
CSL *COORDINATED SCIENCE LABORATORY*

ROBUST TECHNIQUES FOR SIGNAL PROCESSING: A SURVEY

SALEEM A. KASSAM
H. VINCENT POOR

APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) R-1021 UILU-ENG 84-2215		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Laboratory	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Office of Naval Research	
6c. ADDRESS (City, State and ZIP Code) University of Illinois 1101 W. Springfield Avenue Urbana, IL 61801		7b. ADDRESS (City, State and ZIP Code) 800 N. Quincy Avenue Arlington, VA 22217	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR/ARO/JSEP/NSF/ONR	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Continued on reverse	
9c. ADDRESS (City, State and ZIP Code) Continued on reverse		10. SOURCE OF FUNDING NOS.	
		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT NO.
11. TITLE (Include Security Classification) Robust Techniques for Signal Processing: A Survey			
12. PERSONAL AUTHOR(S) Saleem A. Kassam and H. Vincent Poor			
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) October 2, 1984	15. PAGE COUNT 161
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	Statistical signal processing; robustness; uncertain statistical environments; detection and estimation; minimax design; non-Gaussian models
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>In recent years there has been much interest in robustness issues in general and in robust signal processing schemes in particular. Robust schemes are useful in situations where imprecise <i>a priori</i> knowledge of input characteristics make the sensitivity of performance to deviations from assumed conditions an important factor in the design of good signal processing schemes. In this survey we discuss the <i>minimax</i> approach for the design of robust methods for signal processing. This has proven to be a very useful approach because it leads to constructive procedures for designing robust schemes. Our emphasis is on the contributions which have been made in robust signal processing, although key results of other robust statistical procedures are also considered. Most of the results we survey have been obtained in the past ten years, although some interesting earlier ideas for minimax signal processing are also mentioned.</p> <p>This survey is organized in five main parts, which deal separately with robust linear filters for signal estimation, robust linear filters for signal detection and related</p> <p>(Continued on reverse)</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL		22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL

8c. AFOSR
Bolling AFB, DC

ARO
Research Triangle Park, NC

JSEP
800 N. Quincy
Arlington, VA 22217

NSF
Washington, DC

ONR
800 N. Quincy
Arlington, VA 22217

9. AFOSR Grant 82-0022
ARO Contract DAAG29-81-K-0062
JSEP Contract N00014-84-C-0149
NSF Grant ECS-8212080
ONR Contracts N00014-80-K-0945
N00014-81-K-0014

19. ABSTRACT (cont.)

related applications, nonlinear methods for robust signal detection, nonlinear methods for robust estimation, and robust data quantization. The inter-relationships among many of these results are also discussed in the survey.

ROBUST TECHNIQUES FOR SIGNAL PROCESSING: A SURVEY

Invited Paper

Saleem A. Kassam, Senior Member, IEEE, and H. Vincent Poor, Senior Member, IEEE

Abstract

In recent years there has been much interest in robustness issues in general and in robust signal processing schemes in particular. Robust schemes are useful in situations where imprecise *a priori* knowledge of input characteristics make the sensitivity of performance to deviations from assumed conditions an important factor in the design of good signal processing schemes. In this survey we discuss the *minimax* approach for the design of robust methods for signal processing. This has proven to be a very useful approach because it leads to constructive procedures for designing robust schemes. Our emphasis is on the contributions which have been made in robust signal processing, although key results of other robust statistical procedures are also considered. Most of the results we survey have been obtained in the past ten years, although some interesting earlier ideas for minimax signal processing are also mentioned.

This survey is organized in five main parts, which deal separately with robust linear filters for signal estimation, robust linear filters for signal detection and related applications, nonlinear methods for robust signal detection, nonlinear methods for robust estimation, and robust data quantization. The inter-relationships among many of these results are also discussed in the survey.

S. A. Kassam is with the Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104.

H. V. Poor is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

This article was prepared under the support of the U.S. Air Force Office of Scientific Research under Grant AFOSR 82-0022, the U.S. Army Research Office under Contract DAAG29-81-K-0062, the Joint Services Electronics Program (U.S. Army, U.S. Navy, U.S. Air Force) under Contract N00014-84-C-0149, the National Science Foundation under Grant ECS-82-12080, and the U.S. Office of Naval Research under Contracts N00014-80-K-0945 and N00014-81-K-0014.

FOREWORD

This report is a preliminary version of a paper that is to appear in the *Proceedings of the IEEE*.

TABLE OF CONTENTS

Chapter	Page
1. INTRODUCTION.....	1
2. ROBUST FILTERS FOR SIGNAL ESTIMATION.....	9
2.1. The Need for Robustness in Signal Estimation.....	8
2.2. Minimax Design of Robust Filters.....	12
2.3. Some Useful Models for Spectral Uncertainty.....	15
2.4. Robust Causal Filtering, Prediction and Smoothing.....	21
2.5. Robust Equalization of an Uncertain Channel.....	29
2.6. Robust Filtering of Signals in Correlated Noise.....	31
2.7. Uncertainty Classes of Spectral Measures.....	34
2.8. Robust Interpolation.....	36
2.9. Robust Linear Filtering for Vector Observations.....	38
3. ROBUST FILTERS FOR SIGNAL DETECTION AND RELATED APPLICATIONS.....	40
3.1. Robust Matched Filters.....	45
3.2. Two Examples of Robust Multi-Sensor Systems.....	56
3.2.1. Robust Spatial Array for Signal Detection and Location.....	57
3.2.2. Robust Eckart Filter for Time-Delay Estimation.....	61
3.3. General Formulation of Robust Matched Filtering Problems in Hilbert Space.....	67
4. NONLINEAR METHODS FOR ROBUST SIGNAL DETECTION.....	76
4.1. Robust Hypothesis Testing.....	77
4.2. Robust Detection of Known Low-Pass Signals.....	83
4.3. Robust Detection of Random Signals.....	96
4.4. Robust Detection of Bandpass Signals in Bandpass Noise.....	101
4.5. Extensions and Other Results.....	107
5. NONLINEAR METHODS FOR ROBUST ESTIMATION.....	118
5.1. Robust Estimation of Signal Amplitude.....	118
5.2. Robust Nonlinear Recursive Filtering and Identification.....	132
6. ROBUST DATA QUANTIZATION.....	137
6.1. Robust Quantization for a Small Number of Levels.....	137
6.2. Asymptotic Robust Quantization.....	141
7. CONCLUSIONS.....	147
REFERENCES.....	148

1. INTRODUCTION

In recent years there has been a resurgence of interest in what are known as robust methods for statistical signal processing. Such methods are applicable wherever schemes are used to carry out functions such as signal detection, estimation, filtering and coding, common examples being in radar and sonar signal processing, communication systems, pattern recognition, and speech and image processing.

In the early days of development of the body of ideas we now possess for statistical signal processing, the emphasis was on the derivation of *optimum* schemes for use in *specified* signal and noise environments. A classic example of this is the matched filter which is optimum for a particular signal and noise model. Because the signals and noise in signal processing applications are usually modeled as random processes and performance measures therefore usually involve probabilistic quantities (such as mean-square error or probability of error), the theory of statistics has played a fundamental role in the development of optimum signal processing techniques.

Suppose a signal processing scheme, say a detector for a signal with known waveform in additive noise, is designed to give optimum performance for noise possessing a specific statistical description. For example, one widespread model for noise is that it is a Gaussian process. One question that arises is, how sensitive is the performance of such an optimum scheme to deviations in the signal and noise process characteristics from those for which the scheme is designed? This is an important question because in practice one rarely has perfect knowledge of, say, the noise characteristics; the Gaussian or any other specific assumption is usually a nominal assumption which may be approximately valid most of the time. Unfortunately, it turns out that in many cases nominally optimum signal processing schemes can suffer a drastic degradation in performance even for apparently small deviations from nominal assumptions. It is this basic observation that motivates the search for *robust* signal processing techniques; that is, techniques with *good* performance under any *nominal* conditions and *acceptable* performance for signal and noise conditions other than the nominal which can range over the whole of allowable *classes* of possible characteristics. Thus in seeking robust methods it is recognized at the outset that a single, precise characterization of signal and noise conditions is unrealistic, and so *classes* of possible sig-

nal and noise characterizations are constructed and considered in the design of such methods.

To illustrate the above observation with a concrete example consider further the detection problem mentioned above in discrete-time. Thus suppose we have scalar observations X_1, X_2, \dots, X_n , forming a vector \underline{X} , which are known either to be noise only or to be a noise plus a known signal sequence s_1, s_2, \dots, s_n with positive amplitude θ . We express this situation as a choice between the two hypotheses

$$H_0: X_i = N_i, \quad i=1,2,\dots,n, \quad (1.1a)$$

and

$$H_1: X_i = \theta s_i + N_i, \quad i=1,2,\dots,n, \quad (1.1b)$$

where the noise components N_i will be assumed to be independent and identically distributed with a common univariate probability density function (pdf) f . The *likelihood ratio* $\Lambda(\underline{X})$ for the observation vector \underline{X} is defined as

$$\Lambda(\underline{X}) = \prod_{i=1}^n \frac{f(X_i - \theta s_i)}{f(X_i)}. \quad (1.2)$$

This ratio can be formed for any particular realization of \underline{X} provided f is known. It is well-known that a test for H_0 versus H_1 based on the comparison of $\Lambda(\underline{X})$ to a threshold is optimum according to several criteria. Such a test is Neyman-Pearson optimum [1.1], yielding maximum detection power (i.e., minimum "miss" probability) subject to a constraint on the maximum value of the false-alarm probability. The structure of such a test is also that of the test minimizing the Bayes risk for a prior set of probabilities for H_0 and H_1 , as well as that of the minimax test for a given loss function or pay-off matrix with unknown priors [1.1].

A test based on the comparison of $\Lambda(\underline{X})$ with a threshold is equivalent to one based on a comparison of the logarithm of $\Lambda(\underline{X})$ with the logarithm of the original threshold. Taking the logarithm on both sides of (1.2), we have

$$\log \Lambda(\underline{X}) = \sum_{i=1}^n L(X_i; s_i, \theta) \quad (1.3)$$

where

$$L(x; s, \theta) = \log \left| \frac{f(x - \theta s)}{f(x)} \right|. \quad (1.4)$$

If $|L(x; s, \theta)|$ is unbounded as a function of x , the value of $\log \Lambda(\underline{X})$ can be influenced heavily by a single observation component X_i for which $|L(X_i; s_i, \theta)|$ is large. Such a component can therefore completely override the weight of a possibly large number of other components in the choice between H_0 and H_1 . While this effect is certainly acceptable if the model for the noise density function is accurate, it may also be observed because of an occasional completely erroneous measurement which the pdf model f does not take into account. In general the assumed probability density function f describes only an approximate or nominal model. Thus, while the *actual* value of $|L(X_i; s_i, \theta)|$ at some observed value $X_i = x_i$ may not be large relative to that obtained at other observation components, for the *assumed* model this may happen. To illustrate this, suppose that f is assumed to be Gaussian, in which case $L(x; s, \theta)$ is *linear* in x and unbounded. If the true density f has exponential, rather than Gaussian, tails, then the true $L(x; s, \theta)$ is a *constant* for x in the tails, and is bounded. For a model specifying exponential tails, an increasing absolute value for an observation component indicates increasing likelihood of one hypothesis over another only up to a "saturation" value; beyond it, larger absolute values do not indicate larger relative likelihood. If the noise density were truly exponential (or some other long-tailed pdf), then the performance of the test which is optimum for Gaussian noise could be very poor, because of the unexpected number of large noise values.

It would appear, then, that to counter the undesirable sensitivity of the test based on $\Lambda(\underline{X})$ one should implement a bounded modification $\tilde{L}(x; s, \theta)$ of the function $L(x; s, \theta)$ of the assumed nominal model. Thus, we are led to consider $\tilde{L}(x; s, \theta)$ of the form

$$\tilde{L}(x; s, \theta) = \begin{cases} b, & L(x; s, \theta) > b, \\ L(x; s, \theta), & -a \leq L(x; s, \theta) \leq b, \\ -a, & L(x; s, \theta) < -a, \end{cases} \quad (1.5)$$

where a and b are constants. One can expect that with a and b not too small test performance should degrade only marginally when the assumed model is accurate. On the other hand, the boundedness of $\tilde{L}(x; s, \theta)$ builds in a robustness against the influence of a small number of spurious observations. The size of the interval $[-a, b]$ apparently controls the trade-off between degree of robustness and performance degradation under the assumed model. It is noteworthy that several analytical considerations of robust detection lead to detectors based on functions with the form (1.5), as will be discussed in Section 4.

Often a class of allowable characteristics, say for a noise probability density function, is constructed by starting with a nominal characteristic and then including in the class all other characteristics that are "close," in some well-defined sense, to this nominal one. Then a signal processing scheme that is robust may have performance at the nominal which is not quite as good as the scheme which is optimum for the nominal case, but its *overall* performance with respect to the defined class of characteristics will be good or acceptable. This loose definition of robustness is perfectly reasonable, but it does not provide a systematic approach to obtaining robust schemes. In order to achieve this we must first specify a measure of "overall" performance of a scheme with respect to a class of allowable conditions at the input. One such measure that has been widely used and which leads to interesting and useful results in many situations is the *worst-case* performance of a scheme over a class of input conditions. Clearly, if its worst-case performance is good we may say that a given scheme is robust. On the other hand, to find such a robust scheme we can look for the scheme which *optimizes worst-case* performance. This approach leads to what we call *minimax*¹ *robust* schemes. Implicit in our association of minimax schemes with robust schemes is the expectation that the worst-case performance of a minimax scheme will be acceptably good, being the best that can be achieved. Another expectation one has in

¹A scheme that minimizes the maximum possible value of a *loss* function is *minimax*; if performance is measured by a *gain* function then a *maximin* scheme would be sought. We shall use the term *minimax* as a general description for such schemes in all cases.

defining robust schemes this way is that at any nominal operating point the performance of the minimax scheme will not be very far below that of the nominally optimum scheme, which on the other hand will have much poorer performance away from the nominal point. Fortunately, it does turn out that minimax schemes for the signal processing applications of interest usually have the above characteristics. They may therefore be said to have a more "stable" performance (in the literature the terms *robust* and *stable* are sometimes used to mean the same thing).

We should emphasize that the classes of allowable characteristics one deals with in robust signal processing are generally nonparametric *function* classes, such as the class of all power spectral density functions with specified total power (area under the function) and which lie between specified upper and lower bounding functions. For uncertainties expressed by parametric classes of allowable values for finite-dimensional parameters (such as the mean and variance of a Gaussian pdf) one can of course use minimax designs as well, although alternative parametric approaches of statistical theory can also be applied in such situations.

In this paper we will concentrate on minimax robust schemes. There are useful formulations of robustness other than the minimax one, most notably the stability or qualitative robustness ideas introduced by Root [1.2] in the context of signal detection and by Hampel [1.3] in the context of parameter estimation. These formulations utilize the idea of robustness as a continuity property of some performance measure as a function of the underlying model, and some brief discussion of these ideas is included here. However, from the viewpoint of design, the minimax approach has had the most impact on robust signal processing schemes. Also we will not survey adaptive procedures, which may also be used as robust schemes when input conditions are not precisely known and may be time-varying. Adaptive procedures, which attempt to learn about input conditions and adjust their specific signal processing structure accordingly to maintain good performance, are generally more complex than fixed minimax schemes. Adaptive schemes are more desirable when the *a priori* uncertainty is so large that the guaranteed level of performance of a minimax scheme would be too poor to be acceptable and when adequate time and data for adapting is available. Conversely, minimax procedures would be more

desirable under more constrained uncertainty classes, and especially as robust procedures to guard against excessive performance degradations of nominally optimum schemes for deviations from nominal assumptions. Minimax schemes may be used in conjunction with an adaptive approach, because the learning mechanism in adaptive schemes can never be expected to perform perfectly given any finite time for adaptation to take place. The application of minimax concepts to obtain robust versions of optimum adaptive procedures has also been considered [1.4]. While our primary concern here is with minimax robust schemes we will mention other specific techniques whenever it is appropriate.

Most of the recent investigations on robust signal processing techniques have been motivated by the works of the statistician Tukey [1.5] and more so by the seminal 1964-1965 results of the statistician Huber [1.6, 1.7] on minimax robust location-parameter estimation and hypothesis testing. There has generally been a tendency to overlook some rather interesting work on minimax procedures which was carried out for signal processing applications in the decade prior to the publication of Huber's results. In 1954 Zadeh [1.8] suggested that minimax solutions are the natural choices to use in filtering noisy signals under *a priori* uncertainties. In [1.9] Root describes the game-theoretic approach and its application to obtain minimax decision rules in some communication problems. These results of Root were originally contained in a 1956 report [1.10]. Early considerations of minimax schemes for signal processing include those of Yovits and Jackson in 1955 [1.11] on signal estimation filters for imprecisely known power spectral density functions and of Nilsson in 1959 [1.12] and Zetterberg in 1962 [1.13] on matched filters. We will mention their results again in the following sections. Other early contributions are the 1957 paper of Blachman [1.14], the 1959 work of Dobrushin [1.15] and the 1961 paper of Gadzhiev [1.16].

The pre-1964 investigations of minimax signal processing schemes tend to be characterized by two attributes. One is that they were generally not concerned directly with probability density function variations but rather with power spectral density function or related variations. Secondly, minimax schemes were advocated simply as reasonable approaches when designing systems for operation under conditions at the inputs which could not be determined *a priori*. Thus, the possible nonrobustness of

optimum schemes for nominal assumptions on the input was not explicitly recognized as an issue.

The term "robust" was first used in describing desirable statistical procedures by Box in 1953 [1.17]. As we have remarked, minimax robustness of estimation and hypothesis testing schemes was considered by Huber in [1.6] and [1.7], and since then a large number of results on minimax and alternative formulations of robustness have been generated in the statistics literature. In a recent paper [1.18] Huber has given a most interesting account of some early concerns about robustness of statistical procedures and specific schemes, some of which date back to the last century. Reviews of the more recent techniques of robust statistics have been given by Huber [1.19, 1.20], Hampel [1.21], Bickel [1.22] and Hogg [1.23, 1.24]. Ershov [1.25] also gives a survey of robust estimation schemes which is quite broad in its scope. A monograph on robust estimation schemes by Andrews *et al.* [1.26] studies the properties of many robust estimates. A collection of chapters edited by Launer and Wilkinson [1.27] contain some useful expositions. A recent book [1.28] may be consulted for a more detailed treatment.

This survey will focus specifically on minimax robust *signal processing* schemes, so that only a small part of the large body of the statistics literature will be mentioned explicitly. Most of the recent developments in robust signal processing have of course been influenced directly by the developments in robust statistics. However, signal processing problems impose their own distinct requirements which are not always standard in problems of statistics. Thus it turns out that some recent developments in robust signal processing have provided new results in robust statistics.

Most of the results we survey here are of the post-1965 period. Two of the earliest papers in the signal processing area from this period are those of Wolff and Gastwirth [1.29] and Martin and Schwartz [1.30], and they have been responsible for driving much of the subsequent work in robust signal processing. Thus a considerable literature has arisen on robust signal processing just in the last ten to fifteen years.

The statistical descriptions of input conditions in signal processing are usually stated in terms of power spectral density or correlation functions and probability density functions. We shall discuss results which have been obtained on minimax robust *linear filtering* for *signal estimation* (e.g., Wiener

filtering) in Section 2 and for *signal detection* (e.g., matched filtering) in Section 3. Here the uncertainty classes are for spectral density or correlation functions. In Section 4 results on minimax robust *nonlinear signal detection* schemes for distributional uncertainties are surveyed. *Nonlinear parameter estimation* schemes are surveyed more briefly in Section 5, since on this topic there is much already available in review form in the statistics literature. Also included in Section 5 is a brief survey of nonlinear modifications of the Kalman filter for robustness against non-Gaussian pdf's for the observation and process noise components. Section 6 treats the problem of robust quantization of data with unknown statistics, and we close with some concluding remarks in Section 7. Although our survey begins with robust linear filtering, studies on this topic are of more recent vintage than those on nonlinear signal detection and estimation. We feel, however, that the very widespread use of schemes such as Wiener and matched filters in signal processing justifies our beginning with robust versions of such linear processing schemes.

Before we begin let us note some other review, tutorial and survey articles which are available in the literature. A tutorial on this subject by the authors has been published recently [1.31]. VandeLinde has given a brief survey in [1.32]. Ershov [1.25] and Krasnenker [1.33] have surveyed nonlinear estimation and detection schemes, respectively. Poor [1.34] has recently given a more mathematically detailed survey of robust detection schemes. Kleiner, Martin and Thomson [1.35] and Martin and Thomson [1.36] treat the robust estimation of power spectral density functions. Robust methods for time series analysis have been considered by Martin in [1.37], and robust methods for system identification have been described by Poljak and Tsyarkin in [1.38].

As a final introductory comment we should note that the literature in the area of robust statistical methods is vast and broad. Thus, although this survey touches on what we feel to be the major contributions in robust signal processing, it is by no means exhaustive. However, the many results and methods that are not discussed here are accessible to the reader through the references provided.

2. ROBUST FILTERS FOR SIGNAL ESTIMATION

One of the most common signal processing tasks arising in applications is that of estimating (e.g., filtering, predicting, or smoothing) a signal waveform from a noisy measurement. This task arises for example in radar and sonar tracking systems, in observers for automatic control systems, in demodulators for analog communication systems, and in medical imaging systems.

Conventional design procedures for optimum signal estimation algorithms often require an exact knowledge of the statistical behavior both of the signal of interest and of the noise corrupting the measurement. For example, in the design of optimum linear estimation algorithms we must know the spectral or autocorrelation properties of the signal and noise in order to specify the optimum procedures, and (as we shall see below) procedures designed to be optimum for a given model can be undesirably sensitive to inaccuracies in the model. As noted in the Introduction, robust procedures can overcome problems arising due to modeling inaccuracy by incorporating modeling uncertainty into the design from the outset.

In this section, we will discuss the design of robust estimation procedures primarily within the context of the stationary linear (i.e., Wiener-Kolmogorov) estimation problem. Several other signal estimation problems have been treated in the context of robust design, including recursive nonlinear filtering and identification. Results on these problems will be discussed briefly in Section 5.

2.1. The Need for Robustness in Signal Estimation

Consider the observation model

$$Y(t) = S(t) + N(t), \quad -\infty < t < \infty, \quad (2.1)$$

where $\{S(t); -\infty < t < \infty\}$ and $\{N(t); -\infty < t < \infty\}$ are real, zero-mean, orthogonal, wide-sense-stationary (w.s.s.) random processes representing signal and noise, respectively. We assume that $\{S(t); -\infty < t < \infty\}$ and $\{N(t); -\infty < t < \infty\}$ have power spectral densities Φ_S and Φ_N , respectively. (Most of these assumptions can be relaxed, as is discussed below.)

Given the observation process $\{Y(t); -\infty < t < \infty\}$ we wish to form an estimate of $S(t)$ of the form

$$\hat{S}(t) = \int_{-\infty}^{\infty} h(t-\tau)Y(\tau)d\tau, \quad (2.2)$$

where h is the impulse response of a time-invariant linear filter. general problem is illustrated in Fig. 2.1. A common performance criterion for signal estimates is the mean-squared error (MSE), which for estimates of the form of (2.2) is given straightforwardly by

$$E\{|S(t) - \hat{S}(t)|^2\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [|1-H(\omega)|^2 \Phi_S(\omega) + |H(\omega)|^2 \Phi_N(\omega)] d\omega$$

$$\triangleq e(\Phi_S, \Phi_N; H), \quad (2.3)$$

where H is the transfer function associated with h (i.e., H is the Fourier transform of h). If Φ_S and Φ_N are known, then the MSE of (2.3) can be minimized over H to find the optimum filter transfer function for linear minimum-MSE estimation. It is straightforward to show (see, e.g., Thomas [2.1]) that the minimizing solution is given by

$$H_W(\omega) = \frac{\Phi_S(\omega)}{\Phi_S(\omega) + \Phi_N(\omega)}, \quad -\infty < \omega < \infty, \quad (2.4)$$

and that the corresponding minimum value of the MSE is given by

$$e(\Phi_S, \Phi_N, H_W) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\Phi_S(\omega)\Phi_N(\omega)}{\Phi_S(\omega) + \Phi_N(\omega)} d\omega$$

$$\triangleq e_W(\Phi_S, \Phi_N). \quad (2.5)$$

Suppose that we design a filter H_0 via (2.4) to be optimum for some nominal signal and noise spectral pair (Φ_S^0, Φ_N^0) , but that the actual spectra Φ_S and Φ_N can range over some classes \mathcal{S} and \mathcal{N} , respectively, of spectra "neighboring" Φ_S^0 and Φ_N^0 . An important question that arises in this situation is:

what is the behavior of the MSE $e(\Phi_S, \Phi_N; H_0)$ as Φ_S and Φ_N range over \mathcal{S} and \mathcal{N} ? For example, it is of interest to know how the quantity

$$\sup_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e(\Phi_S, \Phi_N; H_0) \quad (2.6)$$

compares with the quantity $e(\Phi_S^o, \Phi_N^o; H_0) \triangleq e_w(\Phi_S^o, \Phi_N^o)$. The first of these quantities represents the worst performance of the filter H_0 over the class of possible spectra, whereas the second quantity represents the predicted performance assuming the nominal model to be accurate. A situation in which (2.6) were considerably larger than $e_w(\Phi_S^o, \Phi_N^o)$ would point to a possible inadequacy of the nominal design H_0 .

To illustrate the degree to which modeling uncertainty can affect performance, we consider the following example taken from Vastola and Poor [2.2]. Suppose we have assumed a nominal model (Φ_S^o, Φ_N^o) given by

$$\Phi_S^o(\omega) = \frac{2\nu_S^2}{1 + \omega^2} \text{ and } \Phi_N^o(\omega) = \frac{20\nu_N^2}{100 + \omega^2}, \quad -\infty < \omega < \infty. \quad (2.7)$$

Note that these spectra represent first-order wide-sense Markov processes with 3 dB signal bandwidth equal to 1, 3 dB noise bandwidth equal to 10, signal power $E\{[S(t)]^2\} = \nu_S^2$, and noise power $E\{[N(t)]^2\} = \nu_N^2$. Suppose, however, that all we really know is the total signal power, the total noise power, and the fractional signal and noise powers in the frequency band $|\omega| \leq 1$. This knowledge corresponds to the uncertainty classes¹

$$\mathcal{S} = \{\Phi_S \mid \frac{1}{2\pi} \int_{-1}^1 \Phi_S(\omega) d\omega = p_S \nu_S^2 \text{ and } \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) d\omega = \nu_S^2\} \quad (2.8a)$$

and

$$\mathcal{N} = \{\Phi_N \mid \frac{1}{2\pi} \int_{-1}^1 \Phi_N(\omega) d\omega = p_N \nu_N^2 \text{ and } \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_N(\omega) d\omega = \nu_N^2\} \quad (2.8b)$$

¹Note that rational models (such as (2.7)) are often forced upon estimated power spectra, although the actual data only predicts fractional powers (such as (2.8)) accurately (Marzetta and Lang [2.3]).

where $p_S = \frac{1}{2\pi} \int_{-1}^1 \Phi_S^g(\omega) d\omega/v_S^2$ and $p_N = \frac{1}{2\pi} \int_{-1}^1 \Phi_N^g(\omega) d\omega/v_N^2$.

For a given estimation filter, H , and spectral pair (Φ_S, Φ_N) , the signal-to-noise ratio (SNR) at the output of H can be defined by

$$\text{Output SNR} = 10 \log_{10}(v_S^2/e(\Phi_S, \Phi_N; H)), \quad (2.9)$$

since the output $\hat{S}(t)$ can be written as $\hat{S}(t) = S(t) + (\hat{S}(t) - S(t))$, and $E\{[S(t)]^2\} = v_S^2$ and $E\{[S(t) - \hat{S}(t)]^2\} = e(\Phi_S, \Phi_N; H)$. Also, the input SNR is given simply by

$$\text{Input SNR} = 10 \log_{10}(v_S^2/v_N^2).$$

Using these definitions, Fig. 2.1 depicts the nominal and worst-case performance of the filter H_0 designed to be optimum for the nominal spectral pair of (2.7). Note the considerable performance degradation throughout the given range of input SNR's. Also depicted in Fig. 2.1 is the performance of trivial filtering, which corresponds to all-pass filtering if the input SNR is positive and no-pass filtering if the input SNR is negative. Note that the worst case over (2.8) of the nominally designed filter is uniformly worse than trivial filtering. Thus, the nominal filter can actually make the signal noisier than it originally was!

2.2. Minimax Design of Robust Filters

The above example illustrates the need for an alternate design philosophy for the stationary linear signal estimation problem for applications in which there is some uncertainty regarding the spectra of interest. In particular, in view of the methods described in the Introduction, we consider as a design philosophy the minimization over H of the worst-case performance degradation described by (2.6). That is, we consider the design criterion

$$\min_H \left\{ \sup_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e(\Phi_S, \Phi_N; H) \right\}. \quad (2.10)$$

A solution to (2.10) can be considered to be a *robust* filter for the uncertainty classes \mathcal{S} and \mathcal{N} .

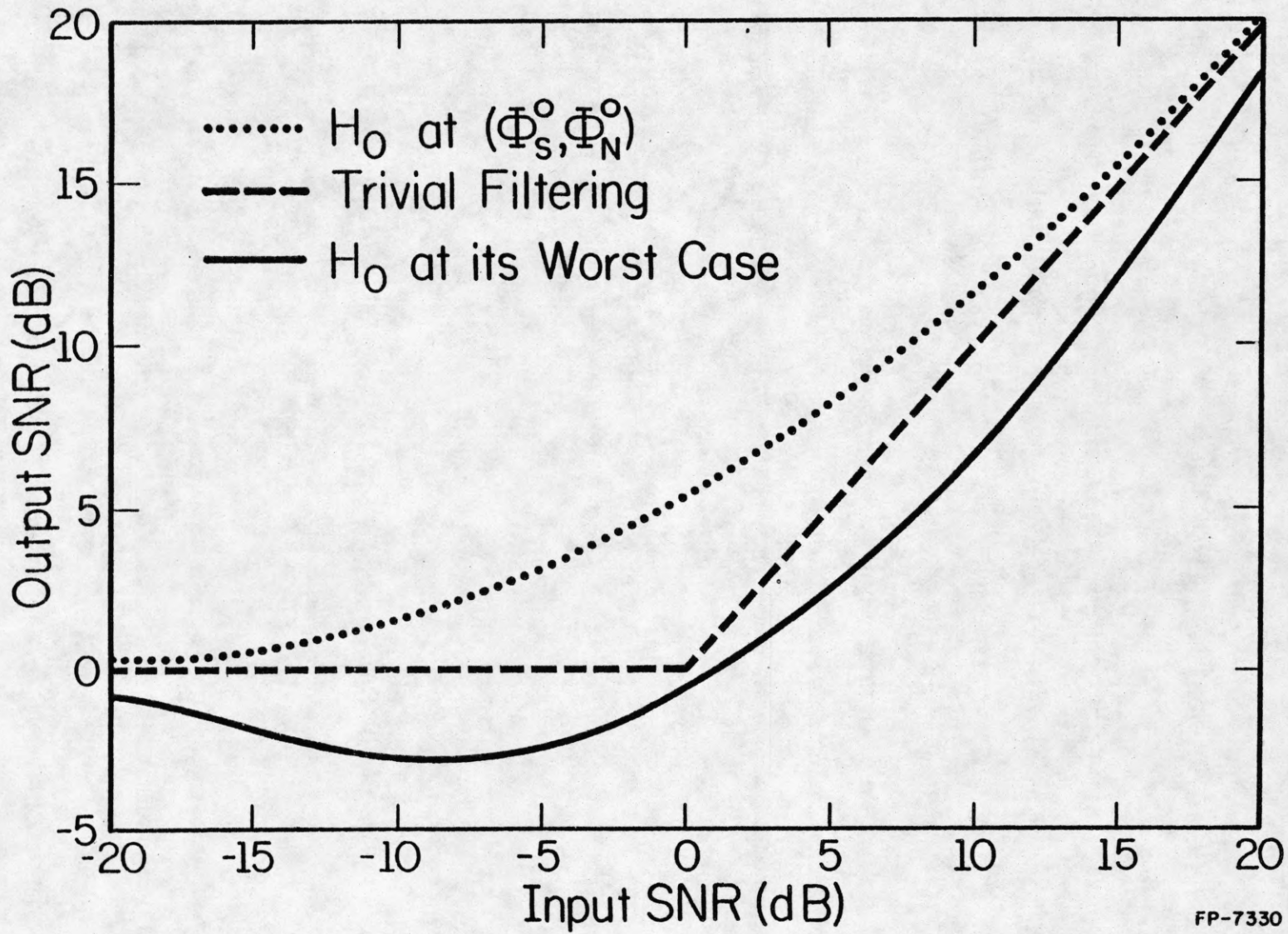


Fig. 2.1: Worst-case and nominal performance of nominal and trivial filtering for example of Section 2.1.

