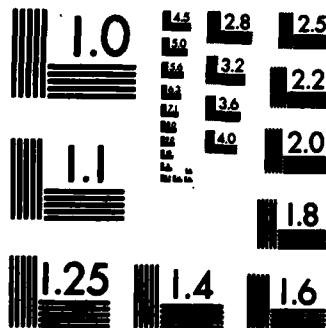
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A 120772

Report Number 6

(12)

NON-GAUSSIAN AND MULTIVARIATE NOISE MODELS FOR SIGNAL DETECTION

A.B. Martinez and J.B. Thomas

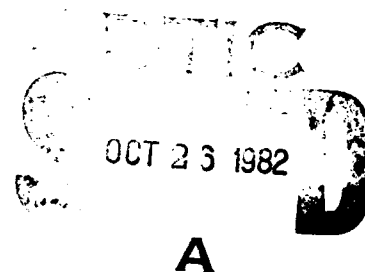
INFORMATION SCIENCES AND SYSTEMS LABORATORY

Department of Electrical Engineering and Computer Science
Princeton University
Princeton, New Jersey 08544

SEPTEMBER 1982

Prepared for

OFFICE OF NAVAL RESEARCH (Code 411SP)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-81-K-0146
S.C. Schwartz, Principal Investigator



Approved for public release; distribution unlimited

82 10 26 006

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO. AD-A120772	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NON-GAUSSIAN AND MULTIVARIATE NOISE MODELS FOR SIGNAL DETECTION		5. TYPE OF REPORT & PERIOD COVERED Technical Report Feb. 1981-July 1982
7. AUTHOR(s) Andrew B. Martinez & John B. Thomas		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Information Sciences & Systems Laboratory Dept. of Electrical Eng. & Computer Sci. Princeton Univ., Princeton, NJ 08544		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0146
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 411 SP) Department of the Navy Arlington, Virginia 22217		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR SRO-103
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE September 1982
		13. NUMBER OF PAGES 170
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Also submitted as the Ph.D. Thesis of A.B. Martinez to the EECS Department, Princeton University, Princeton, NJ 08544, August 1982.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Non-Gaussian Noise Adaptive Detection Multivariate Noise Models Discrete Matched Filters Locally Optimal Detection		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report deals with the problem of detecting a known signal in non-Gaussian or dependent noise. Although likelihood ratio (LR) detectors are discussed, primary attention is paid to asymptotic detector performance, and therefore to maximum efficacy or locally optimal (LO) detectors. The detectors considered con- sist of a nonlinearity followed by a filter and a threshold com- parator. (con't. on next page)		

(abstract con't.)

The asymptotic performance of three common suboptimal detectors is considered for several families of noise densities as the input signal-to-noise ratio (SNR) varies. Contours of equal detector performance are plotted allowing the relative utility of the detectors to be assessed.

A method of designing suboptimal detector nonlinearities is presented. A suboptimal nonlinearity is chosen, and the family of densities for which it is locally optimal is found. A member of this family is then fitted to the observed noise, and the corresponding detector is used. When the nonlinearity is a rational function, the Pearson family of densities results. Not only does this family contain many common univariate densities, but a member density can be fitted to an observed noise using only the first four sample moments. Other possible nonlinearities are considered, including possible multivariate extensions.

A ten minute sample of Arctic under-ice ambient noise is subjected to a moment analysis as suggested above. The noise is found to be nonstationary and nearly Gaussian with sporadic bursts of non-Gaussian noise, a conclusion also reached by other investigators.

The LR and LO detectors for several classes of multivariate densities are given. These classes include closed form, differential equation, spherically symmetric, series expansion, transformation and moving average models. The transformation model, (noise generated by a memoryless, nonlinear transformation of a correlated, Gaussian source) is discussed in some detail. The performance and practical aspects of obtaining the transformation LO detector are considered. Applicability of the noise models and tractability of the resulting detectors are discussed.

The problem of finding the minimum length matched filter in discrete-time to achieve a desired level of performance (SNR) is considered when there is some freedom in choosing a signal shape. Exact and approximate upper and lower bounds on the SNR are given, and the problems of optimal and suboptimal design are discussed.



A

NON-GAUSSIAN AND MULTIVARIATE NOISE MODELS FOR SIGNAL DETECTION

by

**Andrew B. Martinez and John B. Thomas
Department of Electrical Engineering
and Computer Science
Princeton University
Princeton, N.J. 08544**

ABSTRACT

This report deals with the problem of detecting a known signal in non-Gaussian or dependent noise. Although likelihood ratio (LR) detectors are discussed, primary attention is paid to asymptotic detector performance, and therefore to maximum efficacy or locally optimal (LO) detectors. In general, the detectors considered consist of a nonlinearity followed by a filter and a threshold comparator.

The asymptotic performance of three common suboptimal detectors is considered for several families of noise densities as the input signal-to-noise ratio (SNR) varies. Contours of equal detector performance are plotted allowing the relative utility of the detectors to be assessed.

A method of designing suboptimal detector nonlinearities is presented. A suboptimal nonlinearity is chosen, and the family of densities for which it is locally optimal is found. A member of this family is then fitted to the observed noise, and the corresponding detector is used. When the nonlinearity is a rational function, the Pearson family of densities results. Not only does this family contain many common univariate densities, but a member density can be fitted to an observed noise using only the first four sample moments. Other possible nonlinearities are

considered, including possible multivariate extensions.

A ten minute sample of Arctic under-ice ambient noise is subjected to a moment analysis as suggested above. The noise is found to be non-stationary and largely Gaussian or nearly Gaussian with sporadic bursts of non-Gaussian noise.

The LR and LO detectors for several classes of multivariate densities are given. These classes include closed form, differential equation, spherically symmetric, series expansion, transformation and moving average models. The transformation model, (noise generated by a memoryless, nonlinear transformation of a correlated, Gaussian source) is discussed in some detail. The performance and practical aspects of obtaining the subsystems of the transformation LO detector are considered. Applicability of the noise models and tractability of the resulting detectors are discussed.

The problem of finding the minimum length matched filter in discrete-time to achieve a desired level of performance (SNR) is considered when there is some freedom in choosing a signal shape. Exact and approximate upper and lower bounds on the SNR are given, and the problems of optimal and suboptimal signal design are discussed.

TABLE OF CONTENTS

Chapter 1 - Introduction

The Detection Problem	1
Dissertation Outline	7
References	11

Chapter 2 - Asymptotic Performance of Detectors with Non-zero Input SNR

Introduction	12
Detection Problem	14
Comparison of Several Tests	16
Results	18
Conclusions	39
References	40

Chapter 3 - Detector Design Using a Density Fit to Non-Gaussian Noise

Introduction	41
Locally Optimal Detector	43
Pearson Approximation to the ZNL	45
Examples	55
More Complex ZNL's	64
Conclusions	66
Appendix - Method of Moments for the Pearson Family	68
References	70

Chapter 4 - An Example of Moment Analysis of Arctic Under-Ice Data

Introduction	71
Computation of Moments	71
Conclusions	80
Appendix - Moment estimators for Dependent Noise	82
References	84

Chapter 5 - Locally Optimal Detection in Multivariate Non-Gaussian Noise

Introduction	85
Detection Problem	89
Transformation Noise Detector	93
Transformation Examples	98
Moving Average (MA) Noise Detector	110
Conclusions	113
Appendix A - Computation of R_{η}	115
Appendix B - Estimation of Transformation Nonlinearity	117
References	122

Chapter 6 - Noise Models for Detection

Introduction	124
Multivariate Densities	128
Conclusions	142
References	143

Chapter 7 - Finite Length Discrete Matched Filters

Introduction	144
Detection Problem	144
Signal Selection and Bounds on the SNR	147
Levinson Algorithm and Optimal Signal Selection	149
Approximate Bounds on the SNR	150
Examples	152
Conclusions	159
References	160

Chapter 8 - Conclusions

Review and Ideas for Further Research	161
---	-----

CHAPTER 1 - INTRODUCTION

THE DETECTION PROBLEM

The detection problem considered here is basically a simple one. A detector must determine whether a signal or, to generalize slightly, which of a number of possible signals, is present at its input. Were there no noise contaminating the received signal the problem would be moot. However, with the presence of noise, often of an unknown or incompletely known character, the choice of detector structure can be quite difficult. This problem occurs in many areas unrelated to communications, but unlike many of those areas, decisions frequently must be made in "real time" and with guaranteed bounds on the probability of error.

While there are many possible measures of detector performance, asymptotic results are used throughout this dissertation, often without much background or justification. While this makes each chapter more readable (and less repetitious), a fuller development of asymptotic theory can increase understanding. A more complete discussion of asymptotic theory can be found several recent books. Pitman's [1] development is fairly rigorous and complete. Of particular interest in the framework of this dissertation, Pitman derives the efficacy as a measure of asymptotic detector performance, and discusses some of the properties of both efficacy and the locally optimal detector. Huber [2] uses asymptotic results to develop his theory of robustness for both detectors and estimators. A more general background, clearly placing asymptotic results within the broader area of statistical inference theory is given by Bikel and Doksum [3], and by Helstrom [4] who approaches the problem

more from an engineering point of view.

In the sections which follow, some of the basic results of asymptotic theory are presented, both to make the notation clear, and to give some background for the following chapters. Fisher's Information is derived, and shown to be a measure of the sensitivity of a density to differential changes in a parameter ϑ . The efficacy is then shown to be related to Fisher's Information through the Information Inequality. And finally the efficacy is shown to be an indicator of the asymptotic behavior of a detector. The comments made here are not intended to form a rigorous development of asymptotic theory; for that the reader is referred to any of the References listed above.

Detection and estimation problems are generally concerned with a set of probability measures $\{P_\vartheta | \vartheta \in \Theta\}$ indexed by a real parameter (or parameters) ϑ contained in an interval Θ . This set has a corresponding set of densities $\{f_\vartheta | \vartheta \in \Theta\}$ and an associated σ -finite measure μ . In the detection problem considered in this dissertation ϑ is generally taken to be the signal level, and the detector is asked to decide whether $\vartheta = \vartheta_0$ or $\vartheta > \vartheta_0$.

1. Fisher's Information

The sensitivity of a detector or estimator to changes in the parameter ϑ is frequently of interest. This would give an indication of detector or estimator performance for small signal level, $\vartheta \approx \vartheta_0$. As shown below, Fisher's Information can be viewed as the sensitivity of a density to differential changes in ϑ .

As a measure of how two densities differ at a point x , consider the likelihood ratio (LR):

$$\Lambda(x) = f_{\vartheta}(x) / f_{\vartheta_0}(x)$$

or equivalently the log-likelihood ratio:

$$\lambda(x) = \ln f_{\vartheta}(x) - \ln f_{\vartheta_0}(x)$$

Both the LR and log-LR detectors are well known [1-4] to be Neyman-Pearson optimal for deciding between f_{ϑ} and f_{ϑ_0} based on an observation x . Because the LR and log-LR are optimal statistics for distinguishing between two densities, either can also be viewed as a measure of the distinguishability of the two densities at a point x . The derivation which follows would be essentially the same, and would, in fact, yield identical results whether the LR or log-LR is used as a measure of distinguishability. However, the log-LR is used since it has several appealing characteristics. Without sacrificing the optimality of the LR, the log-LR equals zero when two densities are indistinguishable at x , that is when $f_{\vartheta}(x) = f_{\vartheta_0}(x)$. It also has an appealing symmetry since

$$\ln(f_{\vartheta}/f_{\vartheta_0}) = -\ln(f_{\vartheta_0}/f_{\vartheta})$$

The sensitivity of f_{ϑ} to a change in ϑ can be written as the ratio of the distinguishability (i.e. log-LR) to the change in ϑ :

$$S(x) = \frac{\lambda(x)}{\vartheta - \vartheta_0} = \frac{\ln f_{\vartheta}(x) - \ln f_{\vartheta_0}(x)}{(\vartheta - \vartheta_0)}$$

As $\vartheta \rightarrow \vartheta_0$ the result is the rate of discrimination:

$$\lim_{\vartheta \rightarrow \vartheta_0} S(x) = \lim_{\vartheta \rightarrow \vartheta_0} \frac{\ln f_{\vartheta}(x) - \ln f_{\vartheta_0}(x)}{(\vartheta - \vartheta_0)}$$

Assuming that ϑ is an element of a real interval, and that the ϑ derivative

of f_{ϑ} exists at ϑ_0 , then

$$\lim_{\vartheta \rightarrow \vartheta_0} S(x) = \lim_{\vartheta \rightarrow \vartheta_0} \frac{\partial}{\partial \vartheta} \ln f_{\vartheta}(x) = \frac{f'_{\vartheta_0}(x)}{f_{\vartheta_0}(x)}$$

where f'_{ϑ_0} denotes the ϑ derivative of f_{ϑ} at $\vartheta = \vartheta_0$. Denoting the log-likelihood function at ϑ_0 as

$$L_{\vartheta_0}(x) = \ln f_{\vartheta_0}(x)$$

Then the rate of discrimination is the ϑ derivative of the log-likelihood function:

$$\lim_{\vartheta \rightarrow \vartheta_0} S(x) = L'_{\vartheta_0}(x) = f'_{\vartheta_0}(x) / f_{\vartheta_0}(x) \quad (1)$$

The rate of discrimination $L'_{\vartheta_0}(x)$ is a random variable with expected value zero:

$$E_{\vartheta_0} L'_{\vartheta_0} = \int f'_{\vartheta_0}(x) d\mu = 0 \quad (2)$$

As a measure of overall discrimination, consider the variance of L'_{ϑ_0} :

$$V_{\vartheta_0} L'_{\vartheta_0} = E(L'_{\vartheta_0})^2 = \int (f'_{\vartheta_0})^2 / f_{\vartheta_0} d\mu = I_{\vartheta_0} \quad (3)$$

which is seen to equal $I(\vartheta_0)$, Fisher's information at ϑ_0 .

Fisher's information can be interpreted to measure the overall sensitivity of a density to differential changes in a parameter ϑ at ϑ_0 . In another light, it can be viewed as the distinguishability of f_{ϑ} and f_{ϑ_0} as $\vartheta \rightarrow \vartheta_0$.

2. Efficacy and Information Inequality

Fisher's Information can be shown to have a corresponding measure of asymptotic detector (or estimator) performance, the efficacy. Further, the Information Inequality (or Cramer-Rao bound) states that Fisher's Information is an upper bound on efficacy.

Let $T(x)$ a statistic of the observation x with density $f_{\vartheta}(x)$ with expected value given by

$$E_{\vartheta}T = \int T(x)f_{\vartheta}(x)d\mu$$

The density f_{ϑ} is assumed regular in Pitman's sense [1], allowing the ϑ derivative of $E_{\vartheta}T$ to be brought inside the integral:

$$\frac{\partial}{\partial \vartheta}E_{\vartheta}T = E_{\vartheta}' = \int T(x)f_{\vartheta}'(x)d\mu$$

and using Eq (1):

$$E_{\vartheta}'T = \int T(x)L_{\vartheta}'(x)f_{\vartheta}(x)d\mu \quad (4)$$

Since $E_{\vartheta}L_{\vartheta}' = 0$ from Eq. (2), then an examination of Eq. (4) reveals that $E_{\vartheta}'T$ is the covariance of T and L_{ϑ}' . Thus by the correlation inequality:

$$(E_{\vartheta}'T)^2 \leq V_{\vartheta}TV_{\vartheta}L_{\vartheta}'$$

with equality if and only if $T = k L'$ where k is a real constant. From Eq. (3), $V_{\vartheta}L_{\vartheta}' = I(\vartheta)$. The result is the Information Inequality:

$$\frac{E_{\vartheta}'T^2}{V_{\vartheta}T} \leq I(\vartheta) \quad (5)$$

which can be slightly rearranged to yield the Cramer-Rao lower bound. The quantity on the left side of the inequality is defined as the efficacy of the statistic $T(x)$ at ϑ , and is denoted

$$J(T) = E_{\vartheta}'T^2 / V_{\vartheta}T$$

It follows from the correlation inequality that the efficacy is maximized by the test statistic $T_{lo} = L_{\vartheta}'$. This test statistic is referred to as the locally optimal test statistic, and has efficacy equal to Fisher's Information, $J(T_{lo}) = I(\vartheta)$.

3. Asymptotic Power of a Detector

The efficacy can be shown to measure the asymptotic behavior of a detector for large n and small ϑ . A general but loose proof of this is given below. This can, and indeed has been made more rigorous, and a more detailed development is given by Pitman [1].

As stated above, the detection problem considered consists of testing the hypothesis that $\vartheta = \vartheta_0$ against the alternative that $\vartheta > \vartheta_0$, where ϑ and ϑ_0 are contained in the real interval Θ . The detector consists of a test statistic T_n which is compared to a threshold t_n to decide for the hypothesis or alternative. Here n is the observation size, and T_n has expected value $E_{\vartheta} T_n = \mu_{\vartheta}$ and finite variance $V_{\vartheta} T_n = \sigma_{\vartheta}^2$. Then if the Central Limit Theorem holds, as $n \rightarrow \infty$, the random variable $(T_n - \mu_{\vartheta}) / \sigma_{\vartheta}$ is asymptotically normal with distribution Φ for all ϑ . The approximate level α_n of T_n can thus be given, for large n , by:

$$\alpha_n = P_{\vartheta_0}[T_n > t_n] \approx \Phi[(\mu_{\vartheta_0} - t_n) / \sigma_{\vartheta_0}]$$

The threshold can be found in terms of α_n :

$$t_n \approx \mu_{\vartheta_0} - \sigma_{\vartheta_0} \Phi^{-1}(\alpha_n) \quad (6)$$

In a similar fashion, the approximate power of a test is given by

$$\beta_n = P_{\vartheta}[T_n > t_n] \approx \Phi[(\mu_{\vartheta} - t_n) / \sigma_{\vartheta}] \quad (7)$$

Combining Eqs. (6) and (7) yields the approximate power function:

$$\beta(\alpha) = \Phi \left[\frac{\sigma_{\vartheta_0} \Phi^{-1}(\alpha) - \mu_{\vartheta} + \mu_{\vartheta_0}}{\sigma_{\vartheta}} \right]$$

Using the mean value theorem, we can write

$$\mu_{\vartheta} - \mu_{\vartheta_0} = (\vartheta - \vartheta_0) \mu_{\vartheta}'$$

where prime indicates differentiation with respect to ϑ , and ϑ' lies

between ϑ and ϑ_0 . The power function can be written as

$$\beta(\alpha) = \Phi \left[\frac{\sigma_{\vartheta_0}}{\sigma_{\vartheta}} \Phi^{-1}(\alpha) + (\vartheta - \vartheta_0) \frac{\mu_{\vartheta}'}{\sigma_{\vartheta}} \right]$$

For convenience, define $d = \mu_{\vartheta}' / \sigma_{\vartheta}$. Then for large n , d dominates the expression in brackets, and the power function is a monotone increasing function of d . Since $d \geq 0$, maximizing d^2 is equivalent to maximizing the power function. For $\vartheta \approx \vartheta_0$, then $\vartheta' \approx \vartheta_0$, and d^2 is very nearly equal to the efficacy $J(T_n)$:

$$d^2|_{\vartheta \approx \vartheta_0} = \frac{(\mu_{\vartheta}')^2}{\sigma_{\vartheta}^2} \approx \frac{(\mu_{\vartheta_0}')^2}{\sigma_{\vartheta_0}^2} = J(T_n)$$

Thus for large observation size n , and $\vartheta \approx \vartheta_0$, maximizing the efficacy is equivalent to maximizing the power function. It can also be shown that maximizing the efficacy is equivalent to maximizing the slope of the power function at $\vartheta = \vartheta_0$.

DISSERTATION OUTLINE

This dissertation deals with the problem of detecting a known signal in additive noise. Several of the traditional assumptions are relaxed; the noise is assumed to be non-Gaussian and/or dependent. In addition, the statistical model is often assumed to be incompletely known.

In Chapter 2, the performance of several common detectors is considered for several families of noise densities as the input signal-to-noise ratio (SNR) varies. The detector consists of a nonlinearity followed by a filter and a threshold comparator. Three commonly used nonlinearities are investigated: the linear amplifier, the sign detector and the amplifier limiter. Contours of equal detector performance are plotted allowing the

relative utility of the three detectors to be assessed for the noise models considered. The investigation was motivated by the observation that the locally optimal (or small signal) detector is designed for zero input SNR and can perform rather poorly as the input SNR increases. While this observation is born out, some additional conclusions can be drawn. In particular, the performance of the sign detector falls off rather quickly with increases in input SNR.

It is well known that the locally optimal detector for known signal in noise consists of a nonlinearity followed by a matched filter. In Chapter 3, suboptimal nonlinearities are investigated. It is assumed that either the locally optimal nonlinearity is too complex to use or that the noise density is not known precisely. A suboptimal nonlinearity can be chosen, and the family of densities for which it is optimal is found. A member of this family is then fitted to the observed noise, and the corresponding detector is used. When a rational function is chosen for the nonlinearity, the Pearson family is the set of solution densities. This is not only a general family which contains many common univariate densities, but for nearly Gaussian noise the method of moments can be used efficiently to fit a member density to the noise. The coefficients of a ZNL are estimated for several (non-Pearson) densities using the first four noise moments. Performance of the suboptimal detectors is investigated. Other possible nonlinearities are considered, including multivariate extensions.

In Chapter 4, 10 minutes of Arctic under-ice ambient noise is subjected to a moment analysis similar to that suggested in chapter 3. The

noise is found to be nonstationary and largely Gaussian or nearly Gaussian with sporadic bursts of non-Gaussian noise. Unfortunately, insufficient data is available for more extensive moments analysis, but the tentative conclusions drawn about the nature of the noise are interesting.

Chapters 5 and 6 (excluding Appendix B of Chapter 5) were coauthored with P.F. Swaszek and also appear in his dissertation [5].

In solving problems in detection, it is often assumed that the underlying statistical description is independent or Gaussian. Not making these assumptions leads to difficulties in detector design due to problems usually encountered in specifying multivariate noise statistics. In Chapters 5 and 6, several characterizations of multivariate densities are considered within a detection framework. The discussions include specific examples and also some general methods of density generation.

The particular detection problem considered in Chapter 5 is that of a known signal with a vanishingly small amplitude in additive noise. Efficacy is employed as a criterion of detector performance and the maximum efficacy (locally optimal) detector is discussed. The class of multivariate densities generated by a memoryless, nonlinear transformation of a correlated, Gaussian source is discussed in some detail. A member of this class has the advantage of being completely characterized by its marginal density and its covariance matrix. The locally optimal detector structure is derived for this class and the practical aspects of obtaining detector sub-systems are considered. Examples of this detector are presented for noise sources with Laplace and Pearson Type VII marginal

densities, and Monte Carlo simulations are included to aid performance analysis. A second class of multivariate densities generated by a linear transformation of an iid noise source is also considered, and its locally optimal detector is described.

In Chapter 8 the likelihood ratio and locally optimal detectors for several classes of multivariate densities are given. These classes include closed form, differential equation, spherically symmetric, series expansion, transformation and moving average models. Applicability of the noise models and tractability of the resulting detectors are discussed.

The Matched Filter (MF) is well known to be the linear detector that has the maximum output Signal-to-Noise Ratio (SNR). In Chapter 7, the problem of finding the minimum filter length in discrete time to achieve a certain level of performance is considered when there is some freedom in choosing a signal shape. Upper and lower bounds on the SNR are given in terms of the eigenvalues of the noise covariance matrix. Since these bounds are rather difficult to compute, looser, but easier to compute, bounds are given. Several examples are presented which illustrate the exact and approximate bounds.

In Chapter 8, the main results of this work are summerized, and ideas for further research are given.

REFERENCES

1. E.J.G. Pitman, *Some Basic Theory for Statistical Inference*. London: Chapman and Hall, 1979.
2. P.J. Huber, *Robust Statistics* New York: John Wiley & Sons, 1981.
3. P.J. Bikel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, 1977.
4. C.W. Helstrom, *Statistical Theory of Signal Detection*. Oxford: Pergamon Press, 1968.
5. P.F. Swaszek, *Robust Quantization, Vector Quantization and Detection*, PhD dissertation, Princeton University, October 1982.

CHAPTER 2 - ASYMPTOTIC PERFORMANCE OF DETECTORS WITH NON-ZERO INPUT SNR.

INTRODUCTION

Frequently, the performance of locally optimal and suboptimal detection schemes is evaluated using asymptotic measures such as the Asymptotic Relative Efficiency (ARE). Unfortunately, the ARE requires the assumption, not only of a large number of samples (for the central limit theorem to hold), but of a vanishingly small input signal-to-noise ratio (SNR). It follows that any optimal or suboptimal system designed by maximizing ARE may perform poorly for a signal level greater than zero.

From a survey of the literature, it is evident that these assumptions have concerned other researchers in the past. In particular, Miller and Thomas [1] investigated the convergence of Relative Efficiency to ARE for a signal in iid Laplace noise. They found the convergence to be relatively slow, implying that the ARE may not be an accurate measure for small to moderate sample size. Spaulding [2] found that, for small sample size and large signal level, the hard limiter may well outperform the locally optimal detector for a signal in Hall-distributed noise.

These results motivate a further investigation. In this chapter, the asymptotic performance of several simple detector structures is considered when input SNR is not constrained to be zero. Originally, we looked at the single case of Laplace noise and were interested in the relative performance of the locally optimal (sign), the Neyman-Pearson (amplifier limiter), and linear detectors as the input SNR is increased. Since these are frequently encountered suboptimal, robust or

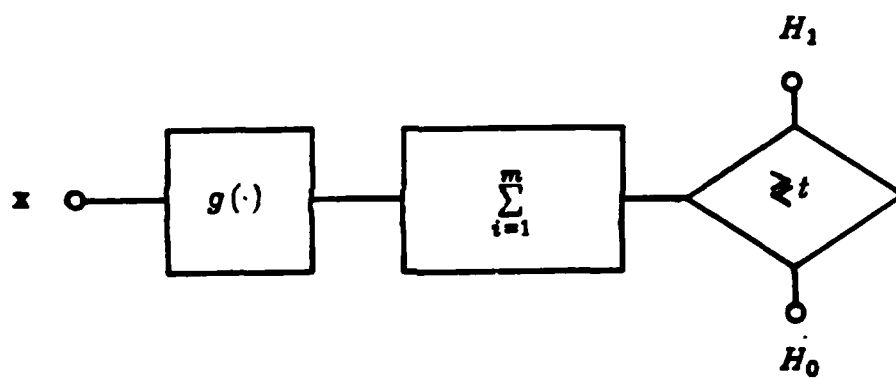


Fig. 1. Typical detector structure.

nonparametric detectors, a comparison of their performance for other families of densities might prove interesting.

The detector structure considered in this chapter consists of a zero memory nonlinearity (ZNL) followed by a summer and a threshold comparator as shown in Fig. 1. Both the likelihood ratio and locally optimal detectors for iid noise and constant signal have this configuration, and for this reason among others, this form has particular appeal not only in optimal but in suboptimal detector design. In this chapter, three detector ZNL's are used: the linear amplifier, the sign detector and the amplifier limiter.

DETECTION PROBLEM

The problem of detecting a constant signal in additive noise, given a sequence of observations, can be expressed as an hypothesis and an alternative:

$$\begin{aligned} H_0: x_i &= n_i \\ H_1: x_i &= n_i + \vartheta \end{aligned}$$

Under the hypothesis H_0 , the observation x_i consists of noise only; under the alternative H_1 it consists of noise plus an additive signal ϑ . When the noise is assumed to be independent and identically distributed (iid) with density $f(n_i)$, both the likelihood ratio detector and the locally optimal detector consist of a ZNL $g(\cdot)$ followed by a linear filter. If the number of samples is m then the test statistic T_m is given by

$$T_m = \sum_{i=1}^m g(x_i)$$

To simplify notation somewhat, let

$$\mu_0 = E_0 T_m \quad \sigma_0^2 = E_0 T_m^2 - E_0^2 T_m$$

and

$$\mu_1 = E_1 T_m \quad \sigma_1^2 = E_1 T_m^2 - E_1^2 T_m$$

where E_0 and E_1 represent the expected value under H_0 and H_1 respectively.

If m is sufficiently large for the central limit theorem to be applied, then under H_0 and H_1 the two random variables

$$\frac{T_m - \mu_0}{\sqrt{m} \sigma_0} \quad \text{and} \quad \frac{T_m - \mu_1}{\sqrt{m} \sigma_1}$$

are both approximately unit normal, and the approximate level α is therefore

$$\alpha = \int_{\frac{t - m\mu_0}{\sqrt{m} \sigma_0}}^{\infty} d\Phi(x) = \Phi\left(\frac{m\mu_0 - t}{\sqrt{m} \sigma_0}\right)$$

where Φ is the Gaussian $N(0,1)$ cdf. The threshold t can be found in terms of α :

$$t = m\mu_0 - \sqrt{m} \sigma_0 \Phi^{-1}(\alpha)$$

Likewise the approximate power can be computed:

$$\beta = \int_{\frac{t - m\mu_1}{\sqrt{m} \sigma_1}}^{\infty} d\Phi(x) = \Phi\left(\frac{m\mu_1 - t}{\sqrt{m} \sigma_1}\right)$$

Using the expression for the threshold yields the power function:

$$\beta(\alpha) = \Phi\left(\frac{\sigma_0}{\sigma_1} \Phi^{-1}(\alpha) + \sqrt{m} \frac{\mu_1 - \mu_0}{\sigma_1}\right)$$

The approximate power of a test for fixed level α and large sample size m is a monotone increasing function of the square root of the detector output SNR, where

$$SNR_{OUT} = \frac{(\mu_0 - \mu_0)^2}{\sigma_0^2}$$

As $\sqrt{m} \rightarrow \infty$, the term containing the square root of the SNR will clearly dominate the other term in the power function, and the SNR is an accurate measure of detector performance. For moderate m , where the Gaussian assumption may still be quite good, the term $\sigma_0 \Phi^{-1}(\alpha) / \sigma_0$ can become important. In this case, using the SNR to compare the performance of two tests may not be enough; σ_0 / σ_0 should ideally equal the same constant for both tests. In the following examples $\sigma_0 = \sigma_0$ for all detectors considered. Asymptotically, the results are equivalent, but this should yield more valid results for moderate m .

COMPARISON OF SEVERAL TESTS

In this section, three tests are considered: the linear detector, sign detector and the amplifier limiter. Under H_0 the observation x has distribution $F(x)$ and under H_1 distribution $F(x - \vartheta)$. Obviously, the noise variance will have no effect on the gain in SNR of each detector (input to output), and therefore it can be set equal to unity with no loss of generality.

1. The linear detector (LD) is simple and frequently used. It is usually justified by making a Gaussian assumption about the noise, and is both the Neyman Pearson and locally optimal detector for Gaussian noise. The LD has the ZNL

$$g(x) = x$$

and so

$$\mu_0 = 0 \quad \mu_0 = \vartheta \quad \text{and} \quad \sigma_0^2 = \sigma_0^2 = 1$$

and the output SNR of this detector is

$$SNR_{LD} = \vartheta^2 \quad (1)$$

2. The sign detector (SD) is probably the simplest ZNL which may be effective in an impulsive environment. It has been shown [4] that the SD frequently performs reasonably well for small signals in impulsive noise.

The SD has the ZNL:

$$g(x) = \text{sgn}(x)$$

Thus

$$\mu_0 = 1 - 2F(0) \quad \mu_\vartheta = 1 - 2F(-\vartheta) \quad \text{and} \quad \sigma_0 = \sigma_\vartheta = 1$$

and the SNR of the SD is

$$SNR_{SD} = 4[F(0) - F(-\vartheta)]^2 \quad (2)$$

3. The form of the amplifier limiter (AL) used here is the Neyman-Pearson optimal test ZNL for Laplace noise, with breakpoints at the signal level ϑ . While this is not necessarily the best breakpoint for all noise environments, it avoids the difficulties inherent in choosing optimal breakpoints [4] while still preserving some of the robust characteristics described by Martin and Schwartz [6].

The AL has the ZNL:

$$g(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq \vartheta \\ \vartheta & \vartheta \leq x \end{cases}$$

Thus

$$\mu_0 = \vartheta[1 - F(\vartheta)] + \int_0^\vartheta x \, dF(x)$$

$$\mu_\vartheta = \vartheta[1 - F(-\vartheta)] + \int_{-\vartheta}^0 x \, dF(x)$$

$$\sigma_0^2 = \int_0^\vartheta x^2 \, dF(x) + \vartheta^2[1 - F(-\vartheta)] - \mu_0^2$$

$$\sigma_0^2 = \int_{-\infty}^0 x^2 dF(x) + 2\vartheta \int_{-\infty}^0 x dF(x) + \vartheta^2 [1 - F(-\vartheta)] - \mu_0^2$$

Denote the i th incomplete moment as I_i , where

$$I_i = \int_0^{\vartheta} x^i dF(x)$$

Then if f is assumed symmetric:

$$\mu_0 = \vartheta[\frac{1}{2} - I_0] + I_1$$

$$\mu_0 = \vartheta[\frac{1}{2} + I_0] - I_1$$

$$\sigma_0^2 = \sigma_{\vartheta}^2 = I_2 + \vartheta^2[\frac{1}{2} + I_0] - 2\vartheta I_1 - \mu_0^2$$

The AL has SNR

$$SNR_{AL} = \frac{4(\vartheta I_0 - I_1)^2}{I_2 + \vartheta^2/4 - \vartheta I_1 - \vartheta^2 I_0^2 - I_1^2 + 2\vartheta I_0 I_1} \quad (3)$$

RESULTS

A. GENERALIZED GAUSSIAN

The first distribution considered is the generalized Gaussian.

$$f(x) = \frac{\alpha^{1/c} c}{2\Gamma(1/c)} \exp(-\alpha |x|^c)$$

For a variance of unity

$$\alpha^{2/c} = \Gamma(3/c)/\Gamma(1/c)$$

In addition to including the Gaussian and Laplace as special cases, the generalized Gaussian offers considerable control over the rate of tail decay through decay parameter c . It has been suggested as a model of impulsive atmospheric noise, particularly for $0.1 < c < 0.6$, and used to demonstrate the locally optimal detector by Miller and Thomas [3], and later by Lu and Eisenstein [5].

The incomplete moments I_i (for $i = 0, 1, 2$) are

$$I_0 = \frac{\gamma(1/c, \alpha \vartheta^c)}{2\Gamma(1/c)}$$

$$I_1 = \frac{\gamma(2/c, \alpha \vartheta^c)}{2[\Gamma(1/c)\Gamma(3/c)]^{1/2}}$$

$$I_2 = \frac{\gamma(3/c, \alpha \vartheta^c)}{2\Gamma(3/c)}$$

where $\gamma(\dots)$ is the incomplete γ function.

Using the incomplete moments and Eqs. (1-3), the output SNR's of the LD, SD and AL detectors are computed and plotted in Figs. 2-6 for several values of decay parameter c . Two values of c are of particular interest.

When $c = 1$ (Fig. 3) the generalized Gaussian reduces to the Laplace density. The SD is then locally optimal and the AL is Neyman-Pearson optimal. As expected, the AL outperforms both the LD and SD for when the input SNR is greater than zero. Perhaps surprisingly, the performance of the locally optimal SD decays significantly as input SNR increases. In fact, the LD outperforms the SD when $SNR_{IN} > 0.272$.

The Gaussian density results when $c = 2$ (Fig. 5.). In this case, the LD is both locally and Neyman-Pearson optimal, and as expected, it outperforms both the AL and the SD.

The points of greatest interest in Figs. 2-6 are the crossover points at which two tests have equal output SNR. In Fig. 7, the contours of $SNR_{SD} = SNR_{LD}$ and $SNR_{AL} = SNR_{LD}$ are plotted in the $c \times SNR_{IN}$ plane. The two contours divide the portion of the plane considered into three regions.

When $SNR_{IN} = 0$, the SD and AL are equivalent; for all $SNR_{IN} > 0$, the AL outperforms the SD. As configured here, the AL requires the signal

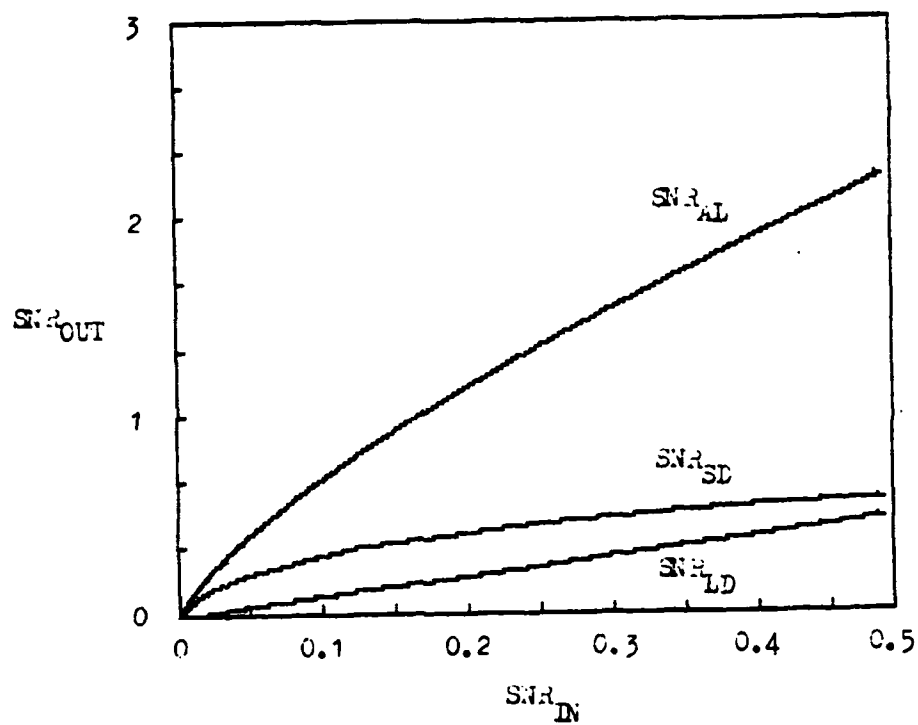


Fig. 2. SNR_{IN} versus SNR_{OUT} for generalized Gaussian, $c = 0.5$.

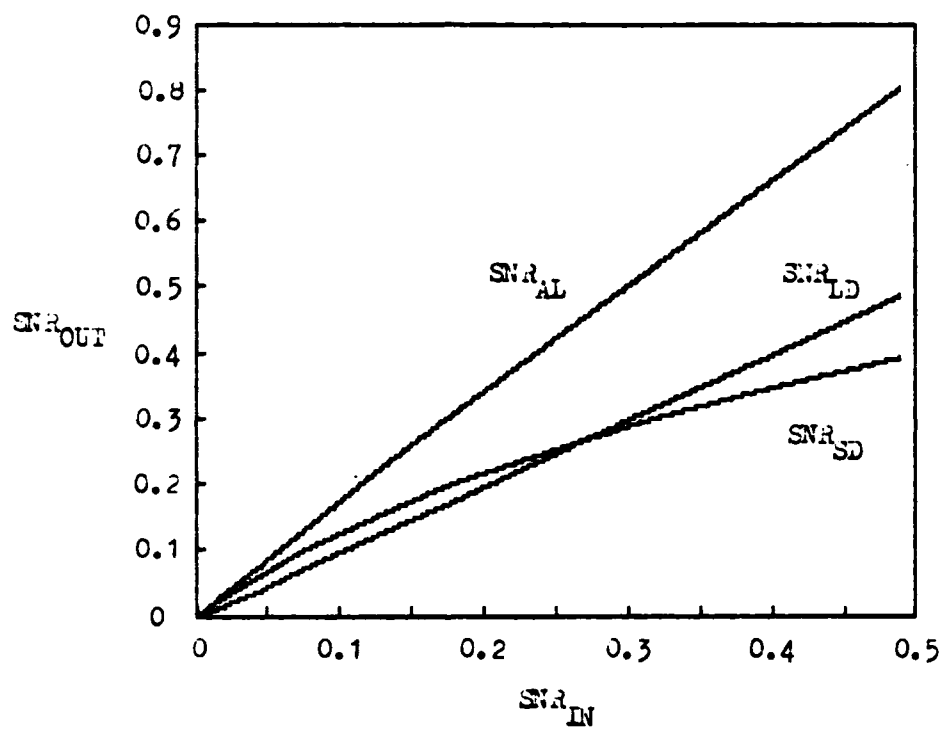


Fig. 3. SNR_{IN} versus SNR_{OUT} for generalized Gaussian, $c = 1.0$.

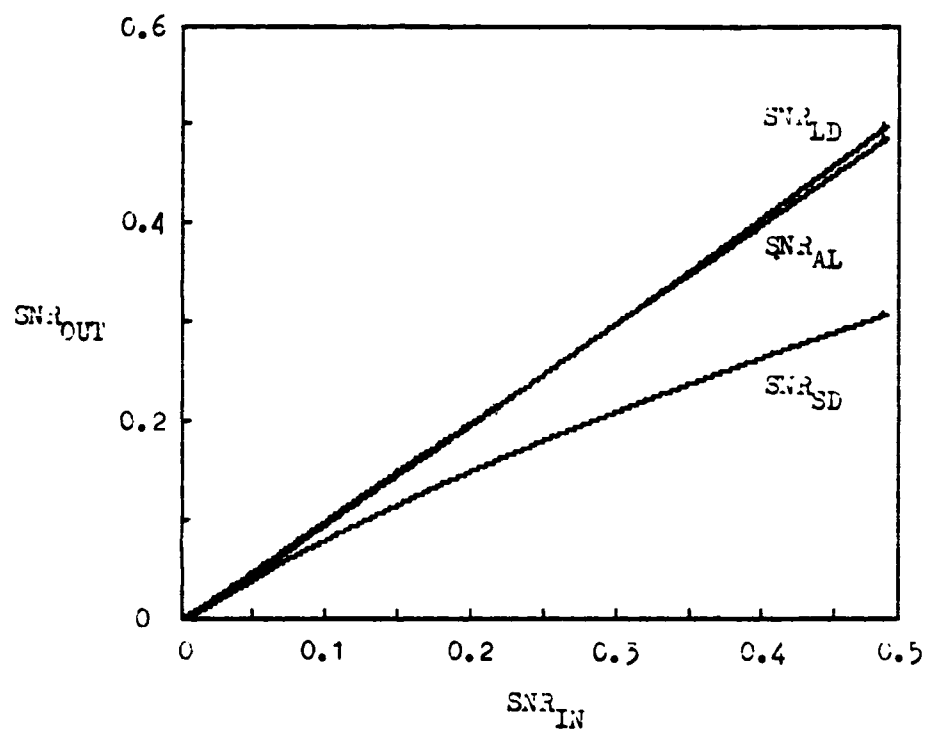


Fig. 4. SNR_{IN} versus SNR_{OUT} for generalized Gaussian, $c = 1.5$.

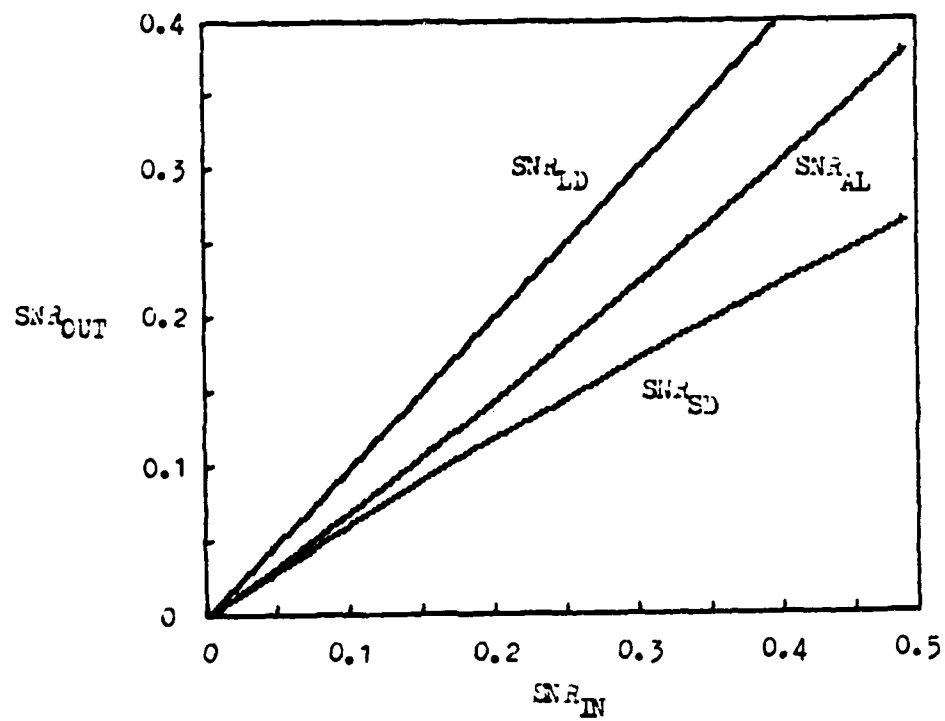


Fig. 5. SNR_{IN} versus SNR_{OUT} for generalized Gaussian, $c = 2.0$.

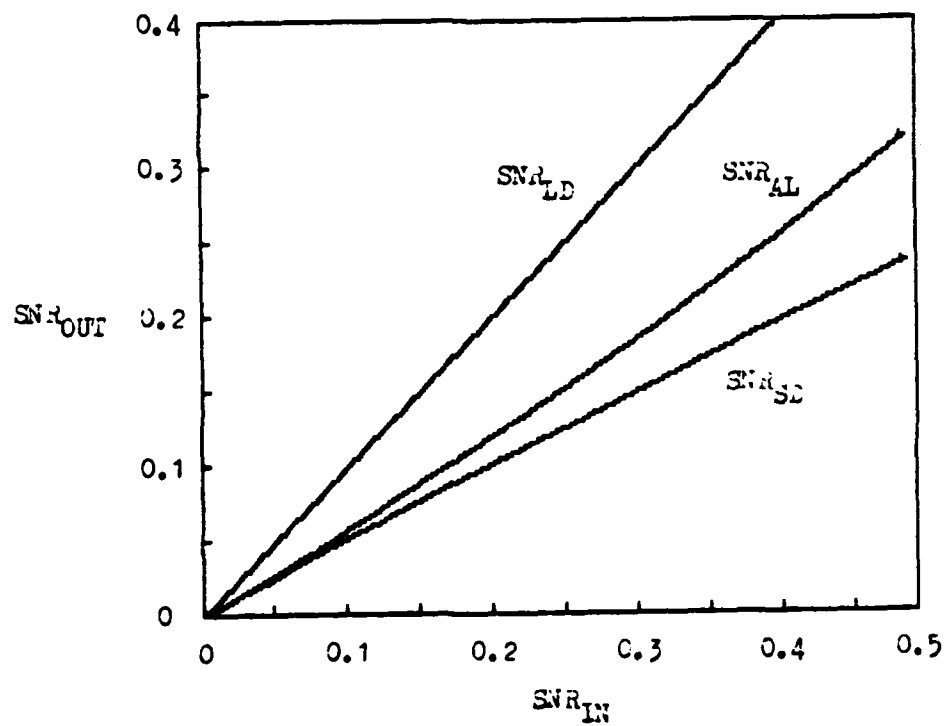


Fig. 6. SNR_{IN} versus SNR_{OUT} for generalized Gaussian, $c = 2.5$.

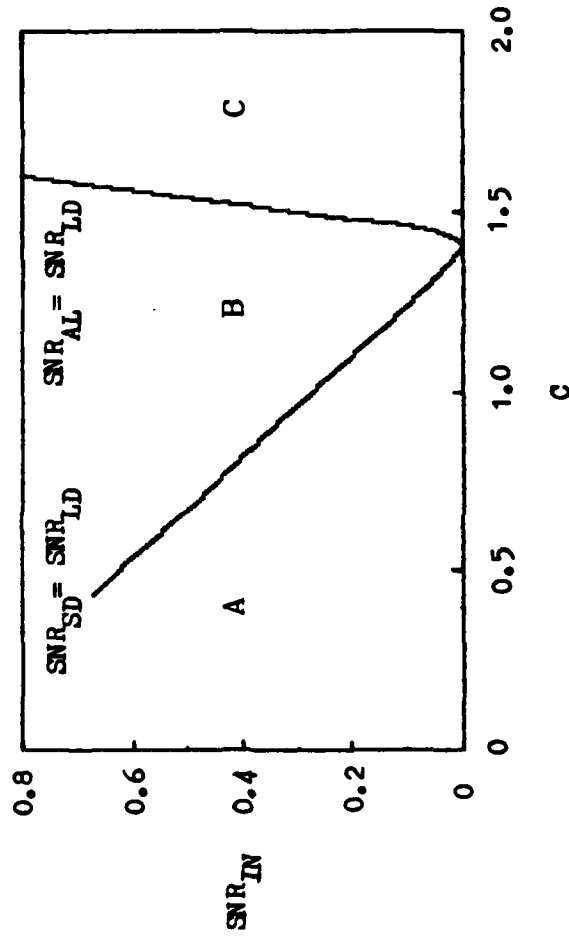


Fig. 7. Contours of $SNR_{SD} = SNR_{LD}$ and $SNR_{AL} = SNR_{LD}$ in the $c \times SNR_{IN}$ plane for generalized Gaussian noise.

- A. $SNR_{AL} > SNR_{SD} > SNR_{LD}$
- B. $SNR_{AL} > SNR_{LD} > SNR_{SD}$
- C. $SNR_{LD} > SNR_{AL} > SNR_{SD}$

level ϑ to be known, and when ϑ is known the AL would normally be used over the SD. The LD is to be preferred over the AL only in the rightmost region, right of the contour $SNR_{AL} = SNR_{LD}$. When ϑ is not known, the choice is between the LD and the SD. The LD outperforms the SD in all but the region left of the contour $SNR_{SD} = SNR_{LD}$.

In Fig. 8, the the relative performance of the SD and the LD is shown as contours of $(SNR_{SD}/SNR_{LD})_{dB}$ in the $c \times SNR_{IN}$ plane. The 0dB line corresponds to the line of equal performance in Fig. 7, and performance of the SD relative to that of the LD can be seen to fall off quickly not only with the transition from Laplace ($c = 1$) to Gaussian ($c = 2$) but with increases in SNR_{IN} .

In practice, one might use single or (especially in nonstationary noise environments) multiple estimates of c and SNR_{IN} and Fig. 7 to choose a detector structure. To simplify the process, since c is rather difficult to estimate, the two contours $SNR_{SD} = SNR_{LD}$ and $SNR_{AL} = SNR_{LD}$ are given in the $\beta_2 \times SNR_{IN}$ plane in Fig. 9. Here β_2 is the normalized fourth moment, or kurtosis of the noise, frequently used as a measure of density "peakedness". The kurtosis is defined to be:

$$\beta_2 = \mu_4 / \mu_2^2$$

where μ_2 and μ_4 are the second and fourth noise moments about the mean. Not only is the kurtosis β_2 easier to estimate, but it is a more general measure than c and can be applied to many other densities.

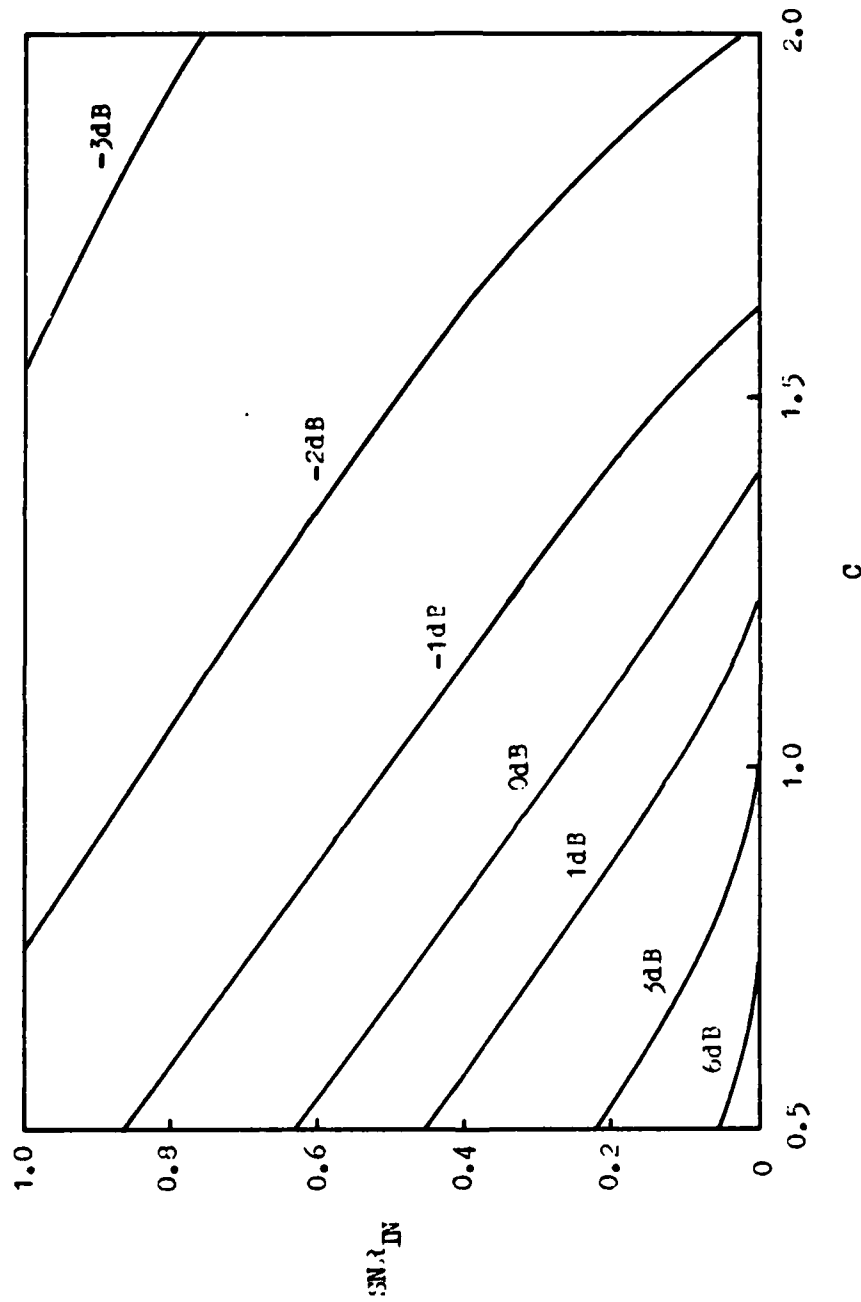


Fig. 8. Contours of $(SNR_{SD}/SNR_{ID})_{dB}$ in $c \times SNR_{IN}$ for generalized Gaussian.

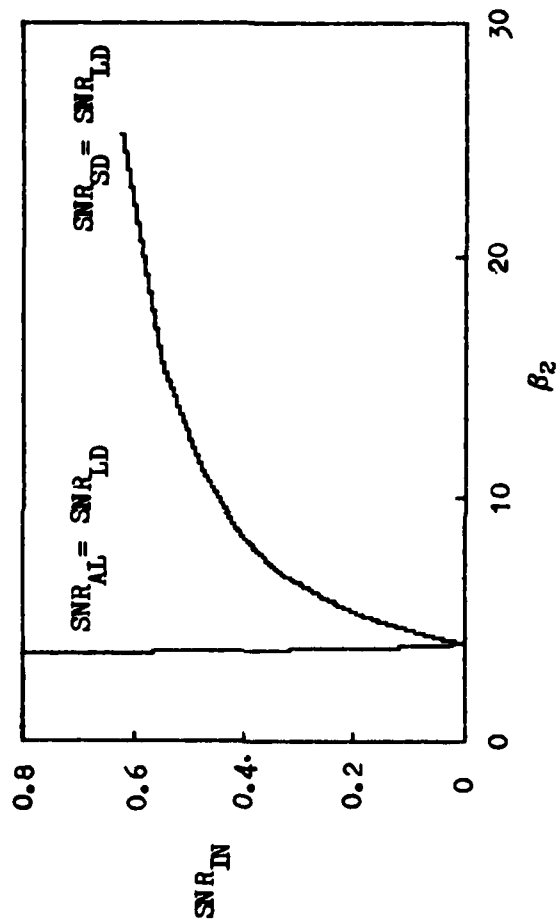


Fig. 9. Contours of $SNR_{SD} = SNR_{LD}$ and $SNR_{AL} = SNR_{LD}$ in the $\beta_2 \times SNR_{IN}$ plane for generalized Gaussian noise.

B. PEARSON VII

The next model considered is the Pearson VII density:

$$F(x) = K(1 + \alpha x^2)^{-m}$$

where

$$K = \frac{\alpha^{1/2} \Gamma(m)}{\Gamma(1/2) \Gamma(m - 1/2)} \quad \alpha = \frac{\beta_2 - 3}{2\mu_2\beta_2} \quad m = \frac{5\beta_2 - 9}{2(\beta_2 - 3)}$$

Rather than exponentially decaying tails as in the previous example, the Pearson VII density has tails which decay algebraically. In the limit as $\beta_2 \rightarrow 3$ the density approaches the Gaussian density.

In this case the incomplete moments I_i (for $i = 0, 1, 2$) are computed numerically. Using these incomplete moments and Eqs. (1-3), the output SNR's of the LD, SD and AL were found for several values of kurtosis, β_2 , and plotted in Figs. (10-13).

At $\beta_2 = 3$, the Gaussian density results, and the performance of the three detectors is described in the last section, and shown in Fig. 5. As β_2 increases, the density becomes heavier tailed and the performance of both the AL and the SD improves relative to that of the LD. As before, the AL consistently outperforms the SD.

In Fig. 14, the contour of equal performance of the SD and LD, $SNR_{SD} = SNR_{LD}$, is shown in the $\beta_2 \times SNR_{IN}$ plane. Considering the results of the last section, it is interesting how poorly the SD performs in this particular noise environment. Comparing Figs. 9 and 14, not only is the SD for Pearson VII noise much more sensitive to increases in SNR_{IN} but the kurtosis must be nearly three times as great (at $SNR_{IN} = 0$) for the SD to outperform the LD.

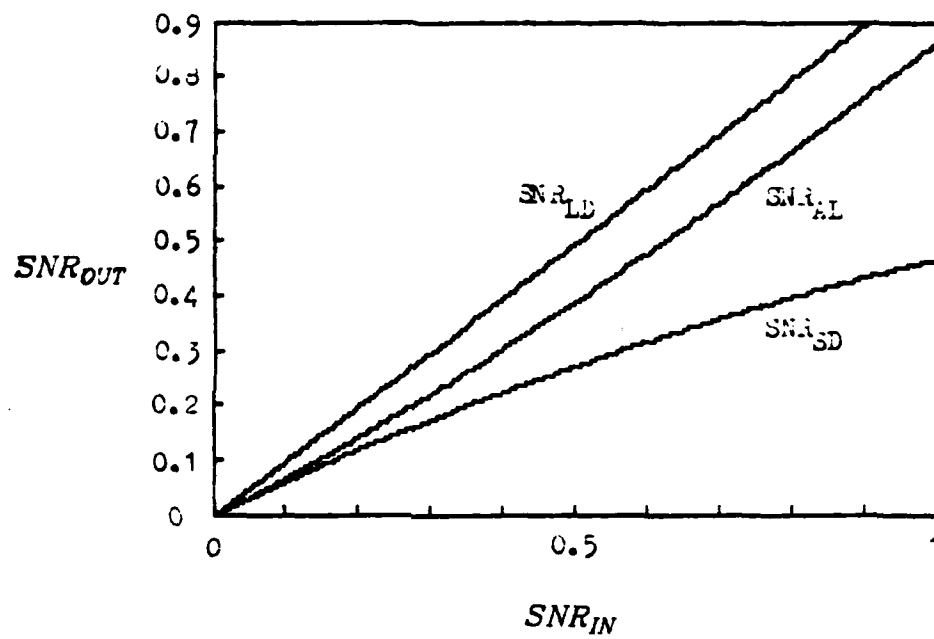


Fig. 10. SNR_{IN} versus SNR_{OUT} for Pearson VII, $\beta_2 = 3.1$.

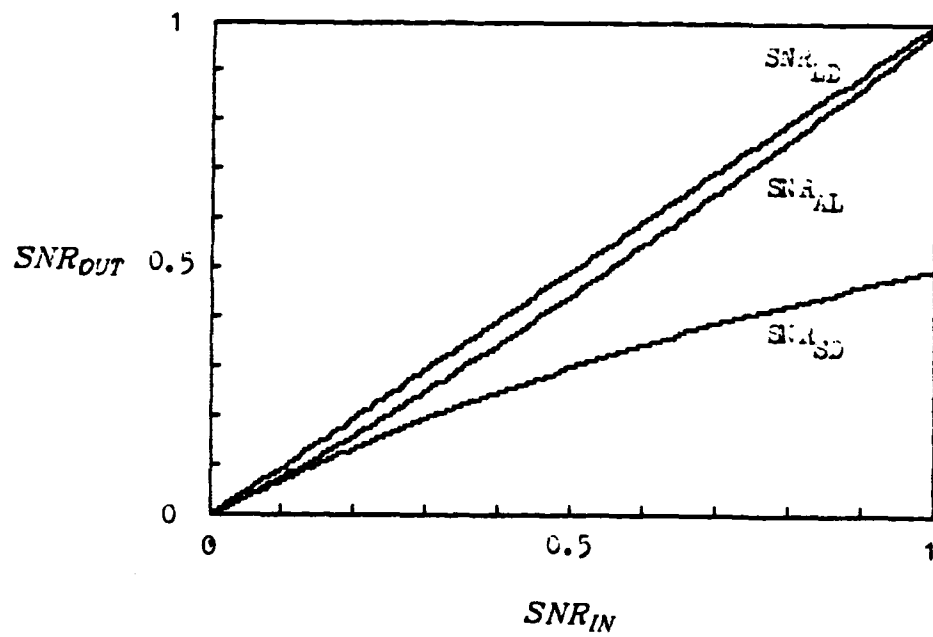


Fig. 11. SNR_{IN} versus SNR_{OUT} for Pearson VII, $\beta_2 = 4.0$.

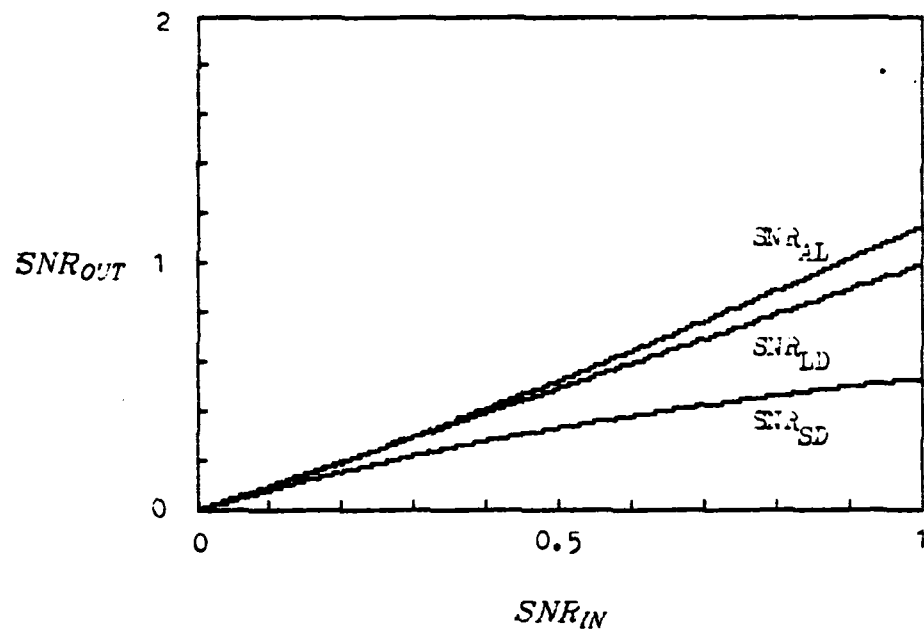


Fig. 12. SNR_{IN} versus SNR_{OUT} for Pearson VII, $\beta_2 = 6.0$.

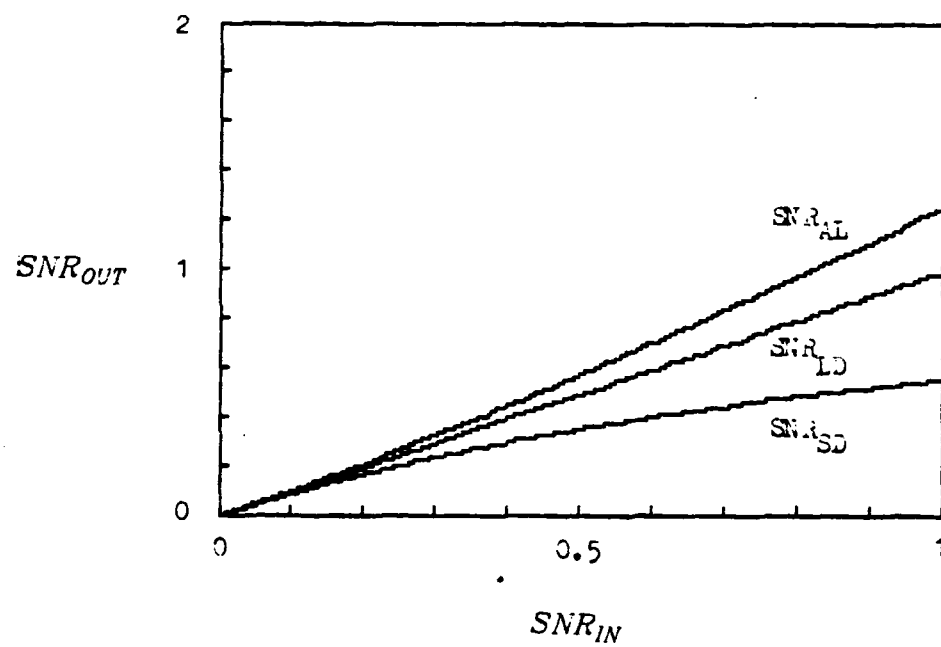


Fig. 13. SNR_{IN} versus SNR_{OUT} for Pearson VII, $\beta_2 = 10$.

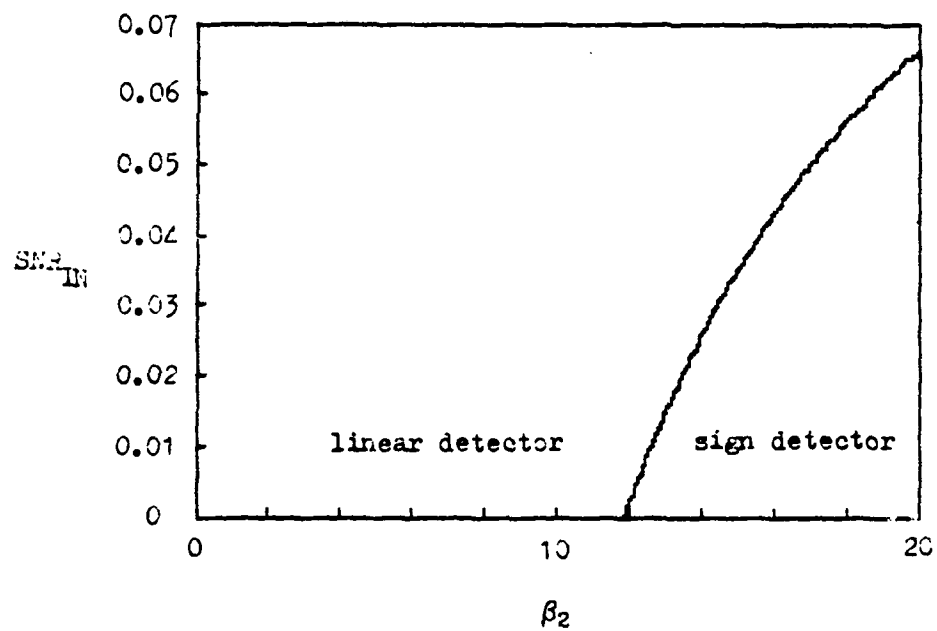


Fig. 14. Contour of $SNR_{SD} = SNR_{LD}$ in $\beta_2 \times SNR_{IN}$ for Pearson VII.

C. MIXTURE MODEL

The next noise model considered is a mixture model. Under certain conditions [4,6,7] a noise process may be well-described by a mixture of the form

$$f(x) = (1 - \epsilon)f_B(x) + \epsilon f_I(x)$$

where f_B describes a background noise process contaminated by bursts of impulsive f_I distributed noise. In the examples which follow, the background noise is assumed Gaussian:

$$f_B(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

The impulsive contaminant is assumed to be Laplace or Gaussian.

1. Gauss-Laplace mixture

The Gauss-Laplace mixture assumes the contaminating impulsive noise to be Laplace:

$$f_I(x) = \frac{\alpha}{2} \exp(-\alpha |x|)$$

The ratio of the variances $\gamma = 2/\alpha^2\sigma^2$ proves to be an important constant.

The incomplete moments, I_i can be computed numerically, and the resulting SNR's can be found from Eqs (1-3). The contours of equal SNR for the linear and sign detector are plotted for several values of γ in Fig. 15. The SD outperforms the LD below the contours, and the LD is superior above. Typically γ will be large and ϵ small corresponding to infrequent large energy bursts. In this case, the sign detector may be of use as shown by the $\gamma = 100$ contour.

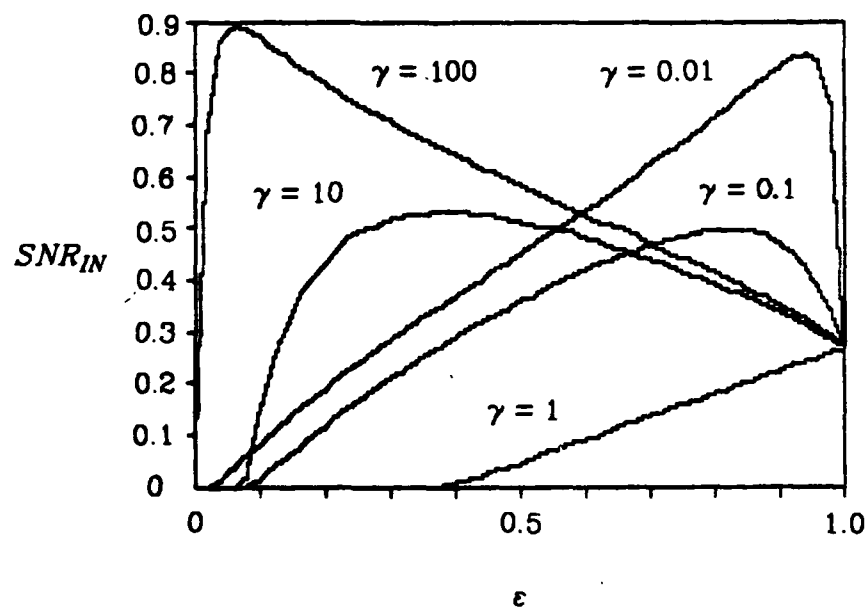


Fig. 15. Contours of $SNR_{SD} = SNR_{LD}$ in $\epsilon \times SNR_{IN}$ for Gauss-Laplace mixture.

2. Gauss-Gauss Mixture

In this case the contaminant is assumed also to be Gaussian, with variance s^2 . The ratio of the variances again proves useful, where $\gamma = s^2/\sigma^2$.

The contours of $SNR_{SD} = SNR_{LD}$ are plotted in Fig. 16 for several values of γ . The SD outperforms the LD below the contours as in the last example. It is interesting to note that while the LD is optimal at $\epsilon = 0$ or $\epsilon = 1$ since f is then purely Gaussian, the performance of the SD may in fact surpass that of the LD for $0 < \epsilon < 1$. As stated above, γ will frequently be large and ϵ small; in this case the SD would appear to perform quite well.