To solve this problem, we would like to find a saddle point for the minimax game of (2.10); i.e., we would like to find a spectral pair $(\Phi_{S,L}, \Phi_{N,L}) \in \mathcal{S} \times \mathcal{N}$ and a filter H_R satisfying

$$\max_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e(\Phi_S, \Phi_N; H_R) = e(\Phi_{S,L}, \Phi_{N,L}; H_R) = \min_H e(\Phi_{S,L}, \Phi_{N,L}; H). \quad (2.11)$$

Note that the right-hand equality in (2.11) implies that H_R is the optimum filter for the pair $(\Phi_{S,L}, \Phi_{N,L})$ [i.e., $H_R(\omega) = \Phi_{S,L}(\omega)/(\Phi_{S,L}(\omega) + \Phi_{N,L}(\omega))$]; thus, the determination of a saddle point involves finding a pair $(\Phi_{S,L}, \Phi_{N,L})$ which satisfies (2.11) with $H_R = \Phi_{S,L}/(\Phi_{S,L} + \Phi_{N,L})$. The left-hand inequality in (2.11) says that H_R achieves its worst performance at the pair of spectra $(\Phi_{S,L}, \Phi_{N,L})$ for which it is optimum. This worst performance, $e(\Phi_{S,L}, \Phi_{N,L}; H_R)$, is the guaranteed level of performance of the filter H_R for the classes of \mathcal{S} and \mathcal{N} .

The problem (2.10) was first posed in the context of robustness by Kassam and Lim in [2.4], wherein a saddle point solution to (2.10) was given for the situation in which the spectra are known only to lie within given spectral bands. The problem of (2.10) was considered for general spectral uncertainty classes by Poor [2.5], and it is shown in [2.5] that for convex \mathcal{S} and \mathcal{N} , a spectral pair $(\Phi_{S,L}, \Phi_{N,L}) \in \mathcal{S} \times \mathcal{N}$ and its optimum filter $H_R = \Phi_{S,L}/(\Phi_{S,L} + \Phi_{N,L})$ form a saddle point for (2.10) if, and only if, the pair $(\Phi_{S,L}, \Phi_{N,L})$ is *least favorable* for $\mathcal{S} \times \mathcal{N}$; i.e., if and only if $(\Phi_{S,L}, \Phi_{N,L})$ solves

$$\max_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e_w(\Phi_S, \Phi_N), \quad (2.12)$$

where e_w is the minimum MSE functional defined by (2.5). The term "least favorable" comes from the fact that $(\Phi_{S,L}, \Phi_{N,L})$ is the pair of spectra in $\mathcal{S} \times \mathcal{N}$ that correspond to the random processes that are hardest to separate by filtering.

Thus, a design procedure for finding a robust filter for given uncertainty classes \mathcal{S} and \mathcal{N} is to solve (2.12) and then to design the optimum filter for the maximizing spectral pair. Since the filter design problem is solved by (2.4), the only possible difficulty is in solving (2.12). This problem, however, is straightforward to solve for many uncertainty classes of interest. In particular, the functional

$e_W(\Phi_S, \Phi_N)$ can be written as

$$e_W(\Phi_S, \Phi_N) = - \int_{-\infty}^{\infty} C \left(\frac{\Phi_S(\omega)}{\Phi_N(\omega)} \right) \Phi_N(\omega) d\omega. \quad (2.13)$$

where C is the convex function $C(x) = -(2\pi)^{-1}x/(1+x)$. Thus, maximizing e_W is equivalent to minimizing the functional $\int C(\Phi_S/\Phi_N)\Phi_N$, which is a special case of a general class of divergences or "distances" between densities (see, Ali and Silvey [2.6]; Csiszar [2.7]). In view of (2.13) least favorables can be interpreted as being the spectra in \mathcal{S} and \mathcal{N} whose shapes are "closest together". Because of this structure, the problem of solving for least-favorable spectra for spectral uncertainty classes in which the total signal and noise powers are known and only the spectral *shapes* are uncertain can be accomplished by analogy with results in robust hypothesis testing. In particular, for a general type of classes with this property, the least-favorable spectra are scaled versions of the least-favorable probability densities for an analogous robust hypothesis testing problem posed by Huber [2.8]. This is a useful result because solutions to the robust hypothesis testing problem are known for many uncertainty models of interest. (See Poor [2.5, 2.9] for further details.)

2.3. Some Useful Models for Spectral Uncertainty

There are a number of useful models for spectral uncertainty for which solutions to the robust stationary linear filtering problem can be obtained straightforwardly. The following examples are typical:

Example 1: ϵ -contaminated models

One very useful spectral uncertainty model is that given by

$$\{\Phi \mid \Phi(\omega) = (1 - \epsilon)\Phi^o(\omega) + \epsilon\sigma(\omega) \text{ and } \int_{-\infty}^{\infty} \Phi(\omega) d\omega = \int_{-\infty}^{\infty} \Phi^o(\omega) d\omega\} \quad (2.14)$$

where Φ^o is a nominal spectrum, σ is an arbitrary and unknown "contaminating" spectrum, and ϵ is a degree of uncertainty (between 0 and 1) placed on the nominal model by the designer. This type of

model allows for a fairly general type of uncertainty in a nominal spectral model, and it is used frequently to model uncertainty in several contexts.

Example 2: variational-neighborhood models

Another useful model for spectral uncertainty arises from allowing all spectra that vary from a nominal spectrum by no more than some given amount. Using a standard measure of "variational distance" this model becomes

$$\left\{ \Phi \mid \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(\omega) - \Phi^{\circ}(\omega)| d\omega \leq \epsilon v^2 \text{ and } \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega) d\omega = v^2 \right\}, \quad (2.15)$$

where Φ° and ϵ play the same roles as in the ϵ -contaminated class and where $v^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi^{\circ}(\omega) d\omega$.

Example 3: p-point models

The classes of (2.8) are particular examples of a more general type of spectral uncertainty class known as p-point classes. These classes are of the form:

$$\left\{ \Phi \mid \int_{\Omega_i} \Phi(\omega) d\omega = \int_{\Omega_i} \Phi^{\circ}(\omega) d\omega, i = 1, \dots, n \right\}, \quad (2.16)$$

where, as before, Φ° is a nominal spectrum and $\Omega_1, \Omega_2, \dots, \Omega_n$ is a partition of the frequency domain. Note that a p-point class consists of all spectra that have a fixed amount of power on each of the spectral regions $\Omega_1, \dots, \Omega_n$. Such a class might arise, for example, when power measurements are taken in a number of frequency bands.

Example 4: band models

The first spectral uncertainty model considered in the context of robust Wiener filtering consists of the class of those spectral densities (with a given amount of power) that lie in a band bounded above and below by two known functions. This class can be written as

$$\{\Phi \mid L(\omega) \leq \Phi(\omega) \leq U(\omega), -\infty < \omega < \infty; \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega) d\omega = \nu^2\} \quad (2.17)$$

where L and U are known functions and where ν^2 is fixed. Note that a model such as (2.17) can be used to describe a "confidence region" around an estimated spectrum. Also note that the ϵ -contaminated model of *Example 1* is a special case of (2.17) with $L = (1 - \epsilon)\Phi^0$ and $U = \infty$.

Example 5: generalized moment-constrained models

As a final example, consider spectral uncertainty classes of the following type:

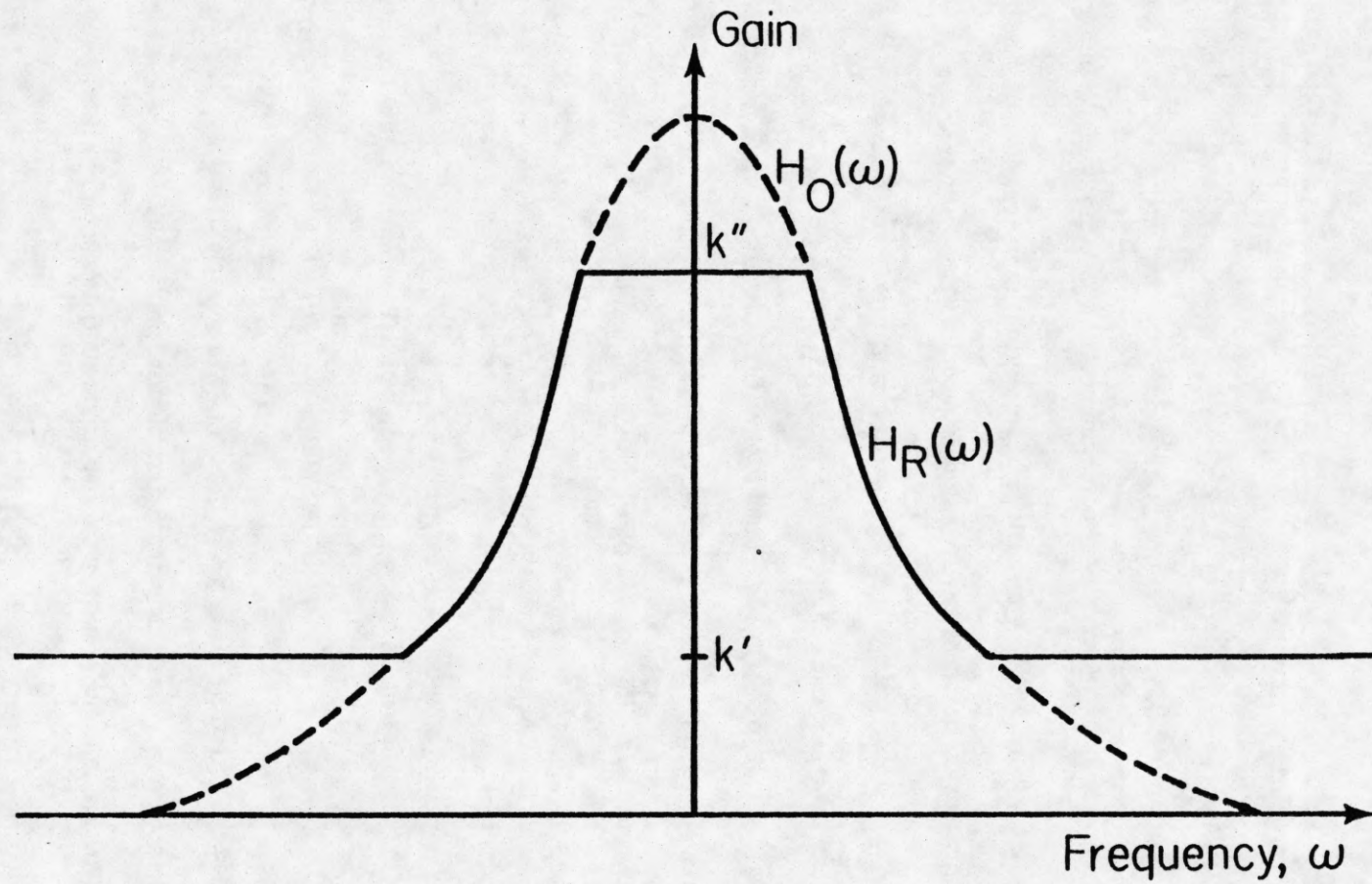
$$\{\Phi \mid \int_{-\infty}^{\infty} f_k(\omega)\Phi(\omega) d\omega = c_k, k = 1, \dots, n\}, \quad (2.18)$$

where f_1, \dots, f_m are known functions and c_1, \dots, c_m are constants. The quantities $\int f_k \Phi, k=1, \dots, n$, are sometimes known as "generalized moments" of the spectrum Φ corresponding to the weightings f_1, \dots, f_n . Note that the p-point class of (2.16) is a special case of (2.18) with $f_k(\omega) = 1$ for $\omega \in \Omega_k$ and $f_k(\omega) = 0$ for $\omega \notin \Omega_k$. More generally (2.18) might represent the set of spectra of all processes that yield output power c_k when input to a filter with transfer function $[2\pi f_k(\omega)]^{1/2}$ for $k=1, \dots, n$. Thus, a model such as (2.18) arises when the available information consists of power measurements taken at the outputs of a filter bank.

If \mathcal{S} and \mathcal{N} are both of the ϵ -contaminated or variational neighborhood type, then it can be shown (see Kassam and Lim [2.4], Poor and Looze [2.10]) that the robust filter is of the form

$$H_R(\omega) = \begin{cases} k', & \text{if } H_0(\omega) < k' \\ H_0(\omega), & \text{if } k' \leq H_0(\omega) \leq k'' \\ k'', & \text{if } H_0(\omega) \geq k'' \end{cases} \quad (2.19)$$

where $H_0 = \Phi_S^0 / (\Phi_S^0 + \Phi_N^0)$ is the nominal filter and where k' and k'' are two constants depending on the value of ϵ and on the particular model used. This robust transfer function is illustrated in Fig. 2.2. Note that the effect of incorporating the uncertainty into the design is a limiting of the gain of the nominal filter both from above and from below. This solution has a nice intuitive interpretation if one



FP-8384

Fig. 2.2: Typical robust filter characteristic for ϵ -contaminated or variational neighborhood models for spectral uncertainty.

considers the action of the nominal filter H_0 . This filter is designed to have near unity gain in spectral regions where the nominal signal-to-noise power density ratio, $\Phi_S^g(\omega)/\Phi_N^g(\omega)$, is large, and to have near zero gain at frequencies where this ratio is small. In other regions, the gain is chosen to balance the effects of signal distortion and noise throughput. The robust filter transfer function reflects similar characteristics but, also, because of the spectral uncertainty, it limits the gain from above to guard against a greater than nominal amount of noise power at the frequencies where $\Phi_S^g/\Phi_N^g \gg 1$ and it limits the gain from below to assure that unexpected signal power at frequencies where $\Phi_S^g/\Phi_N^g \ll 1$ will not lead to undue distortion.

If both signal and noise spectral uncertainty classes are the p-point form of (2.16) with common spectral bands $\Omega_1, \dots, \Omega_n$, then the robust filter can be shown to be given by (see Cimini and Kassam [2.11]; Vastola and Poor [2.12])

$$H_R^*(\omega) = \frac{v_{S_i}^2}{v_{S_i}^2 + v_{N_i}^2}, \text{ if } \omega \in \Omega_i, i = 1, \dots, n \quad (2.20)$$

where $v_{S_i}^2 = \frac{1}{2\pi} \int_{\Omega_i} \Phi_S^g(\omega) d\omega$ and $v_{N_i}^2 = \frac{1}{2\pi} \int_{\Omega_i} \Phi_N^g(\omega) d\omega$ for $i = 1, \dots, n$. A typical filter of this type is depicted in Fig. 2.3. Note that this is a zonal filter that can only be implemented approximately for electrical signals; however, in optical filtering where the variable t in (2.1) is interpreted as a spatial parameter and ω as spatial frequency, this type of filter would be very simple to implement (see Cimini and Kassam [2.11, 2.13]). An interesting feature of this model is that the performance of the robust filter is *constant* over the classes \mathcal{S} and \mathcal{N} and is given by [2.11]

$$e(\Phi_S, \Phi_N; H_R) = \sum_{i=1}^n \frac{v_{S_i}^2 v_{N_i}^2}{v_{S_i}^2 + v_{N_i}^2} \quad (2.21)$$

for all $(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}$.

The robust solution for the band model of *Example 4* is given by Kassam and Lim in [2.4] and its behavior is similar to that for the ϵ -contaminated model (which is a special case). Solutions for generalized moment classes have been given by Breiman in [2.14], and a particular case of this model will be

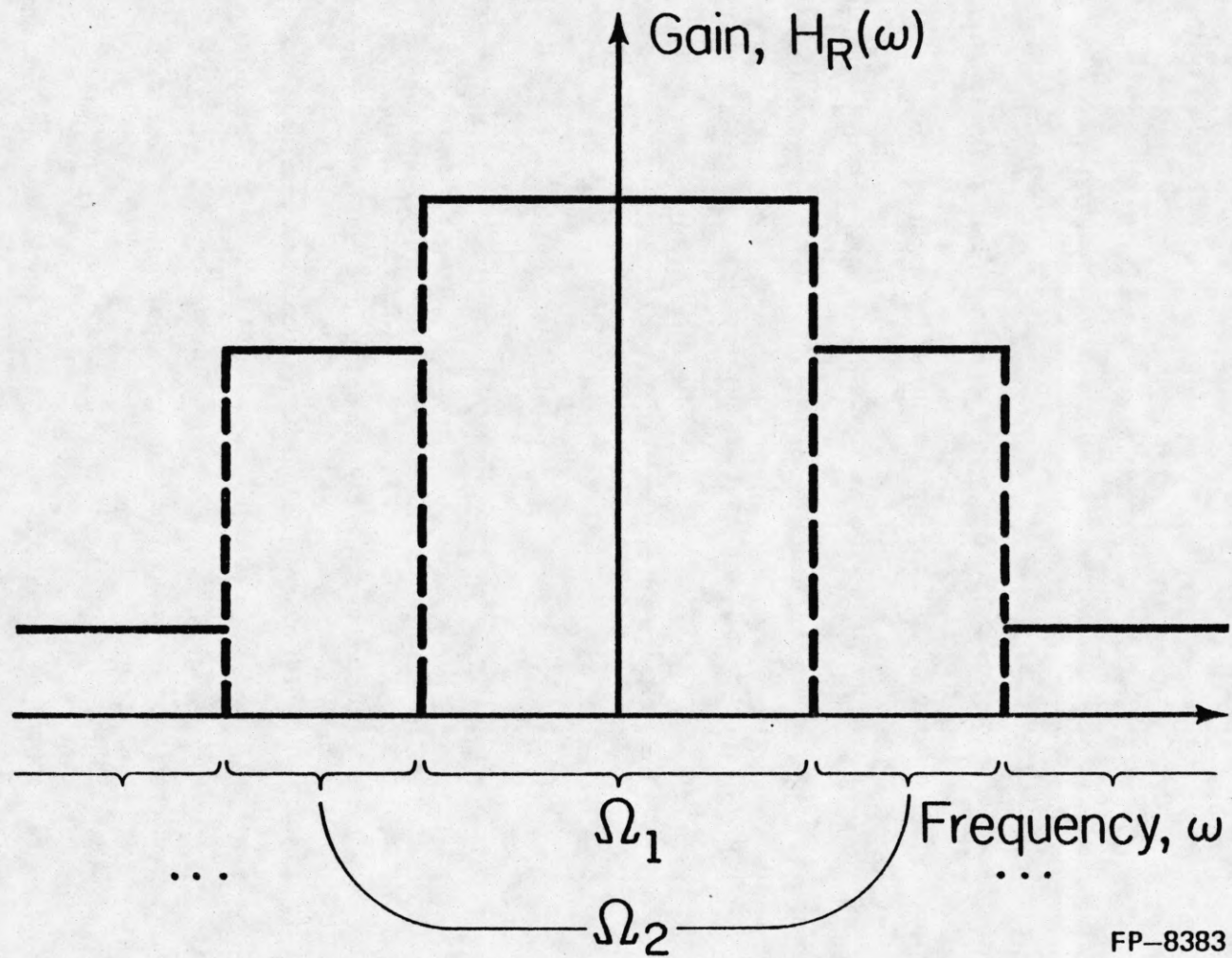


Fig. 2.3: Typical robust filter characteristic for p -point models for spectral uncertainty.

discussed below. (Of course, the p-point model is also a special case.) Other models including combinations of the above models (such as the bounded p-point model [2.15]) and more general models based on Choquet capacities [2.9, 2.16] have also been considered.

To illustrate the potential effectiveness of the robust filter we return to the problem described by the nominals of (2.7) and the uncertainty model of (2.8). This is a p-point model with $\Omega_1 = [-1, 1]$, $\Omega_2 = \Omega_1^c$, $p_S = 1/2$ and $p_N = 0.063$. Figure 2.4 superimposes the (constant) performance of the robust filter (2.20) for this case onto the nominal and worst-case performance curves of the nominal filter from Fig. 2.1. Note that the performance of the robust filter over the entire uncertainty class is only slightly degraded from the nominal performance of the nominal filter and is remarkably improved over the worst-case performance of the nominal filter. This example illustrates fairly dramatically the favorability of the minimax design for filtering in uncertain environments.

2.4. Robust Causal Filtering, Prediction and Smoothing

The results discussed in the above subsections assume implicitly that the class of allowable estimation filters includes noncausal filters; i.e., $h(t - \tau)$ in (2.2) is not necessarily zero for $\tau > t$. While this assumption is not restrictive for many applications such as those involving spatial filtering or enhancement of stored signals, there are also many applications in which causality of the estimation filter is desired for the purposes of real-time processing. To discuss this situation we consider again the observation model given by

$$Y(t) = S(t) + N(t), \quad -\infty < t < \infty, \quad (2.22)$$

where $\{S(t); -\infty < t < \infty\}$ and $\{N(t); -\infty < t < \infty\}$ satisfy the assumptions made below (2.1). We wish to estimate the signal at time $t + \lambda$ for some fixed λ based on observations up to time t ; i.e., we wish to consider estimates of the form

$$\hat{S}(t + \lambda) = \int_{-\infty}^t h(t - \tau) Y(\tau) d\tau. \quad (2.23)$$

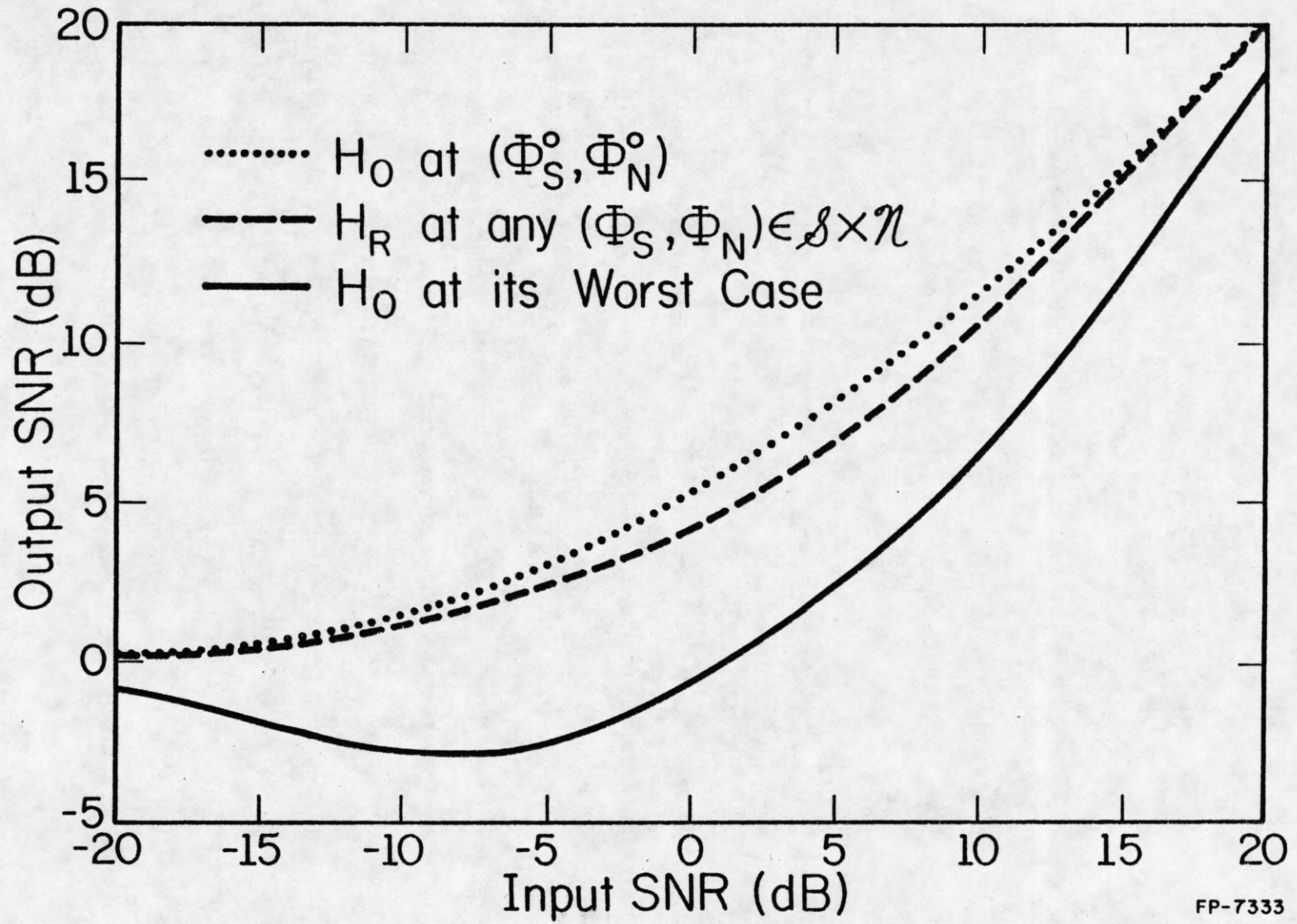


Fig. 2.4: Performance curves depicting the favorability of the robust filter for the example of Section 2.1.

Note that $\lambda < 0$ in (2.23) corresponds to fixed-lag smoothing of the signal, $\lambda = 0$ corresponds to causal filtering, and $\lambda > 0$ corresponds to signal prediction.

The MSE associated with the estimate of (2.23) is given by

$$E \{ |S(t + \lambda) - \hat{S}(t + \lambda)|^2 \} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [|e^{j\lambda\omega} - H(\omega)|^2 \Phi_S(\omega) + |H(\omega)|^2 \Phi_N(\omega)] d\omega$$

$$\triangleq e_\lambda(\Phi_S, \Phi_N; H), \quad (2.24)$$

where H is the transfer function of the filter $\{h(t); t \geq 0\}$. For fixed Φ_S and Φ_N such that the observation spectrum, $(\Phi_S + \Phi_N)$, satisfies the Paley-Wiener condition, the MSE of (2.24) is minimized over all causal filters by the filter with transfer function

$$H_0^+(\omega) = \frac{1}{[\Phi_S(\omega) + \Phi_N(\omega)]^+} \left| \frac{e^{j\lambda\omega} \Phi_S(\omega)}{[\Phi_S(\omega) + \Phi_N(\omega)]^-} \right|_+, \quad (2.25)$$

where the subscript $+$ denotes causal part in an additive spectral decomposition and the superscripts $+$ and $-$ denote causal and anticausal parts, respectively, in a multiplicative spectral decomposition (see, e.g., Wong [2.17]). The minimum MSE is then obtained by combining (2.24) and (2.25) as

$$e_\lambda(\Phi_S, \Phi_N; H^+) \triangleq e_\lambda^+(\Phi_S, \Phi_N). \quad (2.26)$$

As in the noncausal situation discussed above, it is commonly the case in practice that Φ_S and Φ_N are not known exactly but rather are known to lie in some uncertainty classes \mathcal{S} and \mathcal{N} of possible signal and noise spectra. In this case we may seek a robust filter for \mathcal{S} and \mathcal{N} by minimizing the worst-case error,

$$\sup_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e_\lambda(\Phi_S, \Phi_N; H), \quad (2.27)$$

over all causal transfer functions H . Although this problem is analytically more difficult than the corresponding noncausal problem, it can be shown for convex \mathcal{S} and \mathcal{N} that, within mild conditions, a

saddle point solution for this problem is given by the optimum causal filter H_R^+ corresponding to the least-favorable spectral pair $(\Phi_{S,l}, \Phi_{N,l})$ for this causal problem, where $(\Phi_{S,l}, \Phi_{N,l})$ is defined via

$$e_{\lambda}^+(\Phi_{S,l}, \Phi_{N,l}) = \max_{(\Phi_S, \Phi_N) \in \mathcal{S} \times \mathcal{N}} e_{\lambda}^+(\Phi_S, \Phi_N) \quad (2.28)$$

(see Poor [2.5], Vastola and Poor [2.16], Franke [2.18], Vastola [2.19], Franke and Poor [2.20]).

Thus, conceptually, the causal robust signal estimation problem is no more difficult than the non-causal one, since one designs a robust filter by first maximizing $e_{\lambda}^+(\Phi_S, \Phi_N)$ over $\mathcal{S} \times \mathcal{N}$ and then designing an optimum filter for the maximizing pair via (2.25). However, in the noncausal problem, there is a tractable closed-form expression, namely (2.5), to be maximized to find least favorables, whereas no such general expression exists for the causal problem; i.e., there is no general closed-form expression for $e_{\lambda}^+(\Phi_S, \Phi_N)$ of (2.28). On the other hand, there are many specific cases of practical interest for which $e_{\lambda}^+(\Phi_S, \Phi_N)$ is known in closed form (see, e.g., Yao [2.21, 2.22], Snyders [2.23, 2.24]) and, moreover, general (but tedious) methods for finding such expressions are available (e.g., [2.23]). Robustness in several causal filtering problems has been considered using these results. Generally speaking, the phenomena observed are more or less the same as for the noncausal case, although, for a given model, nominal causal filters appear to be somewhat less sensitive to uncertainty than nominal non-causal filters are, due to the relatively lower selectivity of causal filters (see, e.g., Vastola and Poor [2.2]).

Example: Robust Prediction

An interesting example of an application in which the above results can be easily applied is the problem of discrete-time one-step pure prediction.² In particular, suppose we observe a discrete-time signal directly up to some time t ; i.e., we have

$$Y(k) = S(k), \quad k \in \{\dots, t-3, t-2, t-1, t\}. \quad (2.29)$$

Suppose further that we wish to predict the value of $S(k)$ at the next sampling instant $k = (t+1)$

²Note that the discrete-time problem is the special case of the continuous-time one in which the spectra are concentrated in the spectral band $|\omega| \leq \pi$. Thus the above discussion holds for both discrete and continuous time.

with a linear predictor

$$\hat{S}(t+1) = \sum_{k=-\infty}^t h(t-k)S(k). \quad (2.30)$$

This is the problem of (2.23) with $\lambda = 1$ and $\Phi_N(\omega) = 0$ for all ω .

The minimum MSE functional $e_{\lambda}^+(\Phi_S, \Phi_N)$ for this problem is given by the well-known Kolmogorov-Szëgo-Krein formula (Hoffman [2.25]):

$$e_1^+(\Phi_S, 0) = \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\Phi_S(\omega)) d\omega \right\}. \quad (2.31)$$

Thus, in order to design a predictor to be robust over an uncertainty class \mathcal{S} of signal spectra, we choose $\Phi_{S, \lambda}$ via

$$\Phi_{S, \lambda} = \arg \left\{ \max_{\Phi_S \in \mathcal{S}} \int_{-\pi}^{\pi} \log \Phi_S(\omega) d\omega \right\}, \quad (2.32)$$

and we then design the optimum predictor for this spectrum. This problem has been considered by Hosoya [2.26] for the particular case of an ϵ -contaminated spectrum and by Franke [2.18] and Vastola and Poor [2.19] for the general case.

It is interesting to note that the spectrum of (2.32) can be interpreted as being that member of \mathcal{S} which is "closest" to a uniform spectrum (see Poor [2.27]).³ This has a very nice intuitive interpretation since a uniform spectrum corresponds to white noise, which is the universal worst-case type of signal to predict. (I.e., past and present data is useless in predicting future values of white noise.) It is also interesting to note that the quantity $\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \Phi_S(\omega) d\omega$ is the *spectral entropy* of the signal process (see, e.g., [2.28]), and thus the least-favorable spectrum is *maxentropic* in S . Also, since the entropy of a

³This follows since $\int_{-\pi}^{\pi} \log \Phi_S(\omega) d\omega = -\int_{-\pi}^{\pi} C(\Phi_S(\omega)) d\omega$ where C is the convex function $C(x) = -\log x$. Comparing with (2.13) one sees the interpretation of the negative of the spectral entropy of Φ_S as being a measure of the distance of Φ_S from the uniform spectrum.

process is a measure of its indeterminism, the least-favorable spectrum can be thought of as the *most indeterministic*, a term introduced by Franke [2.18].

As a specific example, we consider the particular case of an ϵ -contaminated first-order wide-sense-Markov signal; i.e., we have

$$\mathfrak{S} = \{ \Phi_S \mid \Phi_S(\omega) = (1 - \epsilon)\Phi_S^0(\omega) + \epsilon\sigma(\omega), \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_S(\omega) d\omega = 1 \} \quad (2.33)$$

where

$$\Phi_S^0(\omega) = \frac{1 - r^2}{1 - 2r\cos(\omega) + r^2}, \quad -\pi \leq \omega \leq \pi, \quad (2.34)$$

and where $0 \leq r < 1$. The least-favorable spectrum for this model is given by (see [2.26])

$$\Phi_{S,l}(\omega) = \begin{cases} (1 - \epsilon)\Phi_S^0(\omega) & \text{if } (1 - \epsilon)\Phi_S^0(\omega) > c' \\ c' & \text{if } (1 - \epsilon)\Phi_S^0(\omega) \leq c', \end{cases} \quad (2.35)$$

where c' is chosen so that $\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{S,l}(\omega) d\omega = 1$. This spectrum is illustrated in Fig. 2.5 for the case $r = 0.5$ and $\epsilon = 0.25$. Note that, as ϵ increases, the peak in the center of the frequency band "melts" into a uniform spectrum in the frequency tails.