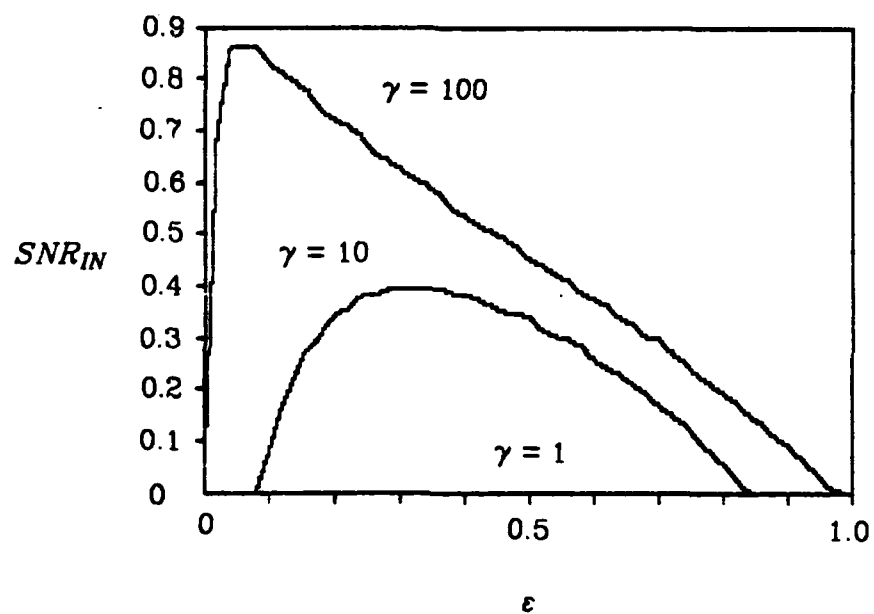


Fig. 16. Contours of $SNR_{SD} = SNR_{LD}$ in $\epsilon \times SNR_{IN}$ for Gauss-Gauss mixture.

CONCLUSIONS

Care should be taken in applying locally optimal results, since the locally optimal detector is designed for zero input SNR. As shown in the Laplace example, the performance can deteriorate rather quickly as SNR_{IN} increases, so quickly in fact that the linear detector may well outperform the locally optimal detector for SNR_{IN} fairly small.

The comparison of generalized Gaussian and Pearson VII results indicates the importance of correctly choosing a noise model. The sign detector performs considerably less well, particularly in sensitivity to increases in SNR_{IN} , for Pearson VII distributed noise. From these results, we can conclude that even two densities which have equal moments up to fourth order can yield decidedly different performance from the same detectors.

In general the performance of the amplifier limiter is quite good for the noise models used. The sign detector on the other hand is at times disappointing. One case in which it does appear to work well was for the mixture model with infrequent but high energy impulsive bursts (γ large, ϵ small). Not only does the sign detector outperform the linear detector, but it is relatively insensitive to the form of the contaminant; similar performance is seen for both Laplace and Gaussian contaminant.

REFERENCES

1. J.H. Miller and J.B. Thomas, "Numerical Results on the Convergence of Relative Efficiencies," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-11, no. 2, March 1975.
2. A.D. Spaulding, "Locally Optimum and Suboptimum Detector Performance in Non-Gaussian Noise," *Proc. of IEEE ICC 1982*, vol. 1, pp. 2H.2.1-7, June 1982.
3. J.H. Miller and J.B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," *IEEE Trans. on Inform. Theory*, vol. IT-18, no. 2, pp. 241-250, March 1977.
4. J.H. Miller and J.B. Thomas, "Robust Detectors for signals in Non-Gaussian Noise," *IEEE Trans. on Comm.* vol. COMM-25, no. 7, pp. 686-690, July 1977.
5. N. H. Lu and B. A. Eisenstein, "Detection of Weak Signals in Non-Gaussian Noise," *IEEE Trans. on Inform. Theory*, vol. IT-27, no. 6, pp. 755-771, November 1981.
6. R.D. Martin and S.C. Schwartz, "Robust detection of a known Signal in Nearly Gaussian Noise," *IEEE Trans. on Inform. Theory*, vol. IT-17, pp.50-56, January 1971.
7. W.J. Richter, Jr. and T.I. Smits, "Signal Design and Error Rate for an Impulsive Noise Channel," *IEEE Trans. on Comm. Tech.*, vol. COMM-19, pp.446-458, August 1971.

CHAPTER 3 - DETECTOR DESIGN USING A DENSITY FIT TO NON-GAUSSIAN NOISE

INTRODUCTION

The locally optimal detector has been suggested for the detection of small known signals in noise. The detector consists of a nonlinearity followed by a linear filter and a threshold comparator (Fig. 1). In this chapter, we consider the selection of suitable nonlinearities when a complete statistical description of the noise is not available.

In practice, the realization of the locally optimal nonlinearity may present some difficulties. The functional form may be unwieldy to implement. Especially for dependent noise, a multivariate density may not be known with precision and it may be difficult to estimate. The noise may not be stationary, making it desirable for the nonlinearity to adapt to changing noise statistics.

For independent, identically distributed (iid) noise, the nonlinearity is memoryless, and suboptimal approximations to this zero memory nonlinearity (ZNL) have received some attention. Kassam and Lim [4] have addressed the problem of finding an optimal quantized version of the ZNL. Miller and Thomas [2,3] use hard limiters, amplifier limiters, multilevel, and piecewise linear suboptimal nonlinearities in place of the optimal ZNL. Frequently these approximations have parameters which are chosen to maximize the detector efficacy. Unfortunately, this usually proves to be a numerical problem involving considerable computations. It is not always clear how to choose or optimize these nonlinearities when

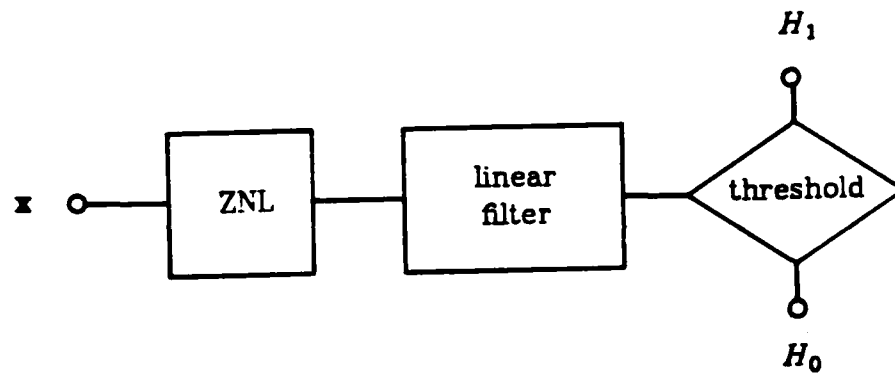


Fig. 1. Typical detector structure.

the noise density is not known exactly.

A second approach is used in this chapter. An approximate non-linearity with a desirable functional form and one or more free parameters is chosen. A family of noise densities with the same free parameters can then be found for which the nonlinearity is optimal. The parameters are evaluated by fitting a member of this family of densities to the noise.

LOCALLY OPTIMAL DETECTOR

The problem of detecting a known signal $\mathbf{s}$ in noise $\mathbf{n}$ can be expressed as an hypothesis H_0 and an alternative H_1 :

$$H_0: \mathbf{x} = \mathbf{n}$$

$$H_1: \mathbf{x} = \mathbf{n} + \vartheta \mathbf{s} \quad \vartheta > 0$$

Under H_0 , the observation vector $\mathbf{x}$ consists of noise $\mathbf{n}$ with density $f(\mathbf{n})$, and under H_1 the observation consists of noise plus known signal $\mathbf{s}$ with amplitude ϑ . The detector can be written as a real-valued functional $T(\mathbf{x})$ on the observation, followed by a threshold comparator to decide for H_0 or H_1 .

When the magnitude ϑ of the signal is unknown, but assumed to be nearly zero, the efficacy

$$J(T) = \lim_{\vartheta \rightarrow 0} \frac{\left[\frac{\partial}{\partial \vartheta} E_{\vartheta} T \right]^2}{V_{\vartheta} T} = \frac{\left[\int T f' d\mathbf{x} \right]^2}{\int T^2 f d\mathbf{x} - \left[\int T f d\mathbf{x} \right]^2} \quad (1)$$

is frequently used as a measure of detector performance [1-8]. It is well known [14] that the statistic T_{lo} which maximizes the efficacy is the ϑ derivation of the log likelihood function at $\vartheta=0$:

$$T_{lo}(\mathbf{x}) = \lim_{\vartheta \rightarrow 0} \frac{\partial}{\partial \vartheta} \ln[f(\mathbf{x}, \vartheta)]$$

known as the locally optimal detector. For a signal in additive noise with amplitude ϑ , the statistic consists of a nonlinearity

$$g_{lo}(\mathbf{x}) = -\nabla f(\mathbf{x}) / f(\mathbf{x})$$

followed by a correlator

$$T_{lo}(\mathbf{x}) = -\frac{\nabla f(\mathbf{x}) \cdot \mathbf{s}}{f(\mathbf{x})} = g_{lo}(\mathbf{x}) \cdot \mathbf{s}$$

where " ∇ " is the gradient and " $\cdot$ " is the vector dot product. The statistic reduces for iid noise to a zero memory nonlinearity (ZNL)

$$g_{lo}(x) = -f'(x) / f(x) \quad (2)$$

followed by a correlator:

$$T_{lo}(\mathbf{x}) = \sum_{i=1}^N -s_i f'(x_i) / f(x_i) = \sum_{i=1}^N s_i g_{lo}(x_i)$$

with efficacy

$$J(T_{lo}) = \int (f'^2 / f) dx$$

Suppose a ZNL g is chosen with free parameters ψ . Then Eq.(2) can be viewed as a differential equation which can be solved under certain regularity conditions. The differential equation is written as

$$g(x) = -f'(x) / f(x)$$

Thus

$$\int g(x) dx = -\int d \ln(f(x)) = -\ln(f(x)) + C$$

and so

$$f(x) = K \exp\left\{-\int g(x) dx\right\}$$

A suitable suboptimal ZNL can now be selected, and the corresponding family of solution densities found for which the ZNL is optimal. Not only should the ZNL have a relatively simple form, but it must lead to a

reasonable set of densities. Since a member of the solution family is to be fitted to the observed noise, it should be general enough to provide a relatively good fit, and the densities should have parameters that are not too difficult to estimate.

PEARSON APPROXIMATION TO THE ZNL

As an approximate ZNL consider a rational function:

$$g(x) = -N(x)/D(x) \quad (3)$$

where N and D are polynomials in x of degree n and d respectively. This is a general form of the Pearson differential equation.

Choosing $n=1$ and $d=2$ yields the classical Pearson family as solutions of the differential equation:

$$\frac{f'(x)}{f(x)} = \frac{a+x}{b_0+b_1x+b_2x^2} \quad (4)$$

The classical Pearson family includes as special cases the Gaussian, Cauchy, t , F , χ^2 , uniform, gamma, exponential and beta densities. Originally proposed by Karl Pearson [11] as a tool for fitting densities to data, it has continued to attract considerable attention.

Solutions of the differential Eq.(4) depend on the roots of the denominator D . There are three possibilities:

1. real roots, same sign
2. real roots, different sign
3. complex conjugate roots

These correspond to the three main Pearson types VI, I and IV respectively. Transition types result when parameters in the main types approach a particular set of limiting values. The Pearson types and

associated common densities are summarized in Table 1.

Maximum likelihood (ML) estimators of the parameters in the Pearson family depend on the density type which is the best fitted to the observations. A simpler approach is to use the method of moments (MM) which requires estimates of the first four noise moments. For nearly Gaussian noise ($\beta_1 \leq 0.1$ and $2.62 \leq \beta_2 \leq 3.42$), the ratio of the variance of MM to ML estimates exceeds 80% [12,13].

The coefficients of the Pearson system can be found in terms of the first four moments (or skewness and kurtosis) if they exist [9,10,Appendix]. Define variance, skewness and kurtosis:

$$\sigma^2 = \mu_2, \quad \beta_1 = \mu_3 / \mu_2^{3/2}, \quad \beta_2 = \mu_4 / \mu_2^2$$

In terms of these quantities the coefficients are:

$$-a = b_1 = \sigma \beta_1^{1/2} (\beta_2 + 3) / D$$

$$b_0 = -\sigma^2 (4\beta_2 - 3\beta_1) / D$$

$$b_2 = (2\beta_2 - 3\beta_1 - 6) / D$$

with the denominator D given by

$$D = 10\beta_2 - 12\beta_1 - 18$$

As an aid to selecting a Pearson type density given the sample values of β_1 and β_2 , Pearson constructed a chart showing the regions occupied by each of the Pearson types in the $\beta_1 \times \beta_2$ plane (Fig. 2).

Although the Pearson family consists of both skewed and symmetric densities, only the symmetric ($\beta_1=0$) are considered below. The symmetric solutions can be seen from Fig. 2 to include the Pearson types II, VII and normal densities. These can be written in a single function form:

$$f(x) = K(1 - \alpha x^2)^m$$

TYPE	EQUATION	RANGE	REMARKS
MAIN			
I	$K(1+x/a_1)^{\nu a_1}(1-x/a_2)^{\nu a_2}$	$[-a_1, a_2]$	beta, F
IV	$K(1+x^2/a^2)^{-m}e^{-\nu \ln^{-1}(x/a)}$	$(-\infty, \infty)$	
VI	$K(x-a)^{m_1}x^{-m_1}$	$[a, \infty)$	Pareto, F, Lomax
TRANSITION			
NORMAL	$K \exp[-x^2/2\sigma^2]$	$(-\infty, \infty)$	
II(I)	$K(1-x^2/a^2)^m$	$[-a, a]$	beta
VII(IV)	$K(1+x^2/a^2)^{-m}$	$(-\infty, \infty)$	t, Cauchy
III	$K(1+x/a)^{\nu a}e^{-\gamma a}$	$[-a, \infty)$	χ^2 , gamma
V	$Kx^{-m}e^{-a/x}$	$[0, \infty)$	inverse Gaussian
VIII(I)	$K(1+x/a)^{-m}$	$[-a, 0]$	
IX(I)	$K(1+x/a)^m$	$[-a, 0]$	
X(III)	$Ke^{-x/\sigma}$	$[0, \infty)$	exponential
XI(VI)	Kx^{-m}	$[a, \infty)$	
XII(I)	$K(x+a_1)^m(x-a_2)^{-m}$	$[-a_1, a_2]$	

Table 1. Pearson Family.

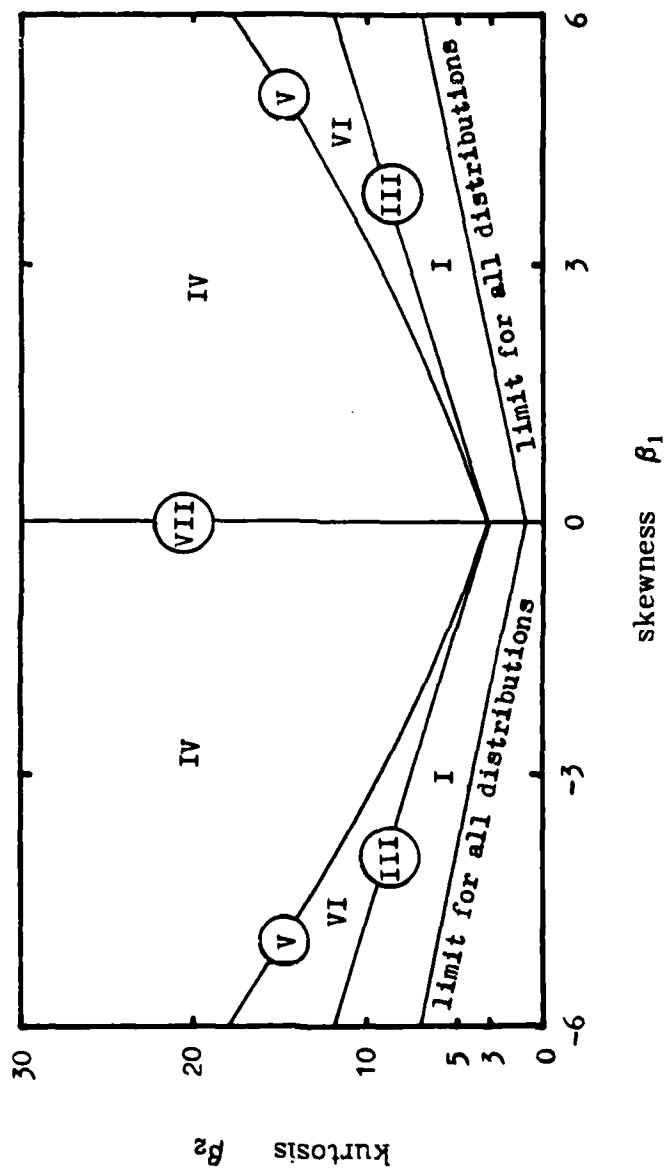


Fig. 2. Pearson densities in $\beta_1 \times \beta_2$ plane.

The constant K can be written in terms of the incomplete beta function $B(\dots)$:

$$K = \begin{cases} \alpha^{\frac{1}{2}} B(\frac{1}{2}, m+1) & m > -1 \quad II \\ \alpha^{\frac{1}{2}} B(\frac{1}{2}, -m-\frac{1}{2}) & m < 1 \quad VII \end{cases}$$

When the first four moments exist:

$$\begin{aligned} \alpha &= (3-\beta_2)/2\sigma^2\beta_2 \\ m &= (5\beta_2-9)/2(3-\beta_2) \end{aligned} \quad (5)$$

Note that for $\beta_2 < 3$, $f(x)$ has finite support, $|x| < \alpha^{-\frac{1}{2}}$. For $\beta_2 \geq 3$, $f(x)$ has infinite support, and at $\beta_2 = 3$, the density is normal. Fig. 3 shows $f(x)$ for several values of β_2 , and properties of $f(x)$ are summarized in Table 2.

Restricting the density to be symmetric simplifies the ZNL since, from Eq.(11), when $\beta_1 = 0$, then $a = b_1 = 0$ and the resulting ZNL is odd symmetric. The locally optimal ZNL is found from Eq.(2) to be:

$$g_{lo}(x) = \frac{x}{1-\alpha x^2} \quad (6)$$

with range $-\infty < x < \infty$ for $\alpha \leq 0$ and $|x| \leq \alpha^{-\frac{1}{2}}$ for $\alpha > 0$. The parameter α in this nonlinearity can be estimated from the variance and kurtosis of the observed noise using the expression in Eq.(5). The II/VII ZNL, shown in Fig. 4 for several values of β_2 , can take three important forms: expander ($\beta_2 < 3$), linear ($\beta_2 = 3$), and contractor ($\beta_2 > 3$). A fourth type, the limiter, is not included, causing some problems when fitting to a density with exponential tails, for which a limiter is optimal.

The Asymptotic Relative Efficiency (ARE) is defined, under certain regularity conditions, to be the ratio of the efficacies of two detectors:

$$ARE_{1,2} = J(T_1)/J(T_2)$$

and is a frequently used measure of relative performance of the detectors. A linear detector for any density is easily seen from Eq.(1) to have the efficacy

$$J(T_{ld}) = 1/\mu_2$$

The efficacy of the locally optimal detector of the Pearson II/VII density is

$$J(T_{lo}) = \frac{5(\beta_2-3)(2\beta_2-3)}{\mu_2\beta_2(7\beta_2-15)}$$

The $ARE_{lo,ld}$ is plotted in Fig. 5. The $ARE_{lo,ld}$ is a minimum of unity at $\beta_2 = 3$ corresponding to the Gaussian density. As β_2 decreases, resulting in densities with increasingly lighter tails, the ARE increases asymptotically to infinity at $\beta_2 = 15/7$. For heavier tailed densities, the improvement is less marked. As $\beta_2 \rightarrow \infty$ the ARE approaches the limiting value $10/7$.

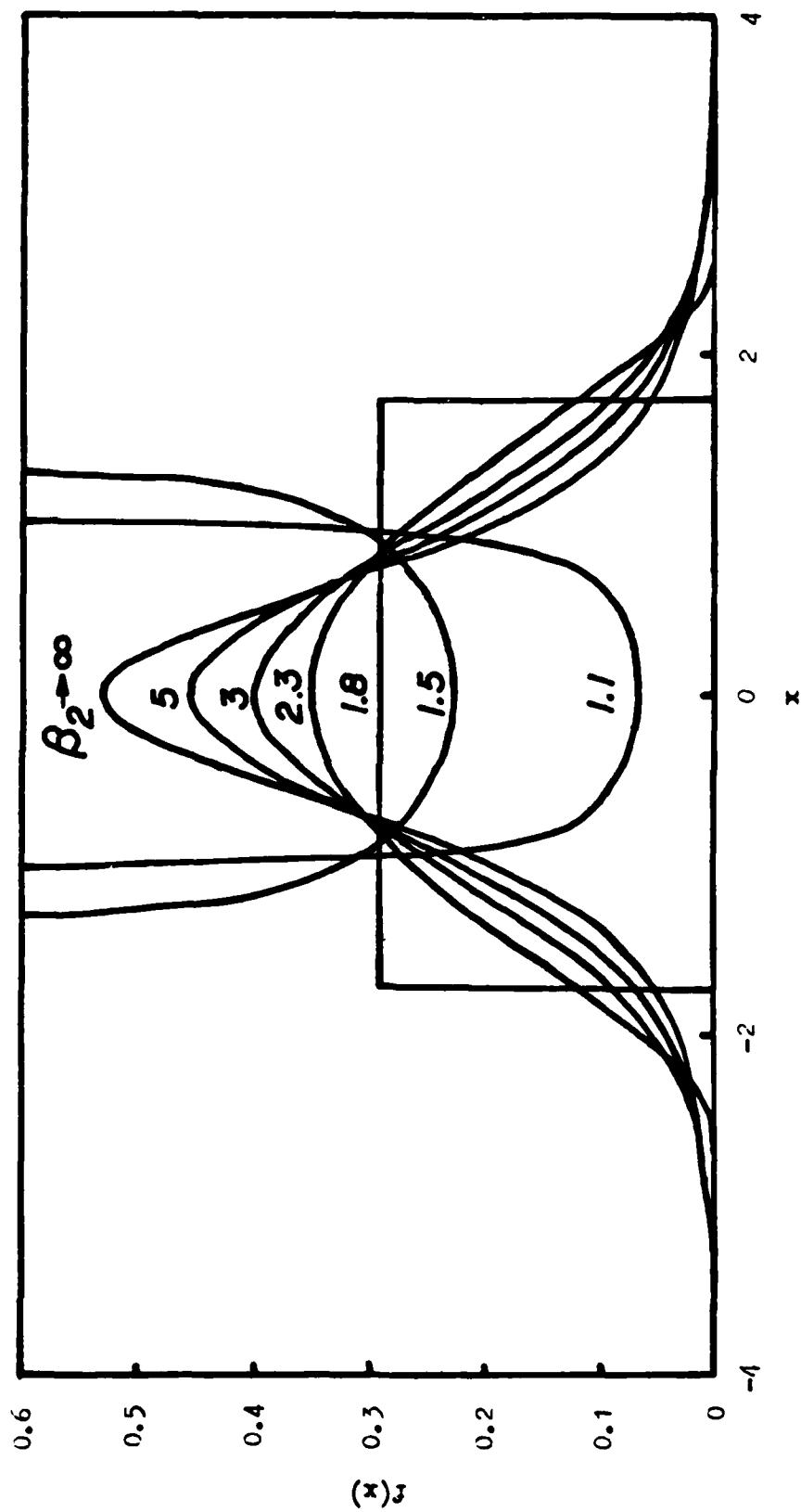


Fig. 3. Pearson II/VII density

D roots	TYPE II			TYPE VII	
	Real, same magnitude, opposite sign	uniform	opposite sign	∞	complex conjugate
Shape				normal	1
β_j	1	(1,1,8)	1.8	(1.8,3)	(3, ∞)
m	-1	(-1,0)	0	(0, ∞)	($-\infty$, -2.5)
$\sigma^2 \alpha$	1	(1,1/3)	1/3	(1/3,0)	(0, -1/2)

Table 2. Pearson Type II/VII density

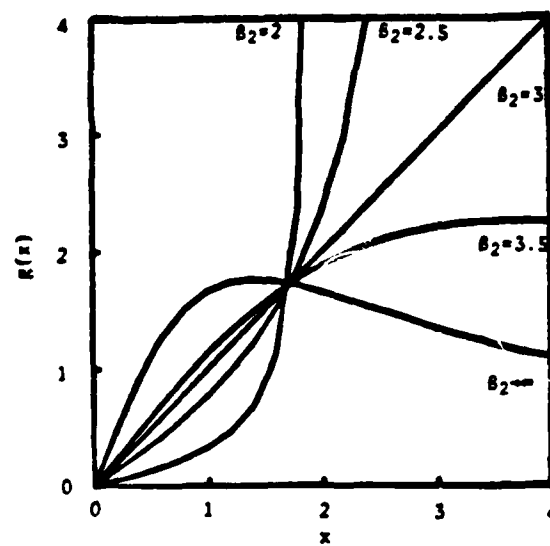


Fig. 4. Pearson locally optimal ZNL's.

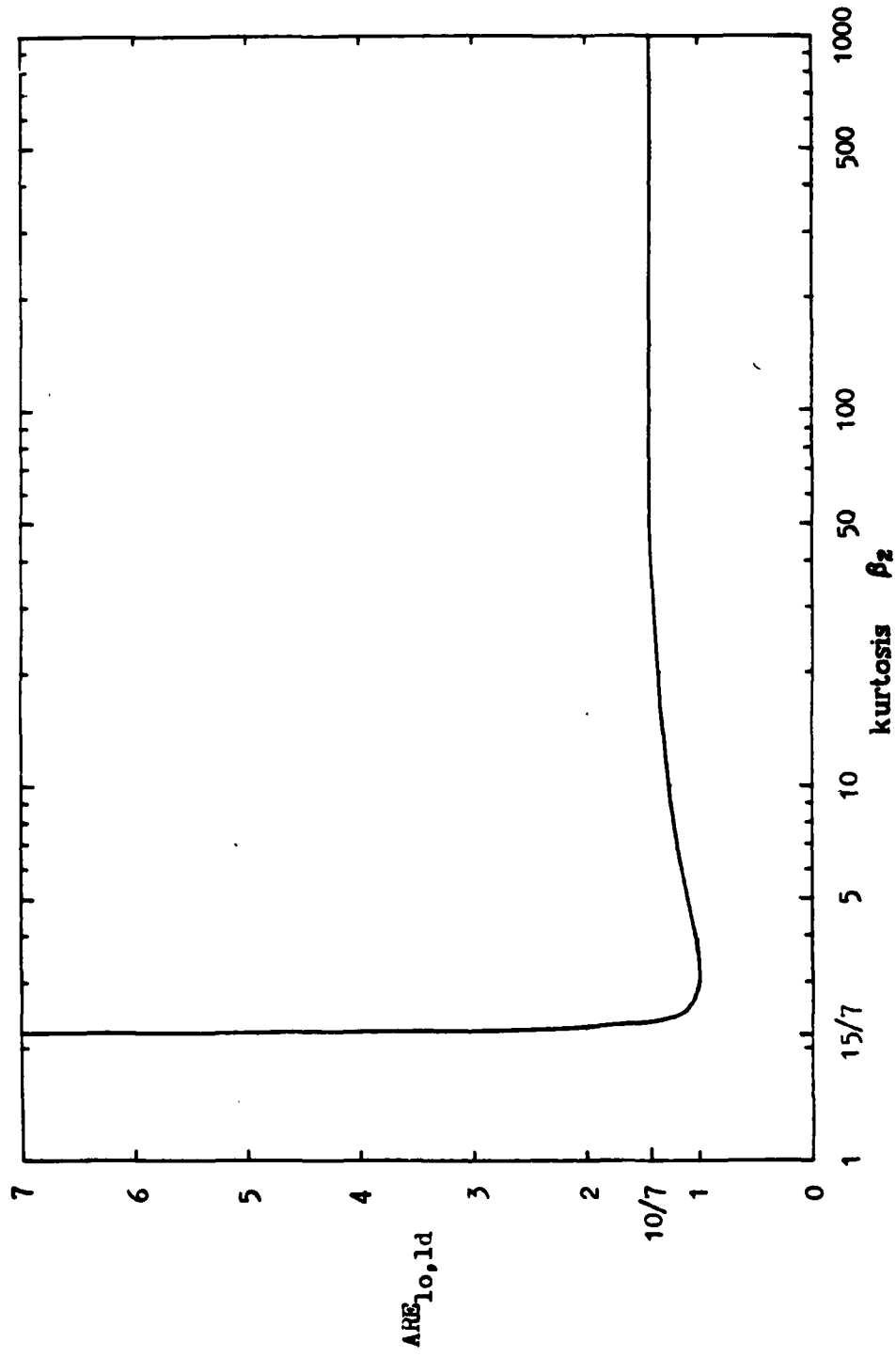


Fig. 5. $ARE_{10,ld}$ for Pearson II/VII density

EXAMPLES

In the last section, a method of finding a suboptimal ZNL using the first four sample moments was developed. In practice we would use only these moments to fit our suboptimal ZNL. The problem of estimating these moments efficiently for an unknown density will not be considered here. Instead, in our examples, we assume two classes of non-Pearson noise densities and use analytically derived moments to fit the Pearson ZNL. Then the efficacies of the suboptimal and optimal detectors are compared as a measure of detector performance. The two classes of densities chosen are the generalized Gaussian and a Gauss-Laplace mixture density.

A. Gauss-Laplace Mixture Noise

The Gauss-Laplace density is given by:

$$f(x) = \frac{(1-\epsilon)}{\sigma^2\sqrt{2\pi}} \exp\left\{\frac{-x^2}{2\sigma^2}\right\} + \frac{\epsilon\vartheta}{2} \exp(-\vartheta|x|)$$

where $0 \leq \epsilon \leq 1$ and $0 < \sigma^2, \vartheta$. One justification of this density is to assume a Gaussian noise with bursts of Laplace (impulsive) noise $100\epsilon\%$ of the time [2]. An additional quantity which proves useful is the ratio of the variances:

$$\gamma = (\vartheta\sigma)^2/2$$

Typically ϵ and γ are assumed small. The moments are:

$$\mu_1 = \mu_3 = 0$$

$$\mu_2 = (1-\epsilon)\sigma^2 + \epsilon(2/\vartheta^2)$$

$$\mu_4 = (1-\epsilon)\sigma^4 + \epsilon(2/\vartheta^2)$$

with skewness and kurtosis:

$$\beta_1 = 0 \quad \beta_2 = 3 \left[\frac{(1-\epsilon)\gamma^2 + 2\epsilon}{[(1-\epsilon)\gamma + \epsilon]^2} \right]$$

The Pearson ZNL is given by Eq.(6):

$$g(x) = x / (1 - \alpha x^2)$$

where α can be found from Eq.(5):

$$\alpha = (\beta_2 - 3) / (2\mu_2\beta_2)$$

Notice that the Pearson ZNL is designed using estimates of the noise moments, and not estimates of the quantities ϵ or γ which are more difficult to compute. Since the moments are relatively simple to estimate, this nonlinearity is by nature adaptive.

$ARE_{g,ld}$ is computed numerically and shown with $ARE_{lo,ld}$ in Figs. 6-10, for several values of γ . Reasonable performance is seen for ϵ small (nearly Gaussian noise) even though the Gaussian noise is contaminated with a non-Pearson density. In addition the performance in all cases equals or exceeds that of the linear detector, $ARE_{g,ld} \geq 1$.

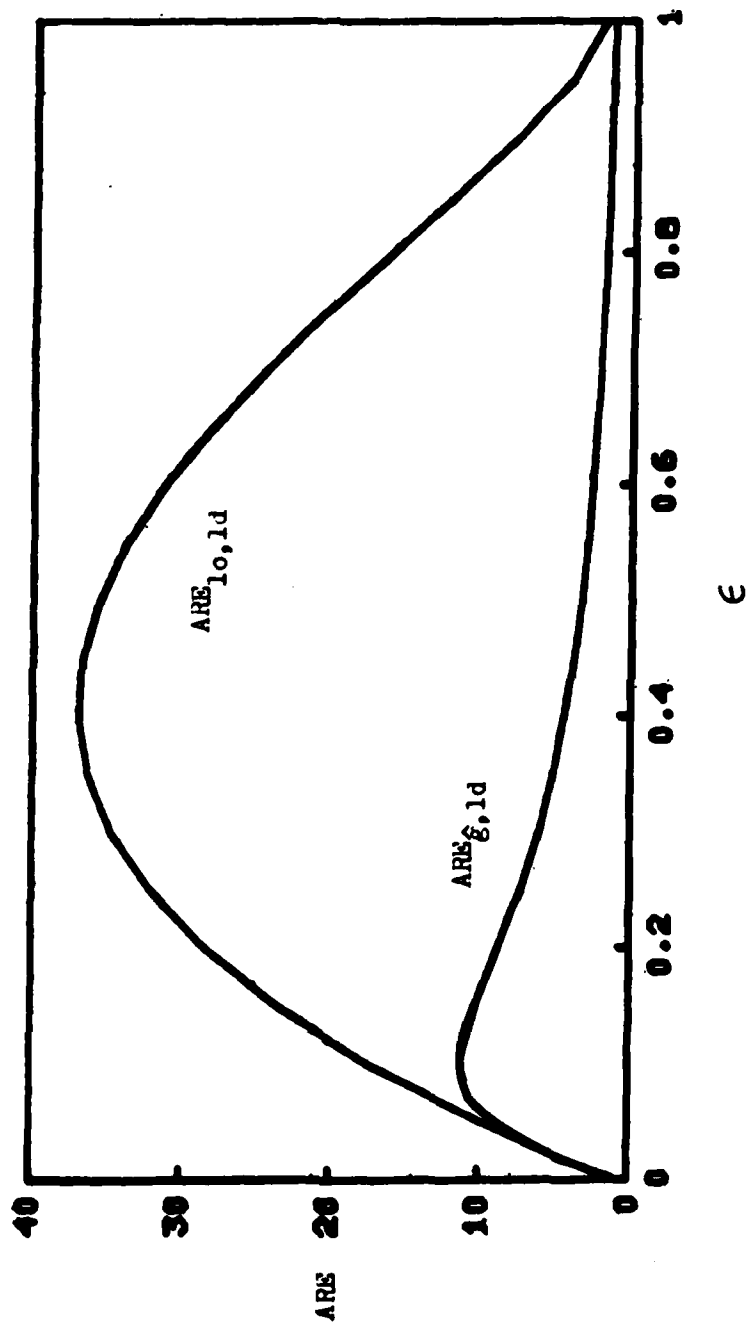


Fig. 6. $ARE_{g,ld}$ and $ARE_{l0,ld}$ versus ϵ for Gauss-Laplace mixture, $\gamma = 0.01$.

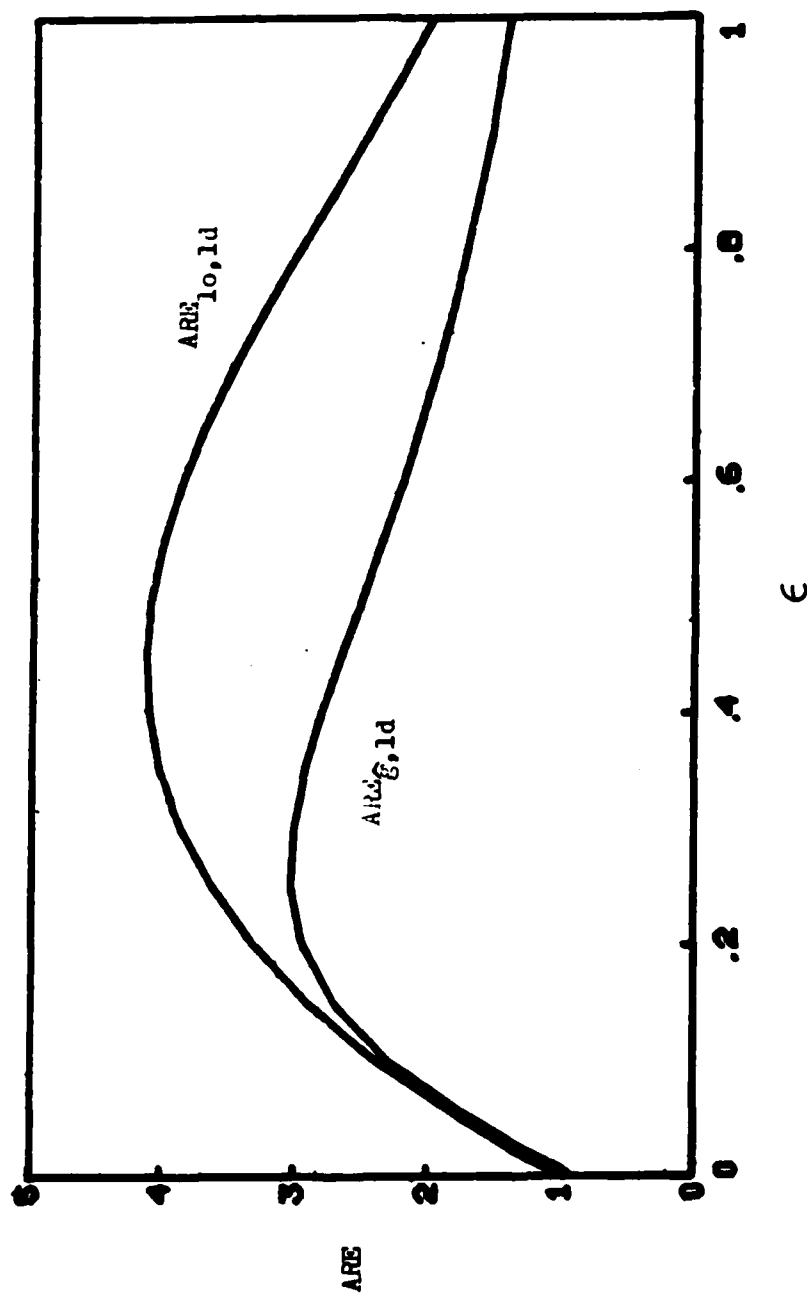


Fig. 7. $ARE_{g,ld}$ and $ARE_{l0,ld}$ versus ϵ for Gauss-Laplace mixture, $\gamma = 0.1$.

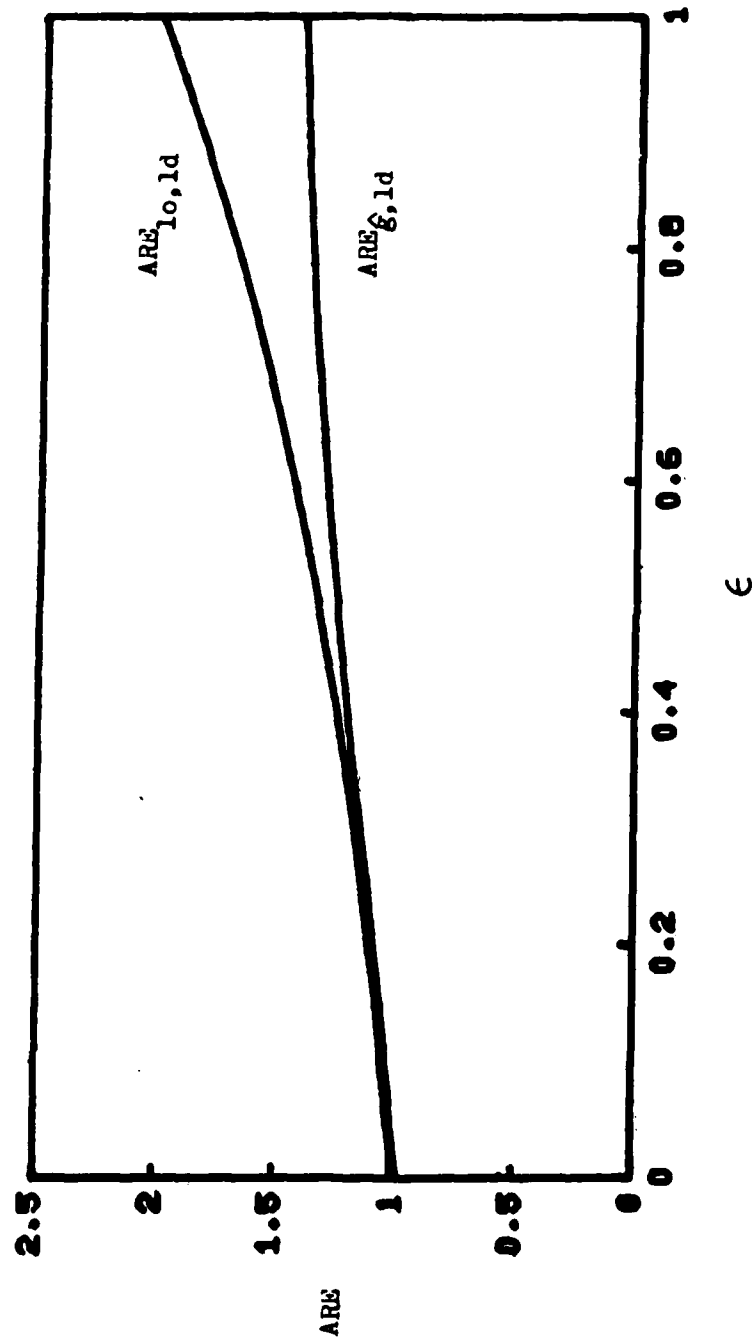


Fig. 8. $ARE_{g,ld}$ and $ARE_{lo,ld}$ versus ϵ for Gauss-Laplace mixture, $\gamma = 1$.

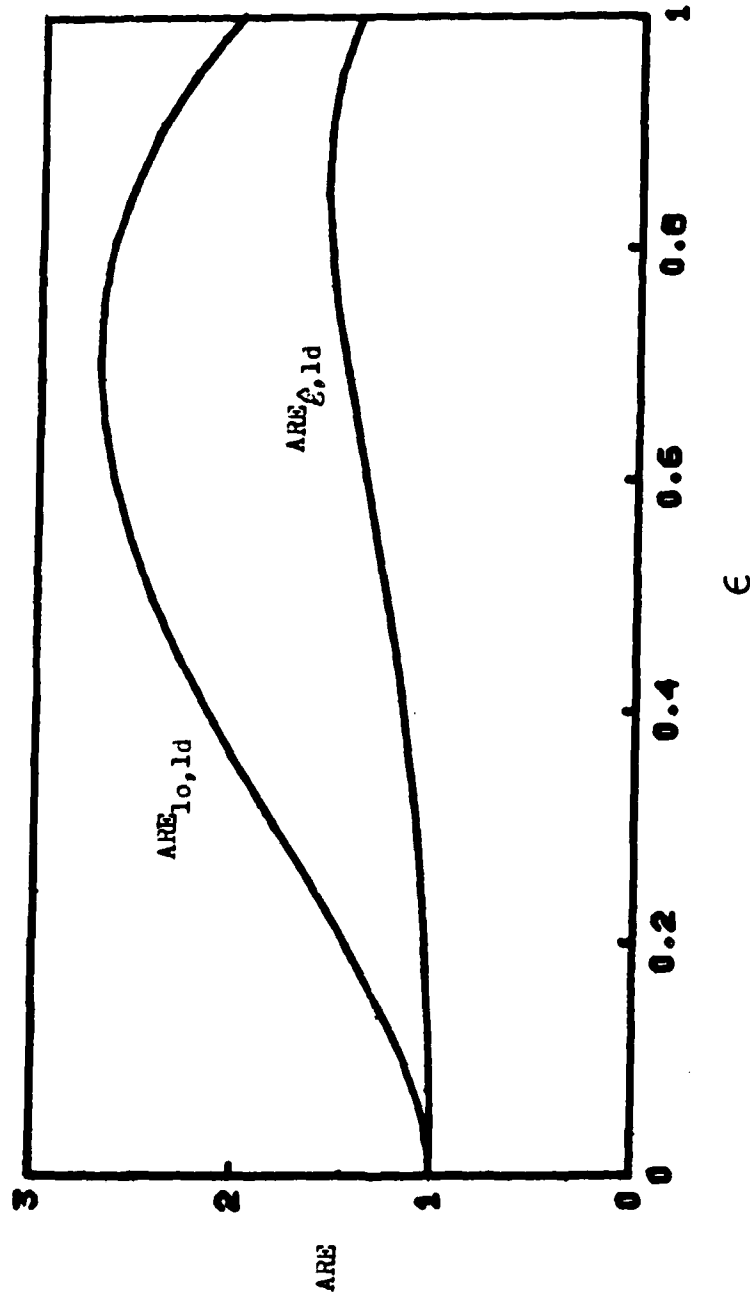


Fig 9. $ARE_{\gamma,ld}$ and $ARE_{\epsilon,ld}$ versus ϵ for Gauss-Laplace mixture, $\gamma = 10$.

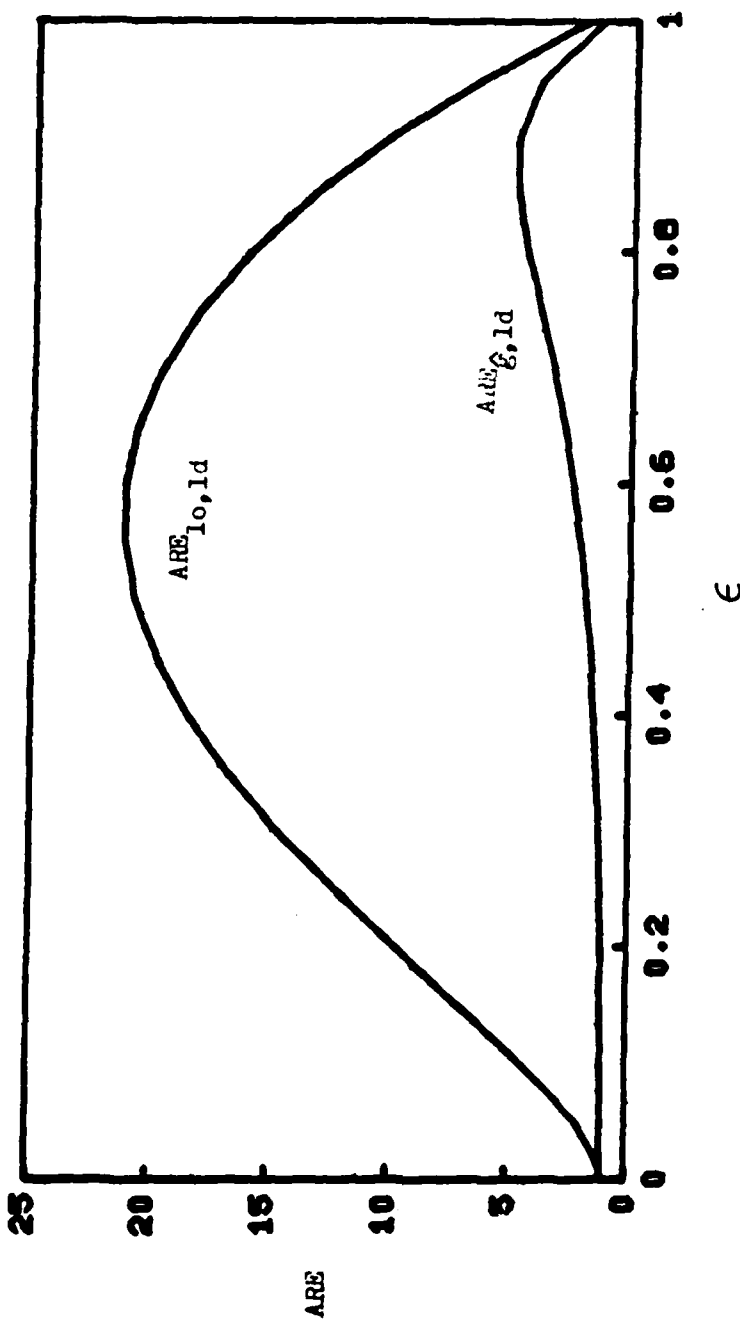


Fig. 10. $ARE_{g,ld}$ and $ARE_{\epsilon,ld}$ versus ϵ for Gauss-Laplace mixture, $\gamma = 100$.

B. Generalized Gaussian Noise

The Gaussian density can be generalized by providing for a variable rate of exponential decay [1]:

$$f(x) = K \exp(-|\zeta x|^c)$$

where

$$\zeta = \left[\frac{\Gamma(3/c)}{\sigma^2 \Gamma(1/c)} \right]^{1/2} \quad \text{and} \quad K = c \zeta / 2 \Gamma(1/c)$$

This class of densities is of interest because it contains both the Gaussian ($c=2$) and Laplace ($c=1$) densities as special cases, and it allows considerable control over the rate of tail decay.

The moments are zero for n odd and for n even:

$$\mu_n = \frac{\sigma^n \Gamma((n+1)/c) \Gamma(1/c)^{\frac{n}{2}-1}}{\Gamma(3/n)^{n/c}}$$

with skewness and kurtosis:

$$\beta_1 = 0 \quad \text{and} \quad \beta_2 = \frac{\Gamma(5/c) \Gamma(1/c)}{\Gamma(3/c)^2}$$

The Pearson ZNL is given in Eq (6) and α in (5). Once again, the moments are estimated, and not the decay parameter c , which is rather difficult to estimate. Since α is easily estimated, the Pearson ZNL is potentially adaptive.

$ARE_{g,ld}$ is computed numerically and shown in Fig. 11. $ARE_{lo,ld}$ and $ARE_{sd,ld}$ are also shown for comparison. $ARE_{sd,ld}$ is the ARE of the sign detector (hard limiter) vs the linear detector. The sign detector is optimal for Laplace noise ($c=1$) and nearly optimal for $c \approx 1$. However, for nearly Gaussian noise $c \approx 2$, the Pearson detector is better. This suggests that a system including both the sign detector and the Pearson detector

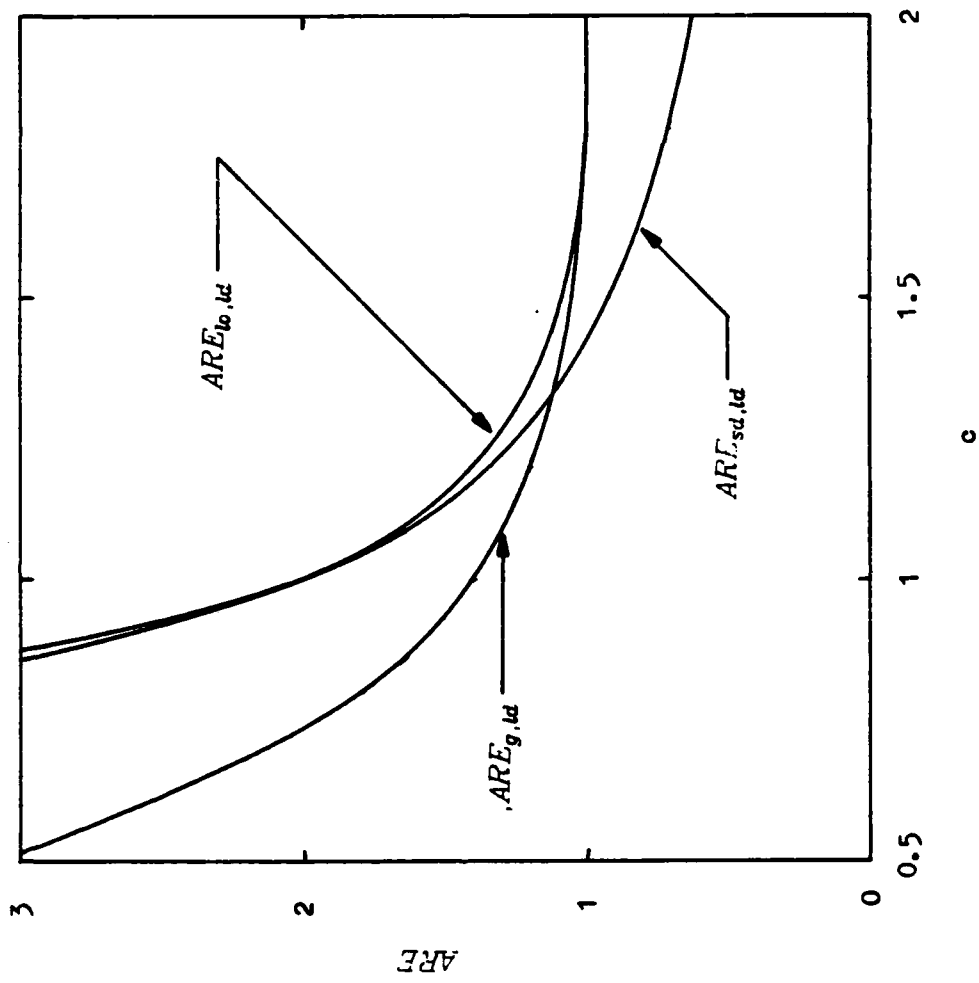


Fig. 11. $ARE_{lo,ld}$, $ARE_{g,ld}$ and $ARE_{sd,ld}$ for generalized Gaussian noise.

might well surpass the performance of either one alone.

MORE COMPLEX ZNL's

Values of n and d can be chosen other than $n=1$ and $d=2$ of the classical family. The resulting families may contain multimodal curves (for instance $n=2$, $d=3$ results in a family including bimodal curves). Higher values of β_1 and β_2 may also be admitted. The major disadvantage of increasing n and d is that high order moments are required. Since the number of parameters to be found is $n+d+1$, using the method of moments requires $n+d+1$ moments. As the moments become increasingly difficult to estimate with increasing order, increasing n and d significantly may not be possible or desirable.

A single or multiple discontinuities can be allowed in the nonlinearity. The result is a family of curves which consist of segments of Pearson curves between the points of discontinuity. For a single discontinuity at $x=0$:

$$\frac{f'(x)}{f(x)} = \frac{(x+a)(1-c+du(x))}{b_0+b_1x+b_2x^2}$$

where

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Solving this differential equation yields a family of Pearson curves consisting of two half curves for $x \geq 0$ and $x < 0$. Each of the two halves are of the same Pearson type but with different parameters. Among the possible solutions, two of particular interest are the Laplace or double exponential and the double Gaussian densities.

As an example where this method of curve fitting fails, consider the nonlinearity on the open interval (x_1, x_2) :

$$g(x) = \alpha \operatorname{sgn}(x) |x|^c$$

Then for $c \neq 1$ and $x_1 < x < x_2$:

$$f(x) = K \exp \left\{ - \left(\frac{\alpha}{c+1} \right) |x|^{c+1} \right\}$$

This density is commonly called the generalized Gaussian with Laplace ($c=0$) and Gaussian ($c=1$) as special cases. Unfortunately there is no good method for estimating the exponential parameter c , and thus the ZNL cannot be fitted in the manner used above.

Several multivariate extensions to the Pearson family have been considered [10,15]. Following the same line of reasoning which led to the univariate Pearson family, K. Pearson wrote the difference equations describing the bivariate hypergeometric density as a pair of differential equations:

$$\frac{\partial f(x,y)}{\partial x} = \frac{\text{cubic in } x,y}{\text{quartic in } x,y} f(x,y) \quad (7)$$

and

$$\frac{\partial f(x,y)}{\partial y} = \frac{\text{another cubic in } x,y}{\text{quartic in } x,y} f(x,y) \quad (8)$$

Other nonlinearities are possible, but all suffer from the same drawbacks. It is extremely tedious to solve the simultaneous equations if the method of moments is used, and the number of coefficients needed for a complete fit increases extremely rapidly. In the bivariate case alone, there are 15 moments up to fourth order [15]:

1. 1-total mass.
2. 2-position of mean.

3. 2-variate standard deviations.
4. 1-coefficient of correlation.
5. 4-marginal β 's ($\beta_1, \beta_2, \beta_1', \beta_2'$).
6. 2-third order product moment coefficients.
7. 3-fourth order product moment coefficients.

Thus a complete fit up to fourth order moments requires 15 coefficients in the nonlinearities. Since there are fewer than 15 in Eqs.(7-8), and in fact in most reasonable nonlinearities, some relationship among the higher moments must be tolerated to reduce the number of coefficients required. It then becomes less clear that the fit will be a good one.

CONCLUSION

The optimal ZNL in a locally optimal detector may be cumbersome to implement or even impossible to find when noise statistics are not known exactly. It is possible to choose a practical form for the nonlinearity, and solve the locally optimal differential equation yielding a family of solution densities for which the nonlinearity is optimal. One of these densities can then be fitted to the noise, and the corresponding detector used. If the family of densities can provide a good fit to the noise, it is likely that the detector so derived will have nearly optimal performance.

Choosing a rational function for the ZNL in a locally optimal detector results in a particularly useful family of solution densities, the Pearson family. Not only does the Pearson family contain many common densities, including the Gaussian, but for nearly Gaussian noise the method of moments can be used efficiently to compute the coefficients of the ZNL. This method produces a detector which performs at least as well as the

linear detector for the two classes of densities considered, and nearly as well as the optimal detectors for nearly Gaussian noise.

APPENDIX - Method of Moments for Pearson Family

The coefficients of the Pearson system can be found in terms of the first four moments if they exist [9,10]. Cross multiply the Pearson differential equation, Eq.(12), multiply both sides by x^n and integrate to obtain

$$\int x^n (b_0 + b_1 x + b_2 x^2) f'(x) dx = \int x^n (a + x) f(x) dx$$

The left side can be integrated by parts:

$$\begin{aligned} & x^n (b_0 + b_1 x + b_2 x^2) \int f'(x) dx - \\ & \int (n b_0 x^{n-1} + (n+1) b_1 x^n + (n+2) b_2 x^{n+1}) f(x) dx \\ & = a \int x^n f(x) dx + \int x^{n+1} f(x) dx \end{aligned}$$

but $\int f'(x) dx = 0$, and μ_i' is the i -th moment about zero. This gives four equations (for $n=0, 1, 2, 3$):

$$-\mu_{n+1}' = a \mu_n' + b_0 n \mu_{n-1}' + b_1 (n+1) \mu_n' + b_2 (n+2) \mu_{n+1}'$$

Shift the moments to moments about the mean by setting $\mu_1 = 0$ and $\mu_i' = \mu_i$ for $i = 2, 3, 4$. Thus

$$\begin{aligned} b_1 + a &= 0 \\ b_0 + 3\mu_2 b_2 &= -\mu_2 \\ 3\mu_2 b_1 + 4\mu_3 b_2 + \mu_2 a &= -\mu_3 \\ 3\mu_2 b_0 + 4\mu_3 b_1 + 5\mu_4 b_2 + \mu_3 a &= -\mu_4 \end{aligned}$$

Solving gives:

$$\begin{aligned} -a &= b_1 = \mu_3(\mu_4 + 3\mu_2^2) / D \\ b_0 &= \mu_2(3\mu_3^2 - 4\mu_2\mu_4) / D \\ b_2 &= (-2\mu_2\mu_4 + 3\mu_3^2 + 6\mu_2^3) / D \\ D &= 10\mu_2\mu_4 - 18\mu_2^3 - 12\mu_3^2 \end{aligned}$$

Define variance, skewness and kurtosis:

$$\sigma^2 = \mu_2, \quad \beta_1 = \mu_3^2 / \mu_2^3, \quad \beta_2 = \mu_4 / \mu_2^2$$

In terms of these quantities the coefficients are:

$$-a = b_1 = \sigma \beta_1^{1/2} (\beta_2 + 3) / D$$

$$b_0 = -\sigma^2 (4\beta_2 - 3\beta_1) / D$$

$$b_2 = (2\beta_2 - 3\beta_1 - 6) / D$$

$$D = 10\beta_2 - 12\beta_1 - 18$$

REFERENCES

- [1]. J.H. Miller and J.B. Thomas, "Detectors for discrete time signals in non-Gaussian noise," *IEEE Trans. on Info. Theory*, vol. IT-18, pp. 241-250, March 1972.
- [2]. J.H. Miller and J.B. Thomas, "The detection of signals in impulsive noise modeled as a mixture process," *IEEE Trans. Comm.*, vol. COM-24, pp. 559-562, May 1976.
- [3]. J.H. Miller and J.B. Thomas, "Robust detectors for signals in non-Gaussian noise," *IEEE Trans. Comm.*, vol. COM-25, pp. 686-690, July 1977.
- [4]. S.A. Kassam and T.L. Lim, "Coefficient and data quantization in matched filters for detection," *IEEE Trans. Comm.*, vol. COM-26, pp. 124-127, Jan. 1978.
- [5]. J.J. Sheehy, "Optimum detection of Signals in non-Gaussian Noise," *J. Acoust. Soc. Amer.*, Vol. 63, Jan. 1978, pp. 81-90.
- [6]. H.V. Poor and J.B. Thomas, "Memoryless Discrete-Time Detection of a Constant Signal in m-Dependent noise," *IEEE Trans. on Inform. Theory*, Vol. IT-25, No. 1, Jan. 1979, pp.54-61.
- [7]. H.V. Poor and J.B. Thomas, "Locally Optimum Detection of Discrete-Time and Stochastic Signals in non-Gaussian Noise," *J. Acoust. Soc. Amer.*, Vol. 63, Jan. 1978, pp. 81-90.
- [8]. J.W. Modestino and A.Y. Ningo, "Detection of Weak Signals in Narrowband non-Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-25, Sept. 1979, pp.592-600.
- [9]. W.P. Elderton and N.L. Johnson, *Systems of Frequency Curves*. London: Cambridge Univ. Press, 1969.
- [10]. J.K. Ord, *Families of Frequency Distributions*. New York: Hafner Publ. Co., 1972.
- [11]. E.S. Pearson, *Karl Pearson's Early Statistical Papers*. London: Cambridge Univ. Press, 1948.
- [12]. E.S. Pearson, "Some problems arising in approximating to probability distributions using moments," *Biometrika*, vol. 50, pp.95-111, 1963.
- [13]. K. Pearson, "The method of moments and the method of maximum likelihood," *Biometrika* vol. 28, pp. 34-59, 1936.
- [14]. E.J.G. Pitman, *Some Basic Theory for Statistical Inference*. London: Chapman and Hall, 1979.
- [15]. K. Pearson, "The fifteen constant bivariate frequency surface," *Biometrika*, vol. 17, p. 268, 1925.

CHAPTER 4 - AN EXAMPLE OF MOMENT ANALYSIS ON ARCTIC UNDER-ICE DATA

INTRODUCTION

As an example of data subjected to moment analysis, we consider samples of under-ice ambient noise. The data were collected on April 23, 1980 under Arctic pack ice at 86°N latitude and 25°W longitude by a multi-institutional experimental group. The raw samples were gathered using an omnidirectional hydrophone suspended 91 meters below the ice, and they represent approximately 10 minutes of noise. The noise was prefiltered with a low-pass filter, cutoff frequency at 2600Hz and 96dB/octave roll-off. The result was sampled at 10kHz and stored as 6006 records of 1024 samples, each record representing about 0.1 seconds. For further information, the reader is referred to References [1 and 2].

COMPUTATION OF SAMPLE MOMENTS

In this chapter, as in [1], the Arctic under-ice samples are assumed independent and stationary. The first four sample moments are computed for each record of 1024 samples (≈ 0.1 seconds). As an estimator of the k th central moment, μ_k , consider the k th central sample moment, m_k :

$$m_k = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^k$$

where $\bar{x}$ is the sample mean:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

In the previous chapter, the skewness β_1 and kurtosis β_2 were defined in terms of the variance σ^2 and the third and fourth central moments μ_3 and μ_4 :

$$\beta_1 = \mu_3 / \sigma^3 \quad \text{and} \quad \beta_2 = \mu_4 / \sigma^4$$

As estimators of σ^2 , β_1 and β_2 , the sample variance s^2 , skewness b_1 and kurtosis b_2 can be used, where

$$s^2 = m_2 \quad b_1 = m_3^2 / s^3 \quad b_2 = m_4 / s^4$$

The sample variance, skewness and kurtosis are plotted versus record number in Figs. 1-3 for the Arctic under-ice ambient noise data. These plots are essentially the same as those in [1], except that here the moments are centered on the sample mean of each record, rather than on the mean of the entire 6006 records. From these figures it can be seen that the noise under study is not only nonstationary, but, at times, non-Gaussian.

Both sample skewness and kurtosis have been suggested [2,4] as measures of Gaussianness. The 1% and 5% confidence intervals for b_1 and b_2 given a sample size 1000 are given in Table 1. as computed in [4]. Applying either of these bounds to Fig. 2 or 3 gives a clear indication that the data consist of Gaussian or nearly Gaussian noise with sporadic, highly non-Gaussian bursts.

If skewness and kurtosis are taken to be measures of Gaussianness, a convenient tool for viewing these statistics is the $\beta_1 \times \beta_2$ plot. Originally suggested by K. Pearson [5,6] as a tool for curve fitting, it is discussed at length in the last chapter. Its most useful property is that each density or family of densities corresponds to a specific region of the $\beta_1 \times \beta_2$ plane.

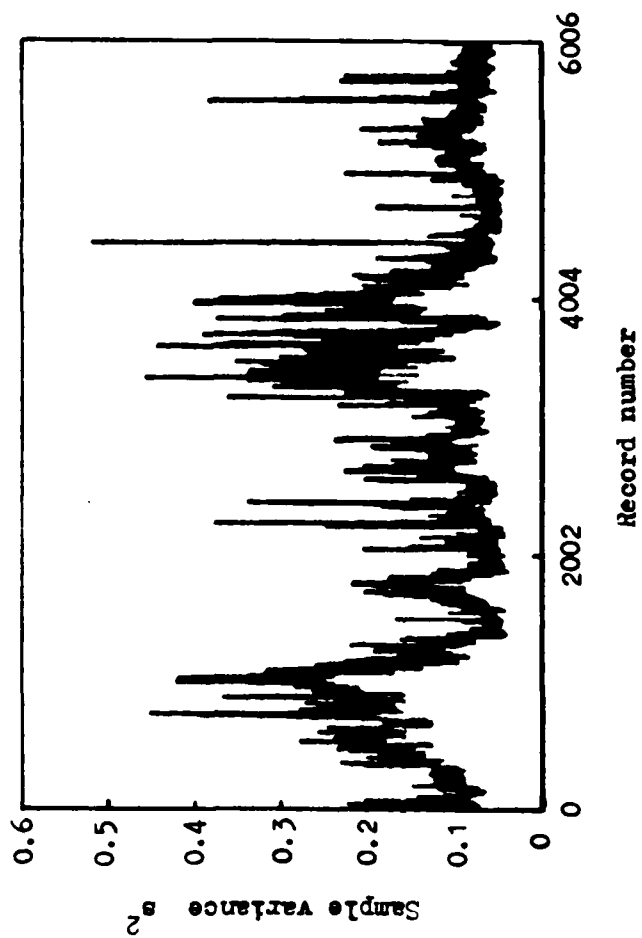


Fig. 1. Sample variance versus record number.

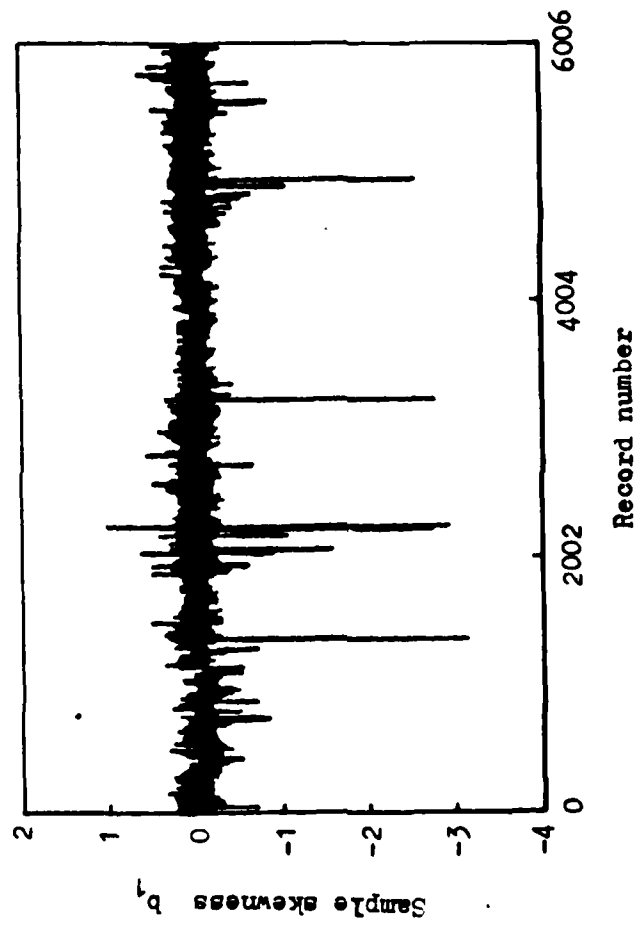


Fig. 2. Sample skewness versus record number.

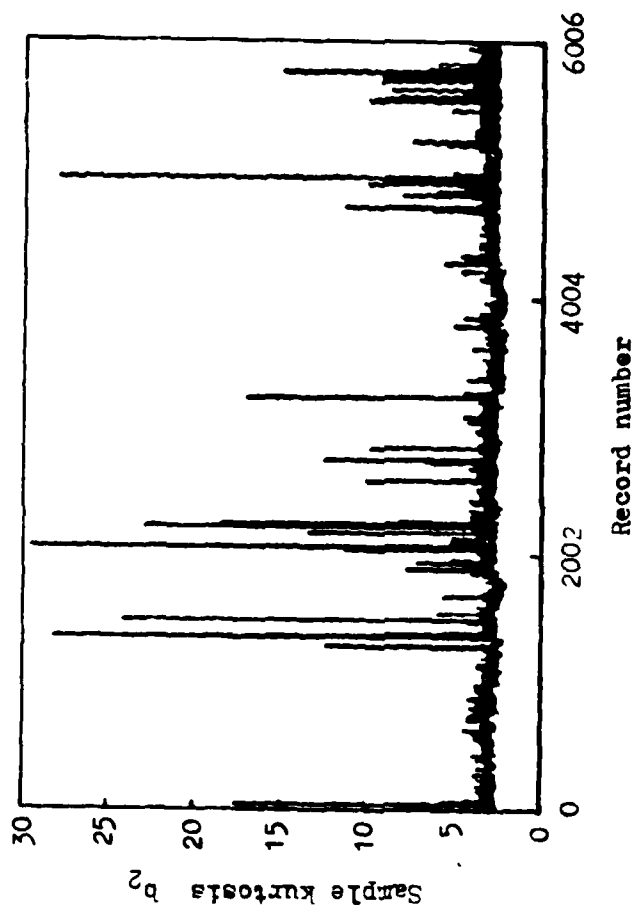


Fig. 3. Sample kurtosis versus record number.

	Confidence Interval	
	5%	1%
b_1	± 0.127	± 0.180
b_2	(2.76, 3.26)	(2.68, 3.41)

Table 1. 1% and 5% Confidence intervals on b_1 and b_2
for 1000 Gaussian samples.

For example the Gaussian density corresponds to a single point, $(\beta_1, \beta_2) = (0, 3)$. While a (β_1, β_2) pair does not uniquely specify a density, two densities at the same point can be made to have equal moments up to fourth order. The Pearson and Johnson families are of particular interest since they both cover the entire possible region of the $\beta_1 \times \beta_2$ plane.

The Pearson family is described in the last chapter and in [5,6]. It contains the Cauchy, t, F, χ^2 , uniform, gamma, exponential and beta densities as special cases. Fig. 4 is a $\beta_1 \times \beta_2$ plot with the regions corresponding to the Pearson types on it. Superimposed on this figure is a scatterplot of the (b_1, b_2) pairs generated for each of the 6006 records.

The Johnson family is described in [5,6] and consists of three types. In the Johnson system, the transformation of a non-Gaussian variate is assumed to be unit normal. In the most common of the Johnson types, the lognormal, S_L , the transformation is

$$g_L(x) = c + \log[(x - a)/b]$$

Thus if x is lognormal, $g_L(x)$ is unit normal. In a similar fashion, S_U and S_B are defined by the transformations:

$$g_U(x) = c + d \sinh^{-1}[(x - a)/b]$$

and

$$g_B(x) = c + d \log[(x - a)/(b - x)]$$

In Fig. 5 the sample (b_1, b_2) pairs are plotted on the $\beta_1 \times \beta_2$ plane for the Johnson family.