Another interesting example of robust prediction comes from consideration of the following signal spectral uncertainty classes (see Franke [2.18]):

$$\mathfrak{S} = \{ \Phi_S \mid \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega k} \Phi_S(\omega) d\omega = c_k, \quad |k| = 0, 1, \dots, m \} \quad (2.36)$$

where c_0, \dots, c_m is a set of constants. Since $\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega k} \Phi_S(\omega) d\omega$ is the k^{th} -lag autocorrelation ($E\{S_l S_{l+k}\}$) of the signal, (2.36) corresponds to the set of all spectra whose first m lag autocorrelations take the given values c_0, \dots, c_m . Such a model is applicable, for example, when we have measurements of a finite number of autocorrelations of the signal process. Note that (2.36) is an example of the generalized moment constrained model from *Example 5* of the preceding subsection.

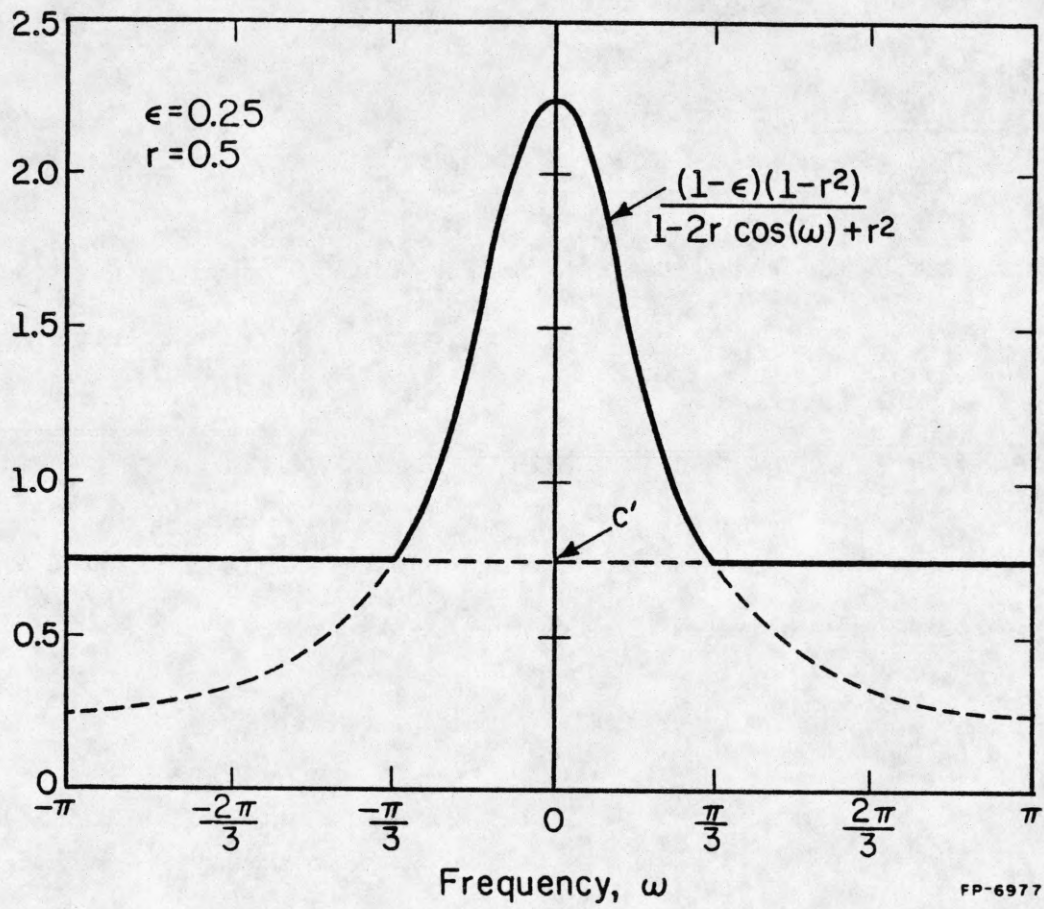


Fig. 2.5: Least-favorable spectrum for predicting an ϵ -contaminated wide-sense Markov signal.

To find a robust predictor for the class \mathcal{S} of (2.36) we first look for the least-favorable signal spectrum by solving the constrained optimization problem

$$\max_{\Phi_S \in \mathcal{S}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \Phi_S(\omega) d\omega \text{ subject to } \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\omega k} \Phi_S(\omega) d\omega = c_k; |k| = 0, \dots, m; \quad (2.37)$$

(2.37) will be recognized as the well-known maximum entropy spectrum fitting problem, and its solution is straightforwardly seen to be given by

$$\Phi_{S, \lambda}(\omega) = \frac{\sigma_0^2}{\left| \sum_{k=0}^m a_k e^{-j\omega k} \right|^2} \quad (2.38)$$

with $a_0 = 1$ and with a_1, \dots, a_m and σ_0^2 satisfying the Yule-Walker equations for the correlations c_0, \dots, c_m (see, e.g., [2.29]). The spectrum $\Phi_{S, \lambda}$ of (2.38) is the spectrum of the m^{th} -order autoregression

$$S_t = \sum_{k=1}^m a_k S_{t-k} + \sigma_0 \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.39)$$

where $\{\epsilon_t\}_{t=-\infty}^{\infty}$ is a sequence of orthogonal, zero-mean and unit-variance random variables. Thus the optimum one-step predictor for the least-favorable spectrum (2.38) is given by

$$\hat{S}_{t+1} = \sum_{k=1}^m a_k S_{t+1-k}, \quad (2.40)$$

and so the robust one-step predictor for (2.36) is a simple finite-length predictor with coefficients determined from the Yule-Walker equations. That this predictor uses only m past samples is an intuitively pleasing result, since we have no knowledge of the correlation structure beyond lags of length m . This result has been generalized to p -step prediction by Franke and Poor in [2.21].

2.5. Robust Equalization of an Uncertain Channel

In the observation model of (2.1) and (2.18) it is assumed that the signal we are interested in estimating is corrupted only by the additive orthogonal noise process $\{N(t); -\infty < t < \infty\}$. However, in many situations of practical interest we also have linear distortion or spreading of the signal by the observation channel. This situation can be described by an observation model of the form

$$Y(t) = \int_{-\infty}^{\infty} k(t-\tau) S(\tau) d\tau + N(t), \quad -\infty < t < \infty, \quad (2.41)$$

where the noise and signal processes satisfy the assumptions made after (2.1) and where $k(t)$ is a channel spread function. The problem of estimating $S(t)$ from the observation of (2.41) is the general problem of channel equalization (or deconvolution) plus filtering, which arises in many applications such as communications, sonar, seismology, and image processing.

If, as before, we consider signal estimates of the form

$$\hat{S}(t) = \int_{-\infty}^{\infty} h(t-\tau) Y(\tau) d\tau, \quad (2.42)$$

then with known Φ_S, Φ_N , and k , the optimum (minimum-MSE) equalization filter is given straightforwardly by the transfer function

$$H_E(\omega) = \frac{K_c^*(\omega)\Phi_S(\omega)}{|K_c(\omega)|^2\Phi_S(\omega) + \Phi_N(\omega)}, \quad (2.43)$$

where K_c is the transfer function of the channel and where K_c^* denotes its complex conjugate.

In practice, the transfer function of the channel is rarely known precisely. However, to design the optimum equalization filter of (2.43) one needs exact knowledge of this channel characteristic. Thus, as in the case in which the signal and noise spectra are uncertain, it is necessary to seek an alternative design objective for situations in which the channel characteristic is uncertain. In particular, if we can model the channel as having a transfer function which lies in an uncertainty class \mathbf{K} , then an

appropriate design criterion might be a minimax MSE formulation where the maximization is taken over all channels in the class \mathcal{K} .

This minimax formulation for equalization of uncertain channels has been proposed and investigated by Moustakides and Kassam in [2.30]. To illustrate the nature of solutions to this problem we suppose, for example, that the channel is known to have a linear phase characteristic and that the channel gain $|K_c(\omega)|$ is known only to lie between known upper and lower limits $L_c(\omega)$ and $U_c(\omega)$, respectively. That is, suppose we have

$$L_c(\omega) \leq |K_c(\omega)| \leq U_c(\omega) \quad (2.44)$$

for all frequencies ω , but that $|K_c(\omega)|$ is otherwise unknown. Then, assuming $U_c(\omega) > 0$ and $\Phi_N(\omega) > 0$, it can be shown [2.30] that the magnitude of the robust (minimax) equalization filter $H_{E,R}(\omega)$ is given by

$$H_{E,R}(\omega) = \begin{cases} \frac{L_c(\omega)\Phi_S(\omega)}{L_c^2(\omega)\Phi_S(\omega) + \Phi_N(\omega)}, & \text{if } \frac{\Phi_N(\omega)}{L_c^2(\omega)\Phi_S(\omega)} \geq A(\omega) \\ \frac{2}{L_c(\omega) + U_c(\omega)}, & \text{if } \frac{\Phi_N(\omega)}{L_c^2(\omega)\Phi_S(\omega)} < A(\omega) \end{cases} \quad (2.45)$$

where

$$A(\omega) = \frac{U_c(\omega) - L_c(\omega)}{2L_c(\omega)}. \quad (2.46)$$

Note that the quantity $\Phi_N(\omega)/L_c^2(\omega)\Phi_S(\omega)$ is a measure of the maximum possible noise-to-signal ratio at frequency ω , and $A(\omega)$ is a measure of the uncertainty in the knowledge about the channel. Thus, (2.45) implies that if the maximum noise-to-signal ratio at a given frequency is larger than the uncertainty in the channel model, then we use the optimum gain for the lower channel, $L_c(\omega)$, at that frequency. Alternatively, if the reverse is true for a given frequency, then we simply ignore the noise at that frequency and use the gain prescribed by the inverse average channel. A similar result can be obtained for situations in which the phase of $K_c(\omega)$ is also unknown.

2.6. Robust Filtering of Signals in Correlated Noise

Another aspect of the assumptions made above on the model (2.1) and (2.18) that is sometimes violated in practice is that of no correlation between the signal and noise processes. Signal-dependent noise arises in many applications such as radar or sonar, for example, due to phenomena such as multipath and clutter. Often the correlation between signal and noise in such applications is not well modeled, so that robust techniques are useful.

If we assume that the signal and noise processes of (2.1) are jointly wide-sense stationary, then their total correlation picture can be described by the spectral density matrix D given by

$$D(\omega) = \begin{bmatrix} \Phi_S(\omega) & \Phi_{SN}(\omega) \\ \Phi_{SN}^*(\omega) & \Phi_N(\omega) \end{bmatrix} \quad (2.47)$$

where Φ_S and Φ_N are as before, Φ_{SN} is the cross spectrum between $\{S(t); -\infty < t < \infty\}$ and $\{N(t); -\infty < t < \infty\}$, and where, as before, the asterisk denotes complex conjugation.

For a signal estimate of the form

$$\hat{S}(t) = \int_{-\infty}^{\infty} h(t-\tau) Y(\tau) d\tau, \quad (2.48)$$

we have that the MSE is given straightforwardly in this case by

$$\begin{aligned} E\{|S(t) - \hat{S}(t)|^2\} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [|1-H(\omega)|^2 \Phi_S(\omega) + |H(\omega)|^2 \Phi_N(\omega) \\ &\quad - 2H(\omega)\Phi_{SN}(\omega) + 2|H(\omega)|^2 \operatorname{Re}\{\Phi_{SN}(\omega)\}] d\omega \\ &\triangleq e(\Phi_S, \Phi_N, \Phi_{SN}; H). \end{aligned} \quad (2.49)$$

The minimum-MSE filter is given by

$$H_W(\omega) = \frac{\Phi_S(\omega) + \Phi_N(\omega)}{\Phi_S(\omega) + 2 \operatorname{Re} \{ \Phi_{SN}(\omega) \} + \Phi_N(\omega)}, \quad (2.50)$$

and the corresponding minimum value of MSE is given by

$$e(\Phi_S, \Phi_N, \Phi_{SN}; H_W) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\Phi_S(\omega) - \frac{|\Phi_S(\omega) + \Phi_{SN}(\omega)|^2}{\Phi_S(\omega) + 2 \operatorname{Re} \{ \Phi_{SN}(\omega) \} + \Phi_N(\omega)} \right] d\omega. \quad (2.51)$$

$$\triangleq e_W(\Phi_S, \Phi_N, \Phi_{SN}).$$

Of course, with $\Phi_{SN}(\omega)$ identically zero the expressions (2.49)-(2.51) reduce to (2.3)-(2.5).

If the correlation matrix D of (2.47) is not known precisely but, rather, is known only to lie in some class \mathcal{D} of spectral density matrices, then to seek a robust alternative to the optimum filter of (2.50) we can replace the objective function $e(\Phi_S, \Phi_N, \Phi_{SN}; H)$ with its supremum over \mathcal{D} . As in the uncorrelated signal and noise case, this yields a minimax game for designing a robust filter H_R . The solution to this problem has been considered by Moustakides and Kassam [2.31, 2.32]. Within mild conditions it can be shown that, for convex classes \mathcal{D} , a spectral density matrix $D_L \in \mathcal{D}$ and its corresponding optimum filter from (2.50) (when it is uniquely defined) will be a saddle-point solution to this game if and only if, D_L is least favorable; i.e., if and only if

$$D_L = \arg \left\{ \min_{D \in \mathcal{D}} e_W(\Phi_S, \Phi_N, \Phi_{SN}) \right\}. \quad (2.52)$$

To illustrate the possible structure of the least-favorable spectral density matrix, it is interesting to consider the case in which the signal and noise spectra are known but the cross spectrum is not known precisely. In particular, suppose we can establish the fact that the cross spectrum satisfies the conditions

$$0 \leq L(\omega) \leq |\Phi_{SN}(\omega)| \leq U(\omega), \quad -\infty < \omega < \infty, \quad (2.53)$$

where L and U are given functions. (Such a model might arise, for example, if a confidence band for the cross spectrum could be determined via spectrum estimation.) For this model it can be shown (see

Moustakides and Kassam [2.32] that the least-favorable cross spectrum is given by

$$\Phi_{SN,L}(\omega) = \begin{cases} -L(\omega) & \text{if } B(\omega) \leq L(\omega) \\ -B(\omega) & \text{if } L(\omega) \leq B(\omega) \leq U(\omega) \\ -U(\omega) & \text{if } U(\omega) \leq B(\omega), \end{cases} \quad (2.54)$$

where B is defined by

$$B(\omega) \triangleq \min \{ \Phi_S(\omega), \Phi_N(\omega) \}. \quad (2.55)$$

Thus, at a given frequency, whether the worst case involves minimum cross spectral density, maximum cross spectral density, or something in between, depends on the relationship among L , B , and U at that frequency. If, for example, nothing is known about the cross spectrum, then all we can say is that

$$0 \leq | \Phi_{SN}(\omega) | \leq [\Phi_S(\omega) \Phi_N(\omega)]^{1/2}, \quad -\infty < \omega < \infty, \quad (2.56)$$

where the right-hand inequality follows from the required nonnegative-definiteness of $D(\omega)$. Since $\min\{a, b\} \leq [ab]^{1/2}$ for all $a \geq 0$ and $b \geq 0$, it follows straightforwardly from (2.54) that in this case we have

$$\Phi_{SN,L}(\omega) = -\min \{ \Phi_S(\omega), \Phi_N(\omega) \}, \quad -\infty < \omega < \infty. \quad (2.57)$$

Equation (2.57) together with the optimum-filter expression of (2.50) gives that the robust filter for completely unknown cross correlation is given by the somewhat surprising result,

$$H_R(\omega) = \begin{cases} 1, & \text{if } \Phi_S(\omega) > \Phi_N(\omega) \\ 0, & \text{if } \Phi_S(\omega) < \Phi_N(\omega). \end{cases} \quad (2.58)$$

Other results for different bounded uncertainty classes are given in [2.31].

2.7. Uncertainty Classes of Spectral Measures

In the above discussion, we have considered several aspects of the basic problem of robust linear filtering of stationary signals in additive stationary noise. In particular, we have discussed the basic robust filtering problem in a noncausal framework, and we have also discussed the treatment of filter causality, equalization, and cross-correlation between signal and noise. One issue which has not been discussed is that of how one can deal with stationary processes which do not necessarily have spectral densities but, rather, have associated spectral measures (or, equivalently, spectral distributions). This situation arises in practice primarily when there are pure harmonics in the signal and/or noise. For example, in a communications receiver, one might have processes that nominally have spectral densities but also contain pure-harmonic uncertainties caused by sources such as line hum or tone jamming. To treat the robustness problem in this more general context requires a measure-theoretic formulation of the filtering problem. This issue has been considered by Poor [2.9] for the case of noncausal filtering and by Vastola and Poor [2.16] and Franke and Poor [2.20] for the case of causal filtering. The results obtained for this situation are quite similar to those for the case in which all processes concerned have spectral densities, with the additional advantage that quite general results concerning the existence of least-favorable spectral measures can be obtained.

In particular, suppose we have the observation model of (2.1) in which the signal and noise processes are real, zero-mean, orthogonal, wide-sense stationary, and quadratic-mean continuous. Then for an estimate of the form of (2.2) we can always write the MSE as

$$E \{ | S(t) - \hat{S}(t) |^2 \} = \left[\int_{-\infty}^{\infty} | 1 - H(\omega) |^2 m_S(d\omega) + \int_{-\infty}^{\infty} | H(\omega) |^2 m_N(d\omega) \right] \triangleq e(m_S, m_N; H), \quad (2.59)$$

where m_S and m_N are the spectral measures associated (via Bochner's Theorem [2.16]) with the processes $\{S(t); -\infty < t < \infty\}$ and $\{N(t); -\infty < t < \infty\}$, respectively. The transfer function that minimizes (2.59) for fixed m_S and m_N is given by

$$H_w(\omega) = \frac{dm_S}{d(m_S + m_N)}(\omega), \quad (2.60)$$

where the differentiation in (2.60) denotes the Radon-Nikodym derivative. Alternatively, if m_S and m_N are known only to lie in classes M_S and M_N , respectively, then we would like to solve the minimax problem

$$\min_H \left\{ \sup_{(m_S, m_N) \in M_S \times M_N} e(m_S, m_N; H) \right\}. \quad (2.61)$$

An interesting class of such problems arises when total powers in the signal and noise processes are both known (i.e., when $m_S((-\infty, \infty)) \triangleq 2\pi P_S$ and $m_N((-\infty, \infty)) \triangleq 2\pi P_N$ are constant on m_S and m_N). In this case, solutions to (2.61) can be characterized for a general type of uncertainty class studied by Huber and Strassen in [2.32]. These classes are of the form

$$M_v = \{m \in M \mid m(B) \leq v(B) \text{ for all } B \in \mathcal{B} \text{ and } m(\mathcal{R}) = v(\mathcal{R})\}, \quad (2.62)$$

where \mathcal{R} denotes the set of real numbers, \mathcal{B} denotes the σ -field of Borel sets in \mathcal{R} , M denotes the set of all spectral measures on $(\mathcal{R}, \mathcal{B})$, and v is a 2-alternating (Choquet) capacity. A 2-alternating capacity in this context is a set function mapping \mathcal{B} to \mathcal{R} with the following properties: $v(\emptyset) = 0$, $v(\mathcal{R}) < \infty$, $A \subset B \Rightarrow v(A) \leq v(B)$, v is continuous from below and is continuous from above on closed sets, and $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$ for all A and B in \mathcal{B} . Examples of classes of the type of (2.62) are given in [2.33]-[2.36] and include such common uncertainty models as the ϵ -contaminated class, the variational neighborhood, the band model, and others.

Classes of the form of (2.62) are useful for the problem of (2.61) because the solution to (2.61) for the situation in which $M_S = M_{v_S}$ and $M_N = M_{v_N}$ for two 2-alternating capacities v_S and v_N is given by (see Poor [2.9])

$$H_R(\omega) = \frac{\pi(\omega)}{\pi(\omega) + 1}, \quad -\infty < \omega < \infty \quad (2.63)$$

where $\pi = dv_S/dv_N$ is a Radon-Nikodym derivative between v_S and v_N . (Note that H_w from (2.60)

could be written as $H_w(\omega) = \lambda(\omega)/(\lambda(\omega) + 1)$ where $\lambda = dm_S/dm_N$.) Thus, for uncertainty classes generated by 2-alternating capacities the general solution to (2.61) is characterized. Similar results for causal problems are discussed in [2.16].

2.8. Robust Interpolation

A problem related to robust filtering is that of robust signal interpolation. In the signal interpolation problem we have a discrete time signal $S(k)$, $k = 0, \pm 1, \pm 2, \dots$, which we observe for all $k \neq 0$. Our objective is to estimate $S(0)$ from $\{S(k); k = \pm 1, \pm 2, \dots\}$ with a linear estimate

$$\hat{S}(0) = \sum_{k \neq 0} h(k)S(k). \quad (2.64)$$

The mean-square estimation error incurred by using (2.64) is given straightforwardly by

$$E\{|\hat{S}(0) - S(0)|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |1 - H(\omega)|^2 \Phi_S(\omega) d\omega \quad (2.65)$$

$$\triangleq e(H; \Phi_S),$$

where H is the transfer function of the filter sequence $\dots, h(-2), h(-1), 0, h(+1), h(+2), \dots$ and where Φ_S is the spectrum of the signal. For fixed Φ_S , the MSE of (2.65) is minimized by the interpolator with transfer function

$$H_I(\omega) = 1 - \frac{1}{\Phi_S(\omega)} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (1/\Phi_S(\omega)) d\omega \right)^{-1}. \quad (2.66)$$

(Note that the zeroth Fourier coefficient of H_I equals zero, as is necessary for H_I to be a valid interpolator.) The minimum possible value of the mean-square interpolation error for a given signal spectrum Φ_S is given by

$$e(H_I; \Phi_S) = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} (1/\Phi_S(\omega)) d\omega \right]^{-1} \triangleq e^*(\Phi_S). \quad (2.67)$$

Suppose we design an interpolator based on an assumed signal spectrum Φ_S^g , but that the true spectrum is Φ_S . The resulting mean-square interpolation error is given by inserting the optimum interpolator (2.66) for Φ_S^g into (2.65), in which case we have the mismatch error

$$\frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_S(\omega) / [\Phi_S^g(\omega)]^2 d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} [1/\Phi_S^g(\omega)] d\omega} \quad (2.68)$$

Note that the numerator integrand $\Phi_S / (\Phi_S^g)^2$ implies a high degree of sensitivity of the interpolation error to large spectral components in the actual spectrum that are not predicted by the nominal model. Thus, as with the other linear estimation problems discussed in the preceding sections, it is desirable to robustify an interpolator against uncertainty in Φ_S . This can be done by embedding Φ_S in a spectral uncertainty class and replacing the minimization of $e(H; \Phi_S)$ with the minimization of its worst-case value over \mathfrak{S} . This problem was posed and solved by Taniguchi in [2.37] for the case in which \mathfrak{S} is an ϵ -contaminated model, and the solution to the general case was characterized by Kassam in [2.38].

In [2.38] it is shown that, within mild conditions on \mathfrak{S} , a spectrum $\Phi_{S,L} \in \mathfrak{S}$ and its corresponding optimum interpolator from (2.65) will be minimax robust over \mathfrak{S} if and only if $\Phi_{S,L}$ is least-favorable; i.e., if and only if

$$\Phi_{S,L} = \arg \max_{\Phi_S \in \mathfrak{S}} e^*(\Phi_S), \quad (2.69)$$

where e^* is from (2.67). Moreover, note that $\Phi_{S,L}$ equivalently minimizes $\frac{1}{2\pi} \int_{-\pi}^{\pi} C(\Phi_S(\omega)) d\omega$ where C is the convex function $C(x) = 1/x$. Thus, as with the filtering and prediction problem discussed above, least-favorables for many normalized uncertainty classes can be found by applying analogous results from Huber's robust hypothesis-testing formulation (see Kassam [2.38] for details).

2.9. Robust Linear Filtering for Vector Observations

Many problems of practical interest involve the vector version of the observation model (2.1). Although conceptually similar to the scalar problem, vector filtering problems after present practical difficulties that are usually overcome by the imposition of specific structural models such as a finite-state signal model. These difficulties are not alleviated in the robust versions of vector filtering problem, and so the types of uncertainty models for which minimax robust filters can be developed are usually more structured than in the single variable case. Recent developments for the vector problem analogous to those for the scalar problem discussed in the preceding subsection are found in [2.39-2.42]. As an example of the type of uncertainty class that can be treated in this context, Chen and Kassam [2.39] consider an uncertainty set of spectral density matrices which share a common (constant) eigenstructure but whose eigenspectra lie in band models.

Another set of structural assumptions that allow for the treatment not only of the vector case but also of time-varying situations is the usual state-space signal model

$$\begin{aligned} S(t) &= C(t)X(t) \\ \dot{X}(t) &= A(t)X(t) + v(t); \quad t > t_0 \end{aligned} \tag{2.70}$$

where $E\{v(t)v(s)\} = Q(t)\delta(t-s)$, $\{v(t)\}$ is independent of $\{N(t)\}$, and $A(t)$ and $C(t)$ are matrices of appropriate dimensions. We assume that the observation noise $\{N(t)\}$ has correlation $E\{N(t)N(s)\} = R(t)\delta(t-s)$. The best linear estimator of $X(t)$ from $\{Y(\tau); t_0 \leq \tau \leq t\}$ is given by the well-known Kalman-Bucy filter

$$\dot{\hat{X}}(t) = A(t)\hat{X}(t) + K(t)[Y(t) - C(t)\hat{X}(t)], \tag{2.71}$$

where $K(t)$ is the Kalman gain matrix determined from the error covariance $P(t) = \text{cov}(X(t) - \hat{X}(t))$.

Although uncertainty can arise within the model (2.70) in several ways, there are two basic types of uncertainty that can affect the performance of linear estimators. One of these is uncertainty in the second-order statistical properties of the noise (and initial condition), such as the matrices Q and R and

the assumptions that $\{v(t)\}$ and $\{N(t)\}$ are white and uncorrelated. The other type of uncertainty is uncertainty in the dynamical model itself; i.e., uncertainty in the A and C matrices. Problems of the first type have been studied by a number of investigators including D'Appolito and Hutchinson [2.42], Vandelinde [2.43], Morris [2.44], Poor and Looze [2.45], and Verdú and Poor [2.46]. The basic result for minimax design within this type of uncertainty is that the linear structure is preserved in the minimax solution and the corresponding gain matrix is chosen to be optimum for a least-favorable set of second-order statistics. These results are thus of the same type as those discussed in the preceding subsections. Problems of the second type, however, have received less attention in this context, and have been treated only recently in a paper by Martin and Mintz [2.47]. As one might expect, the effects of uncertainties in the dynamical structure of the model (the A matrix) are quite different than those of uncertainty in the noise statistics. In particular, the minimax solutions for uncertain A matrix are not pure strategies (i.e., they are not simply of the form (2.71)) but rather are mixed strategies — randomizations among several filters of the type (2.71).

3. ROBUST FILTERS FOR SIGNAL DETECTION AND RELATED APPLICATIONS

One of the most pervasive of functions that signal processing schemes are required to carry out is that of detecting a signal of a generally known type in noisy observations. Obvious examples of applications in which signal detection is required are provided by radar (detection of echo pulses) and sonar (detection of a random signal process present in an array of hydrophones). Numerous other applications may be listed, for example detection of specified two-level pulse code sequences in communication systems and detection of abnormal patterns in medical imaging.

In the classical theory of signal detection one starts with specific statistical or deterministic models for the signal and observation process, and proceeds to obtain a detector which has optimum performance under an appropriate detection performance measure such as detection probability or output signal-to-noise ratio. In this section we will restrict attention to the class of *linear* detectors or filters and to the output *signal-to-noise ratio* (SNR) as the measure of a linear filter's detection performance. The design of optimum detectors under these two restrictions has been carried out for a large number of special applications. There are several reasons for the widespread acceptance of these design restrictions. One is that optimization of the linear filter design to maximize output SNR is usually a simple mathematical problem and leads to explicit solutions, and the implementation of a linear filter is usually straightforward. Another reason is that in many cases the statistical models for the noise and any random signal are often stated in terms of only their correlation functions or power spectral densities (that is, *second-order* statistical properties). In such cases one simply does not have enough statistical information, such as the parametric form of the probability density functions, to allow optimization to be attempted over a larger class of detectors. If the noise process can be assumed to be *Gaussian* then it is possible to show that the optimum detector maximizing the detection probability in the detection of a *deterministic* signal is indeed the optimum linear filter maximizing the output SNR; this is the ubiquitous *matched filter* of detection theory.

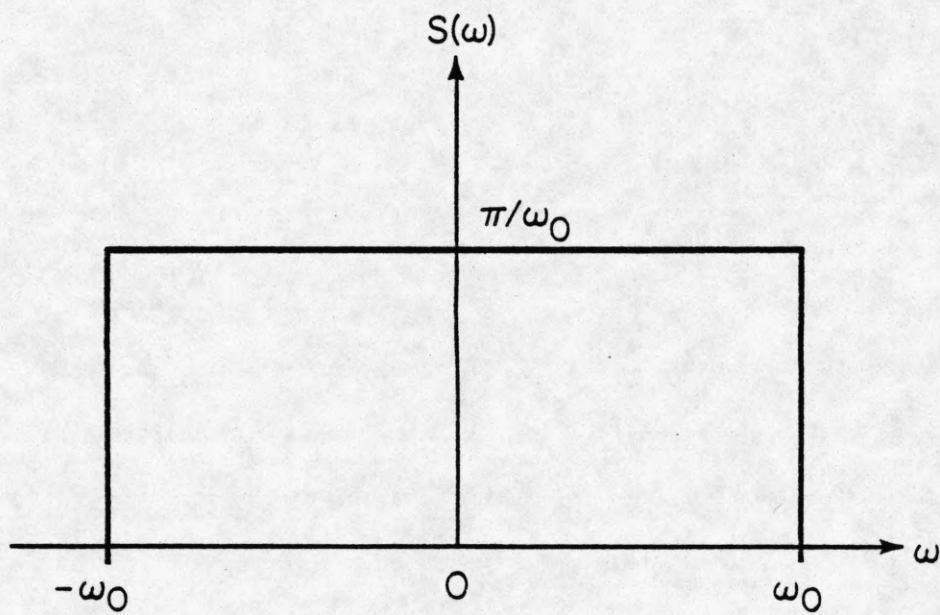
In general the matched filter specification depends on the exact form of the deterministic signal for which it maximizes its output SNR, and on the exact noise autocorrelation function or power spec-

tral density. Since these quantities are rarely known exactly, the need for applying robust techniques may arise naturally in the matched filtering problem. One interesting extreme case occurs when the nominal assumptions on signal and noise characteristics are such as to result in a *singular* detection problem. This means that under nominal conditions the output SNR is infinitely large, implying perfect detection is possible. Consider, for example, an ideal low-pass nominal signal $\text{sinc}(\omega_0 t)$ whose Fourier transform $S(\omega)$ is shown in Fig. 3.1, and let the noise have a nominal power spectral density $\Phi_N(\omega)$ which is the triangular function of Fig. 3.2. The matched filter frequency response $S^*(\omega)/\Phi_N(\omega)$ is shown in Fig. 3.3; it increases without bound as ω approaches $\pm\omega_0$. The output SNR under this situation is

$$SNR_o = \frac{1}{2\pi} \int_{-\omega_0}^{\omega_0} \frac{|S(\omega)|^2}{\Phi_N(\omega)} d\omega = \infty \quad (3.1)$$

But suppose that the signal deviates very slightly from the nominal and becomes $\text{sinc}([\omega_0 - \epsilon] t)$. The output SNR using the original matched filter now drops to zero, or $-\infty$ dB! In this situation where the signal bandwidth may not be precisely ω_0 , it would be better to design the matched filter for the smallest bandwidth that may be encountered. The resulting filter will then perform well for the minimum bandwidth signal and will be fairly insensitive to deviations of the signal bandwidth from Ω_0 . This basic idea is expanded upon in subsections 3.1 and 3.3, where we consider maximin robust matched filters which maximize the worst case performance over a given pair of classes for the allowable signal and noise characteristics. Also a property of such robust matched filters in most cases is that their performance is relatively *stable*, or not too variable, over the allowable classes of characteristics. The issues of stability and singularity in certain matched filtering problems have been considered by several authors, most notably Root [3.1] and Kailath [3.2]. The maximin robust matched filter has also been described as an *optimally stable* matched filter.

While the simple example we have considered above to illustrate the possibility of extreme sensitivity of performance of a nominal matched filter assumed that only the value of one parameter (the signal bandwidth) was variable, in general we deal with nonparametric classes of functions in



FP-8379

Fig. 3.1: Ideal low-pass nominal signal characteristic $S(\omega)$.