In both plots, the vast majority of the points are clustered about the point $(0, 3)$ corresponding to the Gaussian density in both families. While

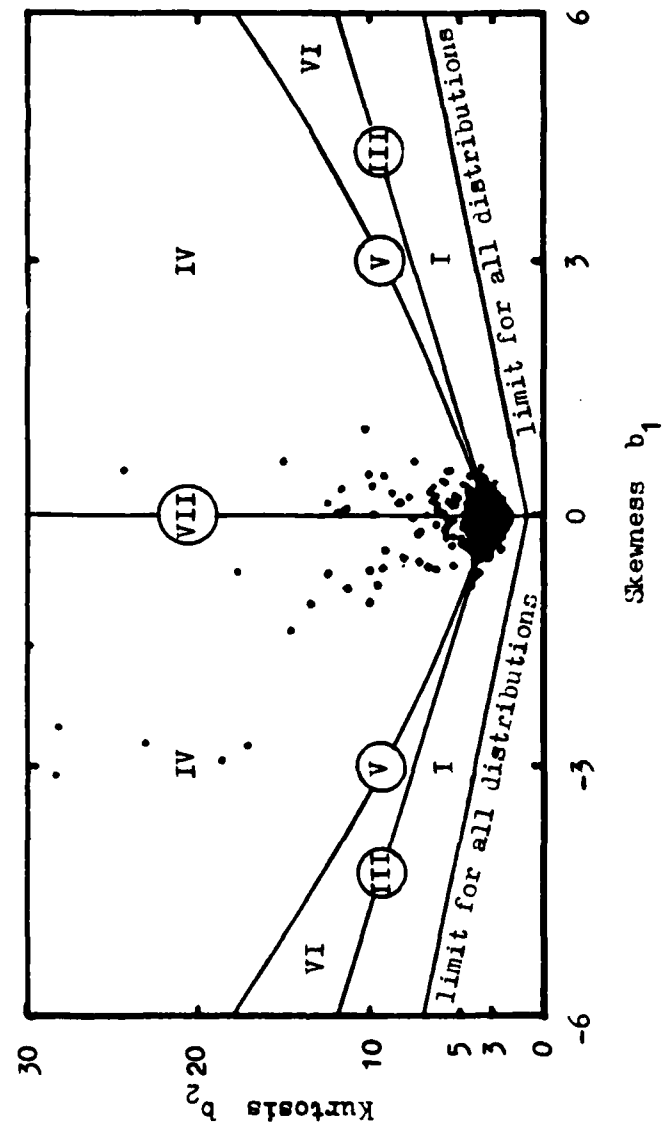


Fig. 4. Scatter-plot of (b_1, b_2) for Pearson family.

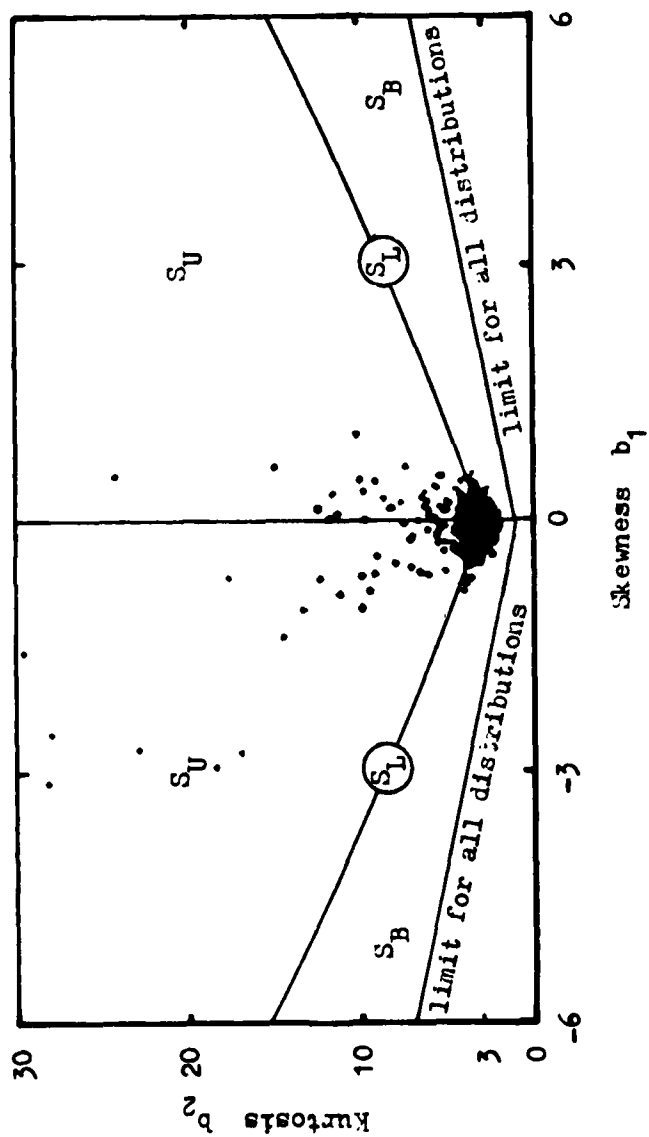


Fig. 5. Scatter-plot of (b_1, b_2) for Johnson family.

confidence regions are extremely difficult to compute, nevertheless these figures show more clearly than Fig. 2 and 3 the essential Gaussian nature of the data.

CONCLUSIONS

The results in Figs. 1-5 indicate that the Arctic under-ice noise is non-stationary and although Gaussian or nearly Gaussian, it is contaminated by sporadic bursts of a highly non-Gaussian nature. Further moment analysis of the non-Gaussian component is of questionable value, due to the small number of highly non-Gaussian records. Perhaps with a larger data set, stronger conclusions could be drawn.

Two assumptions are made earlier in this chapter should be examined further. The data are assumed stationary, but the results obtained do not bear this out. Some investigation of the degree of nonstationarity, perhaps the rate at which the moments change would be of interest. The data are clearly not independent, since, among other sources of dependence, they were prefiltered before being recorded. However, the assumption of independence is made, and the sample moment estimators are perhaps not the best that could be used. For example, consider the sample variance:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N x_i x_j$$

Taking the expected value yields

$$Es^2 = \frac{N-1}{N} \sigma^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N \sigma_{ij}^2$$

where the σ_{ij}^2 are the covariance terms. When the noise is uncorrelated,

the cross terms vanish, but for non-white noise, the cross terms can affect the estimates. For higher order central moment estimates, higher order cross terms result. Several possible moment estimators for dependent data are discussed in [3], but in general, they require knowing or estimating the cross terms.

APPENDIX - Moments Estimators for Dependent Noise.

The following is an example of moment estimators which have been corrected for second order dependence. The prewhitening technique used requires that the correlation either be known or estimated. Note that this corrects only to second order, and the expected value of higher order moment estimates still may contain non-zero cross-moment coefficients of order greater than two.

A data record can be written as a length N vector $\mathbf{x}$ with $N \times N$ covariance matrix Σ . Computing the moments about zero requires that the vector $\mathbf{x}$ be centered; this can be written [4] in terms of the centering matrix $\mathbf{C}$:

$$\mathbf{x} - \bar{\mathbf{x}} = \left[\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{N} \right] \mathbf{x} = \mathbf{C}\mathbf{x}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of 1's.

The centered vector may still be correlated, but it can be prewhitened if it is assumed that the correlation matrix $\mathbf{R} = \Sigma / \sigma^2$ is known. Assuming $\mathbf{R}$ is Toeplitz, it can be factored into an upper and lower triangular matrix:

$$\mathbf{R} = \mathbf{L}\mathbf{U} \quad \text{and} \quad \mathbf{R}^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$$

where $\mathbf{L} = \mathbf{U}^T$ and $\mathbf{L}^{-1} = (\mathbf{U}^{-1})^T$. The matrix $\mathbf{L}^{-1}$ can be used to prewhiten the centered matrix with the result that $\mathbf{L}^{-1}\mathbf{C}\mathbf{x}$ is uncorrelated. An estimate of the i th moment about the mean is given by:

$$m_k = \frac{1}{N} \sum_{i=1}^N (y_i)^k$$

where $\mathbf{y} = \mathbf{L}^{-1}\mathbf{C}\mathbf{x}$ is the centered and prewhitened observation vector. The three parameters of interest in the last chapter are the variance,

skewness and kurtosis. The variance σ^2 has estimate s^2 .

$$s^2 = \frac{1}{N} \mathbf{x}^T \mathbf{U}^{-1} \mathbf{C} \mathbf{L}^{-1} \mathbf{x}$$

The estimates of skewness and kurtosis are

$$b_1 = m_3 / s^3 \quad \text{and} \quad b_2 = m_4 / s^4$$

REFERENCES

1. R.F. Dwyer, "FRAM II Single Channel Ambient Noise Statistics," *NUSC Technical Document 6583*, 25 November 1981.
2. R.F. Dwyer, "Statistical Frequency Domain Signal Processing Method," *Proc. of Princeton CISS*, March 1982.
3. K.V. Mardia, J.T. Kent and J.M. Biddy; *Multivariate Analysis* New York:Academic Press,1979.
4. R.C. Geary and E.S. Pearson, "Tests of Normality," *Biometrika Statistical Tables, Vol I*. Cambridge:Cambridge University Press, 1972.
5. W.P. Elderton and N.L. Johnson, *Systems of Frequency Curves*. London: Cambridge Univ. Press, 1969.
6. J.K. Ord, *Families of Frequency Distributions*. New York: Hafner Publ. Co., 1972.

CHAPTER 5 - LOCALLY OPTIMAL DETECTION IN MULTIVARIATE NON-GAUSSIAN NOISE†

INTRODUCTION

In problems of detection and estimation in discrete time, the observed noise sequence is often assumed to be independent. Under this assumption, various detection schemes have been found and analyzed [1,2,3,4]. Often, however, a strong dependence structure exists suggesting that designs based on the assumption of independence are less than optimum. Previous research on detection in dependent noise is often limited to simple detector structures such as a linear (matched) filter or a ZNL followed by a linear filter [5]. In this chapter, a noise observation consists of a length m vector which is assumed to be no more than m -dependent. Under this model the noise statistics are contained in an m -dimensional multivariate distribution and optimal detectors often contain nonlinearities with memory. Finding a suitable multivariate noise distribution for dependent, non-Gaussian noise is a problem with no single best solution. Instead we discuss several known forms of multivariate densities and draw attention to a characterization entitled *transformation noise* which has several useful properties.

There exist a number of well known, closed form, multivariate densities [6,7] including the Gaussian, Wishart, multivariate Pearson family and multivariate forms of many common univariate densities. The multivariate Gaussian is often employed not only because of its tractability

†This chapter (excluding Appendix B) was co-authored with Peter F. Swaszek, and also appears in his PhD dissertation *Robust Quantization, Vector Quantization and Detection*, Princeton University, October 1982.

but because a preponderance of evidence suggests that it is the natural multivariate density with Gaussian marginals. Unfortunately central limit arguments do not always hold and this approximation can be poor. Additional closed form multivariate densities can be generated by replacing the argument of a univariate density with the square root of a quadratic form [8]. This extends each univariate to an elliptically symmetric density in m -space. These densities all have closed functional forms and the advantage of tractability, but the dependence structure may not be well represented by any of these densities.

A second approach is to use a multivariate series expansion of the Gram-Charlier type [9]. This has a certain theoretical elegance and can be used to derive general properties of densities. For independent data, useful detection procedures have resulted from this method since the nonlinearity is a function of the moments of the noise process [10]. There is still the problem of choosing a proper weighting density on which to base the series. By careful selection, it is often possible to minimize truncation and tail errors.

In a multivariate setting however, it is less clear how well series will work. For one thing, the number of coefficients in a series increases exponentially in m and the expansion requires estimates of the cross moments of the process. It is necessary to truncate the series which may result in a poor representation or in negative values in the tail regions of the pdf. The number of high order, cross moment estimates required for good representation may be prohibitive.

A third class of densities which are often considered are those gen-

erated by transformations from other densities [11]. Common univariate examples are the log-normal and the Johnson family. In a multivariate setting, this method involves a nonlinearity (possibly with memory) operating on a noise sequence with a known density. The intractability of this general class of transformations suggests constraining interest to those invertible transformations with zero memory (ZNL). A block diagram of the system generating this type of noise is given in Fig. 1. Given an identically distributed, m -dependent background noise process ν with a known density $\varphi(\nu)$ and marginals $\varphi_1(\nu_i)$, a ZNL g can be selected to produce a noise $\mathbf{n}$ with the desired marginals $f_1(n_i)$. If φ_1 and g are fixed, the dependence structure of $\mathbf{n}$ is completely determined by the background noise dependency. This lack of flexibility in choosing a dependence structure may be a disadvantage in some cases. Noise generated by this method will be called *transformation noise*.

If the input ν has a multivariate Gaussian distribution, this transformation can be modeled as shown in Fig. 2 where $\mathbf{z}$ is an iid $N(0,1)$ random vector, and $\mathbf{L}$ is a linear operator. It is well known that any symmetric matrix $\mathbf{R}$ can be factored by Crout resolution [12] to give the form:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T$$

where $\mathbf{L}$ is an unique lower triangular (hence causal) matrix. This corresponds to spectral factorization in continuous time which is used to solve the Weiner-Hopf equation. The linear filter $\mathbf{L}$ in Fig. 2 can be therefore chosen to produce a desired covariance in ν :

$$E\{\nu\nu^T\} = E\{(\mathbf{L}\mathbf{z})(\mathbf{L}\mathbf{z})^T\} = \mathbf{L} E\{\mathbf{z}\mathbf{z}^T\} \mathbf{L}^T = \mathbf{L}\mathbf{L}^T = \mathbf{R}$$

This transformation model allows selection of the noise marginals but

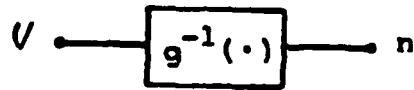


Fig. 1. General transformation noise.

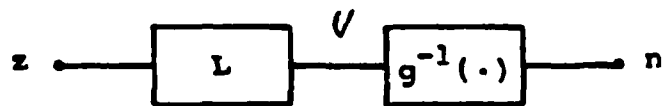


Fig. 2. Transformation noise from a Gaussian background.

increases the complexity of calculating the output dependency. The advantages of the model alluded to above are that the only knowledge of the source needed for its complete statistical specification are the non-Gaussian marginal density and the $m \times m$ covariance matrix.

DETECTION PROBLEM

The detection problem considered is formulated as a hypothesis H_0 and an alternative H_1 :

$$H_0: \mathbf{x} = \mathbf{n}$$

$$H_1: \mathbf{x} = \mathbf{n} + \vartheta \mathbf{s}$$

Under H_0 , the length m observation vector $\mathbf{x}$ consists of m -dependent noise $\mathbf{n}$ only and under H_1 it consists of signal $\mathbf{s}$ with unknown amplitude ϑ plus noise. It is assumed that the signal is known and that the noise has an m -dimensional density $f(\mathbf{n})$. A detector is represented as a functional $\Psi(\mathbf{x})$ operating on the observation $\mathbf{x}$ and this scalar valued test statistic is compared to a threshold to decide for H_0 or H_1 (see Fig.3).

As the signal amplitude approaches zero, a frequently used measure of detector performance is its efficacy:

$$J(\Psi) = \lim_{\vartheta \rightarrow \vartheta_0} \frac{\left[\frac{\partial}{\partial \vartheta} E_{\vartheta} \Psi(\mathbf{x}) \right]^2}{\text{Var}_{\vartheta} \Psi(\mathbf{x})}$$

This is a measure of the asymptotic performance of the detector and is maximized to produce the locally optimal test.

Throughout this discussion we assume that both the family of densities $f(\mathbf{x}, \vartheta)$ and the test statistic $\Psi(\mathbf{x})$ satisfy certain regularity conditions. Let the test statistic have expected value

AD-A120 772 NON-GAUSSIAN AND MULTIVARIATE NOISE MODELS FOR SIGNAL DETECTION IN CORRELATION WITH UTILITY OF ELECTRONIC

2/2

DETECTION(U) PRINCETON UNIV NJ DEPT OF ELECTRICAL
ENGINEERING AND COMPUTER SCIENCE 3 3 MARTINEZ ET AL

UNCLASSIFIED

F/G 9/4

NL

END

FILED

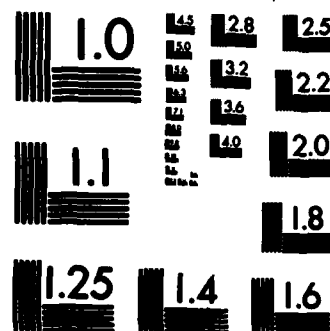
MF



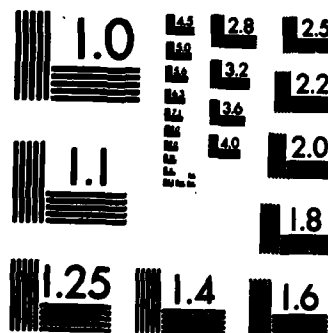
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

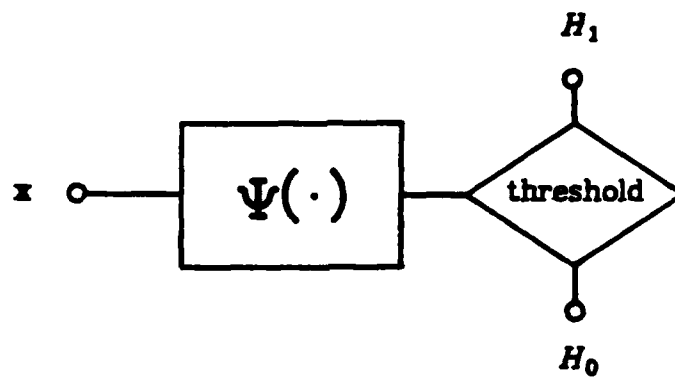


Fig. 3. Typical detector structure.

$$E_{\vartheta} \Psi(\mathbf{x}) = \int \Psi(\mathbf{x}) f(\mathbf{x}, \vartheta) d\mu$$

To be able to move the derivative inside the expectation operator the statistic $\Psi(\mathbf{x})$ is required to be *regular* at ϑ_0 ; that is for all ϑ in an open neighborhood of ϑ_0

$$E_{\vartheta}' \Psi(\mathbf{x}) = \frac{\partial}{\partial \vartheta} \int \Psi(\mathbf{x}) f(\mathbf{x}, \vartheta) d\mu = \int \Psi(\mathbf{x}) f'(\mathbf{x}, \vartheta) d\mu$$

We also desire that $\Psi(\mathbf{x})=1$ be regular:

$$\frac{\partial}{\partial \vartheta} \int f(\mathbf{x}, \vartheta) d\mu = \int f'(\mathbf{x}, \vartheta) d\mu = 0$$

To satisfy these conditions restrictions must be made on the test statistic $\Psi(\mathbf{x})$ or on the family of densities $f(\mathbf{x}, \vartheta)$. Rather than overly limit the class of test statistics, restrictions are applied primarily to the family of densities. It can be shown [13] that all real-valued test statistics with finite second moment are regular in the neighborhood of ϑ_0 if the family of densities f are smooth (in Pitman's sense) at ϑ_0 . For family of densities to be smooth at ϑ_0

- (i) f has a ϑ derivative f' at almost every $\mathbf{x}$ for each value of ϑ in an open neighborhood of ϑ_0 .
- (ii) $(f')^2/f$ is integrable and $\int (f')^2/f d\mu < \infty$ (finite Fisher's Information).

It is well known [13] that under the above regularity conditions the function that maximizes efficacy, the locally optimal test statistic $\Psi_{lo}(\mathbf{x})$, is the ϑ derivative of the log likelihood function at $\vartheta = 0$:

$$\Psi_{lo}(\mathbf{x}) = \lim_{\vartheta \rightarrow 0} \frac{\partial}{\partial \vartheta} \log[f(\mathbf{x}, \vartheta)]$$

In particular, for this detection problem (a shift in mean):

$$\Psi_{lo}(\mathbf{x}) = \left[\frac{-\nabla f(\mathbf{x})}{f(\mathbf{x})} \right]^T \mathbf{s} \quad (1)$$

where ∇f is the gradient of f . The efficacy of this test is given by

$$J(\Psi_{lo}) = E_0(\Psi_{lo}^2)$$

In the independent and identically distributed (iid) case, the multivariate density is the m -fold product of the noise marginal $f_1(n_i)$. Eq. (1) reduces to a ZNL followed by a correlator, a form considered extensively in the literature:

$$\Psi_{lo}(\mathbf{x}) = -\sum_{i=1}^m s_i \frac{f_1'(x_i)}{f_1(x_i)} = \sum_{i=1}^m s_i g_{lo}(x_i)$$

The ZNL g_{lo} is the locally optimal nonlinearity for iid noise, given by

$$g_{lo}(x_i) = -\frac{f_1'(x_i)}{f_1(x_i)} \quad (2)$$

Given a specific, closed form multivariate density of the type considered in Section I, the locally optimal detector can be found from Eq. (1). In the case of the multivariate Gaussian density:

$$f(\mathbf{n}) = \frac{1}{(2\pi)^{m/2} |\mathbf{R}|^{1/2}} \exp(-\mathbf{n}^T \mathbf{R}^{-1} \mathbf{n} / 2)$$

where $\mathbf{R}$ is the covariance matrix, the resulting detector is the matched filter:

$$\Psi_{lo}(\mathbf{x}) = \mathbf{s}^T \mathbf{R}^{-1} \mathbf{x}$$

As a non-Gaussian example, consider a bivariate, elliptically symmetric density with Laplace marginals [8]:

$$f(\mathbf{n}) = \frac{1}{\pi |\mathbf{R}|^{1/2}} K_0(\sqrt{2\mathbf{n}^T \mathbf{R}^{-1} \mathbf{n}})$$

The resulting detector is

$$\Psi_{lo}(\mathbf{x}) = \left[\frac{2}{\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}} \right]^{1/2} \frac{1}{|\mathbf{R}|^{1/2}} \frac{K_1(\sqrt{2\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}})}{K_0(\sqrt{2\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}})} (\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x})$$

where K_0 and K_1 are modified Bessel functions. The test statistic is the product of the outputs of a matched filter and a nonlinearity with memory.

As in the example above, closed form multivariate densities typically yield detectors which include nonlinearities with memory, making these detectors difficult to implement. Employing the transformation noise model (with Gaussian background) simplifies the resulting detector to a combination of ZNL's and a linear filter.

TRANSFORMATION NOISE DETECTOR

As described in the Introduction, the m -dimensional noise $\mathbf{n}$ is generated by passing an identically distributed noise sequence ν with a known multivariate density $\varphi(\nu)$ and marginals $\varphi_1(\nu_i)$ through an invertible ZNL as shown in Fig. 1. The output noise $\mathbf{n}$ has a multivariate density $f(\mathbf{n})$ with a dependency structure and marginals $f_1(n_i)$ determined by φ and g . It is assumed that g is twice differentiable almost everywhere and that both n_i and ν_i have invertible cdf's. The density of $\mathbf{n}$ is found with a change of variables:

$$f(\mathbf{n}) = \varphi[g(\mathbf{n})] \prod_{i=1}^m |g'(n_i)|$$

where

$$g(\mathbf{n}) = [g(n_1), g(n_2), \dots, g(n_m)]^T$$

The locally optimal detector is found from Eq. (1) to be

$$\Psi_{lo}(\mathbf{x}) = - \left[\frac{\nabla \varphi[g(\mathbf{x})]}{\varphi} g'(\mathbf{x}) + \frac{g''(\mathbf{x})}{g'(\mathbf{x})} \right]^T \mathbf{s}$$

A block diagram of Ψ_{lo} is given in Fig. 4 where the symbol $\otimes$ is an element

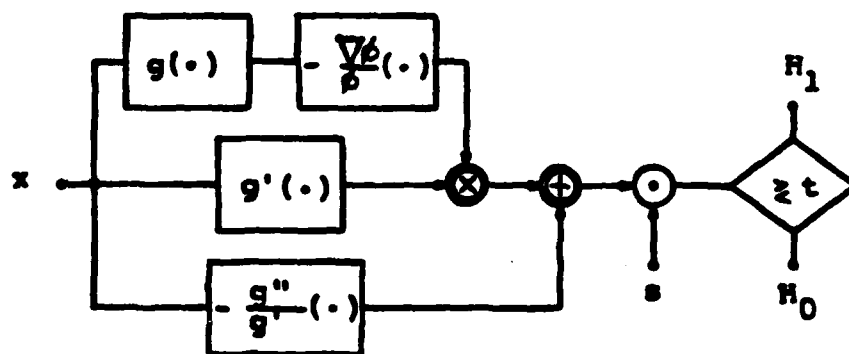


Fig. 4. Transformation noise detector.

by element vector multiplication and $\odot$ is a vector dot product. The system consists of the locally optimal nonlinearity with memory (for signal in φ distributed noise), three ZNL's and a correlator followed by a threshold comparator.

The transformation ZNL g can be found by equating the marginal cdf's of $\mathbf{n}$ and ν ($F_1(n_i)$ and $\Phi_1(\nu_i)$ respectively):

$$g(x_i) = \Phi_1^{-1}[F_1(x_i)] \quad (3)$$

The nonlinearity g' is found by differentiating Eq. (3)

$$g'(x_i) = \frac{f_1(x_i)}{\varphi_1[g(x_i)]} \quad (4)$$

and likewise

$$-\frac{g''(x_i)}{g'(x_i)} = -\frac{f_1'(x_i)}{f_1(x_i)} + \frac{\varphi_1'[g(x_i)]}{\varphi_1[g(x_i)]} g'(x_i)$$

At values where the denominator of these expressions equals zero, the quantity is assumed to equal infinity. This is acceptable since an observation in a region of zero probability of the noise process indicates the sure presence of the signal.

When ν is a multivariate Gaussian random vector with unit variance and covariance matrix $\mathbf{R}$, this third nonlinearity becomes

$$-\frac{g''(x_i)}{g'(x_i)} = -\frac{f_1'(x_i)}{f_1(x_i)} + g'(x_i)g(x_i)$$

The locally optimal nonlinearity, with memory, is a linear filter

$$-\frac{\nabla \varphi(\mathbf{x})}{\varphi} = \mathbf{R}^{-1}\mathbf{x}$$

Therefore, the system can be reconfigured as shown in Fig. 5. The third ZNL g_{10} is the locally optimal nonlinearity for signal in iid, f_1 distributed noise as given in Eq. (2). Thus, under the assumption of transformed unit

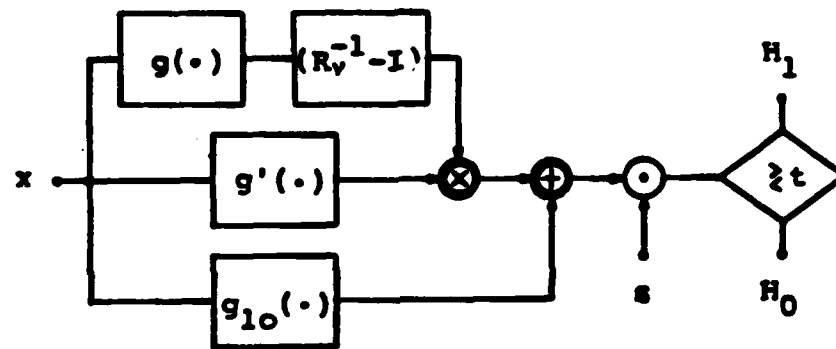


Fig. 5. Locally optimal detector for transformation noise with Gaussian background.

variance Gaussian noise (Fig. 2), the system consists of one linear filter and three ZNL's, one of which is the nonlinearity occurring in the iid locally optimal detector.

Under the assumed noise model, the detector design requires knowledge of g , the nonlinear transformation, and R_v , the covariance matrix of the background Gaussian vector v . In practice, however, the available information is often the noise marginal density and the noise covariance matrix R_n . The transformation nonlinearity g is found from the marginal cdf with Eq.(3). Computing R_v from R_n in general requires numerical integration (Appendix A).

TRANSFORMATION EXAMPLES

Two examples of the proposed detector are considered in this section. The first example employs the Laplace density as the noise marginal and the second example has Pearson Type VII marginals. For both of these examples, Eqs. (3),(4) and (2) yield the ZNL's g , g' and g_{lo} respectively. Covariance mapping described in Appendix A relates the background and noise covariance matrices, R_v and R_n .

The Asymptotic Relative Efficiency (ARE) is often used to compare the performance of two detectors. It is defined as the ratio of the number of samples needed to achieve the same level and power of the two detectors as the number of samples goes to infinity. With the above regularity conditions and the Pitman-Noether theorem, the ARE is equal to the ratio of the efficacies

$$ARE_{1,2} = \frac{J_1 \Psi(\mathbf{x})}{J_2 \Psi(\mathbf{x})}$$

An analytic evaluation of the efficacy of these detectors is generally intractable and depends on the signal $\mathbf{s}$. As a result, the operation of the optimal and several suboptimal detectors was simulated to evaluate the performance in two multivariate transformation noise environments. The simulated structures are numbered as follows:

1. Locally optimal system (Fig. 5)
2. Locally optimal test for iid noise (g_{lo})
3. Matched filter paralleled by g_{lo} and the values summed
4. Matched filter

Detector 3 is included to determine the effects of removing the vector multiplication in the optimal detector, significantly simplifying its structure. A Monte Carlo simulation was employed to compute the efficacy of the four detectors above. These results were then used to compute the ARE for each scheme as compared to the matched filter.

The matched filter has efficacy $\mathbf{s}^T \mathbf{R}^{-1} \mathbf{s}$. In a bivariate noise environment, with correlation coefficient ρ , the efficacy of the matched filter is maximized by $s_1 = -s_2$ when $\rho \geq 0$ and by $s_1 = s_2$ when $\rho < 0$. Conversely, it is minimized by $s_1 = s_2$ for $\rho \geq 0$ and by $s_1 = -s_2$ for $\rho < 0$. Since all four systems are symmetric in s_1 and s_2 , the condition $s_1 = s_2$ maximizes while $s_1 = -s_2$ minimizes efficacy for $\rho \leq 0$ (the extrema are reversed for $\rho \geq 0$). A constant signal with unit power is used in the evaluation of bivariate detectors in this chapter, and therefore, the simulation yields minimum and maximum efficacies for $\rho \geq 0$ and $\rho < 0$ respectively.

The bivariate simulation shows ARE for $-1 < \rho < 1$. However, for m -

dependent noise, this entire range may not be of importance. Since all terms of the m -dependent autocorrelation sequence beyond the first m vanish and the covariance matrix must be positive definite, bounds can be placed on the correlation sequence [14]. For the bivariate case ($m = 2$), the bounds on ρ are $|\rho| \leq \frac{1}{2}$ with the result that the values of ARE where $|\rho| \leq \frac{1}{2}$ are perhaps of the most interest. Although the efficacies are maximized or minimized by the selection of the signal, this does not appear in the plots of ARE. This results because the matched filter is more sensitive than many other detectors to signal selection.

For multivariate simulations of length m , a triangular correlation function sampled at equal intervals is assumed:

$$\rho_n(i) = \begin{cases} 1 - |i/m| & |i| \leq m \\ 0 & |i| > m \end{cases}$$

This triangular function approximates the type of correlation often seen in highly correlated noise sequences. For these examples, the signal to maximize the efficacies is no longer obvious due to the nonlinearities present in the detectors. Rather than find the signal to maximize the efficacy, simulations are presented with a constant ($s_i = k$) and an oscillating ($s = \{k, -k, k, \dots\}$) signal. These were chosen because they yield nearly the worst and best performance from the matched filter respectively for the correlation sequence above.

1. Laplace Results

Impulsive noise environments are often modeled by densities whose tails are heavier than the Gaussian [15,16]. The Laplace density is chosen as a first example since it is a frequently used model of non-Gaussian, impulsive noise [17]. For the marginal density

$$f_1(n_i) = \frac{\alpha}{2} \exp(-\alpha |n_i|)$$

Eqs. (3), (4) and (2) yield the three detector ZNL's

$$g(x_i) = \Phi^{-1} \left\{ \frac{1}{2} \left[1 + \operatorname{sgn}(x_i) [1 - \exp(-\alpha |x_i|)] \right] \right\}$$

$$g'(x_i) = \frac{\alpha \sqrt{2\pi}}{2} \exp(g^2(x_i)/2 - \alpha |x_i|)$$

$$g_{lo}(x_i) = \alpha \operatorname{sgn}(x_i)$$

where Φ^{-1} is the inverse normal cdf. For unit power noise $\alpha = \sqrt{2}$ and correlation coefficient mapping shows that $\rho \approx \rho_L$. Figs. 6, 7 and 8 contain the three nonlinearities g , g' and g_{lo} . The resulting ARE's for the bivariate case are shown in Fig. 9, and for the multivariate case in Figs. 10 and 11.

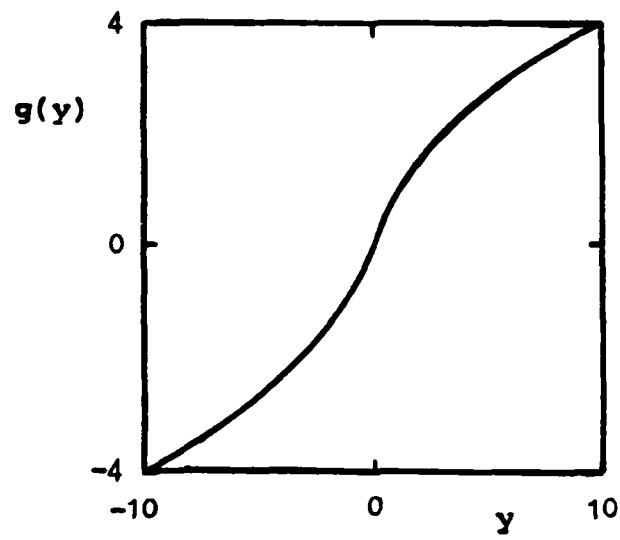


Fig. 6. Laplace ZNL1, $g(y)$.

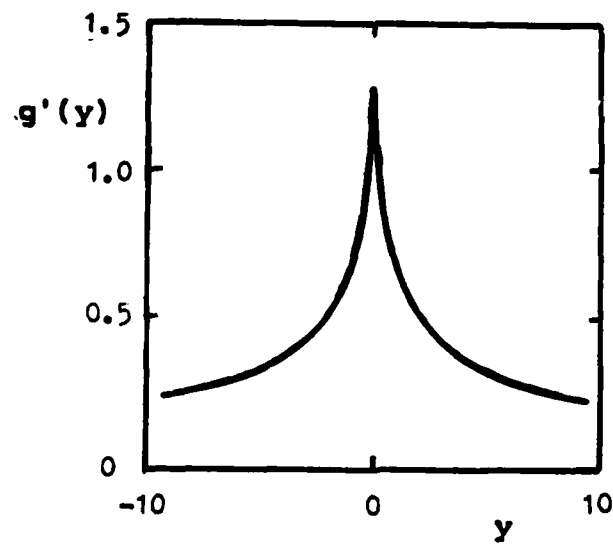


Fig. 7. Laplace ZNL2, $g'(y)$.

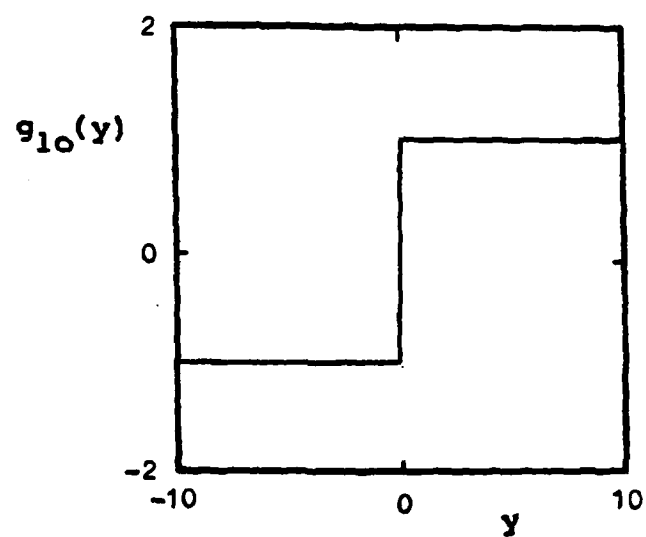


Fig. 8. Laplace ZNL3, $g_{10}(y)$.