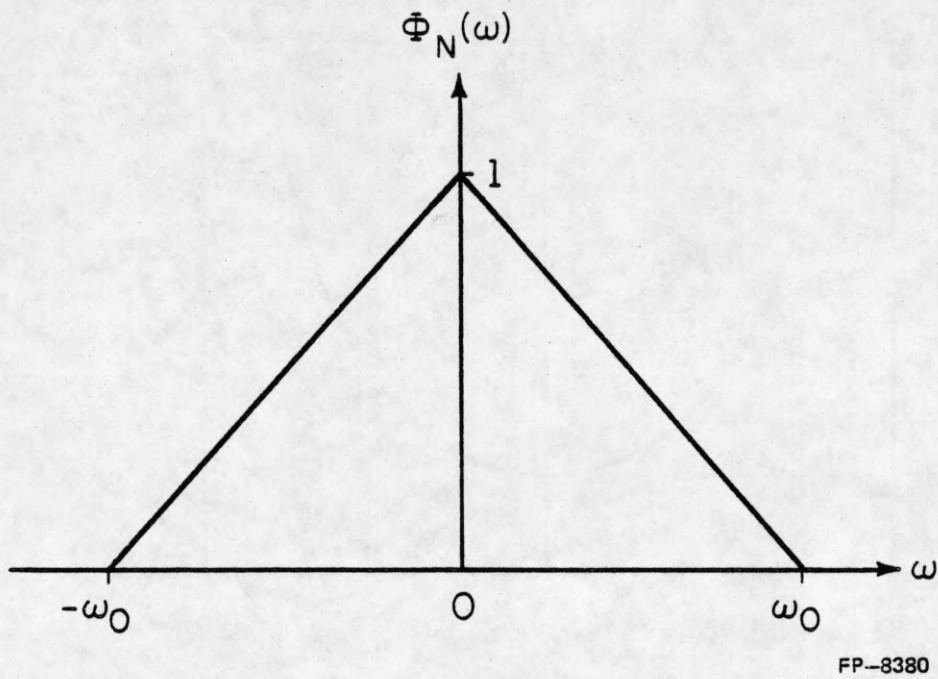
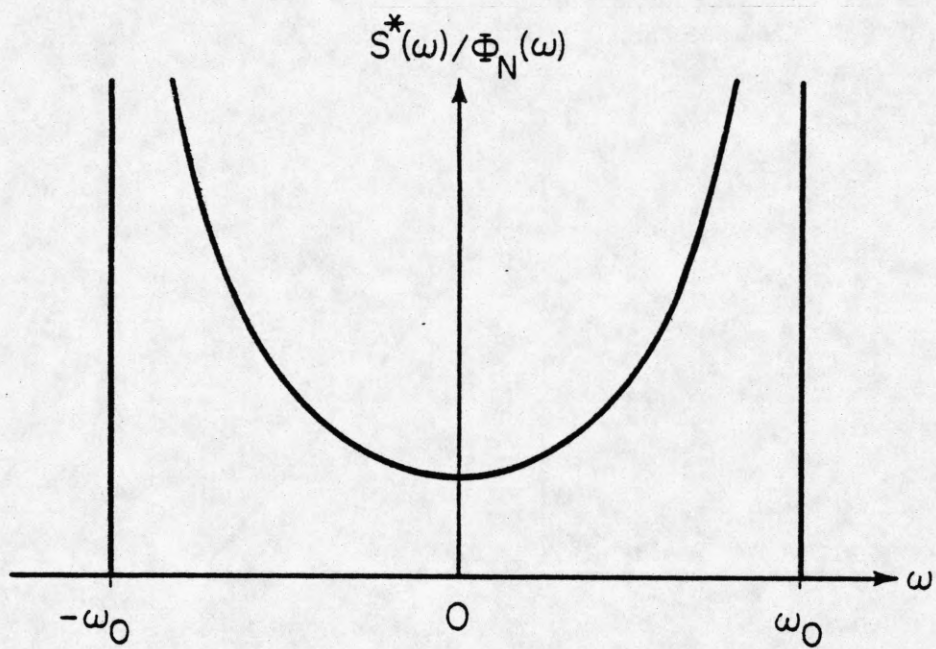


Fig. 3.2: Triangular noise power spectral density $\Phi_N(\omega)$.



FP-8378

Fig. 3.3: Matched filter frequency response for ideal low-pass signal and triangular noise power spectral density.

modeling the uncertainties in signal and noise characteristics. The results and interpretations we survey in this section differ from those in the previous section on robust filters for signal estimation primarily because our criterion of performance here is the output SNR. This leads to mathematical approaches which are conveniently unified using Hilbert spaces. In addition, no direct correspondence exists between these results on maximin robust matched filters and results on robust hypothesis tests, although some indirect connections do exist.

In subsection 3.2 we will give some interesting special applications of the general results on robust matched filters in spatial array processing and time-delay estimation. We will close this section with a more mathematical discussion in subsection 3.3 of the general Hilbert space approach for formulating robust linear detection problems; this approach provides a common framework for a variety of such problems.

3.1. Robust Matched Filters

As an example of a system using a matched filter for signal detection, consider the structure of the receiver for a pulse train in which a given pulse shape $p(t)$ occurs in the i -th position of the train with some amplitude a_i , $i=1,2,\dots,m$. In a pulsed radar system these might represent the echo pulses from a target in some fixed range-gate, with the amplitude factors a_i produced by the beam pattern of the receiving antenna as its main beam scans past the target position. The noisy received waveform may be described by the equation

$$Y(t) = \theta \sum_{i=1}^m a_i \tilde{p}_i(t) + N(t), \quad 0 \leq t \leq mT, \quad (3.2)$$

where $\tilde{p}_i(t) = p(t - [i-1]T)$, the basic pulse $p(t)$ delayed by $(i-1)T$ units of time, T is the pulse repetition period, θ is the overall pulse-train amplitude and $N(t)$ is random noise. Note that the above observation model can also describe the received signal in a binary signaling scheme; we may take $m=1$ and $\theta = \theta_0$, $\theta = \theta_1$ as being two possible values of θ , with $\theta_0 = 0$ and $\theta_1 \neq 0$ corresponding to the case of on-off keying. For the general problem of testing $\theta = \theta_0$ vs. $\theta = \theta_1$ in the observation model

(3.2) when $N(t)$ is *Gaussian* with power spectral density $\Phi_N(\omega)$, an optimum receiver structure is shown in Fig. 3.4. This receiver is an implementation of the likelihood-ratio test for $\theta = \theta_0$ vs. $\theta = \theta_1$ in the case of Gaussian noise. More generally, the output of the matched filter at the end of each pulse interval has the maximum SNR obtainable from a linear filter. The frequency response $H_M(\omega)$ of the matched filter is

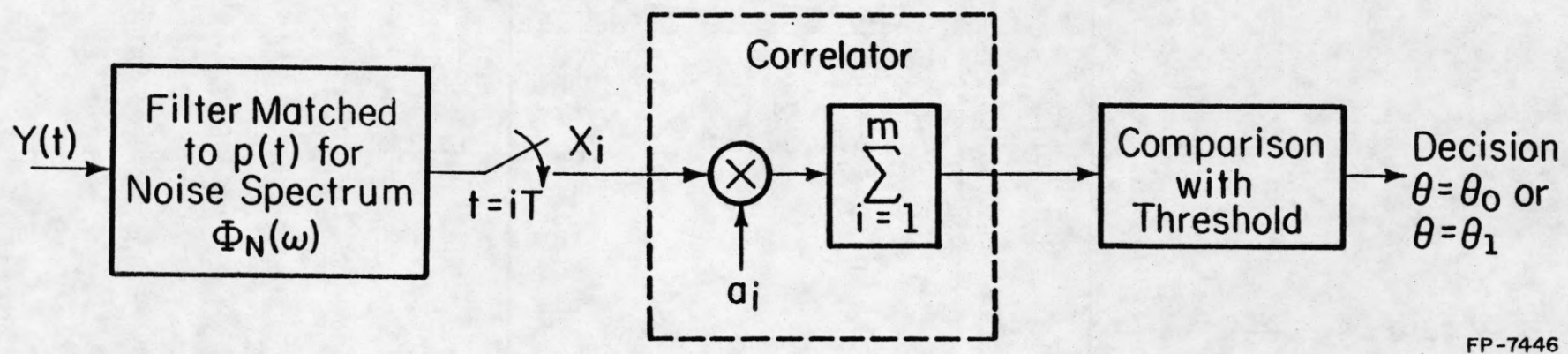
$$H_M(\omega) = \frac{P^*(\omega)}{\Phi_N(\omega)} e^{-j\omega T} \quad (3.3)$$

where $P(\omega)$ is the Fourier transform of the pulse $p(t)$. We shall now focus on the design of the matched filter when $P(\omega)$ and $\Phi_N(\omega)$ are not precisely known. In the next section we will survey techniques for modifying the correlator detector following the matched filter in Fig. 3.4, for situations where the noise at its input cannot be assumed to be Gaussian.

The general expression for the SNR at time T at the output of a filter with frequency response $H(\omega)$, when the input is $\theta p(t) + N(t)$, is

$$SNR = \frac{\left| \frac{\theta}{2\pi} \int_{-\infty}^{\infty} H(\omega) P(\omega) e^{j\omega T} d\omega \right|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 \Phi_N(\omega) d\omega} \quad (3.4)$$

This SNR is maximized by $H(\omega) = H_M(\omega)$ of (3.3). The synthesis of $H_M(\omega)$ in this case requires an exact knowledge of $P(\omega)$ and $\Phi_N(\omega)$. Suppose, however, that the received pulse $p(t)$ is a possibly distorted version of a nominal pulse $p_0(t)$, as illustrated in Fig. 3.5. A good measure of the degree of pulse distortion is the integrated squared difference between $p(t)$ and $p_0(t)$ or, equivalently (via Parseval's relation), the integrated squared difference between their Fourier transforms, $P(\omega)$ and $P_0(\omega)$. Thus, a useful signal uncertainty model for this matched filtering problem is to assume that the received pulse $p(t)$ is known only to lie in the class C_p which is the class of pulses $p(t)$ with Fourier transforms $P(\omega)$ satisfying



FP-7446

Fig. 3.4: Structure of optimum receiver for pulse train in Gaussian noise.

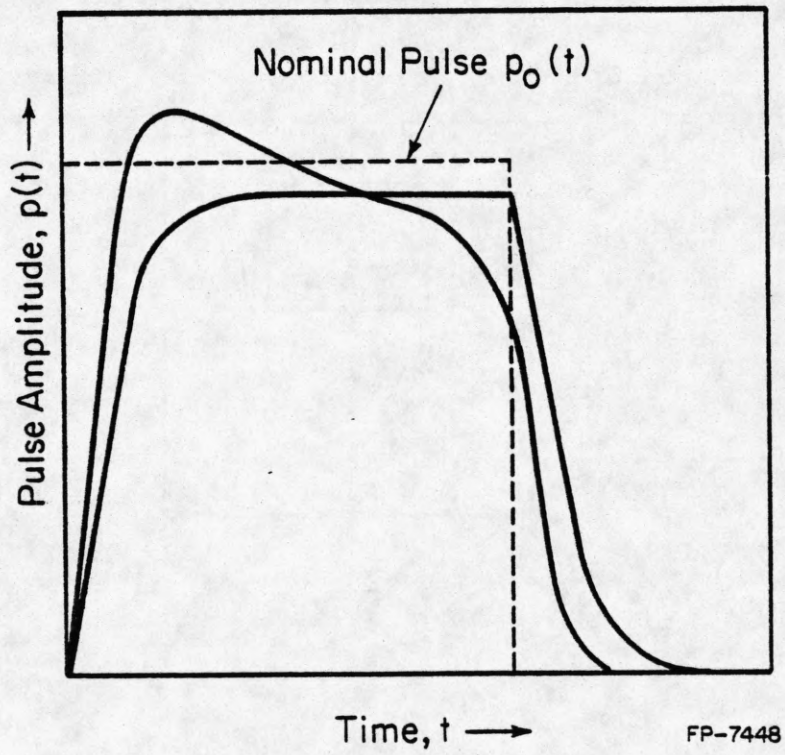


Fig. 3.5: Non-nominal pulse shapes $p(t)$.

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |P(\omega) - P_0(\omega)|^2 d\omega \leq \Delta. \quad (3.5)$$

Here $P_0(\omega)$ is the Fourier transform of the nominal pulse, and Δ determines the degree of uncertainty or possible distortion in $p(t)$.

The solution $H = H_R$ to the problem

$$\max_H \min_{P \in C_p} \text{SNR} \quad (3.6)$$

where SNR is from (3.4) is given by Kuznetsov [3.3] and Kassam, Lim and Cimini [3.4] as

$$H_R(\omega) = \frac{P_0^*(\omega)e^{-j\omega T}}{\Phi_N(\omega) + \sigma_0}, \quad (3.7)$$

where the positive constant σ_0 depends on Δ and is an increasing function of Δ . Equations (3.3) and (3.7) imply that the robust filter for (3.5) is forced to have less gain than it would otherwise have at frequencies where the noise power is low. It is the higher gain of the optimum filter for $P_0(\omega)$ [that is, $P_0^*(\omega)e^{-j\omega T} / \Phi_N(\omega)$] at these frequencies which makes it too sensitive to deviations in the pulse characteristics, the actual pulse possibly having lower energy at these frequencies. Thus, the filter (3.7) is robust from an intuitive viewpoint. Another interpretation of (3.7) is that signal uncertainty is translated into an additional white noise component in the noise spectrum. It is interesting to note that in the case of white noise with the above signal uncertainty the filter matched to the nominal pulse is itself the robust filter, since any absolute gain factor is irrelevant in matched filtering. This result agrees well with behavior observed in practice [3.5]. It is also interesting to note that the additional white noise component which appears under signal uncertainty in this approach ensures that the detection is not singular, regardless of how artificial the nominal model is.

Alternatively, suppose we assume that the pulse shape, say $p_0(t)$, and the total noise power σ_N^2 are known, but that the true noise spectrum is known only to lie in a class C_N of spectra bounded by known upper and lower bounds $U_N(\omega)$ and $L_N(\omega)$, respectively. That is, suppose we assume that the spectral shape of the noise is constrained only by the *band model*

$$L_N(\omega) \leq \Phi_N(\omega) \leq U_N(\omega), \quad -\infty < \omega < \infty \quad (3.8a)$$

and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_N(\omega) d\omega = \sigma_N^2. \quad (3.8b)$$

This class is illustrated in Fig. 3.6. In this case the robust filter can again be given an interesting interpretation. It turns out that the robust filter $H_R(\omega)$ in this case is the matched filter for detecting $p_0(t)$ in noise whose spectrum tends to be as close in its shape as possible to the shape of $|P_0(\omega)|$. Specifically, the results in [3.3, 3.4] show that the robust filter solving

$$\max_H \min_{\Phi_N \in C_N} \text{SNR} \quad (3.9)$$

is given by

$$H_R(\omega) = \frac{P_0^*(\omega)e^{-j\omega T}}{\Phi_{N,L}(\omega)} \quad (3.10)$$

where $\Phi_{N,L}(\omega)$ is a *least-favorable* noise spectrum in C_N given by

$$\Phi_{N,L}(\omega) = \begin{cases} L_N(\omega), & |P_0(\omega)|/k < L_N(\omega) \\ |P_0(\omega)|/k, & L_N(\omega) \leq |P_0(\omega)|/k \leq U_N(\omega) \\ U_N(\omega), & U_N(\omega) < |P_0(\omega)|/k. \end{cases} \quad (3.11)$$

Here k is a constant determined by the requirement that $\Phi_{N,L}(\omega)$ must be in C_N and therefore must have total power σ_N^2 , which is assumed to be known. It is clear that $\Phi_{N,L}(\omega)$ tries to follow as closely as possible a scaled version of $|P_0(\omega)|$. This phenomenon is illustrated in Fig. 3.7. An illustration of the magnitude of $H_R(\omega)$ for the example of Fig. 3.7 is shown in Fig. 3.8. This figure also shows what the optimum filter magnitude would be if a "nominal" noise spectrum having the characteristic $\frac{c}{2} [L_N(\omega) + U_N(\omega)]$ (that is, the normalized center of the band) is used in the design of the filter, where c is a normalizing constant chosen to get the correct power σ_N^2 . The implication for the robust filter is that its magnitude characteristic is flatter than it would be with any other $\Phi_N(\omega)$ in C_N . This

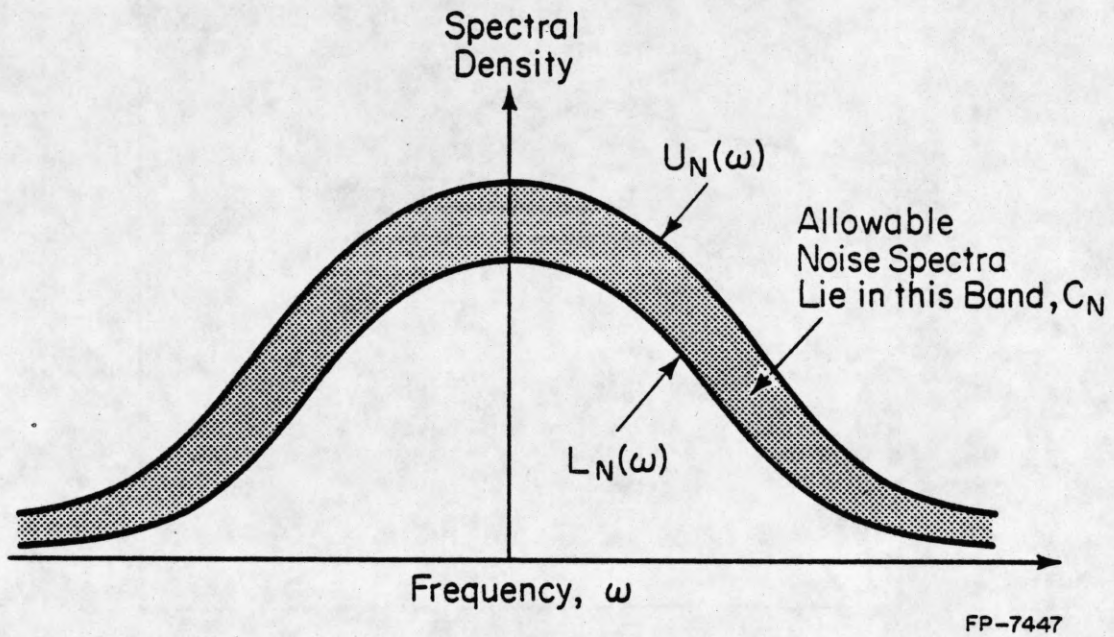


Fig. 3.6: Bounded noise power spectral density class.

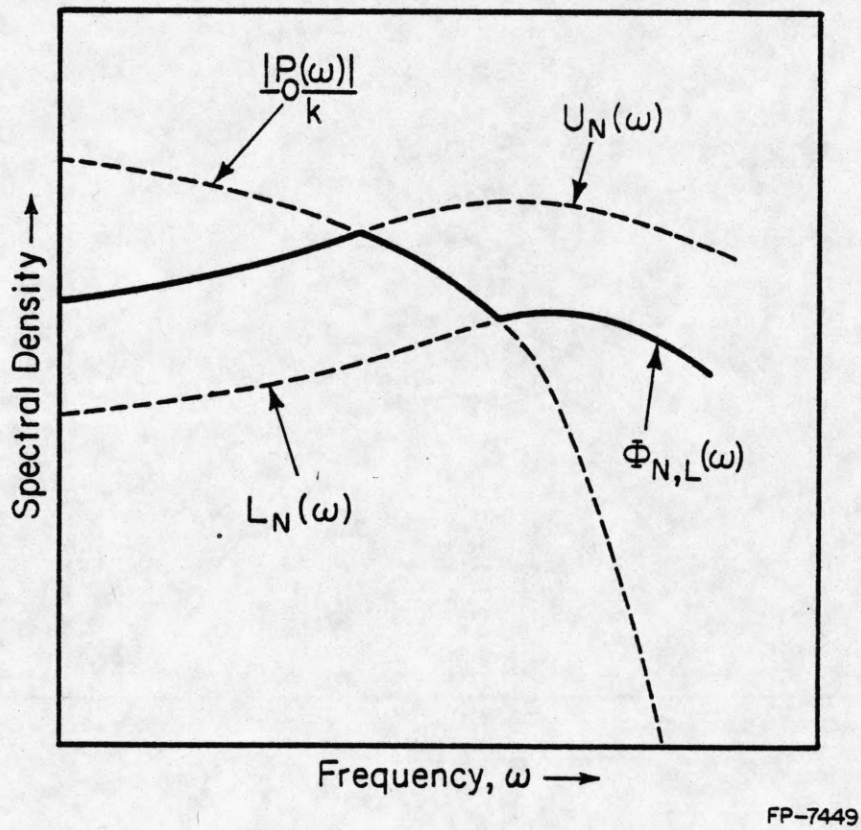


Fig. 3.7: Illustration of least favorable noise spectral density of Eq. (3.11).

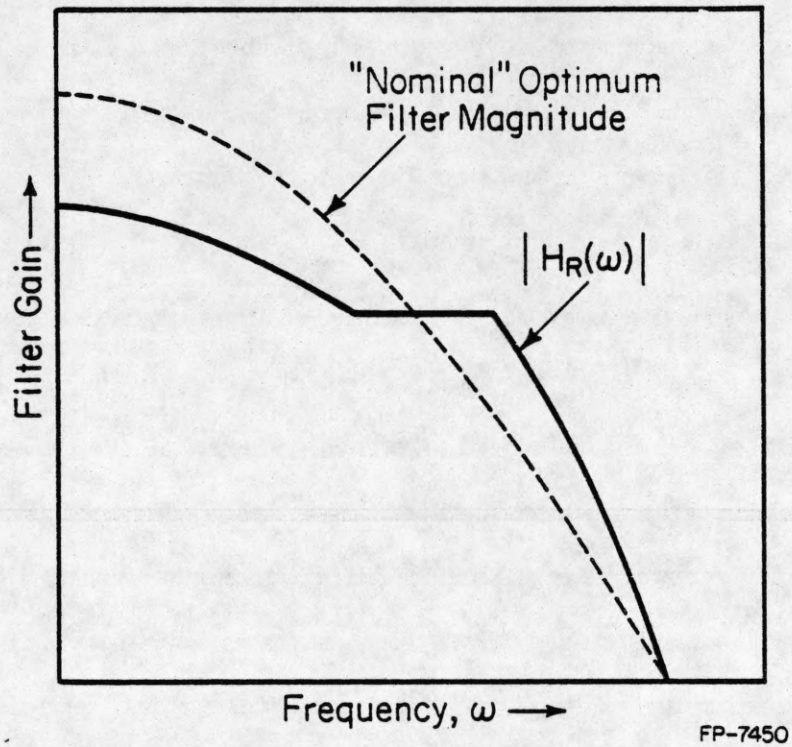


Fig. 3.8: Illustration of $|H_R(\omega)|$, robust matched filter amplitude response, for bounded noise power spectral density class.

means that extremes of gain and attenuation do not appear, making the filter less sensitive to the presence of additional noise power at relatively high-gain frequencies and to a reduction in noise power at relatively high-attenuation frequencies. The performance of the robust filter may thus be said to have been stabilized. Of course, the advantage gained in using a robust filter in any particular case depends on how extreme the frequency response of the nominally optimum filter is.

By combining and extending the above results it is possible to obtain the robust filter for uncertainties *simultaneously* about the pulse shape and the noise spectrum. This has been done in [3.4]. The extension of this spectral domain formulation of the matched filtering problem to the case of discrete-time observation processes is quite straightforward.

It is interesting to note that an explicit consideration of a game-theoretic approach in the design of a matched filter was first considered by Nilsson [3.6] and Zetterberg [3.7]. In [3.7], in particular, Zetterberg considered uncertainty in the noise spectrum which is assumed to have a known white noise component and another component which has fixed total power but is otherwise unknown. He obtained a result which is similar to that of (3.11). Although Zetterberg did not allow uncertainty in the signal, it is noteworthy that he assumed a fixed white component for the noise spectrum. We have seen above that a white noise component for the noise spectrum is prescribed for the L_2 signal uncertainty class of (3.5). Other related studies are also discussed in [3.7]. More recent work of this nature has also been reported by Cahn [3.8] and Turin [3.9].

The *multi-input* matched filtering problem under signal and noise uncertainty in the frequency domain has recently been considered in [3.10, 3.11]. An m -component signal vector $\underline{s}(t)$ may be modeled to belong to a generalized L_2 uncertainty class which is a neighborhood of some nominal signal vector $\underline{s}_0(t)$. In [3.10, 3.11] the class was defined in terms of a set of constants $\Delta_1, \Delta_2, \dots, \Delta_q$; these were now the bounds on the allowable values of q sums of component-wise integrated square differences between the vectors $\underline{s}(t)$ and $\underline{s}_0(t)$. The noise power spectral density matrix $\Phi_N(\omega)$ was assumed to lie in a generalized bounded spectrum class. Specifically, from the decomposition $\Phi_N(\omega) = P_N(\omega)\Lambda_N(\omega)P_N^\dagger(\omega)$ where $P_N(\omega)$ is a unitary matrix and $\Lambda_N(\omega)$ is diagonal, one may obtain an uncer-

tainty class by requiring the components of $\Lambda_N(\omega)$ to have known upper and lower bounds. In addition, a generalized noise power constraint was imposed which required that the sums of integrals of the components of $\Lambda_N(\omega)$ for r sets in some given partition of these components be equal to known values Q_1, Q_2, \dots, Q_r . The results given in [3.11] for this multi-input matched filtering problem form a useful extension of the scalar results we have mentioned above.

The frequency domain formulation of the matched filtering problem is useful for large (infinite) observation intervals. A direct time-domain approach in modeling the signal and noise uncertainties is desirable when the observation interval is some finite or semi-infinite interval $[t_o, t_f]$ with one or both end-points finite. In this case the matched filter weighting function (impulse response) is the solution of an integral equation. In the *finite-interval discrete-time* case the SNR functional and the equation for optimum filter weighting function are given in terms of matrix and vector quantities. Thus for this situation the robust matched filtering problem is amenable to simpler mathematical analysis. By viewing the finite-length discrete-time matched filtering problem as a multi-input (time-domain) problem it is seen that the ideas for the models for signal and noise quantities in the multi-input frequency domain case are directly applicable. This approach has been followed in [3.10, 3.11], where the signal vector $\underline{s} = (s_1, s_2, \dots, s_m)$ of samples s_i and the noise covariance matrix R_N are modeled as belonging to generalized l_2 and bounded eigenvalue uncertainty classes, respectively. In [3.12] a more general approach has been taken, and three signal vector uncertainty classes are considered (l_1, l_2 , and l_∞ classes defined as neighborhoods around a nominal signal vector). A general class of noise covariance matrices, defined in terms of matrix norms, has also been studied in [3.12]. The general approach in [3.12] is based on a Hilbert space formulation of the matched filtering problem, which we shall discuss in subsection 3.3. There we will also see how the time-domain continuous-time version of the problem may be approached. In fact, the finite-time-interval continuous-time robust matched filtering problem was originally considered by Kuznetsov [3.13] even before his frequency domain problem formulated in [3.3]. In [3.13] Kuznetsov used a signal uncertainty class in which the possible deviation $\delta(t)$ of a signal waveform $s(t)$ from a nominal signal $s_o(t)$ is bounded in its integrated squared error over the interval. Further aspects of this problem have been considered more recently by Burnashev [3.14]. A similar

model was also used for the uncertain noise covariance function, and in addition uncertainty in the mean value of the noise was also taken into account. Subsequently Aleyner [3.15] modified the signal uncertainty class by imposing a signal-energy equality constraint.

A variant of the robust matched filtering problem is that of optimal *nominal-signal selection* to obtain the best possible minimax performance. This situation has recently been considered in [3.16] for the finite-length discrete-time situation. There the actual signal \underline{s} is assumed to be a nominal signal \underline{s}_0 with an additive vector $\underline{\delta}$ which can belong to some class of possible vectors. For given \underline{s}_0 the robust matched filter can be obtained and the minimax performance level can be determined. With \underline{s}_0 allowed to be any nominal vector with fixed total energy, the choice of \underline{s}_0 maximizing this guaranteed performance level is obtained in [3.16] for signal uncertainty $\underline{\delta}$ in l_2 , l_1 , and l_∞ neighborhoods of the zero vector. While the result for the l_2 uncertainty class indicates that \underline{s}_0 should be the minimum-eigenvalue eigenvector of the noise covariance matrix (a result obtained independently by Kuznetsov [3.17]), this classical solution does not hold for the other uncertainty classes.

Before considering two specific applications in the next subsection, we mention that recently a maximin sonar system design problem has been considered [3.18], in which the signal and detector pair are picked to maximize the worst-case performance of the system over a class of possible reverberation scattering functions.

3.2. Two Examples of Robust Multi-Sensor Systems

We will now briefly describe two specific applications of these results on robust matched filters. Our first application is a narrowband spatial array system in which the individual sensors or antenna outputs are given complex (amplitude and phase) weights to detect signals arriving from any particular direction. While spatial matched filtering has previously been considered for this case [3.19], here we will be concerned with uncertainties in some of the signal and noise characteristics. It has been found that the use of minimax robust weights leads to another significant advantage in such a system. The second application we will consider is that of a system for estimating the time-delay between the random signal arriving at two sensors, with the signal observed in independent additive noise at each

sensor. While this is not directly a signal detection problem involving matched filters, we find that in at least one approach to optimum system design the mathematical analysis is closely related to that performed in obtaining the matched filter in a deterministic signal detection problem.

3.2.1. Robust Spatial Array for Signal Detection and Location

Consider a J -element narrowband linear array, illustrated in Fig. 3.9. The signals received at the array elements from any source are time-delayed (or phase-shifted) versions of the source signal. Let x_i be the position of the i -th array element, measured from an arbitrary origin. Let λ be the wavelength of the source signal, and suppose a far-field source is in some direction θ from the array broadside.

To detect a signal from the direction θ , a *phased-array* system uses a set of complex weights or phase shifts $h_{pi}(\theta) = \exp(-j 2\pi x_i \sin \theta/\lambda)$ to "line-up" in phase the signals received at each element. More generally, when sources of interference are present in specified directions and the observations at each element are noisy, one may design the weights $h_i(\theta)$ to maximize the SNR at the array output due to a signal in direction θ . Now it can be shown that the generalized $J \times J$ noise covariance matrix R_N for K interfering sources at locations $\theta_1, \theta_2, \dots, \theta_K$ is, element wise,

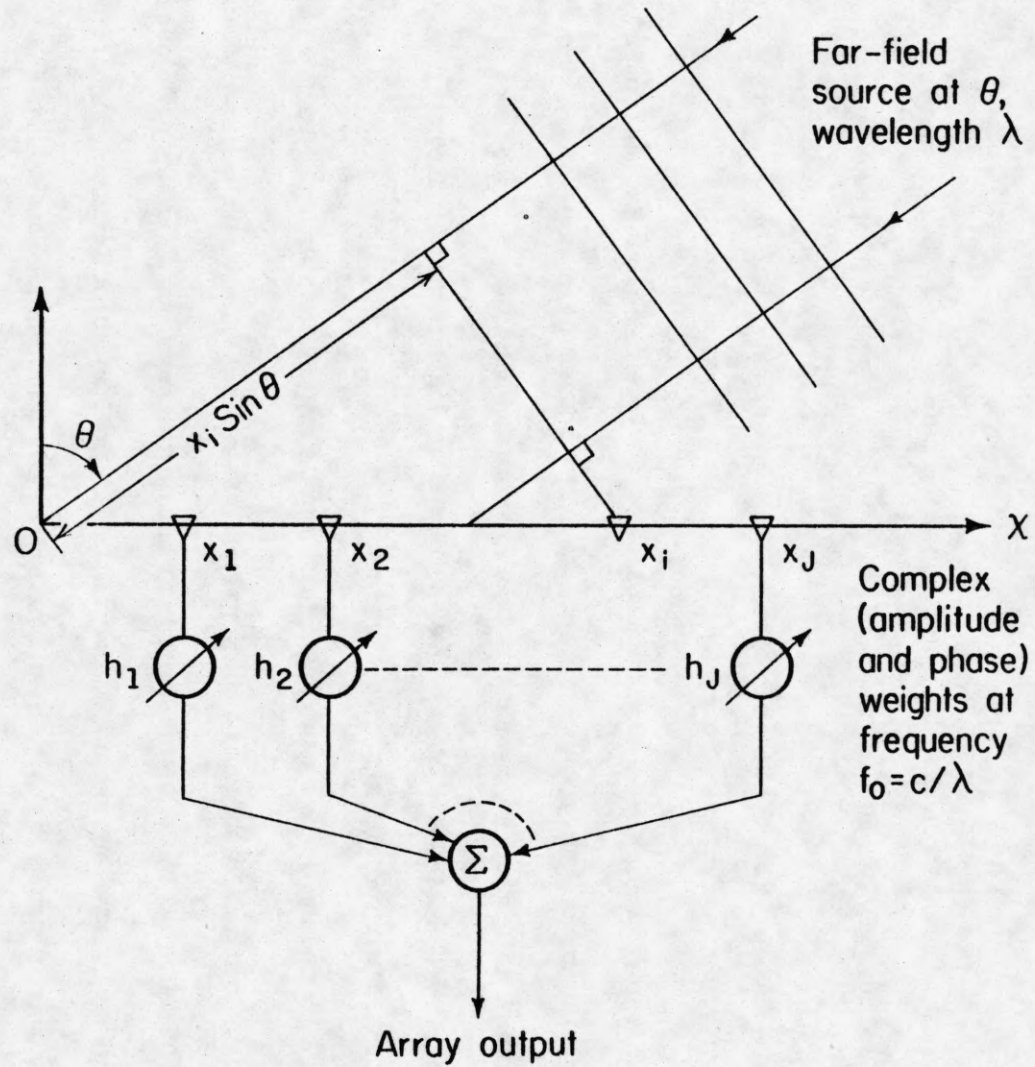
$$R_N(i, l) = W_i \delta_{il} + \sum_{k=1}^K R_k e^{j 2\pi(x_i - x_l) \sin \theta_k / \lambda} \quad (3.12)$$

where W_i is the white noise level at the i -th element, δ_{il} is the Kronecker delta and R_k is the power of the k -th interfering source. For a source in direction θ the nominal signal complex envelope at the i -th element is

$$s_{oi} = e^{j 2\pi x_i \sin \theta / \lambda}, \quad (3.13)$$

assuming a normalization to unit signal amplitude at each array element. The matched-filter array weight vector maximizing output SNR is

$$\underline{h}_0(\theta) = R_N^{-1} \underline{s}_0^*(\theta), \quad (3.14)$$



FP-8382

Fig. 3.9: Narrowband spatial array for signal detection and location.

where $\underline{s}_0(\theta) = [s_{01}(\theta), s_{02}(\theta), \dots, s_{0J}(\theta)]^T$. This weight vector becomes the phased-array weight vector with components $h_{pi}(\theta) = \exp(-j 2\pi x_i \sin \theta/\lambda)$ (with amplitudes normalized to unity) when there are no interfering sources present. For the general case of R_N given by (3.12), the spatial matched filter will have unequal amplitude weights and a set of phase weights different from that for the phased array case.