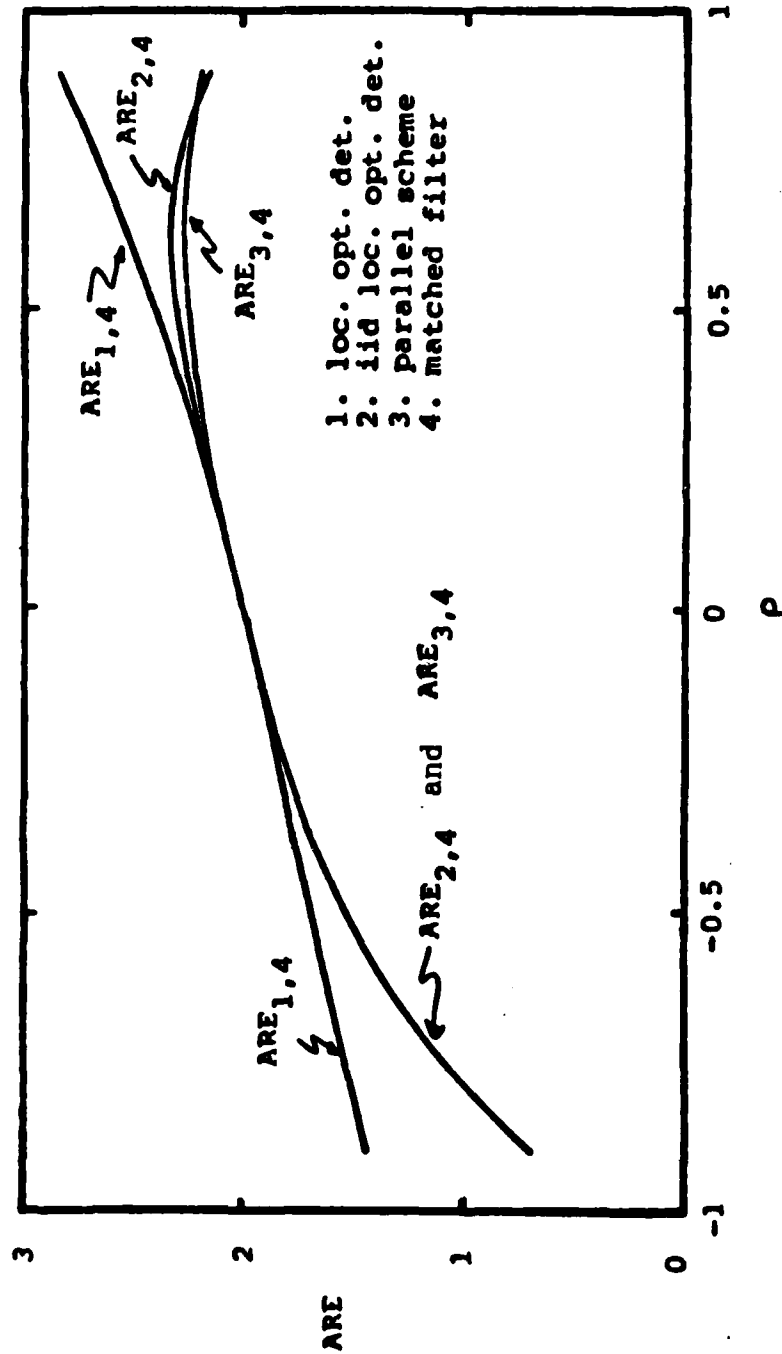


Fig. 9. ARE of three detectors for Laplace transformation noise vs. ρ with a constant signal.

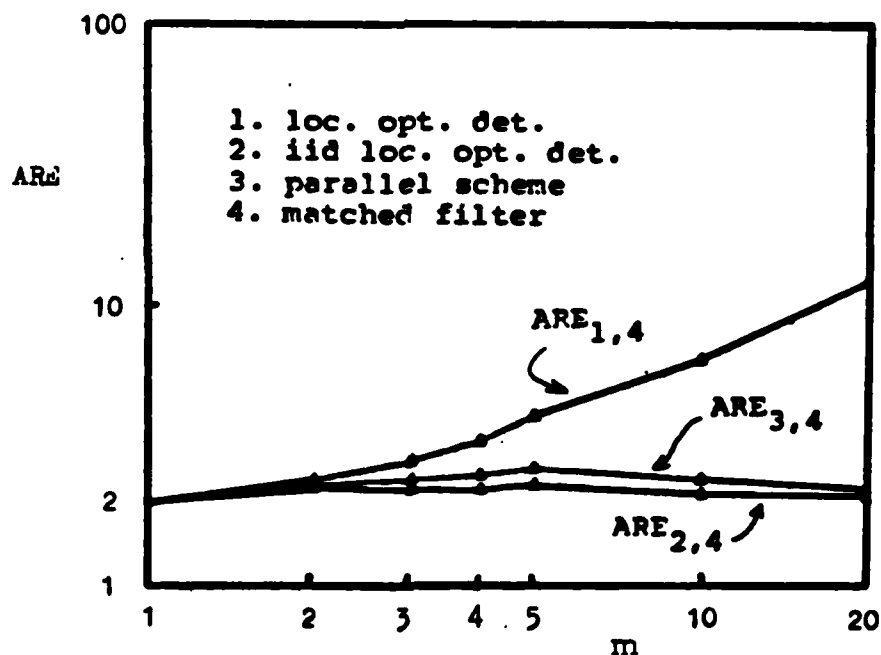


Fig. 10. ARE of Laplace detectors vs. detector length m for triangular correlation and a constant signal.

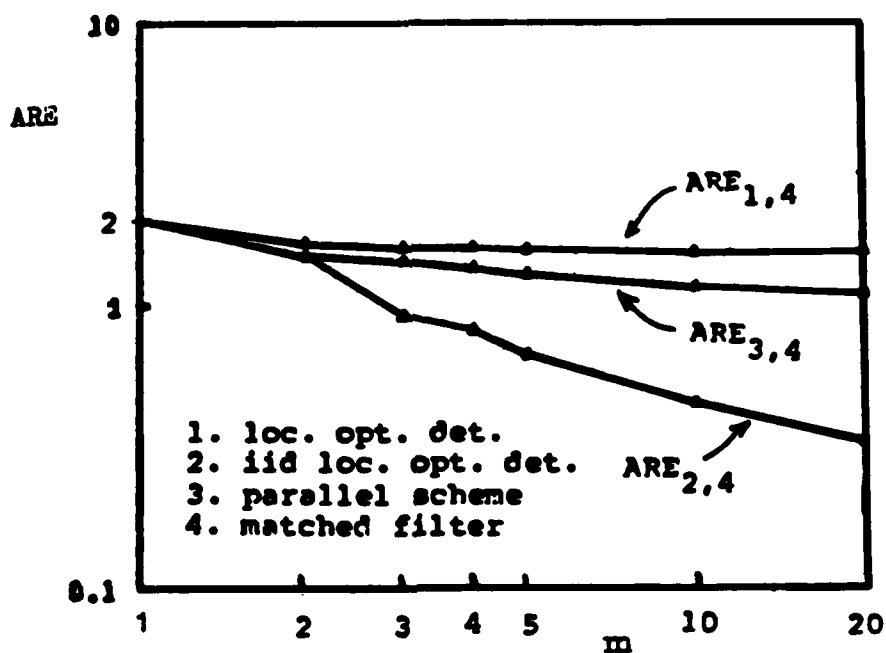


Fig. 11. ARE of Laplace detectors vs. detector length m for triangular correlation and an alternating signal.

2. Pearson Type VII Results

The Pearson VII density is selected as a second example because it can be made arbitrarily close to Gaussian by changing one of its parameters [11]. Here the marginal is

$$f_1(n_i) = K \left[1 + (n_i/a)^2 \right]^{-\alpha}$$

As $\alpha \rightarrow \infty$, the density approaches the Gaussian. For this example, let $\alpha = 7$. This results in a density with a nearly Gaussian body and algebraically decaying tails. The three nonlinearities in the detector are found from Eqs. (3), (4) and (2) to be

$$\begin{aligned} g(x_i) &= \Phi^{-1}[F_1(x_i)] \\ g'(x_i) &= K \sqrt{2\pi} \left[1 + (x_i/a)^2 \right]^7 \exp(g^2(x_i)/2) \\ g_{10}(x_i) &= 14x_i / (a^2 + x_i^2) \end{aligned}$$

where F_1 is the Pearson VII cdf. These ZNL's are depicted in Figs. 12, 13 and 14. In Fig. 15, the ARE's generated by simulation for bivariate noise are plotted. The multivariate simulation results are depicted in Figs. 16 and 17 for constant and alternating signals respectively.

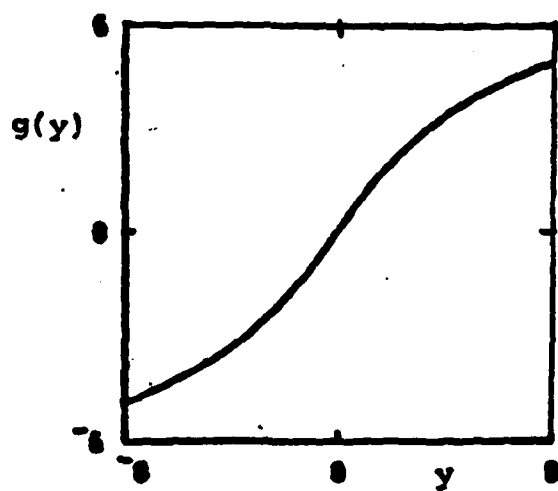


Fig. 12. Pearson VII ZNL1, $g(y)$.

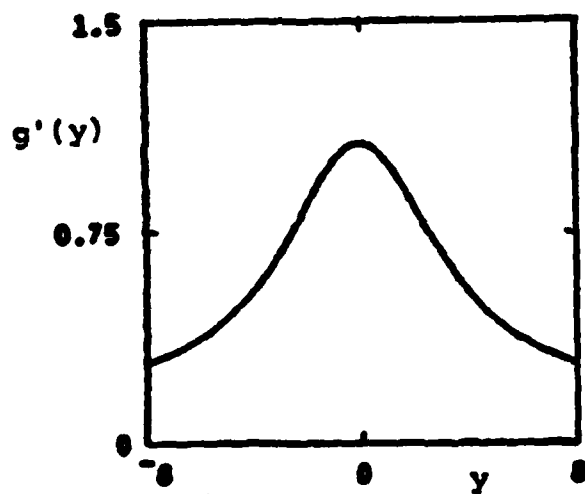


Fig. 13. Pearson VII ZNL2, $g'(y)$.

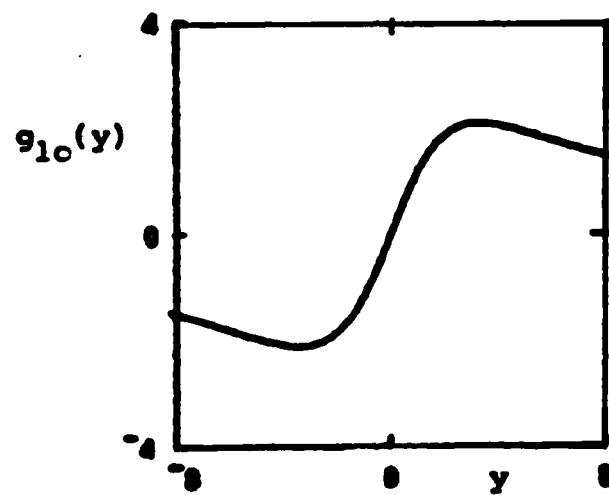


Fig. 14. Pearson VII ZNL3, $g_{10}(y)$.

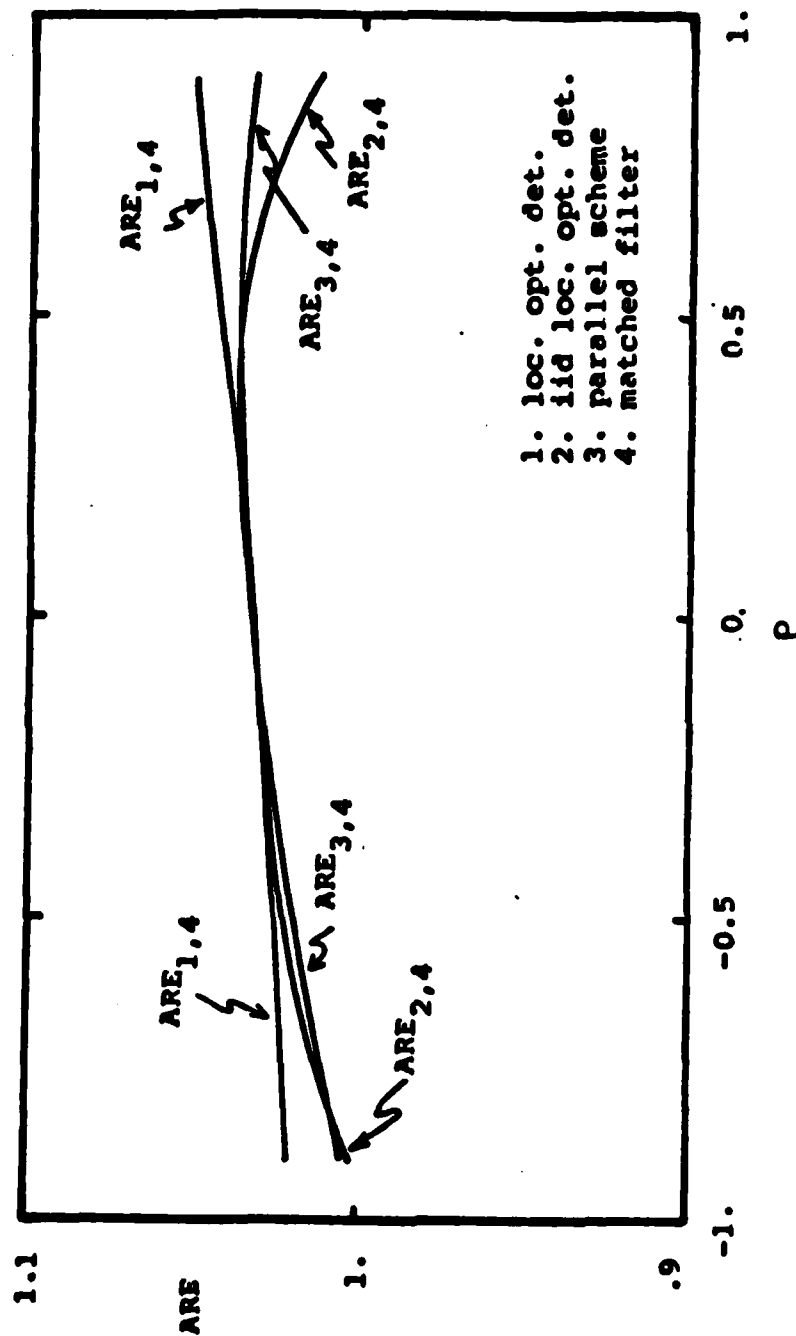


Fig. 15. ARE of three detectors for Pearson VII transformation noise vs. ρ with a constant signal.

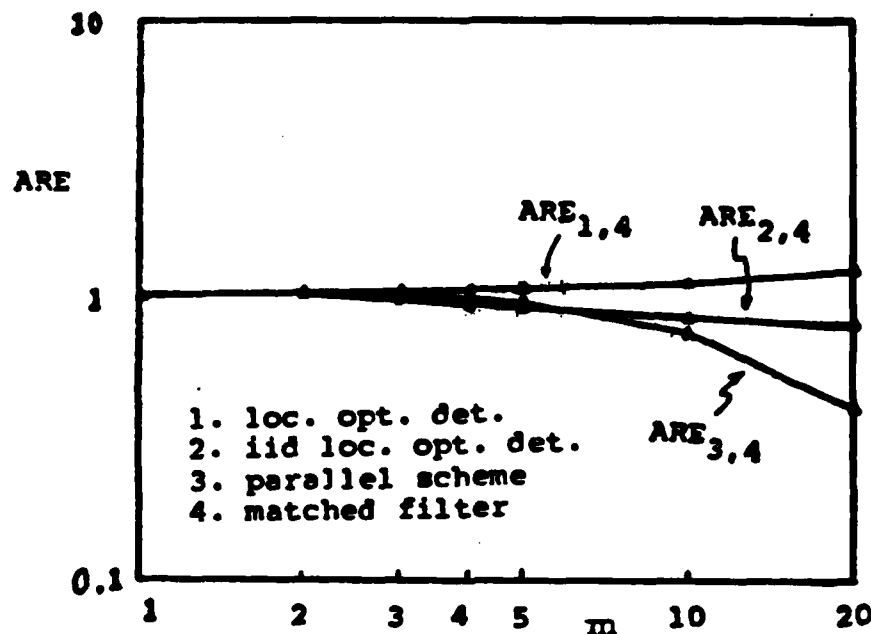


Fig. 16. ARE of Pearson VII detectors vs. detector length m for triangular correlation and a constant signal.

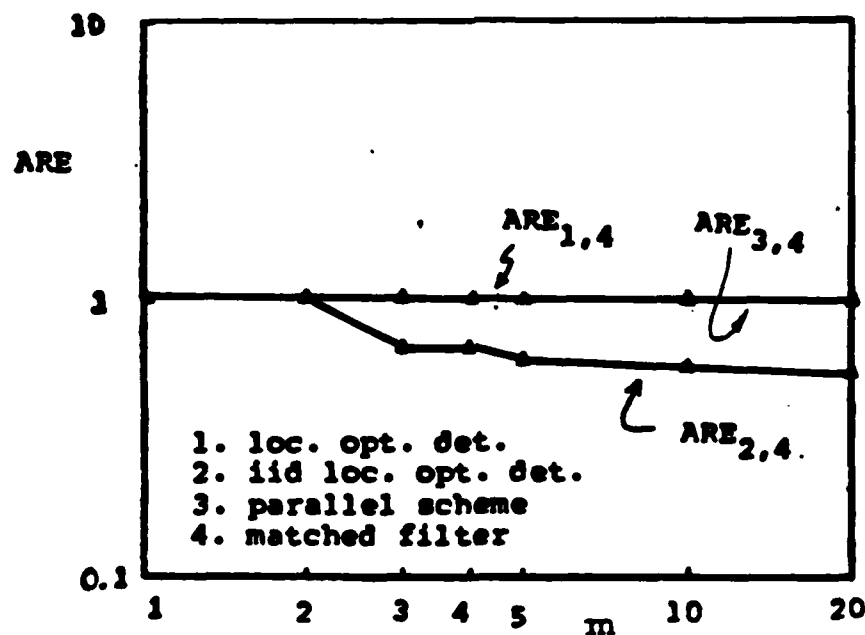


Fig. 17. ARE of Pearson VII detectors vs. detector length m for triangular correlation and an alternating signal.

MOVING AVERAGE (MA) NOISE DETECTOR

Modeling the noise process as in Fig. 2 suggests the reversed model shown in Fig. 18. The input $\mathbf{z}$, iid $N(0,1)$, is passed through the ZNL g^{-1} to produce the iid vector γ with non-Gaussian marginals φ_i and variance σ^2 . Passing γ through a linear filter L of length m produces the sequence $\mathbf{n}$ with a dependence structure generated by L . This model is typically called a moving average process (MA) of order $m-1$ [14].

As in the Introduction, Crout factorization can be used to solve for an unique lower triangular matrix L , given R , since the covariance matrix of $\mathbf{n}$ is given by

$$R_n = E\{\mathbf{n}\mathbf{n}^T\} = E\{(L\gamma)(L\gamma)^T\} = L E\{\gamma\gamma^T\} L^T = \sigma^2 L L^T$$

where σ^2 is the variance of the γ_i . This MA model has a more straightforward dependence structure than the transformation noise, but solving for the marginals is more difficult. Note that these two schemes are equivalent only when g^{-1} is linear ($\mathbf{n}$ is multivariate Gaussian).

Denoting the γ marginal as $\varphi_1(\gamma_i)$, the density of the MA noise $\mathbf{n}$ is found to be

$$f(\mathbf{n}) = |L^{-1}| \prod_{i=1}^m \varphi_1[(L^{-1}\mathbf{n})_i]$$

where φ_i is assumed continuously differentiable and strictly positive. The locally optimal detector $\psi_{lo}(\mathbf{x})$ is found from Eq. (1) to be

$$\psi_{lo}(\mathbf{x}) = (L^{-1}\mathbf{s})^T g_{lo}(L^{-1}\mathbf{x})$$

where g_{lo} is the ZNL in the iid locally optimal detector for the φ_1 marginal:

$$g_{lo}(x_i) = \frac{-\varphi_1'(x_i)}{\varphi_1(x_i)}$$

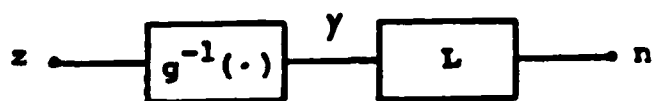


Fig. 18. Moving average noise model.

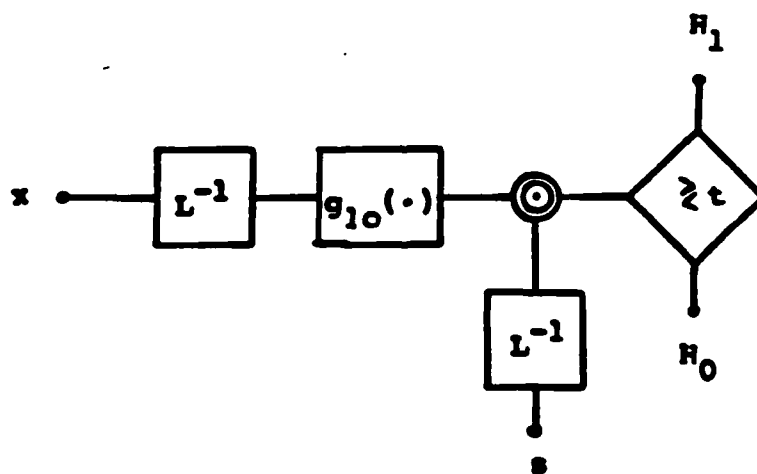


Fig. 19. Moving average noise detector.

A block diagram of this detector is given in Fig. 19. The system is a ZNL embedded in a matched filter. Note that for $\mathbf{n}$ iid, $\mathbf{L}$ is an identity matrix, and the resulting detector is the iid locally optimal detector $\mathbf{s}^T \mathbf{g}_{lo}(\mathbf{x})$. For the Gaussian case [$g^{-1}(\mathbf{x}) = \mathbf{x}$] the result is the matched filter $\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x}$.

The efficacy of the locally optimal test statistic is

$$J(\Psi_{lo}) = E_0\{\Psi_{lo}^2(\mathbf{x})\}$$

where E_0 is the expectation under H_0 . This can be calculated for the moving average model:

$$J(\Psi) = (\mathbf{L}^{-1} \mathbf{s})^T E\{g_{lo}(\mathbf{L}^{-1} \mathbf{x}) g_{lo}^T(\mathbf{L}^{-1} \mathbf{x})\} (\mathbf{L}^{-1} \mathbf{s})$$

where g_{lo} is the ZNL embedded in the detector. The outer product under the expectation can be simplified since $\mathbf{L}^{-1} \mathbf{x} = \gamma$ is an iid random vector. The nonlinearity g_{lo} has expected value zero for any locally optimal detector since

$$E\{g_{lo}\} = - \int_{-\infty}^{\infty} \nabla f(\mathbf{x}) d\mathbf{x} = 0$$

The outer product reduces to a diagonal matrix with diagonal elements:

$$E_0\{g_{lo}^2(x)\} = I_0$$

where I_0 is Fisher's information for the marginal density f_1 . The efficacy of the system becomes

$$J(\Psi_{lo}) = \frac{I_0}{\sigma^2} \mathbf{s}^T \mathbf{R}^{-1} \mathbf{s}$$

and the ARE as compared to the matched filter is

$$ARE_{lo, mf} = I_0 / \sigma^2$$

Fisher's information is known to have a minimum of σ^2 for the Gaussian density; hence, the ARE is always greater than or equal to unity and the locally optimal detector performs at least as well as the matched filter.

CONCLUSIONS

Given a completely specified multivariate density, methods of finding optimal detectors are well known. If the noise is non-iid or non-Gaussian, the complete statistical description, the m -dimensional density, may be an unreasonable amount of information to expect in practice. The transformation and moving average density models, however, are characterized solely by the marginal densities and the covariance matrices. Also, the use of moving average and transformation noise models as structures for a non-Gaussian noise environment simplifies the locally optimal detector structure since no nonlinearities with memory are required.

Three of the four detector structures considered in the simulations are of particular importance. The matched filter uses the covariance matrix only and is optimal when the density is Gaussian. The iid locally optimal detector employs the marginal density of the noise and is optimal if the noise is independent. The multivariate transformation noise detector uses both the correlation and the marginals, and is optimal or nearly optimal when the noise is described well by the transformation noise model.

In the bivariate case the effects of dependence upon performance as indicated in the simulations are minimal (particularly for $|\rho| \leq 1/2$). The iid locally optimal detector performed nearly as well as the multivariate locally optimal detector. Both detectors performed better than the linear detector indicating that in fact the form of the marginal is more important than the dependence structure.

For longer correlation sequences ($m > 2$), the dependence structure becomes increasingly more important. In the simulations, the ARE of the iid locally optimal detector falls well below that of the optimal detector. When the signal is chosen to nearly optimize the matched filter performance, even the matched filter performs better than the iid locally optimal detector. In this case, the dependence structure and the marginals are both important and a detector structure should exploit both when possible.

The moving average model was introduced as a permutation of the transformation formulation. Simulations were not attempted because of the need to factor the marginal characteristic function to arrive at a γ process. However, the detector structure itself is simple and suggests its use as a general suboptimal structure which can be optimized numerically.

APPENDIX A - Computation of R_ν

Implementation of the transformation noise detector requires knowledge of the covariance matrix of the background Gaussian process. Assuming that the input and output processes have unit power, this problem reduces to finding a mapping between the noise correlation coefficient ρ_n and the background process correlation ρ_ν [18]. The output correlation coefficient is given by

$$\rho_n(i) = E\{n_j n_{j+i}\} = E\{g^{-1}(\nu_j) g^{-1}(\nu_{j+i})\}$$

where g^{-1} is defined to be

$$g^{-1}(\nu_i) = F^{-1}[\Phi(\nu_i)]$$

Since the ν_j and ν_{j+i} are unit normal random variables, with correlation coefficient $\rho = \rho_\nu(i)$, the expression for $\rho_n(i)$ becomes

$$\rho_n(i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g^{-1}(x) g^{-1}(y) \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} dx dy \quad (A1)$$

Solution of this integral expression, usually numerically, provides the required mapping of ρ_ν to ρ_n .

In [19], Wise and Thomas express the joint normal density as a series expansion, Mehler's formula:

$$N(0,0,1,1,\rho) = \frac{1}{2\pi} \exp\{-(x^2+y^2)/2\} \sum_{k=0}^{\infty} \frac{\rho^k H_k(x) H_k(y)}{k!}$$

where H_k is the k -th Hermite polynomial. The nonlinearity g^{-1} can be represented by its Hermite expansion:

$$g^{-1}(x) = \sum_{k=0}^{\infty} \frac{b_k H_k(x)}{\sqrt{k!}}$$

with coefficients b_k given by

$$b_k = \int_{-\infty}^{\infty} g^{-1}(x) \frac{1}{2\pi} \exp(-x^2/2) \frac{H_k(x)}{\sqrt{k!}} dx$$

Employing the orthogonality relationship of the Hermite polynomials:

$$\frac{1}{2\pi\sqrt{k!j!}} \int_{-\infty}^{\infty} H_k(x)H_j(x)\exp(-x^2/2) dx = \delta_{jk}$$

the integral in Eq.(A1) reduces to

$$\rho_n(i) = \sum_{k=0}^{\infty} b_k^2 [\rho_v(i)]^k$$

In the noise examples in this chapter, the marginal density is assumed to be symmetric; hence, both g and g^{-1} are odd symmetric. The even Hermite expansion coefficients (b_{2k} , $k=0,1,\dots$) are zero and $\rho_n(i)$ is an increasing, odd symmetric function of $\rho_v(i)$. Therefore in practice, the values of ρ_v can be interpolated from a table of ρ_v versus ρ_n . Also, as an added note, the Gaussian process has a correlation coefficient that is always larger in modulus than the output correlation:

$$|\rho_v| \geq |\rho_n|$$

APPENDIX B - Estimation of Transformation Nonlinearity.

Suppose a particular multivariate non-Gaussian noise process is well described by the transformation model with Gaussian pre-transformation noise. Then it is assumed that the length m non-Gaussian noise vector $\mathbf{n}$ with multivariate density $f_m(\mathbf{n})$ can be produced by passing Gaussian noise ν with density $\varphi_m(\nu)$ through an invertible zero memory nonlinearity (ZNL) as shown in Fig B1. It is assumed that both $\varphi_m(\nu)$ and $f_m(\mathbf{n})$ each have identical marginals ($\varphi(\nu_i)$ and $f(n_i)$ respectively).

If the marginals φ and f are both known, then g can be found by equating marginal cdf's and the problem is solved. Frequently in practice, f may not be known exactly, and an approximation to this ZNL with easily estimated parameters may be desired. For instance, the locally optimal detector for transformation noise derived in this chapter requires g given only the observed post-transformation, non-Gaussian noise $\mathbf{n}$. Under certain regularity conditions on the f , the Cornish-Fisher expansion can be used to generate an approximate transformation of the form [11,20]:

$$\nu = g(\mathbf{n}) = a_0 + a_1 n + a_2 n^2 + \dots$$

Certain regularity conditions on f are required. The Edgeworth expansion (and therefore the cumulants) of f should exist:

$$f(\mathbf{n}) = \exp \left\{ -(\kappa_1 - \mu)D + \frac{(\kappa_2 - \sigma^2)D^2}{2!} - \frac{\kappa_3 D^3}{3!} + \frac{\kappa_4 D^4}{4!} - \dots \right\} \varphi(x)$$

where $\varphi(x)$ is $N(\mu, \sigma^2)$, D is the differential operator and κ_r is the r^{th} cumulant of f . The r^{th} cumulant is required to be of order $O(\vartheta^{1-r})$, where ϑ is chosen so that f approaches Gaussian as $\vartheta \rightarrow \infty$; for example, ϑ

might be sample size. In general these conditions are not too severe, and are satisfied by most common densities.

Having met these conditions, the Cornish-Fisher expansion can be used to estimate the transformation. Using terms up to order $O(v^{-2})$, the approximate transformation is [20]:

$$v \approx g(n) = a_0 + a_1n + a_2n^2 + a_3n^3 + a_4n^4 + a_5n^5 \quad (B1)$$

where

$$\begin{aligned} a_0 &= \frac{\kappa_3}{6} - \frac{\kappa_5}{40} + \frac{5\kappa_3\kappa_4}{48} - \frac{13\kappa_3^3}{162} \\ a_1 &= 1 + \frac{\kappa_4}{8} - \frac{7\kappa_3^2}{36} - \frac{\kappa_6}{48} + \frac{35\kappa_4^2}{384} + \frac{17\kappa_3\kappa_5}{120} - \frac{19\kappa_3^2\kappa_4}{36} + \frac{2473\kappa_3^4}{7776} \\ a_2 &= -\frac{\kappa_3}{6} + \frac{\kappa_5}{20} - \frac{7\kappa_3\kappa_4}{24} + \frac{187\kappa_3^3}{648} \\ a_3 &= -\frac{\kappa_4}{24} + \frac{\kappa_3^2}{9} + \frac{\kappa_6}{72} - \frac{\kappa_4^2}{12} - \frac{2\kappa_3\kappa_5}{15} + \frac{547\kappa_3^2\kappa_4}{864} - \frac{907\kappa_3^4}{1944} \\ a_4 &= -\frac{\kappa_5}{120} + \frac{11\kappa_3\kappa_4}{144} - \frac{23\kappa_3^3}{216} \\ a_5 &= -\frac{\kappa_6}{720} + \frac{5\kappa_4^2}{384} + \frac{7\kappa_3\kappa_5}{360} - \frac{111\kappa_3^2\kappa_4}{864} + \frac{79\kappa_3^4}{648} \end{aligned}$$

EXAMPLES

The two densities used in the examples earlier in this chapter are considered here. To simplify the examples without loss of generality, assume that both f and φ have zero mean and unit variance.

The first density considered is the Laplace:

$$f(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|)$$

The odd cumulants are zero, and the even cumulants are given by

$$\kappa_{2n} = \frac{(2n)!}{n 2^n}$$

Thus

$$\kappa_1 = \kappa_3 = \kappa_5 = 0 \quad \text{and} \quad \kappa_2 = 1 \quad \kappa_4 = 3 \quad \kappa_6 = 30$$

From Eq.(B1) the approximate transformation is

$$g(n) = \frac{201}{128}n - \frac{11}{24}n^3 + \frac{29}{384}n^6$$

The exact transformation is

$$\nu = g(n) = \Phi^{-1} \left[1 - \frac{1}{2} \exp(-\sqrt{2}n) \right]$$

where Φ is the unit normal cdf. The exact and approximate transformations are plotted in Fig B1.

The second example is the Pearson VII:

$$f(x) = K(1 - (x/a)^2)^m$$

As in the examples earlier in this chapter, let $m = 7$. Then $a^2 = -11$ and the cumulants are

$$\kappa_1 = \kappa_3 = \kappa_5 = 0 \quad \text{and} \quad \kappa_2 = 1 \quad \kappa_4 = 2/3 \quad \kappa_6 = 13$$

Then from Eq. B1 the approximate transformation is:

$$g(n) + \frac{737}{864}n + \frac{25}{216}n^3 - \frac{53}{4320}n^5$$

The exact and approximate transformation is plotted in Fig. B2.

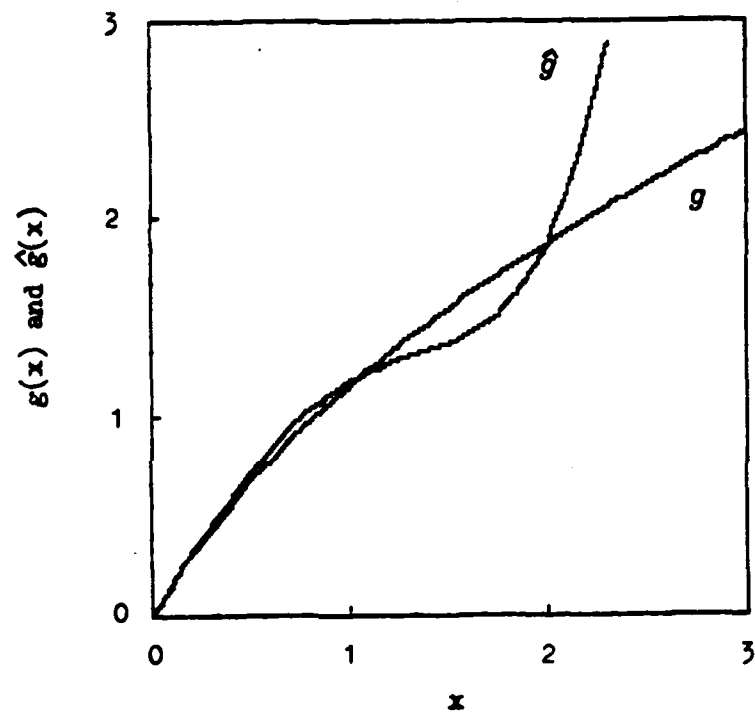


Fig. B1 Exact and approximate transformation for $n \sim \text{Laplace}$.

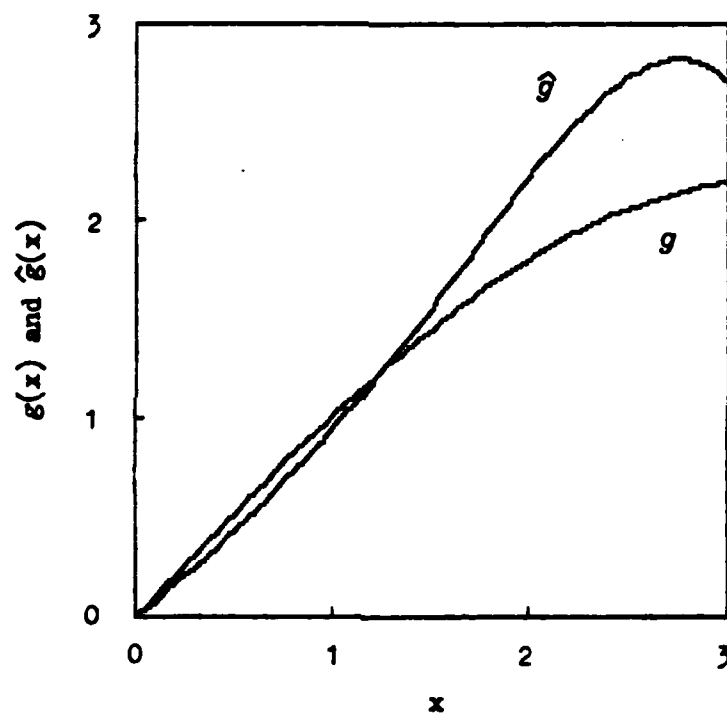


Fig. B2 Exact and approximate transformation for $n \sim$ Pearson VII.

REFERENCES

1. T.S. Ferguson, *Mathematical Statistics: a Decision Theoretic Approach*. New York: Academic Press, 1967.
2. S.A. Kassam, "Optimum Quantization for Signal Detection", *IEEE Trans. Comm.*, Vol. COM-25, May 1977, pp.479-484.
3. J.M. Miller & J.B. Thomas, "Detectors for Discrete time signals in non-Gaussian noise," *IEEE Trans. Inform. Theory*, Vol. IT-18, May 1972, pp.241-250.
4. J.J. Sheehy, "Optimum Detection of Signals in non-Gaussian Noise," *J. Acoust. Soc. Amer*, Vol. 63, Jan. 1978, pp.81-90.
5. H.V. Poor and J.B. Thomas, "Memoryless Discrete-Time Detection of a Constant Signal in m-Dependent Noise," *IEEE Trans. on Inform. Theory*, Vol. IT-25, No. 1, Jan. 1979, pp. 54-61.
6. N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley, 1972.
7. K.V. Mardia, *Families of Bivariate Distributions*. New York: Hafner, 1972.
8. D.K. McGraw and J.F. Wagner, "Elliptically Symmetric Distributions," *IEEE Trans. on Inform. Theory*, Vol. IT-14, Jan. 1968, pp. 110-120.
9. H.O. Lancaster, "The Structure of Bivariate Distribution," *Ann. Math. Stat.*, Vol. 34, 1963, pp. 532-538.
10. S.C. Schwartz, "A Series Technique for the Optimum Detection of Stochastic Signals in Noise," *IEEE Trans. Inform. Theory*, Vol. IT-15, May 1969, pp.362-369.
11. J.K. Ord, *Families of Frequency Distributions*. New York: Hafner, 1977.
12. G. Strang, *Linear Algebra and its Applications*. New York: Academic Press, 1980.
13. E.J.G. Pitman, *Some Basic Theory for Statistical Inference*. London: Chapman and Hall, 1979.
14. G.E. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day, 1970.
15. S.L. Bernstein et. al., "Long Range Communication at extremely low frequencies," *Proc. IEEE*, Vol. 62, March 1974, pp.292-312.
16. P. Mertz, "Model of impulsive noise for data transmission," *IEEE Trans. Comm. Sys.*, Vol. CS-9, June 1961, pp.130-137.
17. R.J. Marks, G.L. Wise, D.G. Haldeman and J.L. Whited, "Detection in Laplace Noise," *IEEE Trans. on Aero. and Elect. Sys.*, Vol. AES-14, No. 6, Nov. 1978, pp. 866-871.
18. S.T. Li and J.L. Hammond, "Generation of Pseudorandom Numbers with Specified Univariate Distributions and Correlation Coefficients," *IEEE Trans. on Sys., Man., and Cybernetics*, Vol. SMC-5, No. 5, Sept. 1975, pp. 557-561.

19. G.L. Wise, A.P. Traganitis, and J.B. Thomas, "The Effect of a Memoryless Nonlinearity on the Spectrum of a Random Process," *IEEE Trans. on Inform. Theory*, Vol. IT-23, Jan. 1977, pp. 84-89.
20. M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 1*. New York: Hafner, 1969.

CHAPTER 6 - NOISE MODELS FOR DETECTION†

INTRODUCTION

Problems in binary detection involve choosing between two statistical environments, described as an hypothesis H_0 and an alternative H_1 :

$$H_0: P_0(\mathbf{x})$$

$$H_1: P_1(\mathbf{x})$$

This decision is based on a statistic of an observation sequence $\mathbf{x}$ of length m . Various functionals of the vector $\mathbf{x}$ are traditionally employed to reduce the data to a scalar statistic easing the resulting detection process. This test statistic is compared to a threshold T to decide for H_0 or H_1 (Fig. 1). For the criteria of detector optimality often considered [1] (Neyman-Pearson, Bayes, probability of error) the Likelihood Ratio (LR) is known to be the optimal statistic of the observation $\mathbf{x}$. A second detector structure often used is the Locally Optimal or small signal detector (LO). Both the LR and LO detectors are considered here for several multivariate noise models.

The class of problems considered here is that of a deterministic signal in additive noise. H_0 and H_1 become

$$H_0: \mathbf{x} = \mathbf{n}$$

$$H_1: \mathbf{x} = \mathbf{n} + \vartheta \mathbf{s} \quad \vartheta > 0$$

Under H_0 the observation consists of noise $\mathbf{n}$ with density $f(\mathbf{n})$, and under H_1 the observation consists of a signal $\mathbf{s}$ with amplitude ϑ plus noise $\mathbf{n}$.

† This chapter was co-authored with Peter F. Swaszek, and also appears in his PhD dissertation *Robust Quantization, Vector Quantization and Detection*, Princeton University, October 1982.

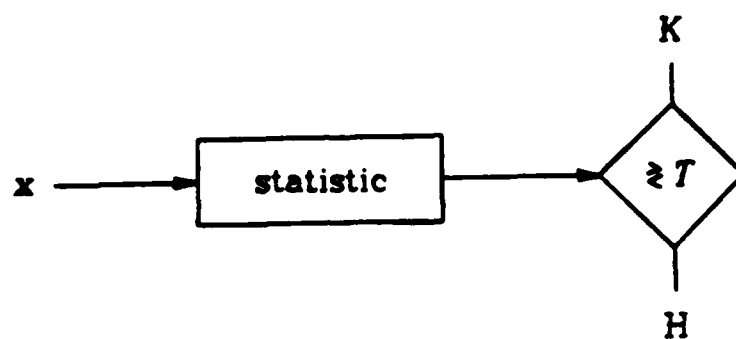


Fig. 1 - General detector structure.

The likelihood ratio is given by

$$LR(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \frac{f(\mathbf{x}-\vartheta \mathbf{s})}{f(\mathbf{x})} \quad (1)$$

When the amplitude constant ϑ is small and unknown, the locally optimal test statistic

$$LO(\mathbf{x}) = - \frac{\nabla f(\mathbf{x})}{f(\mathbf{x})} \cdot \mathbf{s} \quad (2)$$

is frequently employed [2].

Throughout this chapter a monotone function of LR or LO may be used in place of LR or LO when it simplifies a detector structure. This can be done without loss of generality since taking a monotonic function of both the test statistic and the threshold value does not effect the performance of the test. The log function is often used since the test statistic for an independent noise process reduces to a sum of zero memory nonlinear functions of the observations.

MULTIVARIATE DENSITIES

In this section several methods of generating families of multivariate noise densities are considered. Methods that produce large families of densities are of particular interest since they are more variable and may provide a closer fit to the actual noise. For each family of densities, LR and LO detectors are derived and discussed. It should be noted that some noise models produce structures which are more practical than others, and the practicality of each structure is considered.

CLOSED FORMS

There are a number of well known closed form multivariate densities

[3]. In general these are named for their marginals, which are often common univariate densities. Unfortunately, the marginal densities do not uniquely determine a multivariate density, and in many cases it is not even possible to make the conditional densities have the same form as the marginals. Although closed form multivariate densities may result from a particular model, more often they are the result of extending a characteristic of the univariate density. While they have the advantage of tractability, there is usually very little control over the dependency structure. *LR* and *LO* detectors as a rule contain nonlinearities with memory which may prove difficult to implement.

In the case of the multivariate Gaussian density

$$f(\mathbf{n}) = \frac{1}{(2\pi)^{m/2} |\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2} \mathbf{n}^T \mathbf{R}^{-1} \mathbf{n}\right\}$$

where $\mathbf{R}$ is the $m \times m$ covariance matrix, the resulting detectors (*LO* and *LR*) are both the matched filter:

$$LR(\mathbf{x}) = LO(\mathbf{x}) = \mathbf{s}^T \mathbf{R}^{-1} \mathbf{x}$$

The multivariate Cauchy density [4]:

$$f(\mathbf{n}) = \frac{K}{\left[c^2 + \mathbf{n}^T \mathbf{R}^{-1} \mathbf{n}\right]^{3/2}}$$

yields the test statistics:

$$LR(\mathbf{x}) = \frac{c^2 + \mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}}{c^2 + (\mathbf{x} - \mathbf{v}\mathbf{s})^T \mathbf{R}^{-1} (\mathbf{x} - \mathbf{v}\mathbf{s})}$$

and

$$LO(\mathbf{x}) = \frac{\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x}}{c^2 + \mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}}$$

Both contain non-linearities with memory.

DIFFERENTIAL EQUATIONS

One class of noise models is of particular interest in *LO* detection. The detector nonlinearity $g_{lo}(\mathbf{x}, \mathbf{a})$ parameterized by $\mathbf{a}$ is chosen for tractability. Eq.(2) can be viewed as a vector differential equation and solved to give a family of solution densities $f_i(\mathbf{x}, \mathbf{a})$. A member density of this family can then be fitted to the noise by choosing $\mathbf{a}$. This estimate of $\mathbf{a}$ can be inserted in g , and the result is a fitted *LO* nonlinearity.

This method has been applied when g_{lo} is assumed to be a rational function [5]. The resulting family of densities is the Pearson family which is particularly well adapted to using the method of moments to estimate $\mathbf{a}$.

This approach has the advantage of producing an adaptive detection system, but there is no guarantee that any member of the family of solution densities will be a good fit for a given noise process. The detector structure is pictured in Fig. 2.

ELLIPTICALLY SYMMETRIC

Another way to characterize multivariate densities is to require that the pdf have contours of constant height which are ellipsoids in m -space. These elliptically symmetric densities can be generated by replacing the independent variable of a univariate density, say $f_1(n)$, with a quadratic form $\mathbf{n}^T \mathbf{R}^{-1} \mathbf{n}$. The resulting density has the form:

$$f(\mathbf{n}) = K f_1\left[\sqrt{\mathbf{n}^T \mathbf{R}^{-1} \mathbf{n}}\right]$$

where K is a scaling constant and $\mathbf{R}$ is similar to a correlation matrix. However, the resulting multivariate density does not in general have as

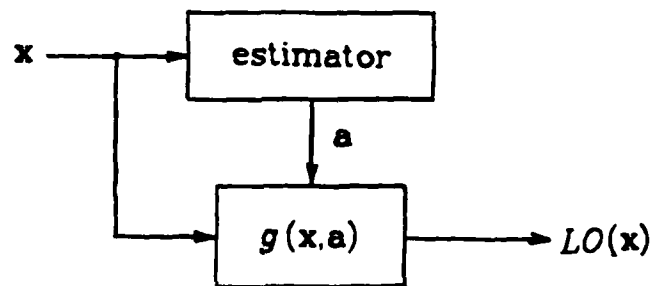


Fig. 2 - Differential equation LO detector

its marginals the univariate density $f_1(n)$.

There is a method of generating elliptically symmetric densities with a specific marginal $f_1(n)$. Assume the marginal has characteristic function $\varphi_1(u)$, and let the m -dimensional characteristic function be

$$\varphi_m(u) = \varphi_1\left[\sqrt{u^T R^{-1} u}\right]$$

Taking the inverse transform yields the m -dimensional density $f(n)$. Care must be taken to ensure that $f(n)$ integrates to unity and is positive for all n . Thus the marginals can be controlled but the dependency structure of n is not related to R in a simple or obvious way.

Since the argument of f is a quadratic form, the detector structures can be written as

$$LR(x) = \frac{f_1\left[\sqrt{(x-\vartheta s)^T R^{-1} (x-\vartheta s)}\right]}{f_1(\sqrt{x^T R^{-1} x})}$$

and

$$LO(x) = \frac{f'_1(\sqrt{x^T R^{-1} x})}{f_1(\sqrt{x^T R^{-1} x})} s^T R^{-1} x$$

Block diagrams of these systems appear in Fig. 3.

As a non-Gaussian example, consider an elliptically symmetric density with Laplace marginals [6]:

$$f(n) = \frac{1}{\pi |R|^{\frac{1}{2}}} K_0\left[\sqrt{2n^T R^{-1} n}\right]$$

The resulting detectors are

$$LR(x) = \frac{K_0\left[\sqrt{2(x-\vartheta s)^T R^{-1} (x-\vartheta s)}\right]}{K_0\left[\sqrt{2x^T R^{-1} x}\right]}$$

and

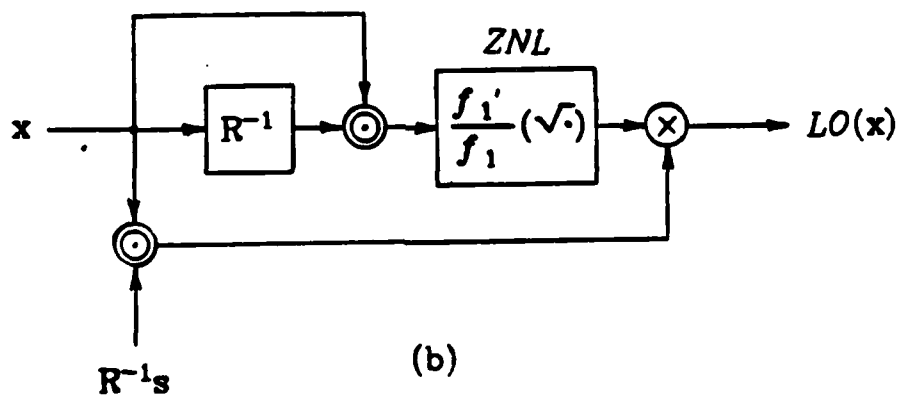
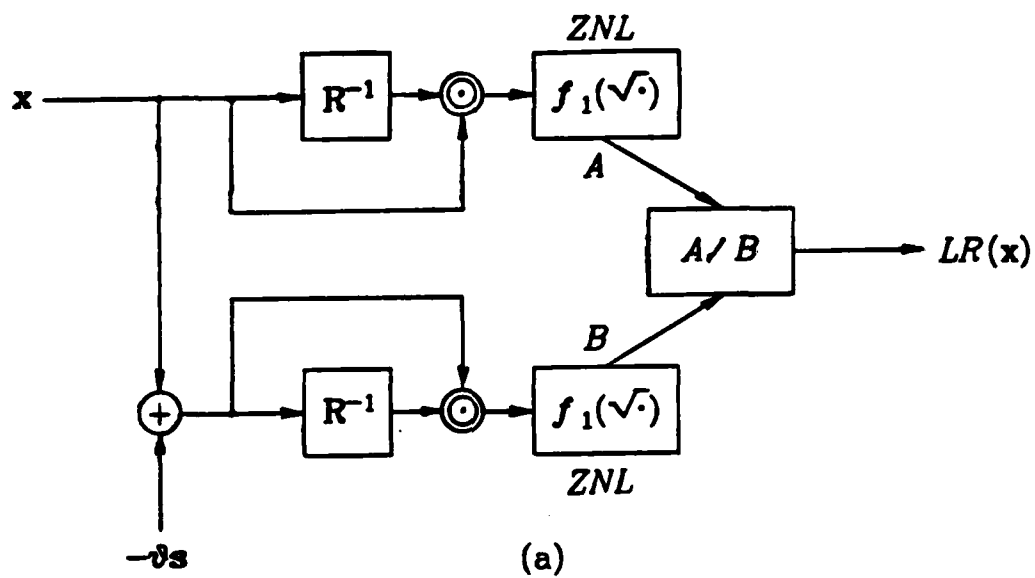


Fig. 3 - Elliptically symmetric LR (a) and LO (b) detectors.

$$LO(\mathbf{x}) = \left[\frac{2}{\mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}} \right]^{\frac{1}{2}} \frac{1}{|\mathbf{R}|} \frac{K_1 \left(\sqrt{2 \mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}} \right)}{K_0 \left(\sqrt{2 \mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}} \right)} (\mathbf{s}^T \mathbf{R}^{-1} \mathbf{x})$$

where K_0 and K_1 are modified Bessel functions. The test statistics both involve nonlinearities with memory.

SERIES

It is well known that under certain regularity conditions a density $f(\mathbf{n})$ with marginals $f_i(n_i)$ can be expanded in series form [7]:

$$f(\mathbf{n}) = \left[\prod_{i=1}^m f_i(n_i) \right] \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} C_{j_1, \dots, j_m} \prod_{k=1}^m \vartheta_{k, j_k}(n_k)$$

where the ϑ_{k, j_k} are sequences of orthonormal functions with generating functions f_j . The coefficients are given by:

$$C_{j_1, \dots, j_m} = \int f(\mathbf{n}) \prod_{k=1}^m \vartheta_{k, j_k}(n_k) d\mathbf{n}$$

Series forms have a certain theoretical elegance and can be used to derive general properties of densities. However, in practice, it is usually necessary to truncate the series, resulting in a poor representation and possibly even negative values in the tail regions of the pdf [8]. Even if this difficulty is over-looked, detector nonlinearities for truncated series are rational forms that are poorly behaved for observations falling in the tail regions.

TRANSFORMATION NOISE

Another class of multivariate densities are those generated by transformations from other densities [8]. Common univariate examples are the log-normal and the Johnson family. In a multivariate setting, this method involves a nonlinearity (possibly with memory) operating on a

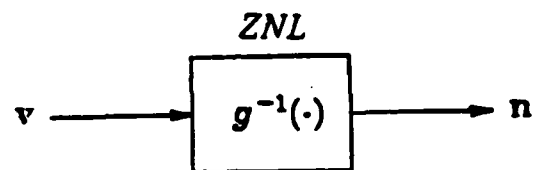


Fig. 4 - Transformation noise generation.

noise sequence with a known density. The intractability of this general class of transformations suggests constraining interest to those invertible transformations with zero memory (ZNL). A block diagram of the system generating this type of noise is given in Fig. 4. Given an identically distributed, m -dependent background noise process $\mathbf{v}$ with a known density $\varphi(\mathbf{v})$ and marginals $\varphi_1(v_i)$, a ZNL g can be selected to produce a noise $\mathbf{n}$ with the desired marginals $f_1(n_i)$. If φ_1 and g are fixed, the dependence structure of $\mathbf{n}$ is completely determined by the background noise dependency. This lack of flexibility in choosing a dependence structure may be a disadvantage in some cases.

The density of $\mathbf{n}$ is found with a change of variables:

$$f(\mathbf{n}) = \varphi[g(\mathbf{n})] \prod_{i=1}^m |g'(n_i)|$$

where

$$g(\mathbf{n}) = [g(n_1), g(n_2), \dots, g(n_m)]^T$$

The transformation ZNL g can be found by equating the marginal cdf's of $\mathbf{n}$ and $\mathbf{v}$ ($F_1(n_i)$ and $\Phi_1(v_i)$ respectively):

$$g(n_i) = \Phi^{-1}[F_1(n_i)] \quad (3)$$

The LR detector is found from Eq.(1):

$$LR(\mathbf{x}) = \log \frac{\varphi[g(\mathbf{x}-\mathbf{v}\mathbf{s})]}{\varphi[g(\mathbf{x})]} + \sum_{i=1}^m \log \frac{g'(x_i - \mathbf{v}s_i)}{g'(x_i)}$$

which is the φ -noise LR detector added to a sum of the observations through a zero memory non-linearity. The nonlinearity g' is found by differentiating Eq.(3)

$$g'(x_i) = f_1(x_i) / \varphi_1[g(x_i)]$$

The block diagram of this detector appears in Fig. 5a.

The locally optimal detector $LO(\mathbf{x})$ is found from Eq.(2) to be

$$LO(\mathbf{x}) = - \left[\frac{\nabla \varphi}{\varphi} [g(\mathbf{x})] g'(\mathbf{x}) + \frac{g''(\mathbf{x})}{g'(\mathbf{x})} \right] \cdot \mathbf{s}$$

A block diagram of $LO(\mathbf{x})$ is also given to Fig. 5b where the symbol $\otimes$ is an element by element vector multiplication and $\odot$ is a vector dot product. The system consists of the locally optimal nonlinearity with memory (for signal in φ distributed noise), three ZNL's and a correlator followed by a threshold comparator.

It can be seen that if the background process, φ is chosen to be Gaussian with correlation matrix R_φ , the LR system reduces to

$$LR(\mathbf{x}) = \mathbf{x}^T R_\varphi^{-1} \mathbf{s} + \sum_{i=1}^m \log \frac{f_1(\mathbf{x}_i - \vartheta \mathbf{s}_i)}{f_1(\mathbf{x}_i)} + \frac{1}{2} \sum_{i=1}^m [g^2(\mathbf{x}_i - \vartheta \mathbf{s}_i) - g^2(\mathbf{x}_i)]$$

which consists of a matched filter and zero memory nonlinearities (Fig. 6a). This can be considerably easier to realize than the nonlinearities with memory required by other schemes.

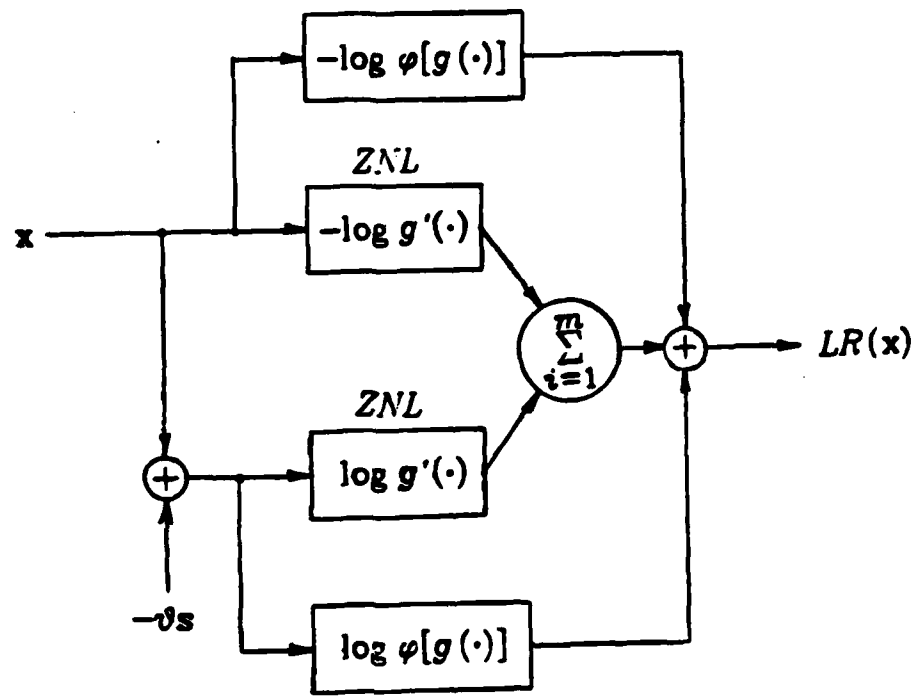
For the LO detector, the third nonlinearity is

$$-\frac{g''}{g'}(\mathbf{x}_i) = -\frac{f_1'}{f_1}(\mathbf{x}_i) + g'(\mathbf{x}_i)g(\mathbf{x}_i)$$

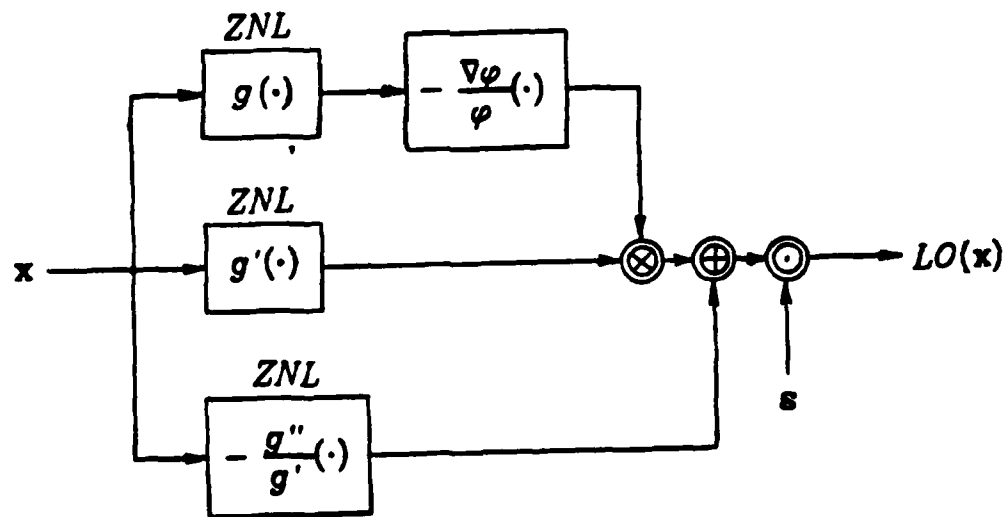
The locally optimal nonlinearity, with memory, is a linear filter

$$-\frac{\nabla \varphi}{\varphi}(\mathbf{x}) = R_\varphi^{-1} \mathbf{x}$$

Therefore, the LO system can be reconfigured as shown in Fig. 6b. The third ZNL g_{lo} is the locally optimal nonlinearity for signal in iid, f_1 distributed noise as given in Eq.(2). Thus, under the assumption of transformed unit variance Gaussian noise, the system consists of one linear filter and three ZNL's, one of which is the nonlinearity occurring in the iid locally optimal detector.



(a)



(b)

Fig. 5 - Transformation noise LR (a) and LO (b) detectors.

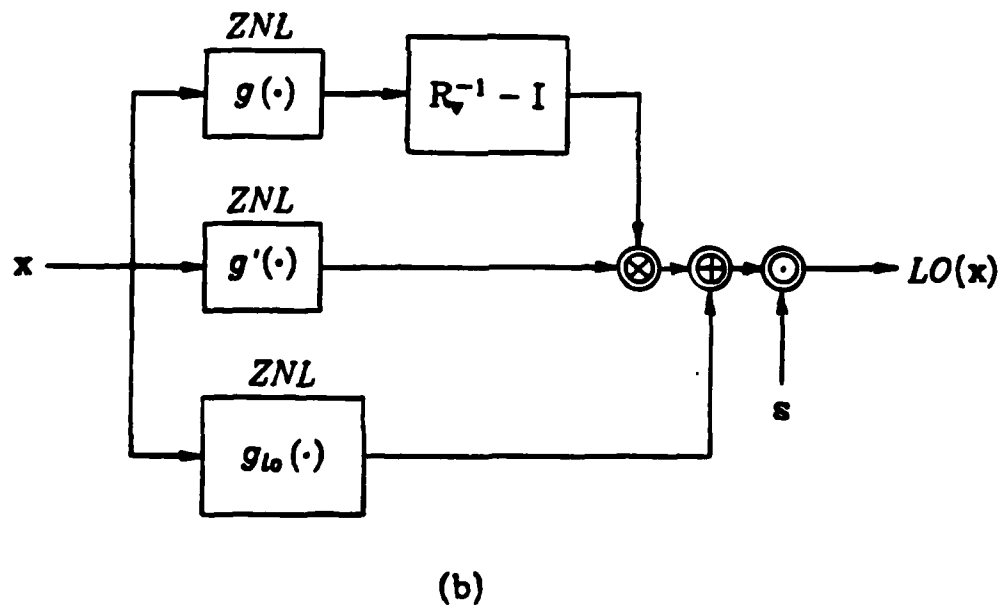
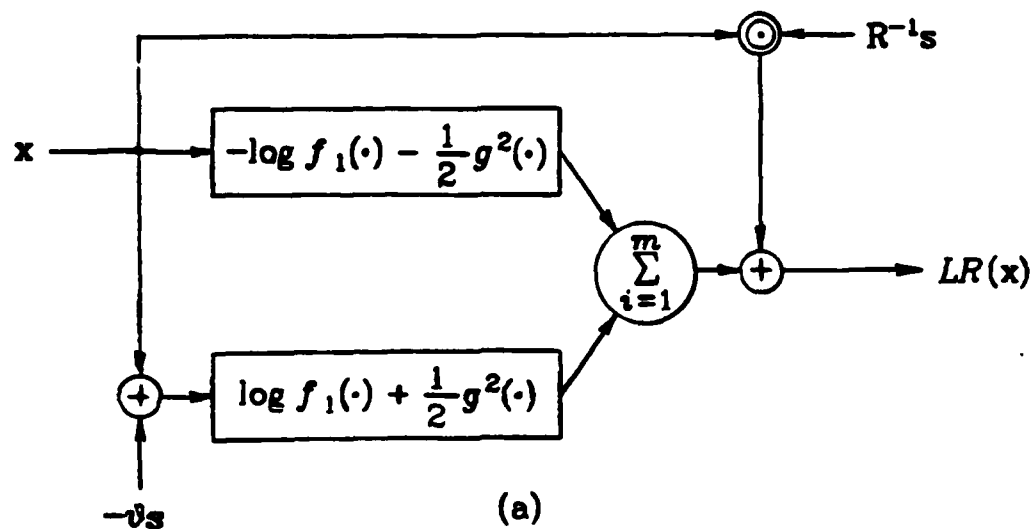


Fig. 6 - Transformed Gaussian LR(a) and LO (b) detectors.

Under the assumed noise model, the detector designs require knowledge of the nonlinear transformation g and R_v , the covariance matrix of the background Gaussian vector v . In practice, however, the available information is often the noise marginal density and the noise covariance matrix R_n . The transformation nonlinearity g is found from the marginal cdf with Eq.(3). Computing R_v from R_n in general requires a numerical integration solution [9,10].

If the input v has a multivariate Gaussian distribution, this transformation can be modeled as shown in Fig.7 where z is an iid $N(0,1)$ random vector, and L is a linear operator. It is well known that any symmetric matrix R_v can be factored by Crout resolution [11] to give the form:

$$R_v = LL^T$$

where L is an unique lower triangular (hence causal) matrix. The linear filter L in Fig.7 can be therefore chosen to produce a desired covariance in v :

$$E[vv^T] = E[(Lz)(Lz)^T] = L E[zz^T] L^T = LL^T = R_v$$

Modeling noise in this way suggests the second model shown in Fig. 8. The input z , iid $N(0,1)$, is passed through the ZNL g^{-1} to produce the iid vector w with non-Gaussian marginals and variance σ^2 . Passing w through a linear filter L of length m produces the sequence n with a dependence structure generated by L . This model is typically called a Moving Average process (MA) of order $m-1$ [12]. As above, Crout factorization can be used to solve for an unique lower triangular matrix L , given R_n , since the covariance matrix of n is given by

$$R_n = E[nn^T] = E[(Lw)(Lw)^T] = L E[ww^T] L = \sigma^2 LL^T$$

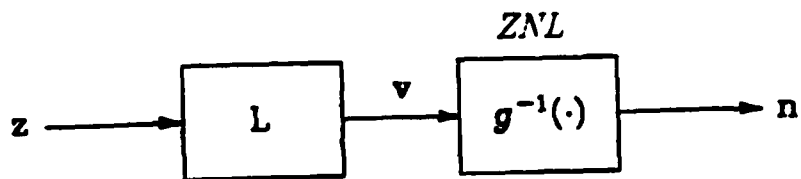


Fig. 7 - Gaussian background model.

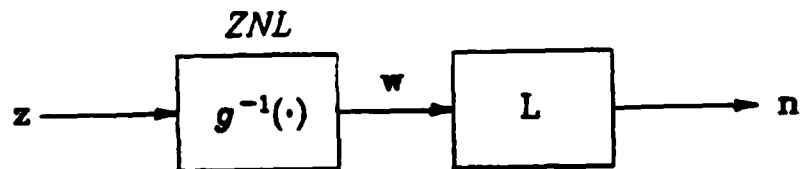


Fig. 8 - Moving average noise model.

The major difficulty in using this model is selecting the marginal density of $\mathbf{w}$ which yields the desired marginal for $\mathbf{n}$. Denoting the $\mathbf{w}$ marginal as $\varphi_1(\mathbf{w}_i)$, the density of the MA noise $\mathbf{n}$ is

$$f(\mathbf{n}) = |L^{-1}| \prod_{i=1}^m \varphi_1[(L^{-1}\mathbf{n})_i]$$

The LR detector is

$$LR(\mathbf{x}) = \sum_{i=1}^m \log \left[\frac{\varphi_1[L^{-1}(\mathbf{x} - \mathbf{0}\mathbf{s})]_i}{\varphi_1[L^{-1}\mathbf{x}]_i} \right]$$

The locally optimal detector $LO(\mathbf{x})$ is found from Eq.(2) to be

$$LO(\mathbf{x}) = (\mathbf{L}^{-1}\mathbf{s})^T g_{lo}(\mathbf{L}^{-1}\mathbf{x})$$

where g_{lo} is the ZNL in the iid locally optimal detector for φ_1 marginal:

$$g_{lo}(x_i) = - \frac{\varphi_1'(x_i)}{\varphi_1(x_i)}$$

Block diagrams of these detectors are given in Fig. 9. The LO system is a ZNL embedded in a matched filter. Note that for $\mathbf{n}$ iid, L is an identity matrix, and the resulting LO detector is the iid locally optimal detector $\mathbf{s}^T g_{lo}(\mathbf{x})$. For the Gaussian case ($g^{-1}(x)=x$) the result is the matched filter (linear detector) $\mathbf{s}^T \mathbf{R}_n^{-1} \mathbf{x}$.

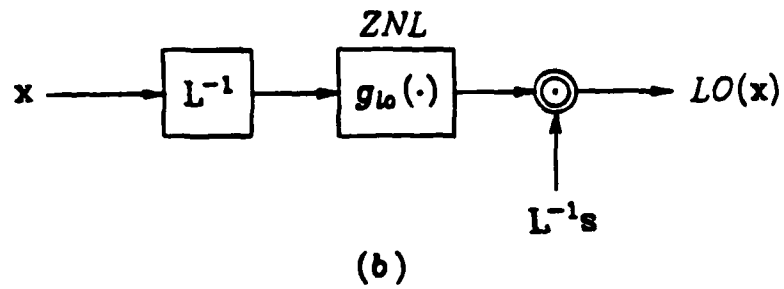
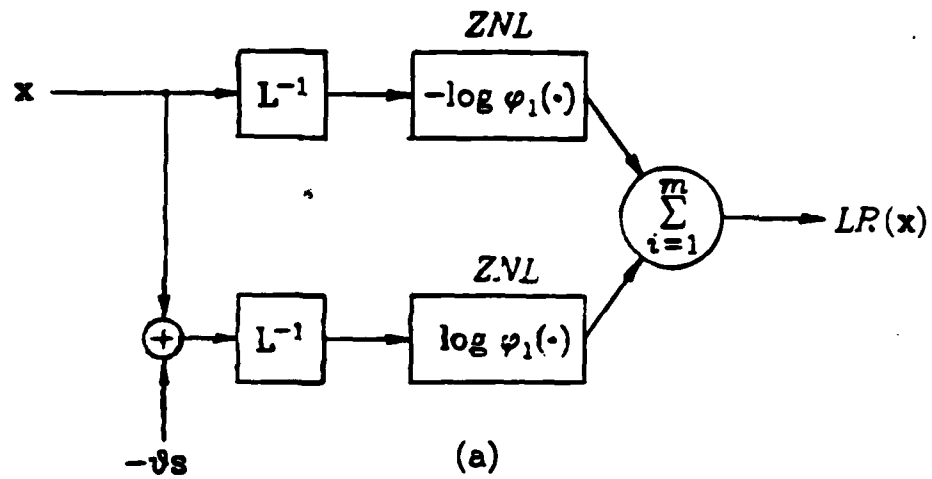


Fig. 9 - Moving average LR (a) and LO (b) detectors.