Our interest is in the case where there is some uncertainty in the characterization (3.13) of the signal components. This may arise because of imperfect propagation characteristics, element position uncertainties or element gain variations. To model such uncertainties we use the l_2 class of possible signal vectors $\underline{s}(\theta)$ which satisfy

$$\sum_{i=1}^J |s_i(\theta) - s_{0i}(\theta)|^2 \leq \Delta. \quad (3.15)$$

While the covariance matrix R_N may also be taken to be uncertain, we will here assume that it is known exactly. If the W_i are all equal to some common value W it can be shown that the eigenvectors of R_N are independent of W . Moreover, $(J - K)$ eigenvalues of R_N are equal to W and the others are of the form $W + \alpha_i$ where the α_i are independent of W . Such properties of R_N may be used in enlarging the class of allowable noise covariance matrices, although we shall not do so here.

The general result for finite-length discrete-time matched filters can be applied here to obtain the robust spatial matched filter weights as

$$\begin{aligned} \underline{h}_R(\theta) &= R_N^{-1} \underline{s}_L^*(\theta) \\ &= (R_N + \sigma_0 I)^{-1} \underline{s}_0^*(\theta); \end{aligned} \quad (3.16)$$

σ_0 is a constant which effectively adds uncorrelated noise components to the array elements, and depends on Δ , and $\underline{s}_L(\theta)$ is the least-favorable signal vector

Table 3.1 shows numerical values for the output SNR's using the nominally optimum weights and the minimax robust weights for $\Delta = 2, 4$ for the following system:

Table 3.1: Signal-to-noise ratio of array output using nominal and robust matched filter weights.

Δ	σ_0	SNR ($\underline{h}_R, \underline{s}_L, R_N$)	SNR ($\underline{h}_R, \underline{s}_0, R_N$)	SNR ($\underline{h}_0, \underline{s}_L, R_N$)
0	0	46.4	46.4	46.4
2	0.19	13.1	42.3	12.0
4	0.43	6.1	39.0	5.1

- θ , signal direction = 0°
 J , number of array elements = 12
 element spacing = $\lambda/2$
 K , number of interferences = 1
 R_1 , power of interference = 1.0
 θ_1 , direction of interference = -60°
 W_i , white noise levels, given by

$$W_i = \begin{cases} 0.25, & i = 1, 2, 8, 12 \\ 0.50, & i = 3, 4, 9, 10, 11 \\ 2.0, & i = 5 \\ 0.1, & i = 6, 7 \end{cases}$$

As expected, the robust weights $\underline{h}_R(\theta)$ give a worst-case performance (for $\underline{s}_L(\theta)$) which is better than the corresponding performance of the nominally optimum weights $\underline{h}_0(\theta) = R_N^{-1} \underline{s}_0^*(\theta)$. Note that the SNR is a function of the filter weights, the signal vector and the noise covariance matrix R_N , and that for $\Delta = 0$ the vector $\underline{h}_R(\theta)$ is the nominally optimum vector $\underline{h}_0(\theta)$.

The performance shown in Table 3.1 does not show a dramatic advantage (or disadvantage) obtained by using the robust spatial filter, in terms of SNR performance. However, there is an important aspect of performance of such an array system which is not captured by examining the output SNR in the correct "look" direction θ only. As the array scans for targets by adjusting its weights for different angles θ the output SNR becomes a function of the "look" angle θ for any fixed source direction $\theta = \theta_0$. Alternatively, a *beam pattern* for the array may be defined as

$$B(\theta, \theta_0) = \left| \sum_{i=1}^J h_i(\theta_0) s_{oi}(\theta) \right|, \quad (3.17)$$

the array output signal magnitude for a nominal signal in direction θ with array weighted to "look" in the direction θ_0 . Ideally $B(\theta, \theta_0)$ should have only a narrow peak at $\theta = \theta_0$, as a function of θ for fixed θ_0 .

Figures 3.10 and 3.11 show (for the numerical example giving Table 3.1) that the robust spatial matched filter weights produce normalized beam patterns that have a distinct advantage over the pattern of the nominally optimum array. The main beams of the robust array weights are narrower and their close-in sidelobes are appreciably below those of the nominally optimum array, a factor important in array design. Thus the use of a robust weighting scheme produces a "side" benefit which probably exceeds in importance its original justification of maintaining the output SNR under uncertainties. The robust weights differ from the nominally optimum weights as a result of the addition of the matrix $\sigma_0 I$ to R_N , and this suggests that as a general procedure a consideration of different values of σ_0 in array weight design can lead to "optimum" trade-offs between output SNR and beam characteristics under nominal conditions.

Notice that the use of Δ values such as $\Delta = 2$ implies that allowance is made for up to two "dead" elements in the array. As an extension of the signal class modeled by (3.15), one can consider a generalized class defined in terms of several tolerances $\Delta_1, \Delta_2, \dots, \Delta_q$ for q groups of sensors. Further aspects of the performance of a robust spatial array system are discussed in [3.10, 3.11]. Before leaving this subject we note that recently Ahmed and Evans [3.20] have considered robust narrowband array processing under uncertainties, although their definition of robustness is based on the notion of an acceptable set of performances rather than on optimizing worst-case performance.

3.2.2. Robust Eckart Filter for Time-Delay Estimation

The estimation of time delay between signals arriving at two spatially separated sensors is of interest in applications such as seismology and sonar. In such systems the time delay measurement gives information about the direction of arrival of a wideband source. Let $Y_1(t)$ and $Y_2(t)$ be the outputs of

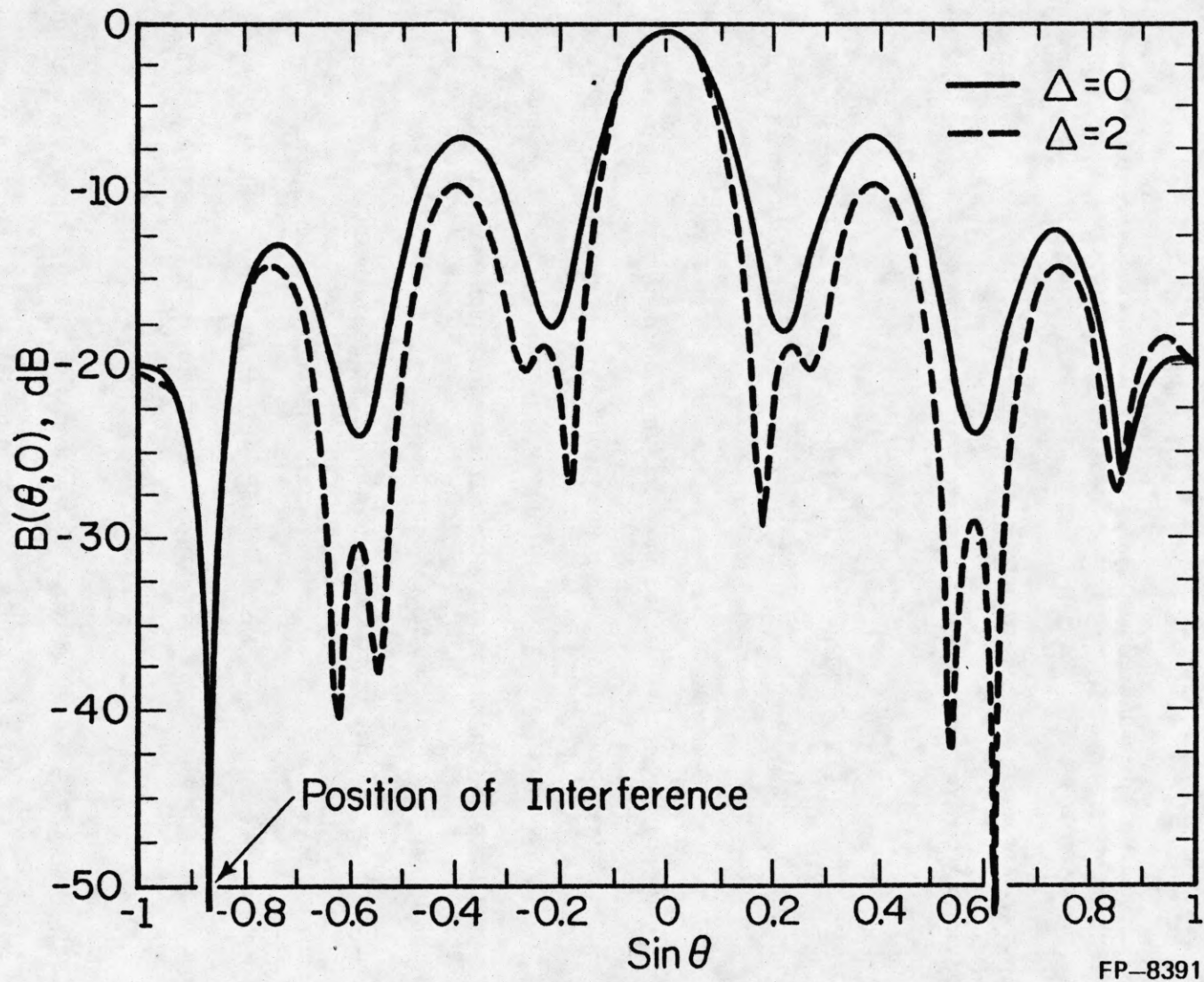


Fig. 3.10: Beam patterns of nominal and robust matched-filter ($\Delta = 2$) spatial arrays.

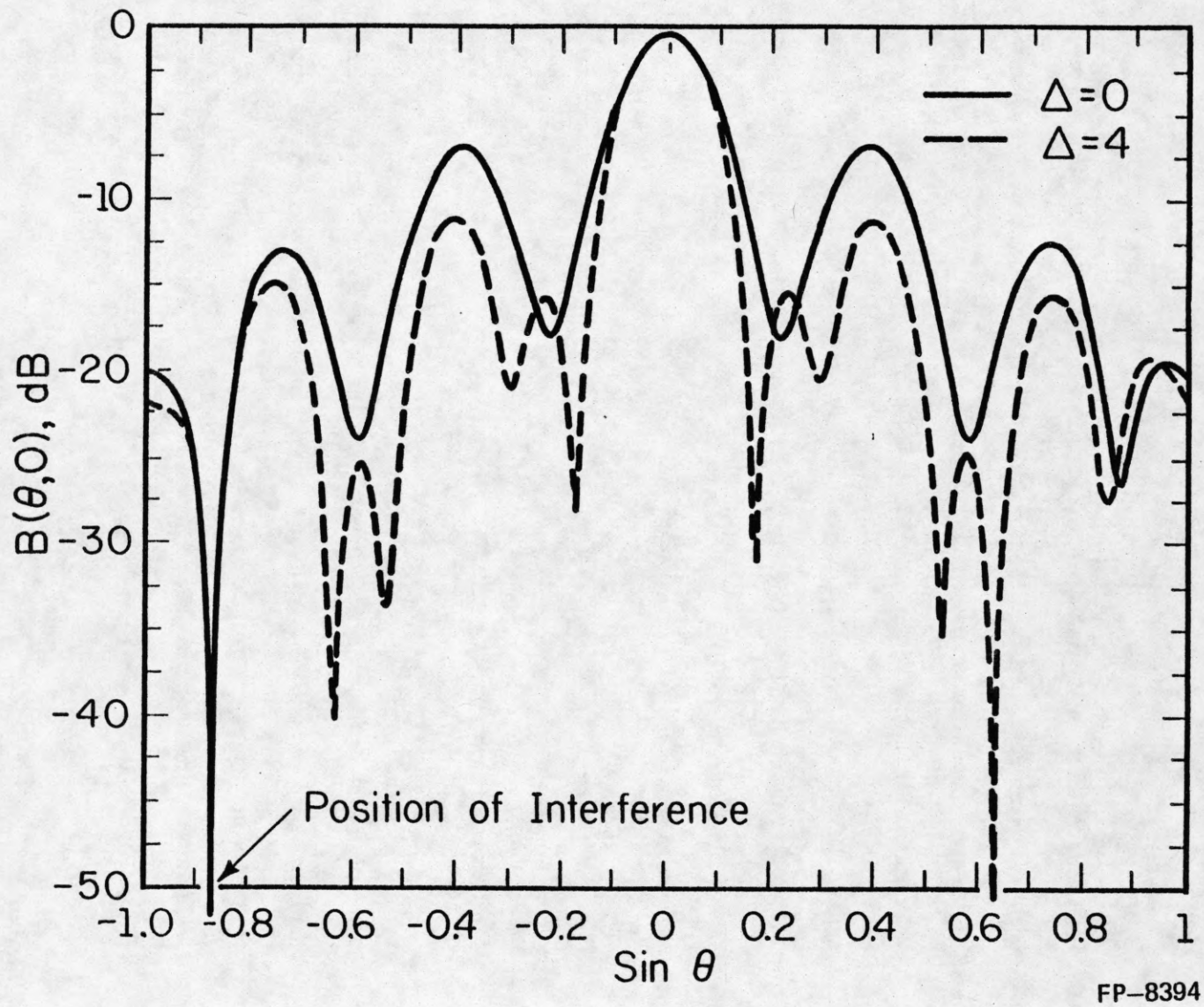


Fig. 3.11: Beam patterns of nominal and robust matched filter ($\Delta = 4$) spatial arrays.

two sensors, these signals being described by

$$Y_1(t) = S(t) + N_1(t) \quad (3.18a)$$

$$Y_2(t) = S(t-D) + N_2(t). \quad (3.18b)$$

Here $S(t)$ is a *random* source signal and $N_1(t)$, $N_2(t)$ are uncorrelated additive noise processes at the sensors. The basic technique for estimating the unknown relative delay D is to cross-correlate $Y_1(t)$ and $Y_2(t)$ and to use as an estimate of D that time argument which gives the maximum value of the cross-correlation function.

To improve the estimation process one can use a filter to weight the cross-spectrum estimate. Various criteria for optimum choice of the filter have been proposed [3.21]. One particular performance measure is the output SNR at the correct time delay; for weak signals (low input SNR) this output SNR for long averaging time is [3.21]

$$\text{SNR} = \frac{\left| \int_{-\infty}^{\infty} W(\omega) \Phi_S(\omega) d\omega \right|^2}{\int_{-\infty}^{\infty} W^2(\omega) Q(\omega) d\omega} \quad (3.19)$$

where $W(\omega)$ is the real filter frequency response, $\Phi_S(\omega)$ is the power spectral density of the zero-mean and stationary signal process and $Q(\omega) = \Phi_N^2(\omega)$, the square of the noise power spectral density at each sensor. The noise processes are here assumed to be zero-mean and stationary with identical power spectral densities, and signal and noise processes are assumed to be uncorrelated. The filter function $W_E(\omega)$ maximizing the above SNR is given by

$$W_E(\omega) = \frac{\Phi_S(\omega)}{Q(\omega)}, \quad (3.20)$$

which is the *Eckart filter*. Comparison of (3.3) and (3.4) with (3.20) and (3.19) shows the correspondence between the matched filter maximizing output SNR in the detection of a deterministic signal and the Eckart filter maximizing output SNR at the correct time-delay for long observation time under

weak signal conditions in time delay estimation. The main difference lies in the fact that $\Phi_S(\omega)$ is a non-negative power spectral density whereas $P(\omega)$ in matched filtering was a Fourier transform of a finite-energy signal.

The implementation of an Eckart filter requires knowledge of $\Phi_S(\omega)$ and $Q(\omega)$. These are often not precisely known but may be modeled as belonging to uncertainty classes such as those considered for power spectral densities in Section 2. As an example, consider the p -point classes for both $\Phi_S(\omega)$ and $Q(\omega)$, which specify the fractions of the total powers of $\Phi_S(\omega)$ and $Q(\omega)$ in specified intervals partitioning the entire frequency spectrum. Such information may be obtained from measurements at the output of a simple cross-correlator when signal is present, and from output power measurements in the frequency bands of interest under noise-only conditions. Since our earlier results on the minimax robust matched filter were obtained specifically for the L_2 signal uncertainty class, these results are not directly applicable here. However, it is quite simple to show [3.22] that for the time delay estimation problem where SNR of (3.19) is the performance measure of interest, the minimax robust Eckart filter for p -point classes for $\Phi_S(\omega)$ and $Q(\omega)$ has a *piecewise-constant* frequency response. The values of the constant levels of the robust Eckart filter frequency response $W_{ER}(\omega)$ are determined simply as the ratios of the given fractional powers in $\Phi_S(\omega)$ and $Q(\omega)$ in the common frequency bands defining the p -point classes.

As an example, assume that the nominal signal power spectral density $\Phi_{S,0}(\omega)$ and the nominal noise power spectral density $\Phi_{N,0}(\omega)$ are given by

$$\Phi_{N,0}(\omega) = \Phi_{S,0}(\omega)/A \quad (3.21)$$

and

$$\Phi_{S,0}(\omega) = \begin{cases} \frac{\alpha^2}{(\alpha^2 + \omega^2)^2}, & |\omega| \leq \pi \\ 0, & |\omega| > \pi. \end{cases} \quad (3.22)$$

For $A=1$ (input SNR = 0 dB) the nominally optimum Eckart filter $W_{E,0}(\omega)$ results in an output SNR of 2π or 8 dB. Now $\Phi_{S,0}(\omega)$ and $\Phi_{N,0}^2(\omega)$ with $\alpha = 1/3$ are members of particular p -point classes with

two frequency bands $[0, 1.61]$ and $[1.61, \infty)$. (The value 1.61 of the frequency band edges is the one which maximizes the performance of a two-level piecewise-constant filter under nominal conditions.) Suppose the noise power spectral density actually is a *constant* c on $[0, \omega_c]$ and zero outside. Now we may pick the values of c and ω_c to make $\Phi_N^2(\omega)$ a member of the p -point class to which $\Phi_{N,0}^2(\omega)$ belongs. In this case we find (with $A=1$, as before) that the input SNR is now -3 dB. The robust Eckart filter for this situation is the optimum two-level piecewise-constant filter for the nominal situation. Its SNR in the nominal case is 2.76 or 4.4 dB, and it maintains this value of the SNR for *all* $Q(\omega)$ in the p -point class defined above. The nominally optimum Eckart filter's output SNR drops to the value of 0.01 or -20 dB when the noise has the above ideal low-pass spectrum in the p -point class. Simulation results [3.22] confirm that the robust two-level piecewise-constant Eckart filter works quite effectively to obtain a good estimate of D for all $Q(\omega)$ in the defined p -point class, while the nominally optimum filter is quite sensitive to deviations from the nominal assumption.

More general results for other convex classes of total-power-constrained allowable power spectral densities $\Phi_S(\omega)$ and $Q(\omega) = \Phi_{N,0}^2(\omega)$ have been discussed in [3.22]. Although the SNR performance criterion of (3.19) is very similar to the SNR measure of (3.4) for deterministic signal detection, the minimax robust Eckart filter is the optimum filter for a least favorable pair $\Phi_{S,L}(\omega)$ and $Q_L(\omega)$ which is found exactly as in the case of robust Wiener filtering in Section 2. This follows from the general observation made in Section 2 concerning least favorable pairs whenever the performance measure is of a type expressible as a "distance" measure between $\Phi_S(\omega)$ and $Q(\omega)$ [3.23], for convex classes of power spectral density or probability density functions.

Before leaving this particular application let us note one further interesting feature of the robust Eckart filter. As in the case of the robust spatial matched filter we have considered above, there is another aspect of the performance of a time-delay estimation system that we have not considered explicitly. This is the variance of the time-delay estimate. It turns out that the robust Eckart filter has the additional advantage of generally producing time delay estimates with considerably *lower* estimation variances than the optimum Eckart filter under non-nominal conditions, with almost similar perfor-

mances for both under nominal conditions. This is discussed further in [3.22].

3.3. General Formulation of Robust Matched Filtering Problems in Hilbert Space

The results we have discussed so far in this section serve to illustrate the ideas behind robust matched filtering and their possible applications, and can be extended in several ways as we have indicated. In this subsection we will consider a general formulation of the robust matched filtering problem which has recently been developed [3.24], using a Hilbert space framework. Most matched-filtering situations (e.g. continuous-time/discrete-time, one-dimensional/multi-dimensional, time-domain/frequency-domain) can be fit into a single general Hilbert space setting which is convenient for studying robustness in all of these problems simultaneously. In particular, suppose \mathbf{H} is a separable Hilbert space (e.g., L_2 or \mathbb{R}^n) with inner product $\langle \cdot, \cdot \rangle$ and let $\bar{\mathbf{H}}$ denote a set of bounded nonnegative linear operators (e.g., integral operators, matrices, or spectral operators) mapping \mathbf{H} to itself. A matched-filtering problem on \mathbf{H} involves three quantities: a signal quantity $s \in \mathbf{H}$ (e.g., a signal spectrum or waveform); a noise quantity $n \in \bar{\mathbf{H}}$ (e.g., a noise spectrum, autocorrelation function, or covariance matrix); and a filter quantity $h \in \mathbf{H}$ (e.g., a filter transfer function or impulse response). The design criterion for the filtering problem is based on a functional $\rho: \mathbf{H} \times \mathbf{H} \times \bar{\mathbf{H}} \rightarrow \mathbb{R}$ defined by

$$\rho(h; s, n) = \frac{|\langle h, s \rangle|^2}{\langle h, nh \rangle}, \quad h \in \mathbf{H}, s \in \mathbf{H}, n \in \bar{\mathbf{H}}, \quad (3.23)$$

and representing a signal-to-noise ratio. Note that, for properly defined \mathbf{H} , most conventional matched filtering formulations fit this model (see, for example, Thomas [3.25]). The example described in subsection 3.1 corresponds to the particular case in which \mathbf{H} is complex $L_2(-\infty, \infty)$ (i.e., \mathbf{H} is the set of complex-valued functions f satisfying $\int_{-\infty}^{\infty} |f(\omega)|^2 d\omega < \infty$) and $\bar{\mathbf{H}}$ corresponds to a set of positive, symmetric, real-valued functions. The signal quantity s is identified with the shifted-pulse Fourier transform $P_0(\omega)e^{j\omega T}$; the filter quantity h is identified with $H^*(\omega)$, the complex conjugate of the filter transfer function; the noise quantity n corresponds to the noise spectrum $\Phi_N(\omega)$; and for an

element $n \in \bar{H}$ and any $f \in H$, the operation nf is defined by $(nf)(\omega) = n(\omega)f(\omega)$. Note that the inner product on $L_2(-\infty, \infty)$ is defined by $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} f^*(\omega)g(\omega)d\omega$. Examples of other matched filtering problems that fit this general Hilbert-space formulation are discussed below.

Within this general Hilbert-space formulation, for fixed signal s and noise quantity R_N the matched filter (i.e., the element $h_0 \in H$ that maximizes ρ) is given by any solution to the equation $nh_0 = s$. If R_N is invertible then h_0 is $n^{-1}s$ and the maximum value of ρ is given by $\langle s, n^{-1}s \rangle$. If, on the other hand, s and n are known only to be within classes \mathcal{S} and \mathcal{N} of signal and noise quantities, respectively, then we can consider the alternative design criterion (as in (3.6) and (3.9))

$$\max_{h \in H} \left\{ \inf_{(s,n) \in \mathcal{S} \times \mathcal{N}} \rho(h; s, n) \right\}. \quad (3.24)$$

It can be shown (see Lemma 1 of [3.24]) that if \mathcal{S} and \mathcal{N} are convex, then a pair $(s_L, n_L) \in \mathcal{S} \times \mathcal{N}$ and its optimum filter h_R (satisfying $n_L h_R = s_L$) is a saddlepoint solution for (3.24) if and only if the following inequality holds for all $(s, n) \in \mathcal{S} \times \mathcal{N}$:

$$2\text{Re}\{\langle s, h_R \rangle\} - \langle h_R, n h_R \rangle \geq \langle s_L, h_R \rangle. \quad (3.25)$$

Moreover, if n_L is invertible then this occurs if and only if (s_L, n_L) is least-favorable for matched filtering for \mathcal{S} and \mathcal{N} , i.e., if and only if

$$\langle s_L, n_L^{-1}s_L \rangle = \min_{(s,n) \in \mathcal{S} \times \mathcal{N}} \left\{ \sup_{h \in H} \rho(h; s, n) \right\}. \quad (3.26)$$

Expression (3.25) provides a means for checking potential solutions to (3.24) and (3.26) provides a means for searching for such solutions since the expression in brackets in (3.26) is known in closed-form for many situations of interest.

By using the above results, solutions to the general robust matched filtering problem of (3.24) have been obtained for generalizations of the uncertainty classes C_P and C_N of (3.5) and (3.8a and b), as well as for other uncertainty classes of interest. For example, consider the situation in which the

noise quantity is known to be some fixed $n_0 \in \bar{\mathbf{H}}$, but the signal quantity s is known only to be in the class $C_s \in \mathbf{H}$ defined by

$$C_s = \{s \in \mathbf{H} \mid \|s - s_0\|^2 \leq \Delta\} \quad (3.27)$$

where s_0 is a known nominal signal, Δ is a fixed positive number representing the degree of uncertainty in the signal, and $\|x\|$ denotes the norm of $x \in \mathbf{H}$ defined by $\|x\| = [\langle x, x \rangle]^{1/2}$. For this problem it can be shown (see Theorem 1 of [3.24]) that a saddle-point solution to (3.24) is given by $(h_R; s_L, n_0)$ where h_R is given by

$$h_R = (n_0 + \sigma_0 I)^{-1} s_0 \quad (3.28)$$

with I being the identity mapping from \mathbf{H} to itself and with σ_0 being the positive solution to

$$\sigma_0^2 \|h_R\|^2 = \Delta; \quad (3.29)$$

and where

$$s_L = s_0 - \sigma_0 h_R. \quad (3.30)$$

Note that, for the specific spectral domain example discussed in subsection 3.1, the identity operator is represented by the unit white-noise spectrum $\Phi_N(\omega) \equiv 1$ and so (3.28) corresponds to (3.7) with n_0 and s_0 being represented by $\Phi_N(\omega)$ and $P_0(\omega)e^{j\omega T}$, respectively. (Recall that, for this case, h is identified with $H^*(\omega)$). It is interesting to note that, as in the spectral case, the identity operator I generally describes a white-noise process so that (3.28) indicates that the type of uncertainty described by C_s has an effect on the design equivalent to that of adding white noise of spectral height σ_0 .

The result regarding the band model of (3.8a and b) can also be extended to general Hilbert spaces. In particular, consider the situation in which the signal is known, say s_0 , and in which all of the noise operators $n \in \mathbf{N}$ can be represented by spectral components $\{n(\omega); \omega \in \Omega\}$ for some set Ω . This situation arises in stationary models such as that yielding (3.8a and b), in which case $n(\omega)$ can be taken to be the power spectral density at the frequency ω and Ω is \mathcal{R} (it could similarly be \mathcal{R}^n).

Alternatively, for other models this situation arises when all of the members of \mathcal{N} share a common eigenstructure and the set $\{n(\omega); \omega \in \Omega\}$ is the eigenspectrum of the operator n . For example, such a model is generated if the all noise autocorrelation functions in \mathcal{N} have Karhunen-Loeve expansions in the same eigenfunctions. In this latter case, Ω would be the set of positive integers, and $n(\omega)$ would be the ω -th eigenvalue of n .

In this general setting the band model (3.8a and b) becomes

$$C_n = \{n \mid n'(\omega) \leq n(\omega) \leq n^*(\omega), \omega \in \Omega \text{ and } \text{tr}\{n\} = c\} \quad (3.31)$$

where n' and n^* are known functions and $\text{tr}\{\cdot\}$ denotes trace. Note that for power spectral densities the trace is just $\frac{1}{2\pi} \int_{-\infty}^{\infty} n(\omega) d\omega$ and for discrete eigenspectra the trace is $\sum_{i=1}^{\infty} n(i)$. (More generally, we can write $\text{tr}\{n\} = \int_{\Omega} n(\omega) \mu(d\omega)$ for some measure μ on Ω .) Assuming that the signal s_0 also has the spectral representation $\{S_0(\omega); \omega \in \Omega\}$ in terms of the same eigenstructure as the members of C_n , it can be shown using (3.25) (see [3.24]) that the robust filter for C_n of (3.31) is the optimum filter for S_0 and a least-favorable noise operator whose spectrum is

$$n_L(\omega) = \max \{n'(\omega), \min \{k^{-1} |S_0(\omega)|, n^*(\omega)\}\} \quad (3.32)$$

where k is chosen so that $\text{tr}\{n_L\} = c$. Note that (3.32) is identical to (3.11) with $n_L = \Phi_{N,L}$, $L_N = n'$, $U_N = n^*$, and $P_0 = S_0$. This result can be combined straightforwardly with (3.28)–(3.30) to give the robust solution for uncertainty in both signal and noise (see Kassam, Lim, and Cimini [3.4] and Poor [3.24]).

Example: Binary Communication

To illustrate the generality of the above result consider the problem of antipodal signaling in additive Gaussian noise as described by the following pair of statistical hypotheses:

$$H_0: Y(t) = N(t) - s(t), \quad 0 \leq t \leq T$$

versus

(3.33)

$$H_1: Y(t) = N(t) + s(t), \quad 0 \leq t \leq T$$

where $\{s(t); 0 \leq t \leq T\}$ is a known (i.e., deterministic) square-integrable signal waveform and $\{N(t); 0 \leq t \leq T\}$ is a zero-mean Gaussian noise process with autocorrelation function $\{R_N(t, u); 0 \leq t \leq T, 0 \leq u \leq T\}$. Assuming that H_0 and H_1 are equally likely, the Bayes optimum (i.e., minimum-probability-of-error) receiver for (3.33) is of the form (see, e.g., Helstrom [3.26])

$$\phi(y) = \text{sgn} \left\{ \int_0^T h(T, t) y(t) dt \right\} \quad (3.34)$$

where $\text{sgn}\{\cdot\}$ denotes the algebraic sign of the argument; $\phi = +1$ and $\phi = -1$ denote the acceptance of hypotheses H_1 and H_0 , respectively; $y \equiv \{y(t); 0 \leq t \leq T\}$ denotes the observed realization of the random process $\{Y(t); 0 \leq t \leq T\}$, and where $\{h(T, t); 0 \leq t \leq T\}$ denotes the impulse response of a linear filter. The probability of error associated with the receiver of (3.34) is given by

$$P_e = 1 - \Phi([\text{SNR}]^{1/2}), \quad (3.35)$$

where SNR denotes the signal-to-noise ratio at the output of the filter at time T and is given by

$$\text{SNR} = \frac{\left| \int_0^T h(T, t) s(t) dt \right|^2}{\int_0^T \int_0^T h(T, t) R_N(t, u) h(T, u) dt du} \quad (3.36)$$

Here Φ denotes the standard (unit) Gaussian probability distribution function. The optimum receiver for (3.33) is given by (3.34) with $h(T, t)$ being the solution to the integral equation

$$\int_0^T R_N(t, u) h(T, u) du = s(t), \quad 0 \leq t \leq T. \quad (3.37)$$

This problem fits within the above Hilbert-space formulation with \mathbf{H} being real $L_2[0, T]$; $s = \{s(t); 0 \leq t \leq T\}$, $h = \{h(T, t); 0 \leq t \leq T\}$, and $n = \{R_N(t, u); (t, u) \in [0, T]^2\}$.

With $a = \{a(t); 0 \leq t \leq T\}$ and $b = \{b(t); 0 \leq t \leq T\}$ in \mathbf{H} , we have $\langle a, b \rangle = \int_0^T a(t)b(t)dt$

and the operation na is defined by $(na)(t) = \int_0^T R_N(t, u)a(u)du$. So, $\rho(h; s, n)$ is given by (3.36), and

(3.37) is the equation $nh = s$. For this model, the signal uncertainty class C_s of (3.27) is given by

$$C_s = \{s \in L_2[0, T] \mid \int_0^T |s(t) - s_0(t)|^2 dt \leq \Delta\}, \quad (3.38)$$

where $s_0 \equiv \{s_0(t); 0 \leq t \leq T\}$ is the known transmitted signal. This model represents a general model for types of receiver and channel distortion that are difficult to model parametrically. Further justification for this model is found in Slepian's notion of indistinguishable signals [3.20].

From (3.28)–(3.29) it follows that the robust receiver of the form (3.34) for uncertain signal distortion described by (3.38) is given by the impulse response solving the equation

$$\int_0^T R_N(t, u)h_R(T, u)du + \sigma_0 h_R(T, t) = s_0(t), \quad 0 \leq t \leq T, \quad (3.39)$$

which is a Fredholm equation of the second kind. The constant σ_0 is specified by (3.39) and the condition

$\sigma_0^2 \int_0^T |h_R(T, t)|^2 dt = \Delta$. The performance of specific solutions to (3.39) is discussed in [3.24]. It

is interesting to note again that the effect of the uncertainty modeled by (3.38) is to introduce a white

noise "floor" of height σ_0 . This device is used in standard treatments (see, e.g., Van Trees [3.28]) in order

to circumvent singularity problems arising in the solution to (3.37) for the case of continuous R_N .

That this phenomenon arises naturally here gives additional justification for considering minimax

design within (3.38). Equation (3.39) and the condition following it for the robust filter was originally

derived in [3.13]. A modified condition for a signal class with an energy equality constraint is given in

[3.15].

To further illustrate the general Hilbert-space results, consider the discrete-time observation models under the two hypotheses

$$H_0: Y_i = N_i - s_i, \quad i=1,2,\dots,m$$

and

$$H_1: Y_i = N_i + s_i, \quad i=1,2,\dots,m$$

(3.40)

where $\underline{N} = (N_1, \dots, N_m)^T$ is a random vector having zero mean and nonsingular covariance matrix Σ and where $\underline{s} = (s_1, \dots, s_m)^T$ is a known signal vector. Here, the output at time m of a linear filter with impulse response $\{\tilde{h}_i; i = 0, 1, \dots, m-1\}$ is given by $\underline{h}^T \underline{s}$ where $h_i = \tilde{h}_{m-1}$, and the output SNR is $(\underline{h}^T \underline{s})^2 / \underline{h}^T \Sigma \underline{h}$. The matched filter for known Σ and \underline{s} is given by $\Sigma^{-1} \underline{s}$. Of course, this fits the above Hilbert space formulation with $\mathbf{H} = \mathbb{R}^m$ (i.e., $\langle \underline{a}, \underline{b} \rangle = \underline{a}^T \underline{b}$), $s = \underline{s}$, $h = \underline{h}$, and with the operator n represented by the matrix Σ . A band model class \mathcal{N} of operators such as (3.31) occurs here in the case in which all members of \mathcal{N} have the same orthonormal eigenvectors $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_m$ in which case (3.31) is given by

$$C_n = \{ \Sigma \mid \Sigma \underline{v}_i = \lambda_i \underline{v}_i \text{ and } \lambda_i' \leq \lambda_i \leq \lambda_i'' \}$$

$$\text{for all } i = 1, \dots, m, \text{ and } \sum_{i=1}^m \lambda_i = c \} \quad (3.41)$$

Thus, applying (3.32), the least-favorable covariance matrix is given (using its spectral representation) by

$$\Sigma_L = \sum_{i=1}^m \lambda_{L,i} \underline{v}_i \underline{v}_i^T \quad (3.42)$$

where

$$\lambda_{L,i} = \begin{cases} \lambda_i', & k^{-1}p_i < \lambda_i' \\ k^{-1}p_i, & \lambda_i' \leq k^{-1}p_i \leq \lambda_i'' \\ \lambda_i'', & \lambda_i'' < k^{-1}p_i \end{cases} \quad (3.43)$$

where $p_i = |\underline{s}^T \underline{v}_i|$ and where k is chosen so that $\sum_{i=1}^m \lambda_{L,i} = c$. Note that (3.42) must be nonsingular and thus the robust filter is described by $\underline{h}_R = \Sigma_L^{-1} \underline{s}$. This result is a special case of more general results found in Chen and Kassam [3.10, 3.11].

The discrete-time model of (3.40) can also be used to illustrate least-favorable signals and noise operators for uncertainty models of interest other than those of (3.27) and (3.31). This problem has been treated in [3.12] in some detail. For example, suppose the noise covariance is known to be given by $\text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\}$ (corresponding to uncorrelated samples), and the signal is known only to differ in total *absolute* distortion by no more than Δ from some nominal signal \underline{s}_0 , i.e., we assume the signal lies in the class

$$C_s^1 = \{ \underline{s} \in \mathbb{R}^m \mid \sum_{i=1}^m |s_i - s_{0,i}| \leq \Delta \} . \quad (3.44)$$

Then, it can be shown [3.12] using (3.25) that the robust filter \underline{h}_R

$$h_{R,i} = \begin{cases} h_{0,i}, & \text{if } |h_{0,i}| \leq c \\ \text{csgn}\{h_{0,i}\}, & \text{if } |h_{0,i}| > c \end{cases} \quad (3.45)$$

where \underline{h}_0 is the nominal filter and where $c > 0$ satisfies the equation $\sum_{i=1}^m \sigma_i^2 \min\{0, |h_{0,i}| - c\} = \Delta$.

Thus, in this case the robust matched filter is a *clipped* version of the nominal filter. Similarly if Σ is diagonal and \underline{s} lies in a class

$$C_s^\infty = \{ \underline{s} \in \mathbb{R}^m \mid \max_{i=1, \dots, m} |s_i - s_{0,i}| \leq \Delta \}, \quad (3.46)$$

then the robust filter becomes (see [3.12])

$$h_{Rj} = \begin{cases} (s_{0j} - \Delta)/\sigma_i^2, & \Delta < s_{0j} \\ 0, & -\Delta \leq s_{0j} \leq \Delta \\ (s_{0j} + \Delta)/\sigma_i^2, & s_{0j} < -\Delta. \end{cases} \quad (3.47)$$

This filter is the optimum filter for a least-favorable signal in C_s^∞ that is as near zero as possible (in terms of the norm $\max_{i=1,2,\dots,m} |s_i|$). Extensions of these results can be found in [3.12].

The above treatment makes clear the rather general applicability of the Hilbert space framework for formulating and solving robust linear detection problems. Recently it has been shown [3.29] that a Hilbert space framework can also be used for robust linear estimation problems (which we surveyed in Section 2); for details the reader is referred to [3.29]. We shall now proceed, in the next section, to consider robust nonlinear signal detection schemes primarily designed to protect against uncertainties in noise probability density functions.

4. NONLINEAR METHODS FOR ROBUST SIGNAL DETECTION

In the previous two sections we focused on the design of robust *linear* filters for signal estimation and detection, for situations in which there was uncertainty about the *spectral densities* or *correlation functions* of the signal and noise. The performance measures considered there were the mean-squared error and the signal-to-noise ratio, which did not involve the exact functional forms of the signal and noise probability density functions (pdf's).

In this section we will survey results on robust signal detection which pertain specifically to robustness when the detection performance is measured by characteristics directly related to the *probability of detection* or *error probabilities* instead of signal-to-noise ratios. Although in most situations an assumption that the noise is Gaussian gives a direct relationship between the SNR and such detection performance measures, this is not true in the case of non-Gaussian noise. In almost all such cases it is possible to obtain explicit results on the structures of minimax robust detectors (generally nonlinear) only when the noise processes are sequences of *independent* random variables so that only univariate probability density function uncertainties need be considered. In the first two sections we were able to avoid considerations of pdf uncertainties because the performance measures (MSE and SNR) depended only on second-moment characteristics. For the detection performance measures we use here, consideration of the correlation functions and more generally of pdf's beyond those of first order is avoided by the assumption of independence. For the pulse-train detection problem discussed at the beginning of Section 3, this means that a robust detection scheme can be arrived at by *decoupling* the treatments of spectral density and probability density uncertainties. That this approach is necessary is not surprising, considering the well-recognized difficulties of dealing in general with non-Gaussian random processes. There are available, nonetheless, some recent results on robust detection in correlated noise, under certain constraints, and we will discuss these later in this section.

For the most part, then, we will be considering here various nonparametric classes of univariate noise probability density functions expressing different types and extents of uncertainty about the exact noise pdf. We will consider in particular the canonical detection problems of known low-pass signals

in additive noise, deterministic bandpass signals in bandpass additive noise, and random signals in additive noise. Even with the simplification due to the white noise assumption, explicit solutions for minimax robust detectors are generally possible only under a further restriction to consideration of the *local* or *weak-signal* case. We will begin, however, with a description of the results of Huber, and other related results, for a general hypothesis testing problem. As we have mentioned in the Introduction, the 1964 and 1965 results of Huber [4,1, 4.2] have greatly influenced and motivated much of the subsequent work on robust signal processing schemes.

4.1. Robust Hypothesis Testing

In 1965 Huber [4.2] published an explicit solution for the robust test for a binary hypothesis testing problem related to the signal detection problem we discussed in the Introduction. The significance of this result lies not so much in the solution it provided for the hypothesis testing problem considered by Huber, as much as in the mathematical justification it provided of the use of detectors based on bounded functions such as $\tilde{L}(x; s, \theta)$ of (1.5).

Let $\underline{X} = (X_1, X_2, \dots, X_n)$ be a vector of independent and identically distributed (i.i.d.) observation components X_i , $1 \leq i \leq n$. Under a hypothesis H_0 (null hypothesis) let the common pdf of the X_i be f_0 , and under an alternative hypothesis H_1 let the common pdf be f_1 . The requirement is to construct a test for H_0 vs. H_1 based on the observations \underline{X} , when f_0 and f_1 are not specified completely. The approach taken by Huber in [4.2] was to define first *classes* of allowable probability density functions under the null and alternative hypotheses. The classes considered in [4.2] were obtained as neighborhood classes containing in each case a nominal density function and density functions in its vicinity. One such pair of neighborhood classes which is popular in robustness studies is the pair of *ε-contamination* classes, for which under H_j , $j=0,1$, the class of allowable density functions is

$$F_j(f_j^0, \epsilon_j) = \{f \mid f = (1-\epsilon_j)f_j^0 + \epsilon_j h_j\}, j=0,1. \quad (4.1)$$

Here f_j^0 is the nominal density function under hypothesis H_j , quantity ϵ_j in $[0,1)$ is the maximum degree of contamination for f_j^0 , and h_j is any density function. These classes are, of course, the same

(aside from normalization) as the spectral uncertainty classes of Example 1 in subsection 2.3. The hypothesis testing problem is therefore that of choosing between the two hypotheses:¹

$$H_0: \text{common univariate pdf of the } X_i \text{ is any pdf } f_0 \text{ in } \mathcal{F}_0, \quad (4.2a)$$

$$H_1: \text{common univariate pdf of the } X_i \text{ is any pdf } f_1 \text{ in } \mathcal{F}_1. \quad (4.2b)$$

Huber then sought the *least-favorable pair* of probability densities in $\mathcal{F}_0 \times \mathcal{F}_1$, which is defined on the basis of a risk function $R(f, \phi)$. In the risk function, ϕ denotes the test for H_0 versus H_1 which rejects H_j in favor of H_{1-j} with probability $\phi^j(\underline{X})$ when $\underline{X} = (X_1, X_2, \dots, X_n)$ is observed,² the X_i being i.i.d. with density function f in \mathcal{F}_0 or \mathcal{F}_1 . Consider, for example, the minimum probability of error criterion for which we would have

$$R(f, \phi) = E_f \{ \phi^j(\underline{X}) \} \text{ if } f \in \mathcal{F}_j. \quad (4.3)$$

Then for $(f_0, f_1) \in \mathcal{F}_0 \times \mathcal{F}_1$ the probability of error (assuming equally likely priors) is

$$P_e(f_0, f_1; \phi) = \frac{[R(f_0, \phi) + R(f_1, \phi)]}{2}, \quad (4.4)$$

and this is minimized when ϕ is a test based on a comparison of the likelihood ratio $\Lambda(\underline{X}) = \prod_{i=1}^n f_1(X_i)/f_0(X_i)$ with a threshold value unity. The *least-favorable pair* (q_0, q_1) in $\mathcal{F}_0 \times \mathcal{F}_1$ for this problem is that pair for which the corresponding test ϕ_q based on its likelihood ratio satisfies, for all $(f_0, f_1) \in \mathcal{F}_0 \times \mathcal{F}_1$,

$$P_e(f_0, f_1; \phi_q) \leq P_e(q_0, q_1; \phi_q). \quad (4.5)$$

In general we may define $R(f, \phi)$ to be $L_j E_f \{ \phi^j(\underline{X}) \}$ when $f \in \mathcal{F}_j$, the L_j being constant weights, and we may consider other risk-based criteria [4.2]. One of the major results in [4.2] is that for the ϵ -contamination classes a least-favorable pair (q_0, q_1) exists satisfying

¹The arguments for $\mathcal{F}_0, \mathcal{F}_1$ will be dropped when no confusion can result.

²Note that $\phi^0(\underline{X}) + \phi^1(\underline{X}) = 1$.

$$R(f, \phi_q) \leq R(q_j, \phi_q) \text{ for } f \in \mathcal{F}_j, j=0,1, \quad (4.6)$$

where ϕ_q is any test for q_0 versus q_1 based on a comparison of the likelihood ratio with a threshold (a "probability ratio test"). This clearly implies that (4.5) is true. The pair (q_0, q_1) and ϕ_q for minimum error probability form a saddlepoint for the error probability functional, that is

$$P_e(f_0, f_1; \phi_q) \leq P_e(q_0, q_1; \phi_q) \leq P_e(q_0, q_1; \phi) \quad (4.7)$$

for any $(f_0, f_1) \in \mathcal{F}_0 \times \mathcal{F}_1$ and any test ϕ . The test ϕ_q is called a *robust* test for $f \in \mathcal{F}_0$ versus $f \in \mathcal{F}_1$; it minimizes over all tests the supremum (least upper bound) of the error probability over all pairs in $\mathcal{F}_0 \times \mathcal{F}_1$. Indeed ϕ_q is robust, in view of (4.6), for other risk-based criteria such as the Neyman-Pearson criterion. For this criterion we can define the threshold for ϕ_q in such a way that for a design value α of the false alarm probability

$$R(f, \phi_q) \leq R(q_0, \phi_q) = \alpha, \quad f \in \mathcal{F}_0 \quad (4.8)$$

with $R(f, \phi)$ as in (4.3).

For the ϵ -contamination class Huber's solution for the least favorable pair turns out to be

$$q_0(x) = \begin{cases} (1-\epsilon_0)f_0^0(x), & \frac{f_1^0(x)}{f_0^0(x)} < c' \\ \frac{1}{c'}(1-\epsilon_0)f_1^0(x), & \text{otherwise} \end{cases} \quad (4.9a)$$

$$q_1(x) = \begin{cases} (1-\epsilon_1)f_1^0(x), & c' < \frac{f_1^0(x)}{f_0^0(x)} \\ c'(1-\epsilon_1)f_0^0(x), & \text{otherwise} \end{cases} \quad (4.9b)$$

where $c' < c^*$ are non-negative numbers such that q_0 and q_1 are probability density functions. The proof of the existence of such a pair in $\mathcal{F}_0 \times \mathcal{F}_1$ where \mathcal{F}_0 and \mathcal{F}_1 are disjoint (the case of interest, since otherwise $q_0 = q_1$), and the proof that (4.6) holds, can be found in [4.2]. Note that the likelihood ratio $\lambda_q(x) = q_1(x)/q_0(x)$ for a single observation for the least favorable pair is a "censored" version of

the nominal one,

$$\lambda_q(x) = \begin{cases} bc^*, & c^* \leq \lambda_0(x) \\ b\lambda_0(x), & c' < \lambda_0(x) < c^* \\ bc', & \lambda_0(x) \leq c' \end{cases}, \quad (4.10)$$

where $b = (1-\epsilon_1)/(1-\epsilon_0)$ and $\lambda_0(x) = f_1^0(x)/f_0^0(x)$.

In [4.2] the least favorable pair of density functions was also obtained for another uncertainty model for the density functions under H_0 and H_1 . This was defined by the *total-variation* classes of probability densities which may be expressed as

$$\mathcal{F}_j(f_j^0, \epsilon) = \left\{ f \mid \int \left| f(x) - f_j^0(x) \right| dx \leq \epsilon \right\}. \quad (4.11)$$

This is analogous to Example 2 of subsection 2.3. In a later paper [4.3] it was shown that least favorable pairs of probability measures can be found for probability measure classes which are defined as being bounded by *2-alternating capacities*, which are generalized notions of measure as discussed in subsection 2.7. The ϵ -contaminated classes of nominal densities have remained the most widely used uncertainty models, partly because of their earlier introduction, but primarily because it is possible to justify their use, in many cases, from physical considerations. In modeling impulsive noise, for example, the presence of a small proportion (whose maximum value is ϵ) of impulsive components in a background of, say, Gaussian noise can be modeled to have an ϵ -contaminated nominal Gaussian pdf. Any uncertainty about the exact pdf of the impulsive components then leads directly to use of an ϵ -contaminated uncertainty class, perhaps with a side constraint on the nature of the contaminating pdf h .

One other specific uncertainty class of pdf's that has been used more recently is that analogous to the spectral band model discussed in subsection 2.3; this model is justifiable from physical considerations in some applications. It may be viewed as being more general than the ϵ -contamination class and allows the results obtained for the ϵ -contamination classes to be extended. Specifically the *bounded* classes $\mathcal{F}_j(f_{jL}, f_{jU})$ considered in [4.4] are defined by

$$\mathcal{F}_j(f_{jL}, f_{jU}) = \{f \mid f_{jL} \leq f \leq f_{jU}\}, \quad j=0,1, \quad (4.12)$$

so that allowable density functions in \mathcal{F}_j are those bounded by given non-negative functions f_{jL} and f_{jU} which make the \mathcal{F}_j non-empty. These classes reduce to the ϵ -contamination classes of (4.1) for $f_{jL} = (1-\epsilon_j)f_j^0$ and $f_{jU} = \infty$. It is not true, however, that the classes (4.12) are obtained from (4.1) by imposing an upper bound on the contamination densities h_j . This is clear from the fact that even if we set $f_{jL} = (1-\epsilon_j)f_j^0$, the resulting "nominal" density f_j^0 may not belong to $\mathcal{F}_j(f_{jL}, f_{jU})$. In [4.4] the least favorable pair in the above classes has been found, and the likelihood ratio for the least favorable pair is also essentially a censored version of either f_{1L}/f_{0L} or f_{1U}/f_{0U} . The bounded classes of pdf's defined by the *band-model* in (4.12) can be viewed as arising naturally in situations where the pdf's under the hypotheses are *estimated* from training data, in which case the bands (f_{jL}, f_{jU}) are confidence bands.

Now it should be clear that in the various uncertainty models considered above for the densities f of each independent component X_i of the observation sequence \underline{X} , there was no necessary restriction that X_i be one-dimensional and that f be univariate. It is possible, for instance, to treat an observation sequence of n one-dimensional components as a single n -dimensional observation, and characterize its multivariate density by one of the above classes. Essentially this approach was taken in [4.5] and [4.6] by Kuznetsov, who developed results for the characteristics of the robust test for hypotheses described by the bounded classes of (4.12). Since the likelihood ratio in a threshold comparison test may be replaced by any function which yields the same critical region and threshold equality region, simplifications may be made to results of the type in (4.10). In (4.10), for example, for a *single* observation test with the threshold τ in between bc'' and bc' , one gets an equivalent test if simply $b\lambda_0(x)$ is compared to τ . Note that $b\lambda_0(x) = f_{1L}(x)/f_{0L}(x)$ when the ϵ -contamination class is viewed as a special case of the bounded class.

In [4.5] and [4.6] Kuznetsov shows that the robust tests are obtained by threshold comparisons using one of the functions f_{1U}/f_{0L} , f_{1L}/f_{0U} , f_{1L}/f_{0L} , or f_{1U}/f_{0U} . As an interesting application, consider the detection problem (1.1) where $\theta_{\underline{S}} = \theta(s_1, s_2, \dots, s_n)$ is not precisely known, but is

known to belong to a neighborhood of a nominal sequence $\theta_{0\underline{s}_0}$. Since \underline{s}_0 is in \mathbb{R}^n one might define the neighborhood as a spheroid of radius Δ centered on $\theta_{0\underline{s}_0}$. Let the N_i be i.i.d. and zero-mean, unit-variance Gaussian random variables. The bounded class of density functions for \underline{X} under H_1 is then defined in terms of the maximum-likelihood and minimum-likelihood estimates of $\theta_{0\underline{s}_0}$. The robust detector is based on a combination of square-law and linear processing of \underline{X} ; details are given in [4.6]. The results in [4.5] and [4.6] on binary hypotheses have recently been extended to multiple hypothesis testing problems in [4.7].

We have remarked that it is possible in general to solve for least favorable pairs of probability measures for other pairs of probability measure classes which are defined to contain measures bounded by *2-alternating capacities* [4.3]. Specific examples of other pairs of classes for which explicit least favorable pairs are available are the p -point classes (discussed in subsection 2.3 in defining spectral density classes; these become classes of pdf's when the power is fixed at unity), and the bounded p -point classes [4.8]. The generalized moment classes of pdf's have been considered in [4.9]. Let us also reiterate the interesting connection between the robust binary hypothesis testing problems for such classes of pdf's and the corresponding robust linear filtering problems with mean-squared errors as performance criteria. For many of the robust linear filtering problems with fixed-power spectral density uncertainty classes, the least-favorable pairs of pdf's for the robust hypothesis testing problems defined on corresponding unit-power normalized spectral density classes produce directly the least-favorable spectra [4.10]. In general one may interpret the least favorable pair of pdf's in $\mathcal{F}_0 \times \mathcal{F}_1$ as being that pair which has the minimum distance between its components. Indeed, strong connections exist between distance minimization and least favorable pdf pairs [4.4, 4.11] under some general restrictions. It is this which results in the close relationship between the solutions for the least favorable pairs in the robust linear filtering and corresponding hypothesis testing problems.

The problem of robust *sequential* hypothesis testing was also considered by Huber in his first paper on robust hypothesis testing [4.2]. More recently this work was extended in [4.12], which also dealt with the ϵ -contamination model for noise pdf classes under the null and alternative hypotheses.

Numerical performance results from simulation experiments and computations are also given in [4.12].

4.2. Robust Detection of Known Low-Pass Signals

Consider again the signal detection problem described by the hypothesis testing situation of (1.1). Here the signal sequence is known to within an overall amplitude factor, that is, the $s_i, i=1,2,\dots,n$ are known. The observations X_i described by H_1 in (1.1b) may be considered to arise as a result of sampling a continuous-time additive mixture of a low-pass or baseband signal waveform $\theta_s(t)$ and a stationary noise process. If the noise bandwidth is large relative to that of the signal in such a situation it becomes reasonable to assume a sampling rate which results in the noise components $N_i, i=1,2,\dots,n$, under H_0 or H_1 being statistically independent random variables with some common univariate pdf f . Another common situation in which the above observation model is appropriate is that arising in the detection of a pulse-train, as described in Section 3.

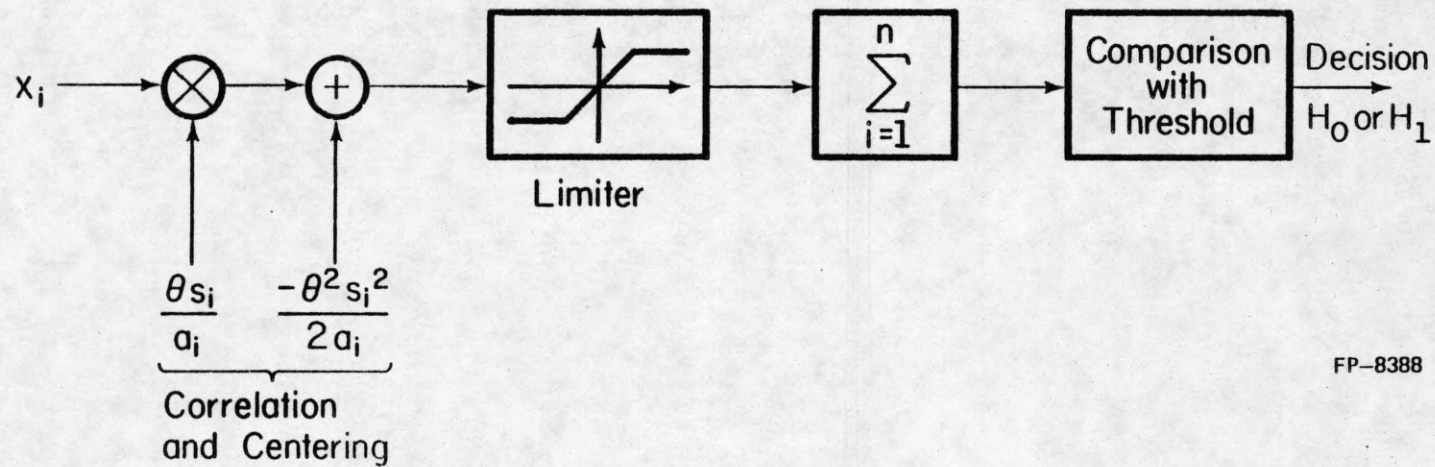
In the absence of precise knowledge about the noise pdf f , one can now attempt to obtain a *robust* detection scheme for a class of possible noise pdf's generated by a model such as the ϵ -contamination model used by Huber. But it should now be apparent that Huber's solution which we discussed in subsection 4.1 above does not directly apply to our known-signal in additive noise detection problem. The least-favorable pair of pdf's given by (4.9a) and (4.9b) was obtained under the assumption that the pdf's $f_0 \in \mathcal{F}_0$ and $f_1 \in \mathcal{F}_1$ are chosen independently; this means that the contaminating pdf's h_0 and h_1 in the ϵ -contamination models were not constrained to be related to each other in any particular way. In the known-signal detection problem we should require h_1 to be a translated or shifted version of h_0 . On the other hand, Huber's approach in which independent contaminating densities are allowed under the two hypotheses can be used to obtain a conservative solution to the robust detection problem.

One of the first attempts to adapt Huber's approach and results on robust hypothesis testing for signal detection problems was reported in 1971 by Martin and Schwartz [4.13]. Martin and Schwartz were interested in the signal detection problem described by (1.1), in which the observations X_i under H_1 are independent but not necessarily identically distributed. They first showed that Huber's result

for the ϵ -contaminated classes extends directly to the time-varying problem, where for each component X_i of \underline{X} the nominal densities f_{0i}^0, f_{1i}^0 and contamination degrees $\epsilon_{0i}, \epsilon_{1i}$ may be different. For the detection of a known signal in nearly Gaussian noise they set $f_{0i}^0 = \phi$, the zero-mean Gaussian density with unit-variance (without loss of generality), and $f_{1i}^0(x) = \phi(x - \theta s_i)$. They took $\epsilon_{ji} = \epsilon$ to be sufficiently small for a given θ so that the resulting ϵ -contamination classes \mathcal{F}_0 and \mathcal{F}_1 are disjoint for each i , which yielded a symmetric censoring or limiter characteristic. The structure of the resulting *correlator-limiter* detector is shown in Fig. 4.1. The quantities a_i in Fig. 4.1 are related to the degree of limiting and can be solved for from an implicit equation. Although this detector requires knowledge of the value of θ for implementation, a lower bound is shown in [4.13] for the detector power function when it is designed for a specific set of values of the parameters.

In addition to the fact that there is only *one* independent class of pdf's, the noise pdf's, in the known-signal detection problem, a further limitation of the above approach to obtaining a robust detector is that the signal *amplitude* θ needs to be known. In Huber's approach the uncertainty classes of pdf's are usually defined as expansions of corresponding *single* nominal pdf's, so that the resulting robust test does not generally give a test which is *uniformly* robust in the strict minimax sense for, say, a *range* of nominal alternative hypotheses. As a specific example, consider the likelihood ratio of the robust test given by (4.10). Here c' and c'' depend on the density f_{1i}^0 . Thus in [4.13] where these results were applied to the known-signal detection problem, the f_{1i}^0 were defined for a particular signal strength θ . While for this problem it is possible to find a lower bound for the robust detector power function when it is designed for a particular θ , only for that θ is the detector implementing a "minimax" test for the detection problem. Finally, the robust tests and least-favorable pairs were necessarily obtainable only when $\mathcal{F}_0, \mathcal{F}_1$ were not only disjoint but had a finite "separation"; thus the case of vanishing signal strength in weak-signal detection cannot be considered.

These considerations point to the desirability of another formulation for robust hypothesis testing problems of signal detection. One such alternative approach which has been quite fruitful considers the asymptotic case of weak signals and large sample sizes ($\theta \rightarrow 0, n \rightarrow \infty$). It is interesting to note that the



FP-8388

Fig. 4.1: Correlator-limiter robust detector for known signal in ϵ -contaminated Gaussian noise.

basis for this alternative approach has been the theory of robust *estimation* of a location parameter which will be discussed briefly in Section 5 below.

Asymptotically Robust Detection

For the known-signal detection problem of (1.1) consider the log-likelihood ratio defined by (1.3) and (1.4). This generally depends on θ , but for the *locally optimum* (LO) detector which maximizes the *slope* of the power (detection probability) function at $\theta = 0$, subject to a false-alarm probability bound, it is well-known that the test statistic for given noise pdf f should be [4.14, 4.15]

$$T_{\text{LO}}(\underline{X}) = \sum_{i=1}^n -s_i \frac{f'(X_i)}{f(X_i)}. \quad (4.13)$$

This LO test statistic is a special case of the *generalized correlator* (GC) statistic

$$T_{\text{GC}}(\underline{X}) = \sum_{i=1}^n a_i l(X_i). \quad (4.14)$$

The class of GC statistics with $a_i = s_i$ is then a natural class of candidate test statistics to which attention may be restricted in obtaining a robust detector for known s_i but for noise pdf f not precisely known.

For a class \mathcal{F}_0 of noise pdf's one would like to be able to obtain the minimax robust detector from amongst the class of GC detectors (with $a_i = s_i$), with slope of the detector power function as the performance criterion. For fixed finite sample size n this is, unfortunately, extremely difficult in general. However, in the asymptotic case $n \rightarrow \infty$ (and $\theta \rightarrow 0$), under mild regularity conditions yielding asymptotically normal distributions for the test statistics, the problem reduces to a consideration of a more tractable performance measure called the *efficacy*. The efficacy E of a detector based on a test statistic $T(\underline{X})$ may be defined as [4.14]

$$E = \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\left| \frac{d}{d\theta} E \{T(\underline{X}) | \theta\} \Big|_{\theta=0} \right|^2}{\text{var} \{T(\underline{X}) | \theta = 0\}}. \quad (4.15)$$

It also turns out that maximizing the efficacy with $T(\underline{X})$ of the form $\sum_{i=1}^n L(X_i; s_i)$, obtained by dropping the undesirable θ -dependence in (1.3), leads one to the LO statistic of (4.13). Without the limit ($n \rightarrow \infty$) in (4.15) the quantity is sometimes called the differential or incremental signal-to-noise ratio.

For the test statistic of (4.14) with $a_i = s_i$ the efficacy becomes

$$\epsilon = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n s_i^2 \frac{\left[\int_{-\infty}^{\infty} l(x) f'(x) dx \right]^2}{\int_{-\infty}^{\infty} l^2(x) f(x) dx} \quad (4.16)$$

under some regularity conditions on f and l and with the assumption that $l(X_i)$ has zero mean value under the noise-only hypothesis. We may also assume without loss of generality that $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n s_i^2 = 1$, that is that the signal has unit average power when its amplitude is unity. A significant observation about the efficacy of (4.16) with this normalization is that the reciprocal of this normalized efficacy is exactly the asymptotic variance of an M -estimate for the signal amplitude θ of a constant signal ($s_i = 1$). An M -estimate $\hat{\theta}$ of θ is in general that quantity solving

$$\sum_{i=1}^n l(X_i - \hat{\theta}) = 0; \quad (4.17)$$

note that $\hat{\theta}$ is the sample mean when $l(x) = x$. The variance of an M -estimate using l is the reciprocal of the normalized efficacy of (4.16), under some regularity conditions.