CONCLUSIONS

This chapter presents several multivariate noise densities and their associated *LO* and *LR* detector structures.

- 1- Closed form densities are tractable but are not general in any sense, and there is little or no control over the dependency structure.
- 2- The differential equation method applies only in the *LO* case. There is a great deal of control over the functional form of the detector nonlinearity, but there is no guarantee that the corresponding family of solution densities provides a good fit to a given noise process.
- 3- The elliptically symmetric densities, although yielding detectors containing nonlinearities with memory, are not difficult to implement. However, the meaning of the matrix *R* in the quadratic form is unclear and the dependence is difficult to control.
- 4- The series forms have difficulties in implementation due to the necessity of truncation. The generally slow convergence of series, and the poor tail behavior usually noted when the series is truncated result in very poor behavior of the detector nonlinearity in the tail region.
- 5- The transformation model allows selection of the noise marginals but increases the complexity of calculating the output dependency. The MA model has a more straightforward dependence structure but solving for the marginals is more difficult. Note that these two schemes are equivalent only when g^{-1} is linear (*n* is multivariate Gaussian).

REFERENCES

1. T.S. Ferguson, *Mathematical Statistics: a Decision Theoretic Approach*. New York: Academic Press, 1967.
2. E.J.G. Pitman, *Some Basic Theory for Statistical Inference*. London: Chapman and Hall, 1979.
3. N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley, 1972.
4. K.V. Mardia, *Families of Bivariate Distributions*. New York: Hafner, 1972.
5. A.B. Martinez & J.B. Thomas, "Robust Detection Using a Pearson Fit to Nearly Gaussian Noise," *Proceedings of the 15-th Johns Hopkins CISS*, March 1981, pp.125-130.
6. D.K. McGraw and J.F. Wagner, "Elliptically Symmetric Distributions," *IEEE Trans. on Inform. Theory*, Vol. IT-14, Jan. 1968, pp. 110-120.
7. H.O. Lancaster, "The Structure of Bivariate Distribution," *Ann. Math. Stat.*, Vol. 34, 1963, pp. 532-538.
8. J.K. Ord, *Families of Frequency Distributions*. New York: Hafner, 1977.
9. G.L. Wise, A.P. Traganitis, and J.B. Thomas, "The Effect of a Memoryless Nonlinearity on the Spectrum of a Random Process," *IEEE Trans. on Inform. Theory*, Vol. IT-23, Jan. 1977, pp. 84-89.
10. S.T. Li and J.L. Hammond, "Generation of Pseudorandom Numbers with Specified Univariate Distributions and Correlation Coefficients," *IEEE Trans. on Sys., Man., and Cybernetics*, Vol. SMC-5, No. 5, Sept. 1975, pp. 557-561.
11. G. Strang, *Linear Algebra and its Applications*. New York: Academic Press, 1980.
12. G.E. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day, 1970.

CHAPTER 7 - FINITE LENGTH DISCRETE MATCHED FILTERS

INTRODUCTION

The design and implementation of the Matched Filter (MF) has received considerable attention [4-8]. As a detector it has the advantage of linearity, and since it is based only on easily estimated 2nd order noise statistics, the MF is simple to optimize. The performance criterion, the Signal-to-Noise Ratio (SNR) is tractable, and intuitively appealing.

For a fixed signal in discrete time, Levinson [8] has presented a simple and efficient algorithm to solve the MF equation. Since the MF impulse response and the SNR are computed iteratively, the algorithm can be terminated when a filter with desired performance is found. Unfortunately, when there is some freedom in choosing a signal, the choice of signal plays an important part in optimizing the detector. Because the optimal signal of length M is a truncated version of the optimal length $M+1$ signal under only very special conditions, the Levinson algorithm must be repeated N times, and thus loses its computational advantage. In this chapter, easily computed bounds on the performance of the MF as a function of length are found. Then, before any attempt is made to solve the MF equation, an estimate of the filter length can be found from these bounds.

DETECTION PROBLEM

The detection problem considered in this chapter is one of finding a linear detector that discriminates between an hypothesis H_0 and an

alternative H_1 . The decision is based on a discrete length N observation vector $\mathbf{x}$ composed under H_0 of noise $\mathbf{n}$ with density f and under H_1 of a known signal $\mathbf{s}$ in noise:

$$H_0: \mathbf{x} = \mathbf{n}$$

$$H_1: \mathbf{x} = \mathbf{n} + \mathbf{s}$$

The detector consists of a real scalar test statistic $T(\mathbf{x})$, a functional of the observation $\mathbf{x}$, which is compared to a scalar threshold to decide for H_0 or H_1 .

The criterion of detector optimality used in this chapter is a SNR measure often called the *deflection*:

$$\text{SNR} = \frac{[E_1(T) - E_0(T)]^2}{\text{Var}_0(T)} \quad (1)$$

where E_0 and E_1 are the expectation under H_0 and H_1 , and Var_0 is the variance under H_0 .

It is well known that the log likelihood ratio detector for Gaussian noise is linear, the matched filter. Since the detector power is a monotone increasing function of the SNR of T , the SNR is frequently used as a measure of detector performance. The SNR, outside of its intuitive appeal, is often justified by making a Gaussian assumption about $\mathbf{n}$ or applying the central limit theorem to T .

Using the MF as a detector for non-Gaussian noise is more difficult to justify. In general, the likelihood ratio detector maximizes the SNR [1], and by a simple calculus of variations argument, maximizing the SNR (as defined above) with no restriction on the linearity of the detector can be shown to yield a linear function of the likelihood ratio. The MF is the

linear filter which maximizes output SNR, but the likelihood ratio is generally nonlinear. Therefore in making a restriction of linearity on T , it is tacitly assumed that the noise is Gaussian or nearly Gaussian in the sense that the MF performs reasonably well and that any loss of optimality is compensated for by the simplicity and linearity of the MF.

Under the assumption of linearity, the test statistic $T(\mathbf{x})$ is equal to the output at time N of a linear filter with impulse response $\mathbf{h}$. As a convenience, the pseudo-signal is defined to be a length N vector with elements $u_i = h_{N+1-i}$, the filter impulse response in reverse order. The output SNR of the linear detector is found from Eq. (1) to be

$$\text{SNR}_o = \frac{\langle \mathbf{u} | \mathbf{s} \rangle^2}{\langle \mathbf{u} | \mathbf{R} \mathbf{u} \rangle} \leq \lambda$$

where $\langle | \rangle$ is standard inner product notation, $\mathbf{R}$ is the noise covariance matrix, and λ is the maximum value of SNR for the optimal pseudo-signal. Cross-multiplication yields

$$L(\mathbf{u}) = \langle \mathbf{u} | \mathbf{s} \rangle^2 - \lambda \langle \mathbf{u} | \mathbf{R} \mathbf{u} \rangle \leq 0$$

This can be maximized in the usual way by setting its gradient equal to zero:

$$\nabla L(\mathbf{u}) = 2\langle \mathbf{u} | \mathbf{s} \rangle \mathbf{s} - 2\lambda \mathbf{R} \mathbf{u} = 0$$

Rearranging and noting that $\lambda / \langle \mathbf{u} | \mathbf{s} \rangle$ is a constant and can be set equal to unity with no loss of generality, the result is the well known MF equation:

$$\mathbf{s} = (\lambda / \langle \mathbf{u} | \mathbf{s} \rangle) \mathbf{R} \mathbf{u} = \mathbf{R} \mathbf{u} \quad (2)$$

The solution of Eq. (2) is the pseudo-signal of the MF:

$$\mathbf{u} = \mathbf{R}^{-1} \mathbf{s}$$

with output SNR given by

$$\text{SNR}_0 = \langle u | s \rangle = \langle s | R^{-1} s \rangle$$

In discrete time and with a fixed signal, the MF matrix equation can be solved quite efficiently using the Levinson algorithm. In continuous time, the classical method of solution is to use spectral factorization to solve the equation on an infinite interval; this (possibly) non-causal solution is then projected onto a causal space [3]. In discrete time, there is a parallel spectral approach using the eigenvectors and values of R .

SIGNAL SELECTION AND BOUNDS ON THE SNR

It is well known that the MF is the linear filter with the maximum SNR_0 for a given signal in noise. In addition for non-white noise, the SNR_0 of the MF can be maximized by proper choice of signal shape. Because of this, for signals of constant energy, the SNR_0 of the MF has a range of possible values.

Since the $N \times N$ covariance matrix R is positive definite and Hermitian, it has positive, real eigenvalues:

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$$

and a corresponding set of orthonormal eigenvectors:

$$e_1, e_2, \dots, e_N$$

The matrix R can be diagonalized:

$$R = E \Lambda E^{-1}$$

where E is the eigenvector matrix:

$$E = [e_1, e_2, \dots, e_N]$$

and Λ is a matrix of eigenvalues:

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix}$$

Likewise R^{-1} has the diagonal form:

$$R^{-1} = E\Lambda^{-1}E^{-1}$$

where Λ^{-1} has as its diagonal elements the eigenvalues of R^{-1} :

$$\frac{1}{\lambda_1} \geq \frac{1}{\lambda_2} \geq \dots \geq \frac{1}{\lambda_N}$$

Thus the MF equation has the solution:

$$u = E\Lambda^{-1}E^{-1}s \quad (3)$$

The signal s can be expanded in terms of the eigenvectors:

$$s = Ec \quad (4)$$

where c is the coordinate of s under the basis formed by the orthonormal eigenvectors of R . From Eqs. (3) and (4) the pseudo-signal is

$$u = E\Lambda^{-1}c \quad (5)$$

The SNR_o of the MF becomes

$$SNR_o = c^T \Lambda^{-1} c$$

If the signal is chosen to be in the eigenspace of the i th eigenvector ($s = k e_i$), then the MF is a simple correlator ($u = s$) and

$$SNR_o = k / \lambda_i$$

The Rayleigh quotient theorem [2] states that

$$\frac{1}{\lambda_N} \leq \frac{\langle s | R^{-1} s \rangle}{\|s\|^2} \leq \frac{1}{\lambda_1}$$

where the upper and lower bounds are achieved for a signals in the eigenspace of e_1 and e_N respectively. Thus the SNR of the MF is bounded:

$$\frac{\|s\|^2}{\lambda_N} \leq SNR_o \leq \frac{\|s\|^2}{\lambda_1} \quad (6)$$

The best choice of signal is e_1 , the eigenvector of R with the smallest eigenvalue. This is equivalent to putting the signal in that part of the spectrum of R where the noise has the smallest magnitude.

Grettenberg [7] has taken the logical step of using M eigenvectors as an M character alphabet of signals. By choosing the eigenvectors of R corresponding to the smallest M eigenvalues, not only is the set orthogonal, but it achieves the greatest minimum SNR_0 of any such M character set. This also has an advantage of simplicity, since, when the signal is chosen to be an eigenvector of R , from Eq. (9) the pseudo-signal equals the signal, and the MF reduces to a simple correlator.

A minimax strategy is used by Turin [5] to find the worst-case noise, and the corresponding best signals in continuous time. He shows that the best signal spectrum should consist of the noise spectral components with the smallest magnitude. As a consequence, the worst spectra and the best signal both have flat spectra.

LEVINSON ALGORITHM AND OPTIMAL SIGNAL SELECTION

For the Levinson algorithm to produce the s -optimal MF on each iteration, the length N optimal eigenvector $e^{(N)}$ has to be a truncated version of the length $N+1$ eigenvector $e^{(N+1)}$.

$$e^{(N+1)} = \begin{bmatrix} e^{(N)} \\ e_{N+1} \end{bmatrix}$$

Let $R^{(N+1)}$ be the $(N+1) \times (N+1)$ covariance matrix with elements $r_{|i-j|}$; then

$$R^{(N+1)} e^{(N+1)} = \lambda^{(N+1)} e^{(N+1)}$$

where $\lambda^{(N+1)}$ is the eigenvalue corresponding to the eigenvector $e^{(N+1)}$.

Noting that the $N \times N$ minor of the covariance matrix $R^{(N+1)}$ is $R^{(N)}$:

$$R^{(N+1)} e^{(N+1)} = \lambda^{(N+1)} e^{(N+1)} = \begin{bmatrix} R^{(N)} & r^{(N)} \\ (r^{(N)})^T & r_0 \end{bmatrix} \begin{bmatrix} e^{(N)} \\ e_{N+1} \end{bmatrix} = \begin{bmatrix} \lambda^{(N)} e^{(N)} \\ (r^{(N)})^T e^{(N)} \end{bmatrix} + e_{N+1} \begin{bmatrix} \lambda^{(N)} \\ r_0 \end{bmatrix}$$

where

$$\mathbf{r}^{(N)} = [\tau_N \ \tau_{N-1} \ \dots \ \tau_1]^T$$

For $\mathbf{e}^{(N)}$ to be a truncated version of $\mathbf{e}^{(N+1)}$, then for all $N \geq 1$:

$$(\lambda^{(N+1)} - \lambda^{(N)})\mathbf{e}^{(N)} = \mathbf{e}_{N+1}\mathbf{r}^{(N)}$$

Solving this equation iteratively yields permissible autocovariance sequences. Let τ_p be the first nonzero term in the covariance sequence after τ_0 . Then every p th term with index less than L is nonzero, and the rest are zero. All nonzero terms have the same magnitude with alternating or constant sign. The covariance sequences have the form:

$$\tau_{ip+k} = \begin{cases} \tau_p [\text{sgn}(\tau_p)]^{i-1} & k = 0 \\ 0 & k = 1, 2, \dots, p-1 \text{ or } ip+k \geq L \end{cases}$$

where $i \geq 0$, $0 \leq k \leq p-1$, $p \geq 0$, and $L \geq 0$. As a special case, if $L \rightarrow 0$ or $p \rightarrow \infty$ the result is white noise, $\tau_i = 0$ for all $i \geq 1$.

This places a severe restriction on the noise autocovariance sequence, and places corresponding limits on the utility of the Levinson algorithm for this particular problem.

APPROXIMATE BOUNDS ON THE SNR

It is impractical to find a suitable filter length N from the bounds in Eq. (6) since they require knowledge of the eigenvalues of each $M \times M$ minor of $\mathbf{R}$. Looser but easier to compute bounds can be found.

The equivalent rectangular time duration ΔT of the noise autocovariance is introduced as a rough measure of correlation [3]:

$$\Delta T = \sum_{i=-\infty}^{\infty} \frac{|\tau_i|}{\sigma^2}$$

where $\sigma^2 = \tau_0$ is the noise variance. The largest eigenvalue of $\mathbf{R}$, denoted

by λ_N is well known to be the smallest norm of R , thus using another norm:

$$\lambda_N \leq \|R\|_\infty = \max_i \sum_j r_{ij} \leq \sigma^2 \Delta T$$

This yields the looser bound:

$$\frac{\|s\|^2}{\sigma^2 \Delta T} \leq \frac{\|s\|^2}{\max_i \sum_j r_{ij}} \leq \frac{\|s\|^2}{\lambda_N} \leq \text{SNR}_0$$

An upper bound can be found. The condition number K of a matrix is defined as:

$$K = \lambda_{\max} / \lambda_{\min} = \|R^{-1}\| \|R\|$$

then from Eq. (6)

$$\text{SNR} \leq \|s\|^2 K / \lambda_{\max}$$

The trace of R equals the sum of its eigenvalues:

$$\text{tr}(R) = N\sigma^2 = \sum_{i=1}^N \lambda_i$$

therefore

$$\sigma^2 \leq \lambda_{\max} \leq N\sigma^2$$

and so

$$\frac{\|s\|^2}{\sigma^2 \Delta T} \leq \text{SNR}_0 \leq \frac{\|s\|^2 K}{\sigma^2} \quad (7)$$

Since the input SNR is given by

$$\text{SNR}_i = \frac{\|s\|^2}{N\sigma^2}$$

the improvement in SNR of the MF is given by

$$\frac{N}{\Delta T} \leq \frac{\text{SNR}_0}{\text{SNR}_i} = \text{SNR}_{MF} \leq NK \quad (8)$$

EXAMPLES

As a first example, consider white noise with an $N \times N$ covariance matrix $R = \sigma^2 I$. The covariance matrix has an N th order eigenvalue σ^2 making the upper and lower bounds in Eq. (6) equal, and the choice of signal arbitrary. Other considerations, such as a ceiling on transmitted signal strength, may still make the spreading of signal energy in time desirable.

For $N = 3$, the autocorrelation sequence is given by

$$r = [1 \ r_1 \ r_2]^T$$

For R to be positive definite, the values that r_1 and r_2 can take are restricted:

$$|r_1|, |r_2| < 1 \quad \text{and} \quad |r_1| < \sqrt{(r_2+1)/2}$$

This region of the $r_1 \times r_2$ plane is shown in Figs. 1 and 2.

The difference in dB between the upper and lower SNR bounds in Eq. (6) is plotted as contours in Fig. 1. Even for a filter this short, the signal selection is shown to be quite important.

In Fig. 2, the SNR of the MF for an alternating signal ($s_i = (-1)^i$) is shown in dB over the lower bound. The alternating signal was chosen as a suboptimal approximation to the optimal signal because of its simplicity, and similarity in shape to the optimal signal for $r_1 \geq 0$. It is readily seen to be nearly optimal in this case. Because of the symmetry of this problem, a constant signal ($s_i = 1$), chosen as a suboptimal signal for $r_1 < 0$ has performance contours which are the mirror image of those in Fig. 2.

Four noise autocorrelation functions were chosen as representative — the exponential:

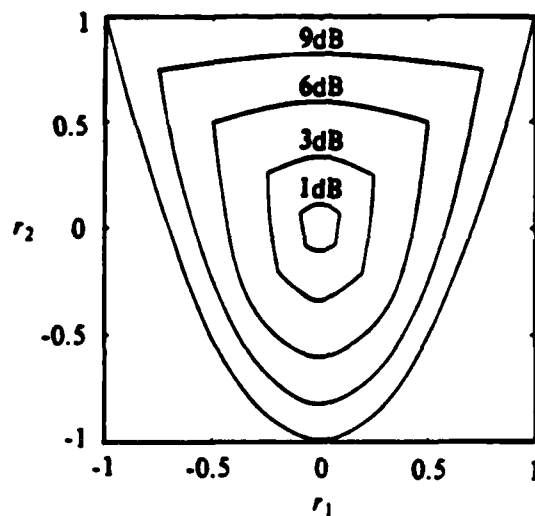


Fig. 1. Contours of $SNR_{u,}$ upper bound for $N=3$ in $r_1 \times r_2$ plane.

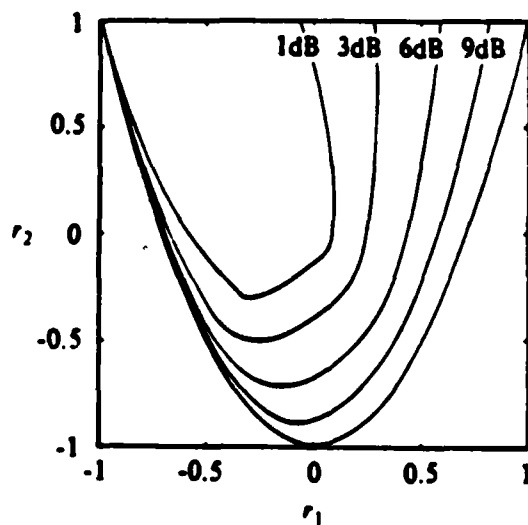


Fig. 2. Contours of $SNR_{u,}$ for alternating signal in $r_1 \times r_2$ plane.

$$\tau_i = \exp(-0.2 |i|)$$

the triangular:

$$\tau_i = \begin{cases} 1 - |i|/10 & |i| \leq 10 \\ 0 & |i| > 10 \end{cases}$$

the Gaussian:

$$\tau_i = \exp(-\pi(i/10)^2)$$

and the hyperbolic secant:

$$\tau_i = \text{sech}(\pi i/10)$$

The exponential is the simplest member of the Markov class; the triangular function has finite support; the Gaussian correlation function has infinite support, yet has tails which fall off faster than the exponential, and the hyperbolic secant has a nearly Gaussian shape at the origin, but exponential tails.

The upper and lower bounds on the SNR_{MF} from Eq. (6) are plotted in dB versus filter length N in Figs. 3-6. Here signal selection is extremely important for all $N > 2$ and increasingly so for increasing N . Even for the length 5 filter, the difference between the best and worse-case SNR_{MF} is at least 15dB for all four cases. At $N = 20$, the difference is at least 19dB.

The parameters of these four correlation functions were chosen so that each has an equivalent rectangular time duration of $\Delta T \approx 10$. Thus the approximate lower bound (Eq. 8) for each function is

$$N/10 \leq \text{SNR}_{MF}$$

The approximate and exact lower bounds for each correlation function is shown in Figs. 7-10.

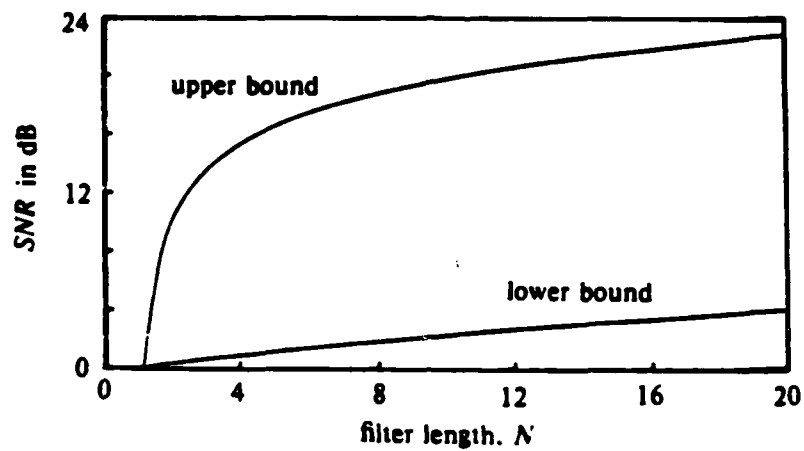


Fig. 3. Upper and lower bounds on SNR_{MF} for exponential correlation.

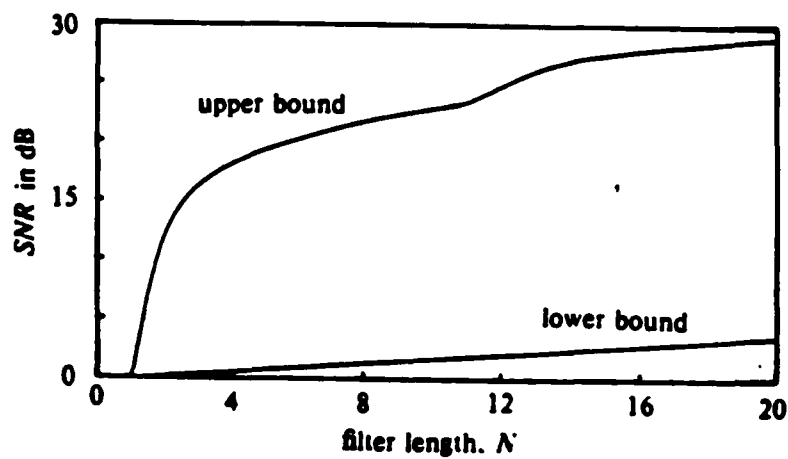


Fig. 4. Upper and lower bounds on SNR_{MF} for triangular correlation.

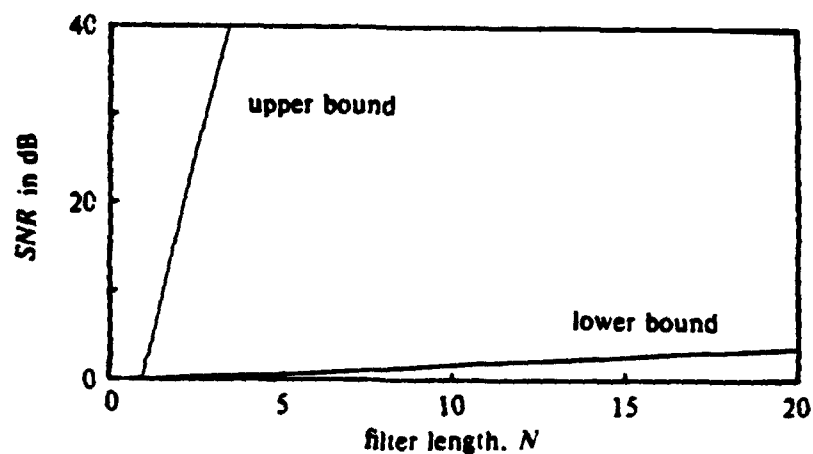


Fig. 5. Upper and lower bounds on SNR_{MF} for Gaussian correlation.

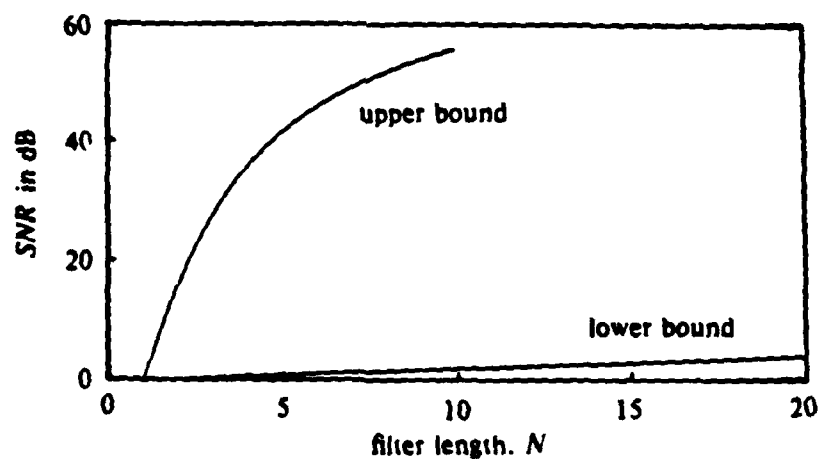


Fig. 6. Upper and lower bounds on SNR_{MF} for hyperbolic secant correlation.

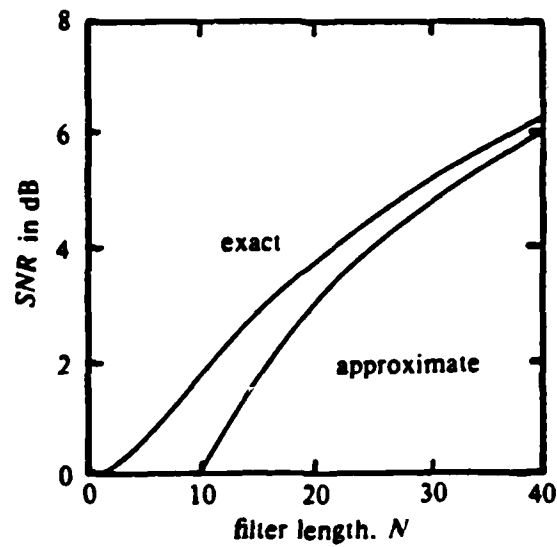


Fig. 7. Approximate and exact lower bound for exponential correlation.

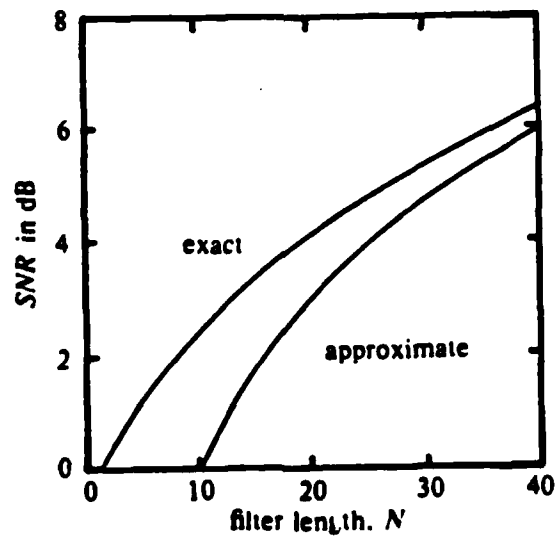


Fig. 8. Approximate and exact lower bound for triangular correlation.

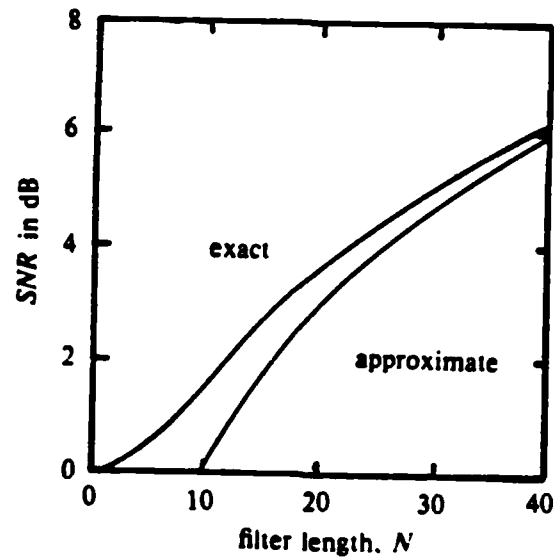


Fig. 9. Approximate and exact lower bound for Gaussian correlation.

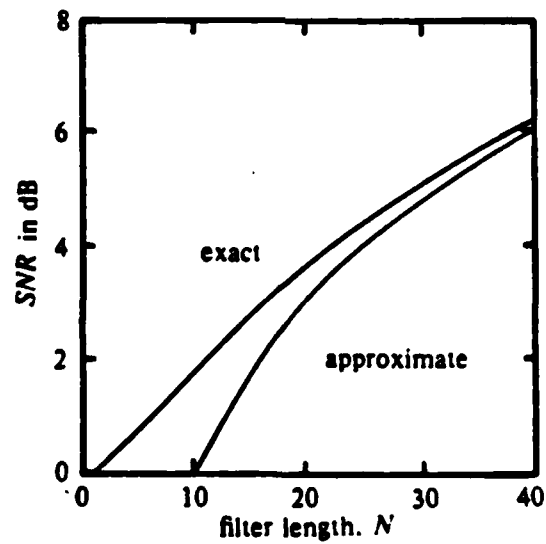


Fig. 10. Approximate and exact lower bound for hyperbolic secant correlation.

CONCLUSIONS

The bounds on SNR_{MF} in Eq. (6) show the selection of the signal to be important for optimal performance of the MF. The selection of a suboptimal signal, if made intelligently, can produce nearly optimal results, and certainly the importance of signal shape should not be overlooked.

The approximate lower bound of Eq. (8) gives a simple although conservative estimate of worse case MF performance. An estimate of filter length can be made with only knowledge of the equivalent rectangular time duration ΔT .

REFERENCES

1. C.W. Helstrom, *Statistical Theory of Signal Detection*. Oxford: Pergamon Press, 1968.
2. H.R. Schwartz, H. Rutishauser, and E. Stiefel, *Numerical Analysis of Symmetric Matrices*. New York: Prentice Hall, 1973.
3. J.B. Thomas, *Introduction to Statistical Communication Theory*. New York: Wiley, 1969.
4. G.L. Turin, "An Introduction to Matched Filters," *IRE Trans. on Information Theory*, Vol. IT-6, June 1960, pp. 311-329.
5. G.L. Turin, "Minimax Strategies for Matched Filter Detection," *IEEE Trans. on Communications*, Vol. COM-23, No. 11, Nov. 1975, pp. 1370-1371.
6. G.L. Turin, "An Introduction to Digital Matched Filters," *Proc. of IEEE*, Vol. 64, No. 7, July 1976, pp. 1092-1112.
7. T.L. Grettenberg, "Signal Selection in Communication and Radar Systems" *IEEE Trans. on Information Theory*, Vol. IT-9, No. 4, Oct. 1963, pp. 265-275.
8. N. Levinson, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," *Journal of Mathematics and Physics*, Vol. XXV, No. 4, Jan. 1947, pp. 261-278.

CHAPTER 8 - CONCLUSIONS

Review and Ideas for Further Research

In this chapter, some of the main points covered in the previous chapters are reviewed briefly, and some ideas for further research are given.

In Chapter 2, the performance of three commonly used detectors is considered for several families of noise densities as the input signal-to-noise ratio (SNR) varies. The three detectors investigated are the linear detector, the sign detector and the amplifier limiter detector. Perhaps the most interesting results of this chapter are the contours of equal detector performance which illustrate the regions of relative superiority for each detector. While the nonlinearities chosen for analysis are interesting, other nonlinearities could be included. In particular, the performance of the locally optimal detector for each specific density would be informative, since it would give some feel for the degradation in performance which can be expected when the input SNR is greater than zero.

In Chapter 3, suboptimal detector nonlinearities are investigated. The design method consists of choosing a suboptimal nonlinearity and finding the family of densities for which the nonlinearity is optimal. A member of this family is then fitted to the observed noise, and the corresponding detector is used. The advantages of this approach are threefold. When the optimal nonlinearity is too complex to use, a simpler, more tractable nonlinearity can be chosen. Because the nonlinearity is in effect fitted to an observed noise using parameter

estimates, the noise statistics need not be known exactly and can be nonstationary. As a result, this procedure can produce simple adaptive detectors. In Chapter 3, the parameters to be estimated are assumed known exactly before the analysis of detector performance is made. One reasonable extension of this work is to study the sensitivity of the detector performance to variations normally experienced in parameter estimates. A more difficult, but possibly quite fruitful area for continued study, is the problem of how best to choose from among suboptimal nonlinearities, or equivalently, how to choose a good noise model when the actual noise statistics are incompletely known.

The moment analysis of Arctic under-ice ambient noise is included in Chapter 4 to give an example of some of the techniques presented in Chapter 3. The results in this chapter suggest that the noise is nonstationary and largely Gaussian or nearly Gaussian with sporadic bursts of non-Gaussian noise. Unfortunately, insufficient data is available for more extensive analysis. With a larger data set, a more complete analysis could be done; in particular, the non-Gaussian bursts merit more attention. Also, the assumptions of stationarity and independence and their effects on moment estimators should be investigated further. Some investigation of the degree of nonstationarity would be of interest, since this would relate directly to the performance of the adaptive detectors proposed in Chapter 3.

In Chapters 5 and 6, it is assumed that a noise process is both significantly dependent and non-Gaussian, and that detector performance would suffer considerably if either Gaussian or independence assump-

tions were made. In this case it is seen that the real problem becomes one of choosing a reasonable noise model. Several multivariate noise models are considered. The value of each is assessed not primarily by the "goodness of fit" of a particular model, but by the tractability of the resulting detector. This may not be too bad an assumption; most of the models considered probably offer a reasonable fit. For example, the transformation model is a perfect fit up to second order moments. However, some investigation of the relative value of each model in fitting a specific observed noise would also be important. A number of appealing detector structures are suggested in these chapters, but very little analysis of their performance is done. Some of these detectors, or slight modifications of these detectors, might well be simple but effective for environments which are both non-Gaussian and dependent.

In Chapter 7, two topics related to discrete time matched filters are considered, bounds on the SNR (used as a measure of filter performance), and signal selection. Exact upper and lower bounds on the SNR are given, and a looser, but easier to compute, lower bound is given. As a possible extension of this work, an approximate upper bound which is easier to compute than the exact bound could be of value. In this chapter it is shown that the selection of a suboptimal signal, if made intelligently, can produce nearly optimal results. The signals used in the examples of Chapter 7 were chosen because of their intuitive appeal. It may be possible to find an orderly procedure for selecting a suboptimal signal.

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

Copies	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217 3	Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522 1
Defense Technical Information Center Cameron Station Alexandria, VA 22314 12	U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706 1
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210 1	Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755 1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101 1	ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002 1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503 1	ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403 1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084 1	Library, Code 1424 Naval Postgraduate School Monterey, CA 93940 1
Commandant of the Marine Corps (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380 1	Technical Information Division Naval Research Laboratory Washington, DC 20375 1
	OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301 1

Copies

Copies

Director
AMSAA
Attn: DRXSYPMP, H. Cohen
Aberdeen Proving Ground, MD 1
21005

Dr. Gerhard Heiche
Naval Air Systems Command
(NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical
Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command
(NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command
(NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies
Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering
Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

Copies

Copies

Mr. F. R. Del Priori
Code 224
Operational Test and Evaluation
Force (OPTEVFOR)
Norfolk, VA 23511

1