A brief discussion of robust M -estimation of a location parameter when the pdf f of the additive noise is not precisely known is given in Section 5. Here we shall note only that in 1964 Huber [4.1] found the least-favorable pdf in the class of ϵ -contaminated noise pdf's for the robust M -estimation problem, and hence obtained the corresponding optimum M -estimate as the robust M -

estimate. It should now be clear (since minimizing the asymptotic variance in M -estimation is equivalent to maximizing the efficacy for our detection problem) that the least-favorable pdf found by Huber will also make its sequence of LO detectors (for $n = 1, 2, \dots$) an asymptotically robust sequence of detectors, providing a saddlepoint value for the game in which the efficacy is the performance function. In fact one can get a stronger result where the false-alarm probability can simultaneously be bounded. These results for asymptotically robust detection were first extended, from Huber's estimation results, by Martin and Schwartz [4.13] and later expanded by Kassam and Thomas [4.16].

The general result in [4.13] and [4.16] may be summarized as follows. Let \mathcal{D} be the class of all detectors of asymptotic size (i.e., false-alarm probability) α for our hypothesis testing problem with $f \in \mathcal{F}_0(g, \epsilon)$ as defined in (4.1). Now g is a *symmetric* nominal density function which is *strongly unimodal* so that $-\log g$ is a convex function, and is twice differentiable, and the contamination h is *symmetric* and bounded but otherwise arbitrary. Let $\beta_D(\theta | f)$ be the power function of a detector D with $f \in \mathcal{F}_0(g, \epsilon)$ the noise density function. Our false-alarm probability constraint is that for each $D \in \mathcal{D}$,

$$\lim_{n \rightarrow \infty} \beta_D(0 | f) \leq \alpha, \quad \text{all } f \in \mathcal{F}_0(g, \epsilon) \quad (4.18)$$

The asymptotically most robust detector³ $D_R \in \mathcal{D}$ is then defined as the locally optimum detector for the least-favorable $f_R \in \mathcal{F}_0(g, \epsilon)$ such that

$$\lim_{n \rightarrow \infty} \frac{\beta_{D_R}'(0 | f)}{\beta_{D_R}'(0 | f_R)} \geq 1, \quad \text{all } f \in \mathcal{F}_0(g, \epsilon), \quad (4.19)$$

in addition to (4.18) and

³More explicitly, each $D \in \mathcal{D}$ is an infinite sequence (D_1, D_2, \dots) of detectors, one for each sample size n . The functional dependence of the test statistics on \underline{X} is the same for all members of the sequence.

$$\frac{\beta_D'(0|f_R)}{\beta_{D_R}'(0|f_R)} \leq 1, \text{ all } D \in \mathcal{D}. \quad (4.20)$$

The D_R satisfying (4.18)-(4.20) exists for $\alpha \geq \alpha(g, \epsilon)$, a lower bound depending on g and ϵ , and is the locally optimum detector of asymptotic size α for $f_R \in \mathcal{F}_0(g, \epsilon)$ given by Huber's exponential-tailed density

$$f_R(x) = \begin{cases} (1-\epsilon)g(x), & |x| < a \\ (1-\epsilon)g(a) \exp[-b(|x| - a)], & |x| \geq a \end{cases} \quad (4.21)$$

where a, b satisfy

$$\int_{-a}^a g(x) dx + 2g(a)/b = (1-\epsilon)^{-1}, \quad (4.22)$$

$$\frac{-g'(a)}{g(a)} = b. \quad (4.23)$$

The lower bound on α is given explicitly as a function of g and ϵ [4.16].⁴ For the unit-variance Gaussian nominal density this bound is no less than 0.158, which is obtained when $\epsilon \rightarrow 0$. Note that the robust detector is based on the test statistic T_{GC} of (4.14) with $l = l_R = -f_R'/f_R$, which is

$$l_R(x) = \begin{cases} -g'(x)/g(x), & |x| < a, \\ -g'(a)/g(a), & |x| \geq a. \end{cases} \quad (4.24)$$

The threshold for the robust detector can be set by considering the normalized test statistic $\frac{1}{\sqrt{n}} \sum_{i=1}^n s_i l_R(X_i)$, and basing the computation of false-alarm probability on the asymptotically normal distribution of this finite-variance statistic.

The resulting generalized correlator detector may be described as a *limiter-correlator* detector, and this is one of the canonical structures of robust detection theory. Note that when g is zero-mean, unit-variance Gaussian we have $-g'(x)/g(x) = x$, and l_R becomes the "amplifier-limiter" or "soft-limiter"

⁴A method for relaxing this restriction is discussed in [4.18].

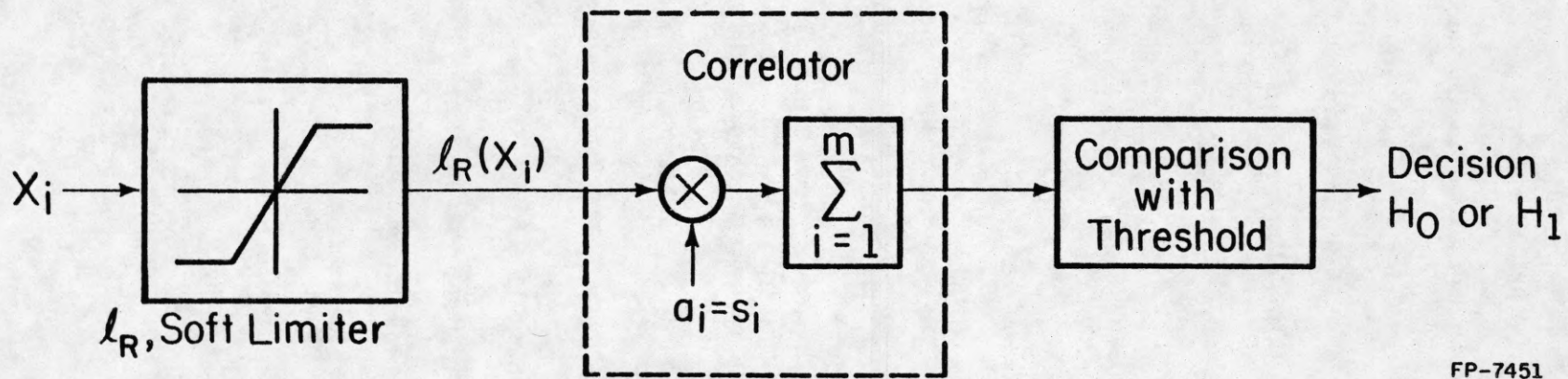
nonlinearity. The structure of the limiter-correlator detector which is asymptotically robust for known-signal detection in ϵ -contaminated Gaussian noise is shown in Fig. 4.2. Note that this particular GC detector can be used in place of the linear correlator detector acting on the matched filter outputs in Fig. 3.4, especially when the matched filter output noise components can be modeled as being ϵ -contaminated Gaussian random variables with unknown contaminating pdf's.

Further details of the above results are given in [4.13] and [4.16], including comparison with other detectors on the basis of asymptotic relative efficiency (ARE), and a discussion of the robustness property of the simple sign detector, which does an extreme form of limiting on the observations. A numerical study has been made in [4.17] of the performance characteristics of various limiter-correlator detector nonlinearities, for the additive known-signal detection problem. The noise density in [4.17] was taken to be the ϵ -contaminated Gaussian nominal with contaminating densities of the impulsive-noise type modeled by exponential (Laplace) and generalized Cauchy functions. The nonlinearities considered were various multi-level approximations of the canonical "amplifier-limiter", including the hard-limiter and the noise-blanker. Performances were characterized in terms of asymptotic relative efficiencies with respect to the linear-correlator detector.

The growing interest in detection systems using quantized data led to a consideration in [4.18] of the above known-signal in additive noise problem with the requirement that the detector characteristic l in (4.14) be an m -level piecewise-constant quantizer characteristic. It has been shown in [4.18] that the most robust quantizer-detector in the sense of (4.18)-(4.20) for the ϵ -contaminated noise density class $\mathcal{F}_0(g, \epsilon)$ of (4.1), for the detection problem (1.1), is again the locally-optimum quantizer-detector for the least-favorable density of (4.21).

Robust Detectors Based on M -Estimators

The detectors we have considered so far in this section have all had the general structure of Neyman-Pearson optimum detectors, which are based on a comparison of the likelihood ratio to a threshold. The test statistics of our detectors have been of the form $\sum_{i=1}^n l_i(X_i)$, which is that of the statistic



FP-7451

Fig. 4.2: Limiter-correlator asymptotically robust detector for weak signal in ϵ -contaminated Gaussian noise.

of (1.3). A different approach for robust detection of additive known signals was proposed by El-Sawy and VandeLinde in [4.19]. In their approach, the test statistic is simply an M -estimate of the signal strength parameter θ , and their robust detector test statistic is the robust M -estimate of θ for the same class of noise densities. To understand the motivation for this approach, and its difference from the previous one and the possible advantages of the resulting detectors, let us re-examine briefly the above results on asymptotically robust known-signal detection. It is clear that, as for any consistent test, the robust detector based on a test statistic of the form of (4.14) has a power function $\beta_D(\theta|f)$ which approaches unity as n approaches ∞ for each $\theta > 0$, assuming that $\sum_{i=1}^n s_i^2/n$ approaches a positive value. Thus, the practical interpretation of (4.19) is that for large enough n for which use of the central limit theorem is reasonably well justified, the slope of the power function at $\theta = 0$ for D_R may be considered to be minimized by f_R in $\mathcal{F}_0(g, \epsilon)$.

In [4.19] it is pointed out that this condition does not guarantee that for each $f \in \mathcal{F}_0(g, \epsilon)$, we will have $\beta_{D_R}(\theta/f) \geq \beta_{D_R}(\theta/f_R)$ in some interval $(0, \theta_f)$. A simple example of a density in $\mathcal{F}_0(g, \epsilon)$ for which $\beta_{D_R}(\theta/f) < \beta_{D_R}(\theta/f_R)$ when $\theta > 0$ proves this. The lack of a strict inequality in (4.19), which allows this to happen, led to the alternative approach in [4.19].

For the hypothesis testing problem of (1.1), note that for a finite-power signal (for which $\lim_{n \rightarrow \infty} \sum_{i=1}^n s_i^2/n = c$, a finite positive value) the signal energy as $n \rightarrow \infty$ becomes infinite under the alternative hypothesis H_1 . If the hypothesis testing problem statement is slightly modified by replacing the amplitude θ with $\nu/n^{1/2}$, the total signal energy in any sample of size n remains finite, and the limiting energy is ν^2 , assuming $c = 1$ without loss of generality. In a practical sense, suppose an observation of length N is given for which the known signal sequence (s_1, s_2, \dots, s_N) has amplitude θ_1 and for which $\sum_{i=1}^N s_i^2/N = C_N$. For N large enough, the results to follow can be applied with $\nu = \theta_1 [NC_N]^{1/2}$.

The minimax robust detector for this known-signal detection problem was obtained from amongst a class of M -detectors in [4.19]. This is a class of detectors for which the test statistic is an M -estimate

(as in (4.17)) of the signal strength θ . Let \mathbf{L} be the class of M -estimator functions L which satisfy convexity, symmetry, monotonicity and differentiability requirements, in addition to mild requirements on the moments of the associated random variables when the noise density function f is a member of a general class \mathcal{F} of symmetric densities. The M -estimator $\hat{\theta}$ of θ , based on n observations and a function L , is the value of θ minimizing $\sum_{i=1}^n L(X_i - \theta s_i)$, so that

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n L(X_i - \theta s_i). \quad (4.25)$$

This is a generalization for non-constant signals of Huber's definition. Note that (4.17) for $s_i = 1$, all i , is obtained from this with $l(x) = dL(x)/dx$.

Let the asymptotic detection power as a function of ν for an M -detector based on a function L be denoted as $\beta_L(\nu | f)$. This power function obviously depends on the noise density function $f \in \mathcal{F}$. The analysis in [4.19] establishes the following results. Let $V^2(f, L)$ be the asymptotic variance of $n^{1/2}(\hat{\theta} - \theta)$, which is asymptotically normally distributed with mean zero for $f \in \mathcal{F}$ and $L \in \mathbf{L}$. Suppose there exists a density $f_0 \in \mathcal{F}$ such that $L_0 = -\log f_0$ is in \mathbf{L} , and $V^2(f, L_0) \leq V^2(f_0, L_0)$ for all $f \in \mathcal{F}$. Let the M -detector based on L_0 have a threshold γ . Then for $\nu \geq \gamma$, we have

$$\beta_{L_0}(0 | f) \leq \beta_{L_0}(0 | f_0), \quad f \in \mathcal{F} \quad (4.26)$$

and

$$\min_{f \in \mathcal{F}} \beta_{L_0}(\nu | f) = \beta_{L_0}(\nu | f_0) = \max_{L \in \mathbf{L}} \beta_L(\nu | f_0) \quad (4.27)$$

In fact, since the M -detector based on $-\log f_0$ is asymptotically equivalent in performance to the detector based on the likelihood-ratio for f_0 when f_0 is the noise density function, the maximum over \mathbf{L} in (4.27) can be replaced by the supremum over all detectors with size equal to that of the M -detector based on L_0 .

The main requirement in the above is that ν not be less than γ . This implies that for a given ν , the design false-alarm probability cannot be too small. The requirement $\nu \geq \gamma$ can be met by making the sample size n sufficiently large in any given situation. Note, however, that the saddlepoint condition (4.27) and (4.26) do not hold together, for a given sample size and false-alarm probability constraint, in an interval around $\nu = 0$. It has been shown in [4.19] that for such cases the robustness criterion based on asymptotic slope of the power function (as in (4.19) and (4.20)) also makes the M -detector based on L_0 the most robust detector, but again subject to the same restrictions on the minimum value for the false-alarm probability.

If the saddlepoint solution (f_0, L_0) exists it may be obtained by first minimizing the *Fisher Information* function for location, $I(f)$, over all $f \in \mathcal{F}$. The Fisher Information function for location is the function

$$I(f) = \int_{-\infty}^{\infty} \left[\frac{f'(x)}{f(x)} \right]^2 f(x) dx . \quad (4.28)$$

It is no accident that $I(f)$ is the maximum value of the normalized efficacy in (4.16), obtained for $l(x) = -f'(x)/f(x)$, the LO nonlinearity. The minimizing density f^* and the function $L^* = -\log f^*$ in \mathcal{L} give the saddlepoint pair $(f_0, L_0) = (f^*, L^*)$. In [4.19] the family of symmetric density functions with a known probability p assigned to an interval $(-a, a)$ was explicitly considered as an example. For this p -point class of densities the least favorable density f_0 was again shown to have exponential behavior outside $(-a, a)$, and the robust M -detector characteristic L_0 had derivative l_0 which was a constant outside $(-a, a)$. Some numerical performance comparisons of asymptotic performance and finite-sample simulation results are also given in [4.19] for a particular example.

The significance of the function $dL(x)/dx = l(x)$ in M -detection is that under appropriate regularity conditions, the test statistic $\hat{\theta}$ must satisfy

$$\sum_{i=1}^n s_i l(X_i - \hat{\theta}_{s_i}) = 0 \quad (4.29)$$

(This is the same as (4.17) for $s_i = 1$, all i). The *variance* of the statistic $\hat{\theta}$ is the reciprocal of the *efficacy* of the generalized correlator detector based on the test statistic of (4.14), with $a_i = s_i$. Note that the mean of $\hat{\theta}$ is θ . Since l is a monotonic increasing function of its argument for L in the class \mathcal{L} , we have a simple way of implementing such a detector when s_i all have the same sign. In this case the sign of $\sum_{i=1}^n s_i l(X_i - \gamma s_i)$ directly indicates if $\hat{\theta}$ is above or below the threshold γ .

The robust detection of known signals using an M -estimate has also been extended in [4.20] to the sequential binary signaling problem. Here the two hypotheses H_0 and H_1 are defined by

$$\begin{aligned} H_0 : X_i &= N_i + \theta_0 \\ H_1 : X_i &= N_i + \theta_1 \end{aligned}, \quad i=1,2,\dots \quad (4.30)$$

where the density function of noise components N_i is a member f of some class \mathcal{F} of symmetric densities. Under the same mild restrictions on the class \mathcal{L} of allowable M -detector characteristics L as in the non-sequential case, a robustness property is established in [4.20] for the sequential M -detector (MS-detector) based on the sequence of robust M -estimates $\hat{\theta}$ defined by the same characteristic L_0 as in the robust non-sequential case. Thus, when $f = f_0$ for this robust scheme, the probabilities of error are upper bounds on the probabilities of error for arbitrary $f \in \mathcal{F}$, and the same holds for "normalized" expected sample sizes, in the limiting case when $|\theta_1 - \theta_0| \rightarrow 0$ and sample sizes are large, for which Gaussian approximations to distributions for $\hat{\theta}$ can be used. In addition, it has been shown that the pair (f_0, L_0) also forms a saddlepoint for performance measured as a risk function which is a linear combination of the error probabilities and the "normalized" expected sample sizes, for one set of weighting or cost coefficients. The "normalization" of the expected sample sizes is required to obtain well-defined quantities, since the results are valid asymptotically when actual sample sizes become infinitely large.

An alternative scheme for robust sequential testing in the situation described by (4.30) was also considered in [4.20]. This scheme is the robust *stochastic approximation* sequential (SAS) receiver. Again the test statistic is directly based on the sequence of robust fixed-sample stochastic approximation estimates of the location parameter $\theta = \theta_0$ or θ_1 ; this estimation scheme will be discussed briefly in Section 5. It has been shown that the robust MS- and SAS-detectors have identical asymptotic characteristics. The SAS-detector does not store past observations, as it uses a recursive algorithm for updating the estimate of θ as new observations arrive. On the other hand, the MS-detector converges to its asymptotic performance faster and does not need an initial estimate to obtain $\hat{\theta}$.

Another class of robust estimates of mathematical statistics is the class of *L-estimates*, which are estimates formed as linear combinations of order statistics. One example is the median, which also happens to be an *M-estimate*. The *α -trimmed mean* is another example of an *L-estimate* and will be discussed in Section 5. The robustness of these *L-estimates* in estimation has prompted investigation of their use in signal detection; references [4.21-4.23] investigate the use of some simple *L-estimates* in signal detection problems.

The results we have discussed in this subsection show quite clearly the central role in robust signal detection of robust *estimation* theory, particularly of the original results of Huber, in addition to that of Huber's results on robust hypothesis testing. We now go on to consider, in somewhat less detail, the other two canonical signal detection problems of random signals and of bandpass signals in additive noise.

4.3. Robust Detection of Random Signals

In many applications involving the detection of random signals the observations are obtained simultaneously from a number of sensors forming an array. The detection problem then often becomes that of detecting a random signal *common* to each sensor in the presence noise processes uncorrelated with each other and with the signal. An example of such an application is a hydrophone array in a passive sonar system used to detect the presence of, and to locate, sources of random signals. For each search direction, relative time-delays are imparted to the sensor outputs to equalize the propagation

delays to each sensor for a potential signal from that direction. This is followed by a detector for a common random signal.

A special case of interest is that in which the sensor array is composed of two elements. In this situation, the sampled observation vectors $\underline{X}_1 = (X_{11}, X_{12}, \dots, X_{1n})$ and $\underline{X}_2 = (X_{21}, X_{22}, \dots, X_{2n})$ can be described by

$$X_{ji} = \theta S_i + N_{ji}, \quad i=1,2,\dots,n, \quad j=1,2, \quad (4.31)$$

where the signal amplitude θ is zero under the noise-only null hypothesis and $\theta \neq 0$ under the alternative hypothesis. We assume here that the $\underline{N}_j = (N_{j1}, N_{j2}, \dots, N_{jn})$ are independent sequences of zero-mean i.i.d. noise components with common pdf f , and $\underline{S} = (S_1, S_2, \dots, S_n)$ is an independent signal sequence of i.i.d. zero-mean components, whose variance may be taken to be unity without loss of generality.

The presence of a signal in the above model ($\theta \neq 0$) causes the output pairs of observations (X_{1i}, X_{2i}) to be positively correlated, in addition to increasing the power level received at each sensor. The increase in the correlation value from zero is in particular due to the presence of the *common* random signal. Thus it is quite reasonable to restrict attention to the class of *generalized cross-correlation* (GCC) detectors based on test statistics of the type

$$T_{\text{GCC}}(\underline{X}_1, \underline{X}_2) = \sum_{i=1}^n l(X_{1i})l(X_{2i}), \quad (4.32)$$

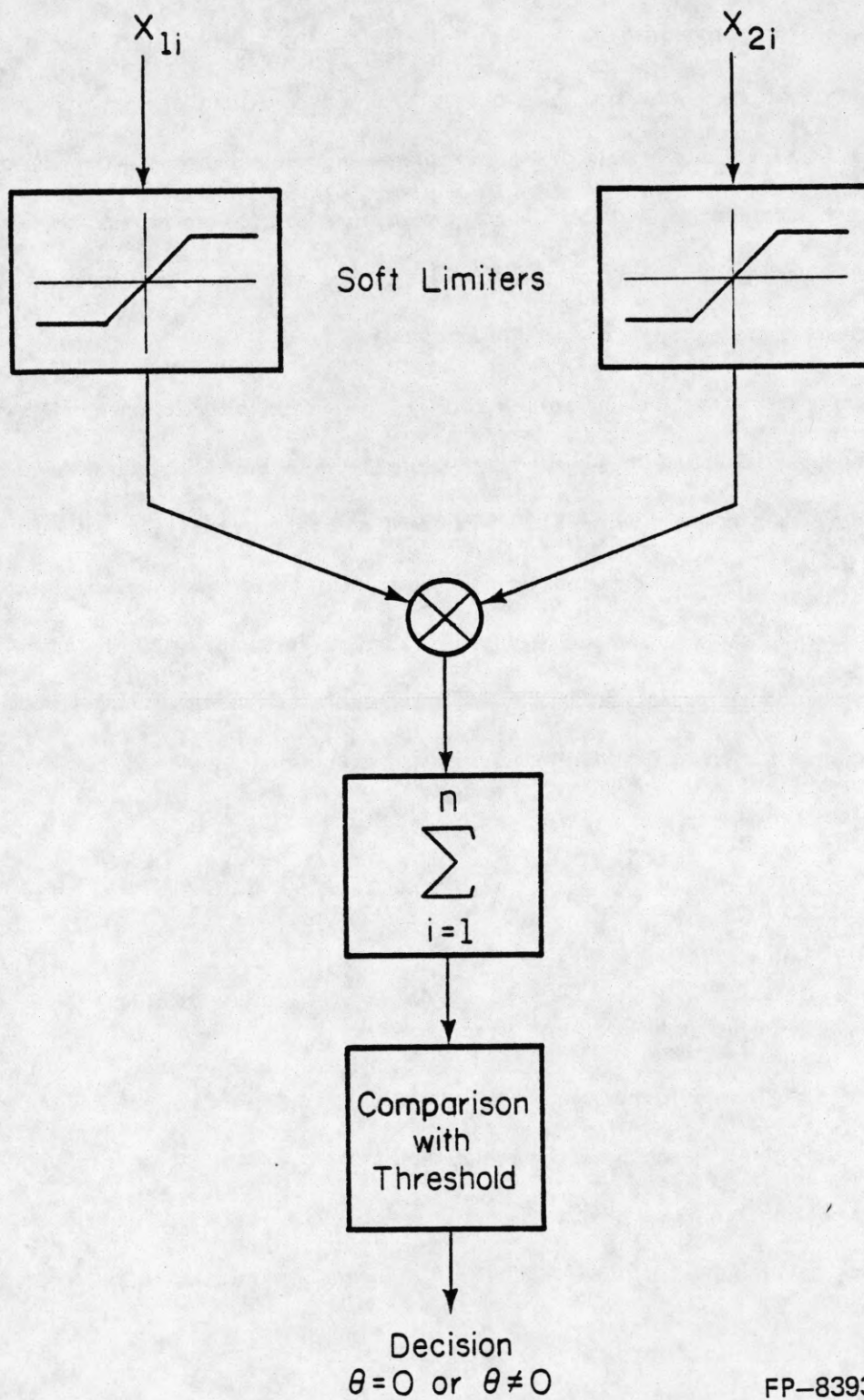
where l is the detector weighting function. For detection of weak signals ($\theta^2 \rightarrow 0$) it can be shown easily that detection efficacy is maximized with $l(x) = -f'(x)/f(x)$, which is also the LO nonlinearity for known-signal detection.

In one of the earliest published studies of robust detection of random signals Wolff and Gastwirth [4.24] examined several specific simple nonlinearities l for robustness of performance when f is not known. A finite class of pdf's comprised of the Gaussian, logistic and the pdf for a particular t -distribution was considered in [4.24]. The nonlinearities l considered in [4.24] were the simple three

and four-level symmetric quantizer and the "amplifier-limiter" or "soft-limiter" continuous function which is linear in an interval including the origin but constant outside that interval. The criterion of performance used in [4.24] was effectively an *asymptotic relative efficiency*; it was the ratio of the efficacy for noise density f obtained with weighting function l and the optimum efficacy for noise density f . In the definition of efficacy, θ^2 is now the signal strength parameter.

In [4.25] the detection of a random signal common to an *array* of receivers was considered as an extension of the two-input case. Again, no attempt was made to include a bound on the false-alarm probability as part of the performance criterion, so that the assumption was that detector thresholds could be varied to get the correct false-alarm probability for any noise density. The performance criterion was the detection efficacy, which as we have remarked before for the known-signal case is directly related to the slope of the power function at $\theta = 0$. For the correlator-array structure (generalization of (4.32) by a second sum over all *pairs* of sensors) an interesting saddlepoint robustness result was shown. This was that for the ϵ -contaminated class $\mathcal{F}_0(g, \epsilon)$ describing the independent noise component at each receiver across the array, with the same restrictions on it as in the known signal in additive noise case, subject to mild smoothness restrictions on allowable l functions, the minimax robust cross-correlator nonlinearity l_R is exactly the same as l_R of (4.24) for the known signal case. Of course, the result here was only possible in terms of the weaker efficacy criterion above. Figure 4.3 shows the structure of a two-element random signal detector using an "amplifier-limiter" GCC nonlinearity l .

In [4.25] other variations of the ϵ -contamination class for this multi-input problem were also considered. Modifications of locally-optimum *detectors* (rather than *cross-correlators*) for common random signal robust detection were also investigated, although explicit minimax robustness results could be deduced only for the special case of contaminated double-exponential noise density, for which a hard-limiter based *polarity coincidence array* detector [4.26] is most robust. The polarity coincidence array detector is based on use of the two-level sign function for l , and is well-known as a non-parametric detector with a fixed false-alarm probability for zero-median noise pdf's.



FP-8393

Fig. 4.3: Generalized cross-correlation asymptotically robust detector for weak random signal.

The class of detectors for single-input additive random signal contains no counterpart of the cross-correlator structure. In [4.27] the optimum quadratic detector for Gaussian statistics was modified, from intuitive considerations, into the limiter-quadratic detector. Numerical asymptotic relative efficiency computations for Gaussian-mixture noise pdf's (ϵ -contaminated nominal Gaussian with larger-variance Gaussian contaminating density) and finite-sample detection power and false-alarm probability computations verify the expected robustness of the limiter structure. The subsequent results in [4.25] extended such structures to the multi-input or array case.

Explicit minimax results for the single-input additive random signal detection problem paralleling those for known-signal detection have been obtained more recently in [4.28]. Two statistical models for this problem have been investigated in [4.28]. In one model the alternative hypothesis of signal presence is defined to produce additive signal and noise components in the observations, (i.e. (4.31) for $j=1$). In the second, scale-change, model the alternative hypothesis is defined by the condition that the pdf of the i.i.d. observation components X_i is $f(x/\sigma)/\sigma$, where f is the null-hypothesis observation (noise) pdf. It had been shown earlier [4.29] that for the additive noise model the locally-optimum detector is based on the *generalized energy* (GE) detector test statistic

$$T_{GE}(\underline{X}_1) = \sum_{i=1}^n \frac{f^*(X_{1i})}{f(X_{1i})}; \quad (4.33)$$

the LO statistic for the scale-change model is also given in [4.29].

From knowledge of the previous results on known-signal robust detection one might conjecture that a minimax result should be obtainable for the robust detector formed by introducing hard-limiting beyond some argument value a into the nonlinearity f^*/f of (4.33). Indeed, consider the exponential-tails least-favorable density of (4.21), with a, b defined as in (4.22) and a modified (4.23), namely

$$\frac{g^*(a)}{g(a)} = b. \quad (4.34)$$

For this noise density the nonlinearity $f''(x)/f'(x)$ becomes

$$\frac{f''_R(x)}{f'_R(x)} = \begin{cases} g''(x)/g'(x), & |x| \leq a \\ b, & |x| > a \end{cases} \quad (4.35)$$

Consider the ϵ -contamination model $\mathcal{F}_0(g, \epsilon)$ for the noise density f , as described in our summary of known-signal results. In addition to requiring a further smoothness property for g (since here g'' is involved), a more restrictive further condition on the allowable contamination h is imposed. This is that allowable contamination densities are zero in $(-a, a)$. With these conditions, results exactly paralleling those for asymptotically robust detection of known signals in additive noise are obtained in [4.28]. An interesting feature of the solution, arising from the restriction on allowable h , is that (4.19) holds with equality over the allowable f and the result is valid for all values of the size, α . Numerical asymptotic relative efficiency comparisons for general limiter nonlinearities are also given in [4.28].

A very similar result is established in [4.28] for the scale-change model. Here again for a similarly restricted version of the ϵ -contamination class, the density of (4.21) is again shown to be least favorable. A noteworthy point about this solution is that as in the case of asymptotically robust known signal detection, this scale-change robust solution is directly related to Huber's results on robust M -estimation of a *scale* parameter [4.1]. Thus we see once again a close relationship between robust estimation and asymptotically robust detection problems.

4.4. Robust Detection of Bandpass Signals in Bandpass Noise

Bandpass signals are commonly encountered in applications such as radar and communication systems, and techniques for their detection in bandpass noise with imprecisely known statistical characterizations is therefore of practical interest.

Let us first consider briefly one bandpass known-signal detection problem for which an asymptotically robust detector may be defined using ideas very similar to those used for low-pass known-signal and completely random signal robust detection problems. An observed continuous-time waveform is now described by

$$X(t) = \theta v(t) \cos [\omega_0 t + \phi(t)] + N(t), \quad (4.36)$$

where the low-pass signal amplitude and phase components $v(t)$ and $\phi(t)$, respectively, and the frequency ω_0 are known, the overall signal amplitude θ being either 0 (noise-only) or having some positive value (alternative hypothesis). The bandpass noise $N(t)$ may also be expressed in terms of its in-phase and quadrature components $N_I(t)$ and $N_Q(t)$. For a detector operating on sampled values of the in-phase and quadrature components of the observation $X(t)$ the input data may then be represented as a vector $\tilde{\mathbf{X}} = \mathbf{X}_I + j\mathbf{X}_Q$ where the components X_{Ii} and X_{Qi} , $i=1,2,\dots,n$, of \mathbf{X}_I and \mathbf{X}_Q , respectively, are

$$X_{Ii} = \theta s_{Ii} + N_{Ii}, \quad (4.37a)$$

$$X_{Qi} = \theta s_{Qi} + N_{Qi}; \quad (4.37b)$$

here the s_{Ii} , s_{Qi} , N_{Ii} and N_{Qi} are samples of $s_I(t) = v(t) \cos \phi(t)$, $s_Q(t) = -v(t) \sin \phi(t)$, $N_I(t)$ and $N_Q(t)$, respectively. We make the assumption that the $\tilde{X}_i = X_{Ii} + jX_{Qi}$ are i.i.d. for $i=1,2,\dots,n$, implying a restriction on the sampling rate.

If the noise components N_{Ii} and N_{Qi} are restricted to be independent then the problem is essentially the same as the low-pass known-signal detection problem. A more general assumption is that the joint pdf f_{IQ} of N_{Ii} and N_{Qi} has *circular symmetry*, so that

$$f_{IQ}(u,v) = c(r) \Big|_r = \sqrt{u^2 + v^2} \quad (4.38)$$

For Gaussian bandpass noise both circular symmetry and independence of the in-phase and quadrature noise components is obtained. Under the circular symmetry assumption the LO detector uses a *generalized narrowband correlator* (GNC) test statistic

$$T_{\text{GNC}}(\tilde{\mathbf{X}}) = \sum_{i=1}^n l(R_i) \text{Re} \{ \tilde{s}_i \tilde{X}_i^* \} \quad (4.39)$$

where the $R_i = |\tilde{X}_i|$ are the observation envelopes with l the LO envelope weighting function [4.30]

$$l(r) = -\frac{c'(r)}{rc(r)}. \quad (4.40)$$

Note that $f(r) = 2\pi rc(r)$ is the pdf of the R_i . For Gaussian noise f is the Rayleigh pdf and the LO function of (4.40) is a *constant* resulting in the *linear* narrowband correlator (LNC) detector. In general if the pdf f is some known function $2\pi rg(r)$ the LO detector can be obtained. In [4.31] the ϵ -contamination model for f was considered, with conditions on the nominal function g similar to those for known-signal asymptotically robust detection. As in the case of random signal detection, however, the class of contaminating pdf's $2\pi rh(r)$ was restricted to produce tail-contamination only, that is, the h were zero in some interval at the origin. With these restrictions the asymptotically robust detector was shown in [4.31] to be a GNC detector which is LO for a least-favorable pdf f_R having an exponentially decaying tail. The weighting function $l = l_R$ in the test statistic $T_{\text{GNC}}(\tilde{\mathbf{X}})$ of this robust detector is

$$l_R(r) = \begin{cases} -\frac{g'(r)}{rg(r)}, & 0 \leq r < a \\ -\frac{g'(a)}{rg(a)}, & r \geq a \end{cases} \quad (4.41)$$

for a some function of ϵ and g . Notice that the function $rl_R(r)$ is a constant for $r \geq a$; this function may be interpreted as providing the weighting applied to the *hard-limiter* narrowband correlator (HNC) terms $\text{Re} \{ \tilde{s}_i(\tilde{X}_i^*/R_i) \}$, the \tilde{X}_i/R_i having unit amplitude value. For a nominal Gaussian case the function $-g'(r)/g(r)$ is *linear* in r .

Figure 4.4 shows the structure of the asymptotically robust detector for known bandpass signal under the above assumptions. Note that the HNC detector is a special case of this structure and may be viewed as the robust detector for a nominal g function which is exponential, or as an extreme case which actually possesses a constant false-alarm probability and functions as a nonparametric detector [4.32]. Further interesting properties of the robust detector employing l_R of (4.41) are mentioned in [4.31].

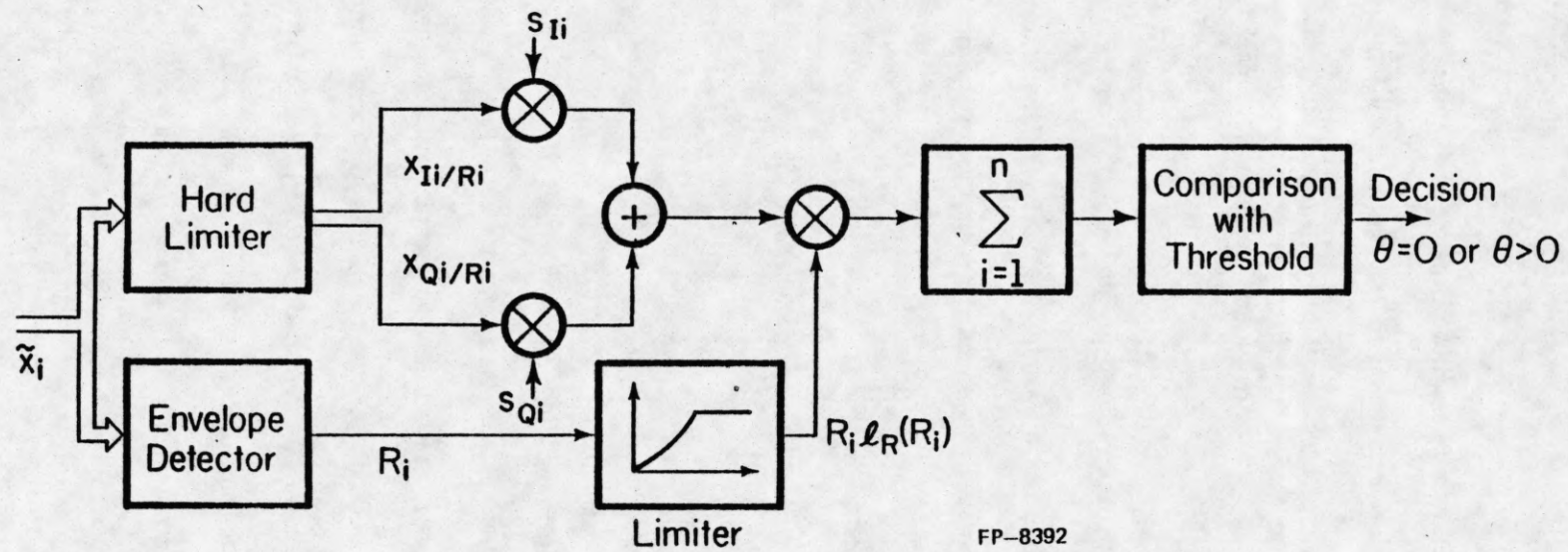


Fig. 4.4: Generalized narrowband correlator asymptotically robust detector for weak bandpass signal.

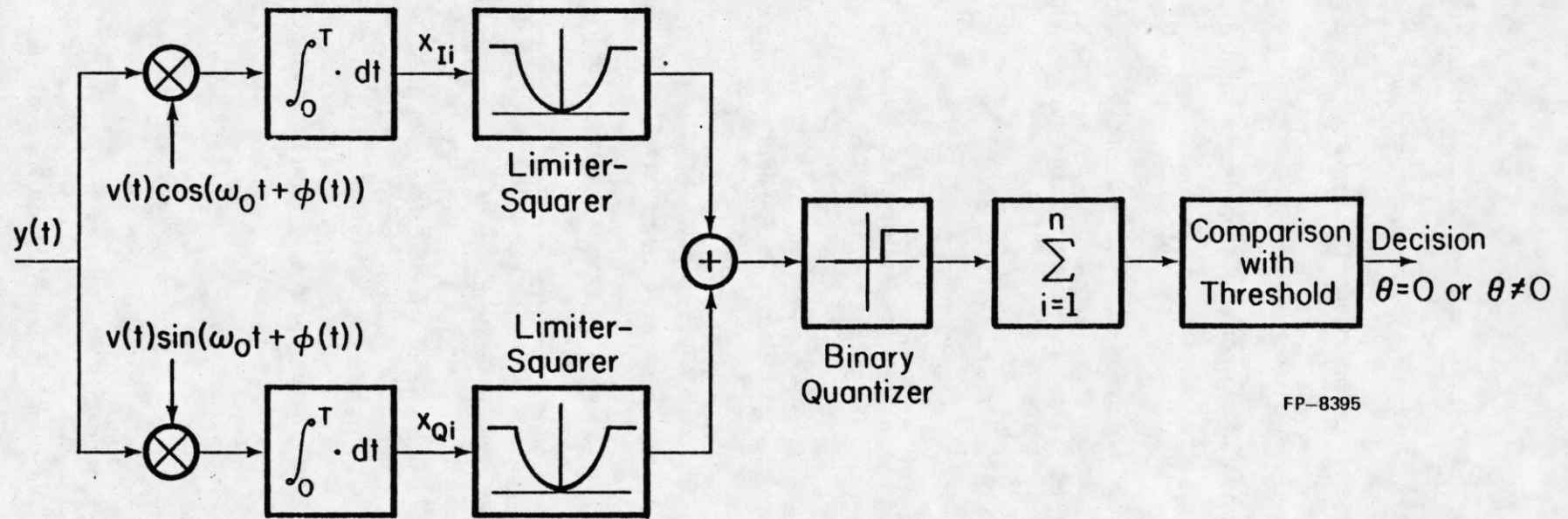
So far we have been considering the case of a coherent bandpass signal. In the incoherent case there is an additional phase term Ψ , uniformly distributed on $[0, 2\pi)$, so that the signal is $v(t) \cos[\omega_0 t + \phi(t) + \Psi]$. The asymptotically optimum detector test statistic now becomes

$$\left| \tilde{T}_{\text{GNC}}(\tilde{\mathbf{X}}) \right|^2 = \left| \sum_{i=1}^n l(R_i) \tilde{s}_i \tilde{X}_i^* \right|^2 \quad (4.42)$$

where l is the optimum nonlinearity of (4.40). It turns out that the efficacy for such a square-law quadrature GNC detector is directly related to that of the GNC detector for coherent signals, and thus $l = l_R$ of (4.41) also results in an asymptotically robust detector in the incoherent case.

In [4.33] a special case of the incoherent signal detection problem was considered in which $v(t)$, the low-pass signal amplitude, was a constant, and $\phi(t)$, the low-pass signal phase term, was $-\pi/2$ (or, equivalently, zero). The in-phase and quadrature noise samples were assumed to be independent and a robust detector was obtained which used the squares of robust M -detector outputs for the separate in-phase and quadrature observations. Specific results were obtained for the p -point classes of univariate noise pdf's. A more general problem involving signals with random parameters was considered in [4.34].

Another important variant of the observation model of (4.36) is obtained by allowing a random signal amplitude $\theta = A$ in addition to the random phase Ψ . The usual assumption for the signal is that A and Ψ are independent, with A being a Rayleigh random variable and Ψ being uniformly distributed on $[0, 2\pi)$. In [4.35] the performance characteristics of a particular modification of the optimum detector for Gaussian noise are considered. The modification consists of replacing the squarers of the optimum detector with limiter-squarers for robust performance. Both the single bandpass pulse detection situation and that of multiple independent bandpass pulses is considered. In the latter case a binary integration or double-threshold detector is considered. The structure of this detector is shown in Fig. 4.5. The numerically computed performance characteristics indicate that, for the independent e -contaminated in-phase and quadrature bandpass matched filter output noise components that were assumed, the limiter-squarer structure is very effective in guarding against drastic performance



FP-8395

Fig. 4.5: Robust detector for independent random amplitude and phase bandpass pulses in additive noise.

deterioration in heavy-tailed non-Gaussian contamination. The structure considered in [4.35] was suggested by the many formal results which we have discussed which show the robustness of limiter-type structures in different situations. In fact, the problem considered in [4.35] may be considered as being a bandpass version of a random signal detection problem, for which robust detectors have been considered in [4.27] and [4.28]. We should note finally that Martin and Schwartz also considered in their study of robust detection [4.13] the detection of multiple independent pulses. The robust structure they suggested used a limiter function on the sampled in-phase and quadrature observations *inside* a digital or discrete-time implementation of the quadrature single-pulse matched filters, followed by a square-law envelope detector.

We have described in these last three subsections the main results on asymptotically minimax robust detection for three canonical signal detection situations. A general theory for robust detectors in the asymptotic case (vanishing signal strengths, infinitely large sample sizes) is discussed in [4.36]. Here asymptotic normality of statistics of the form of (4.14) is studied for four detection problems, namely known signal in additive noise, two random signal problems (one with a scale-change in the noise density, another with an additive random signal, under the alternative hypothesis) and envelope detection of a narrowband signal. In addition, the general characteristics of the robust solutions are considered, which are not restricted only to the univariate case, and a few explicit uncertainty models and robust detector structures are discussed. This work essentially considers only a performance measure related to the efficacy in the known signal case, and defines a general notion of Fisher's Information; simultaneous consideration of a false alarm probability constraint is not attempted in [4.36].

4.5. Extensions and Other Results

While we have surveyed the basic results of robust hypothesis testing and robust signal detection in the above subsections, several further results in this area deserve some comment. We have seen that explicit minimax robust structures can be derived under some rather specific and sometimes restrictive assumptions. In this subsection we will also mention some work that has been directed at easing some of these constraining assumptions.

Signal Uncertainty

All the results that we have discussed in this section were obtained for robust detection when only the noise pdf's are not precisely known. Consider, for example, the known-signal detection problem of subsection 4.2. In our discussion of the asymptotically robust detector of the generalized correlator type with test statistic of the form (4.14), the choice $a_i = s_i$ for the coefficients in the test statistic was the obvious one for known-signal detection. But it is also possible to consider here uncertainty about the exact values of the signal components s_i . Notice that this problem was considered in Section 3 for robust matched filters and SNR performance; here the GC detector is not necessarily linear and we have been concerned with weak-signal detection performance.

In [4.37] Kuznetsov considered this signal uncertainty problem, using the differential signal-to-noise ratio (see discussion following (4.15)) as a performance measure. For known noise pdf and l_2 signal uncertainty class the result is quite similar to that for the robust linear matched filter of Section 3, because the differential signal-to-noise ratio is after all a performance measure derived from the SNR. The random-signal detection problem with signal covariance matrix uncertainty is also considered in [4.37]. For joint uncertainty in signal characteristics and noise pdf's solutions for the minimax robust structures are difficult to obtain in the general case. The special case where the noise is Gaussian with uncertain covariance matrices for the signal and noise has been treated in [4.37], with the detector restricted to employ a combination of linear and quadratic statistics. A similar problem involving deterministic signals and linear detectors has been considered in [4.38], as mentioned in Section 3.

Serial Dependence and Asymmetry

Although a consideration of the asymptotic performance of detectors allows appealing minimax robust detector structures to be derived, a major limitation of the schemes we have described is that the asymptotic robustness property can be attributed to them only under the assumption of independence of the detector input samples. This assumption is generally difficult to get around, serial data dependence introducing considerable complications in the analysis. One of the problems, of course, is the definition of appropriate statistical models and uncertainty classes. Recently, however, some progress

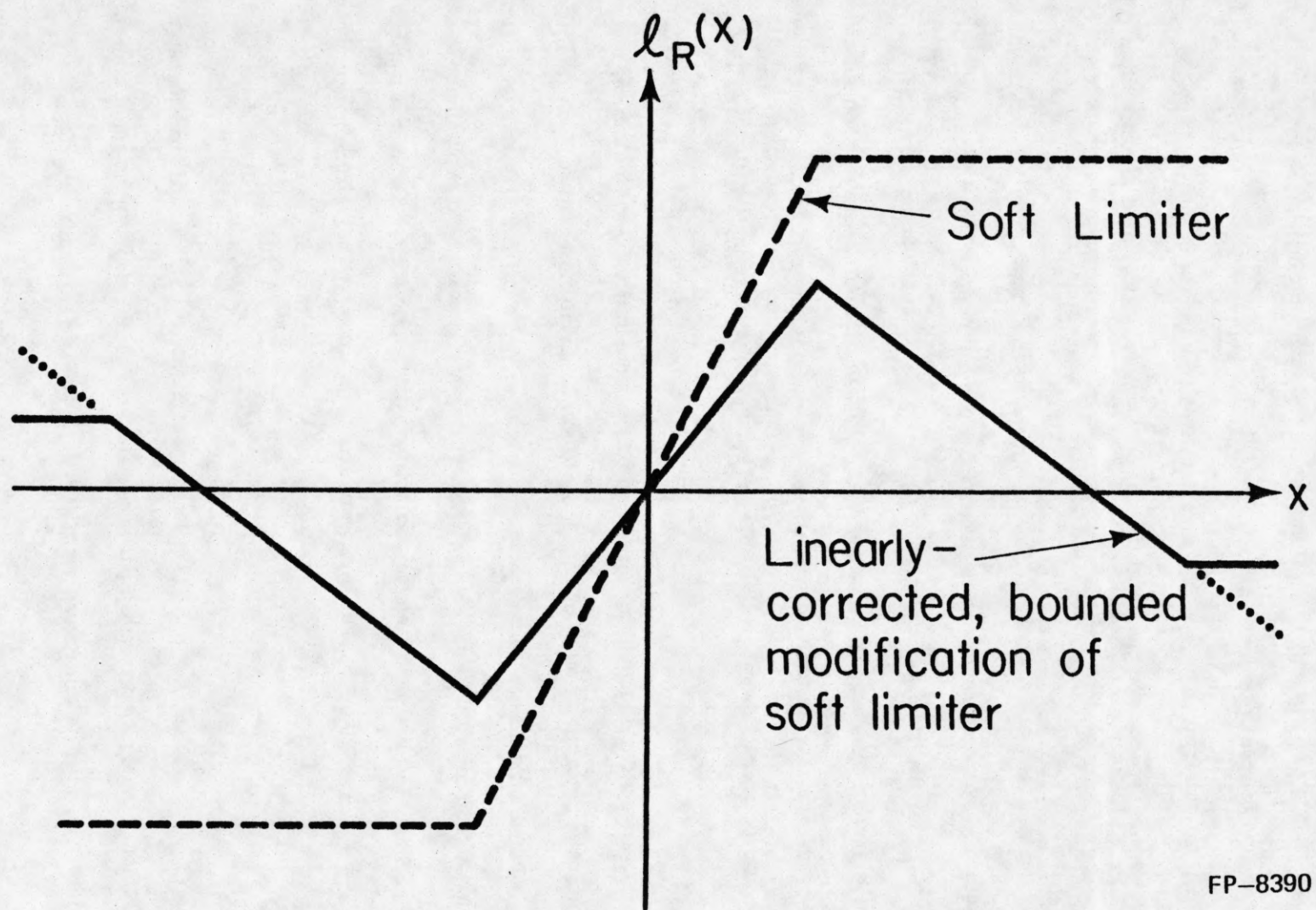
has been made in the specification of asymptotically minimax robust detectors for serially dependent data samples.

In [4.39] Poor considers the case of weak serial dependence of data in a constant-signal M -detection problem. The dependence structure considered is that obtained from a moving-average model for the noise components. In particular, the noise model

$$N_i = \rho W_{i+1} + W_i + \rho W_{i-1} \quad (4.43)$$

is considered where the $\{W_i\}$ are i.i.d. with uncertain marginal pdf's. For $\rho = 0$ the approaches described previously are applicable; in [4.39] the case of weak non-zero correlation is considered. For the ϵ -contaminated pdf classes, it is shown there that the robust M -detector nonlinearity is the limiter function which is robust for $\rho = 0$, corrected by a linear term. Thus for the nominal Gaussian case a nonlinearity l_R of the form illustrated in Fig. 4.6 is suggested; more detailed considerations modify this so that the nonlinearity remains bounded. These results are similar to those derived earlier for robust estimation with dependent data [4.40, 4.41].

More recently Moustakides and Thomas [4.42] considered a less structured dependence assumption for the known-signal detection problem. The additive noise sequence was here assumed to be ϕ -mixing, which includes the case of stationary Markov noise sequences. For the ϵ -contamination univariate noise pdf model, with attention confined to the class of generalized correlator detectors, and with efficacy as the performance criterion and an explicit requirement of false-alarm probability control, the form of the robust detector nonlinearity was derived. The very interesting conclusion was that, subject to some regularity conditions, the robust detector nonlinearity is a null-zone modification of the independent-data robust detector nonlinearity for the same class of univariate noise pdf's. For the nominal Gaussian noise pdf the form of this nonlinearity is shown in Fig. 4.7. It is interesting to note that a simple three-level nonlinearity which approximates this characteristic can be used to provide nonparametric performance for symmetric noise pdf's [4.43]. The result of Moustakides and Thomas represents a major breakthrough in the extension of asymptotically robust correlator-detector structures for dependent



FP-8390

Fig. 4.6: M -detector nonlinearity for asymptotically robust detection of weak signal in ϵ -contaminated, weakly correlated, Gaussian noise.

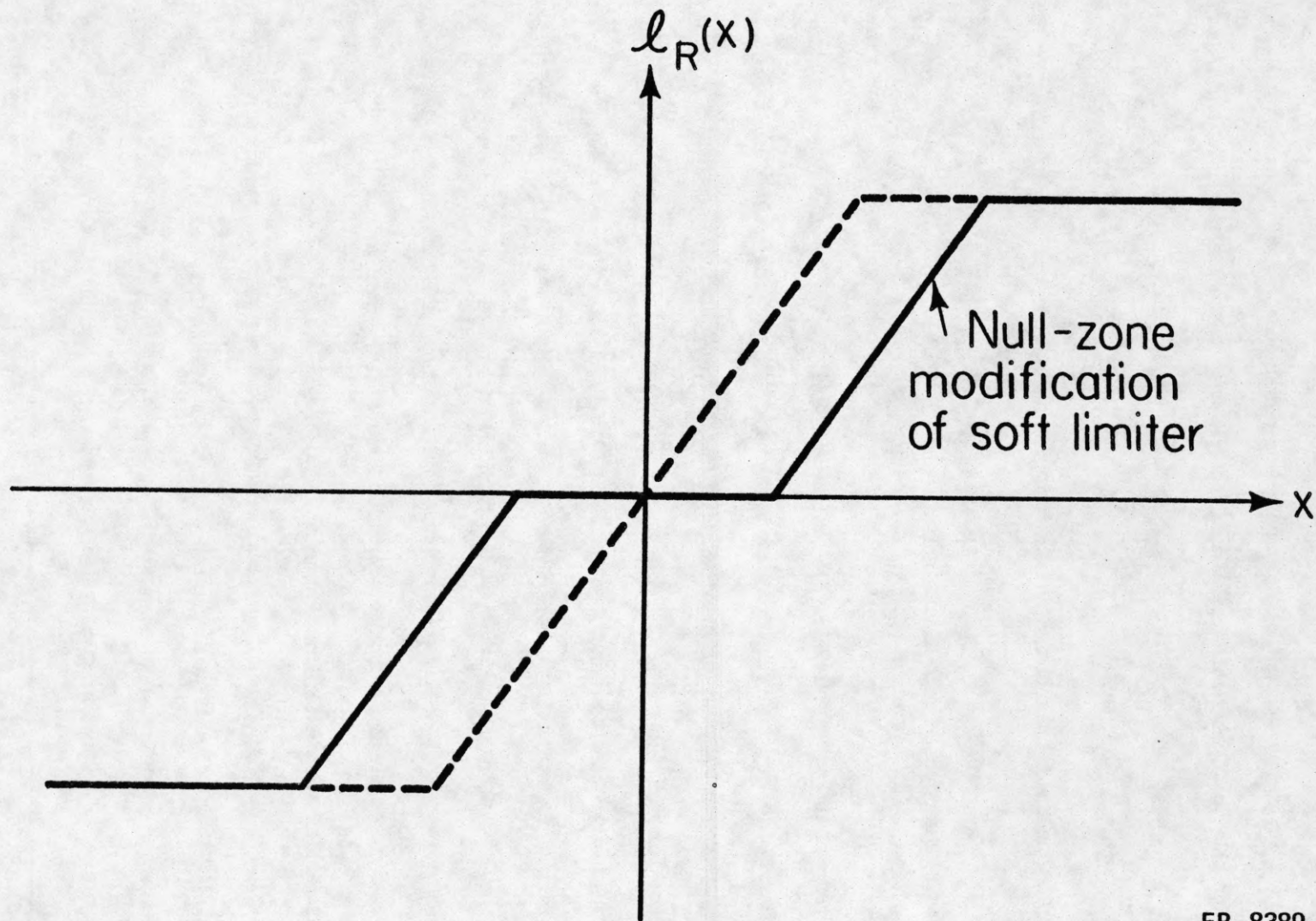


Fig. 4.7: Generalized correlator detector nonlinearity for asymptotically robust detection of weak signal in ϵ -contaminated Gaussian noise.

data, and similar results can now be sought for random and bandpass signal detection problems.

Another aspect of robust detection in dependent data has been considered recently by Martin [4.44], in which autoregressive dependence and a regression signal model which includes the known signal and random-phase bandpass signal cases is assumed. He suggests that, in this situation, robust M -detection is best accomplished by first prewhitening the noise and then applying an M -estimate to obtain the test-statistic.

The asymptotic theory of robust detection, fundamentally related to Huber's results on robust parameter estimation as it is, is limited by the same factors that have constrained the applicability of robust parameter estimation theory. In addition to serial independence of data one other assumption which has been generally required, at least in the widely used ϵ -contamination model for uncertain pdf's, is *symmetry* of the pdf's of the noise. In estimation theory this results in unbiased estimates for which the variances can be written down as second moments. Recently an attempt has been made in [4.45] to apply to robust detection of known signals ideas formulated for robust M -estimation for classes of density functions allowing tail asymmetry. Specifically, [4.45] considers both the generalized correlator structure of (4.14) and the M -estimate structure (4.29) for detection test statistics, and develops the form of the robust detection function l_R in (4.14) and (4.29) for classes of ϵ -contaminated nominal densities with arbitrary behaviors outside a central interval. The development is based on the work on the corresponding M -estimation problem [4.46], and shows the robustness of detector functions which redescend to zero and remain zero outside the interval of symmetry.

Two general limitations of minimax robust detection theory should by now be clear. These are that a considerable part of the theory that has been developed is asymptotic theory, and that it is currently unable to deal directly with continuous-time observation processes. It should be kept in mind that here our concern is with uncertainty in the noise pdf's, and not just with power spectral density uncertainties and linear schemes which we discussed in Section 3. Care has to be taken in applying asymptotic theory in practical situations involving a finite number of observations, since in some cases predicted asymptotic performance is approached rather slowly. The Gaussian approximation

for test statistic distributions in setting threshold values should be applied with care, even if the structure of the robust test statistic appears to have strong justification. In addition, another aspect of asymptotic formulations is that they are based on the assumption that signal strengths are approaching zero. Thus the effect of a non-zero signal and a finite sample size together could result in deviations in performance from theoretical predictions, even for large samples. Since closed-form analytical results are rarely feasible, about the only alternative left is numerical computations and simulations to verify the actual performance of such robust schemes in general.

Continuous-time results are very difficult to come by because of convergence issues and also because of the difficulty in specifying models to define simple classes of allowable random processes which are physically meaningful. Of course, if only second-moment characteristics are modeled and exact probability distribution functions are irrelevant, as in maximizing signal-to-noise ratio at the output of a linear detector, robust filters such as the robust matched filter can be obtained. Although [4.47] introduces classes of "mixture" or contaminated random processes, there has been no application made of this in signal detection with continuous-time observations.

Adaptive and Nonparametric Detectors

In a general sense a detector can be said to be robust if it has good (close-to-optimal) detection performance under nominal conditions and if it also maintains an acceptable level of performance when the noise statistics deviate, within some allowable class, from the nominal. Our focus in this paper has been on *fixed* detectors designed to provide minimax detection schemes, which generally possess the above two characteristics. Another approach which can be taken when the noise environment statistical characterization is imprecisely known, or is largely unknown, is to use adaptive procedures. In most adaptive procedures one generally begins with a specific test statistic structure in which some parameters are free to be set and updated as functions of previous inputs, which might include separate training data. In addition, the threshold is also usually free to be set adaptively. We have mentioned earlier that even for "fixed" minimax robust detectors the thresholds may have to be set adaptively to maintain false-alarm probability requirements whenever the performance criterion does not explicitly include

such requirements. Since our focus here has been on minimax robust fixed-test-statistic structures we shall not try to give here an exhaustive survey of adaptive robust detectors, but will mention only some of the main recent contributions.

One general structure in the known-signal detection problem is obtained by requiring the function l in the generalized correlator test statistic of (4.14) to be an m -level quantizer characteristic. Such a quantizer-correlator detector may be viewed as partitioning the observations X_i into m subsets or intervals to each of which a distinct level is assigned. For a *given* noise pdf the asymptotically optimum quantizer characteristic can be found for this and similar detection problems [4.48]. If, however, the noise pdf is not known one may use estimated values of these optimum parameters. Alternatively, the quantizer breakpoints may be required to fall at the quantiles of the noise distribution, which can be estimated and updated. Adaptive m -interval partition detection schemes have been described by Kurz [4.49] and for sequential detection by Dwyer [4.50]. Other studies of this nature can be found in [4.51-4.55].

Recognizing that contamination of a Gaussian noise density function by an impulsive-noise component, and heavy-tailed densities in general, are reasonable models for random noise and interference in several applications [4.17, 4.56], adaptive structures are considered in [4.57, 4.30] for good performance over such classes of noise. The simple structure considered in [4.57] for the detection of a known signal in additive noise is illustrated in Fig. 4.8. It is motivated by the fact that for Gaussian noise the linear-correlator test statistic is optimum, whereas for heavy-tailed noise pdf's often used to model impulsive noise the sign-correlator test statistic performs very well. The mixture test statistic considered in [4.57] is $T_{GC}(\underline{X})$ of (4.14) with $a_i = s_i$ and

$$l(X_i) = \gamma X_i + (1-\gamma) \operatorname{sgn}(X_i), \quad (4.44)$$

where the free parameter γ is set adaptively. Modestino in [4.57] discusses a stochastic approximation technique maximizing the signal-to-noise ratio for setting the value of γ . A similar robust detector structure based on adaptively forming an optimum test statistic combination is described in [4.30] for

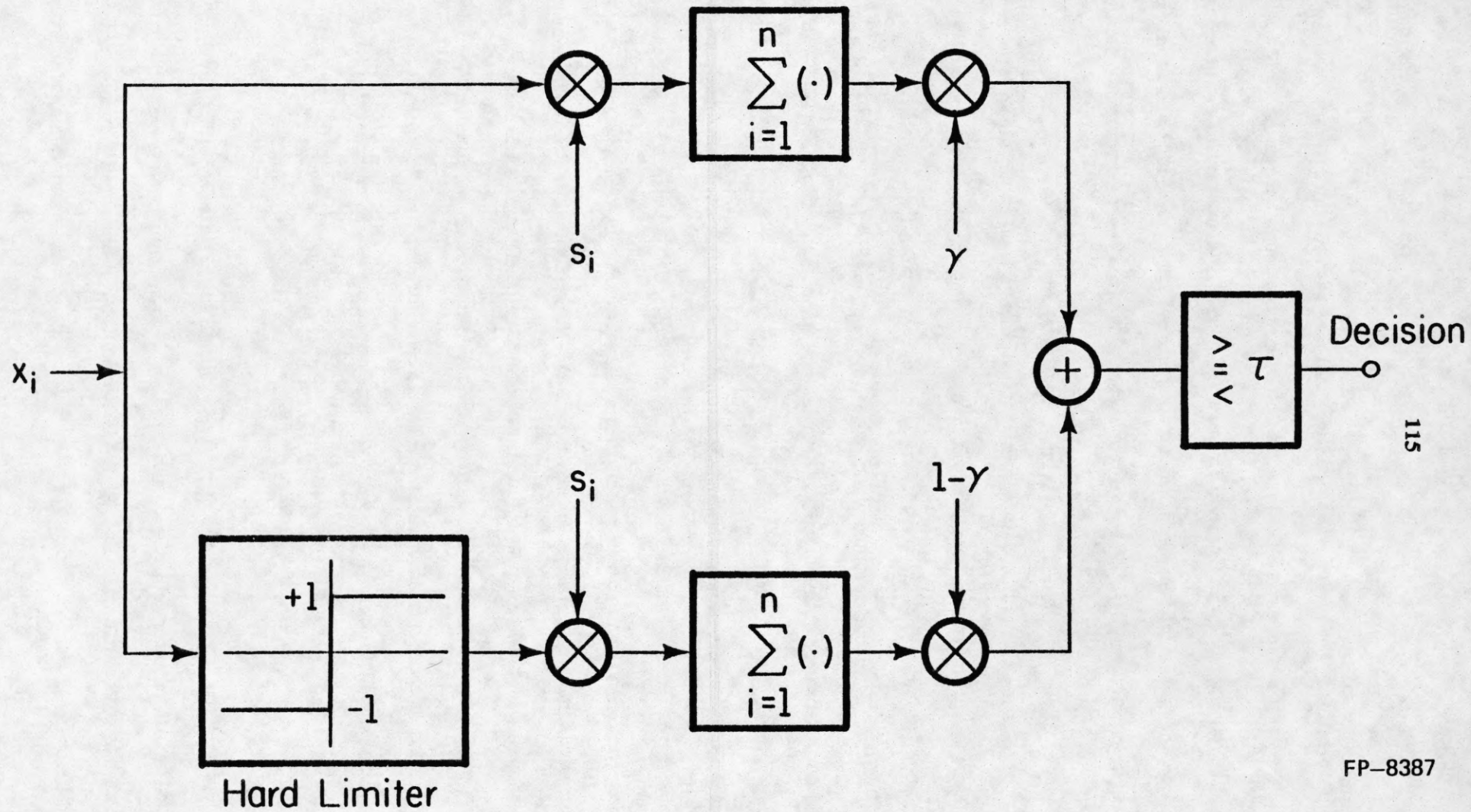


Fig. 4.8: Adaptive detector for signals in a mixture of Gaussian and impulsive noise.

bandpass signal detection. The detector described in [4.58] uses extreme-value theory to obtain the proper threshold setting in an otherwise fixed matched-filter envelope detector structure.

Although adaptive schemes can be useful in situations where the noise statistics are unknown or are nonstationary, the implementation of efficient adaptive schemes can add in a major way to the complexity of a detector. Furthermore, other considerations such as of speed of convergence may limit their applicability.

In nonparametric detection the main concern is that the probability of false alarm remain bounded by some design maximum value for broad classes of noise statistics, for example for all univariate noise pdf's which are symmetric about the origin. Once again it is possible to consider adaptive structures for nonparametric detection, which are mainly of the adaptive threshold type, although fixed-structure and fixed-threshold nonparametric schemes are the most common. For a given class of noise pdf's there generally exist several possible nonparametric detectors, and in choosing between alternatives one usually considers detection performance at some nominal cases within the class. The most common nonparametric detection schemes are those based on signs and ranks of observations, including multilevel (quantization) versions of sign-based schemes. It often turns out that nonparametric detectors are robust, in the more general sense, in their detection performance over the classes of noise pdf's for which they are designed. Note, however, that the primary aim of nonparametric schemes is to keep the false alarm probability bounded by any desired value, whereas robust schemes are required to additionally exhibit good detection performance for the whole class instead of at some nominal operating points only. In this regard the reader is referred to some comments by Huber in [4.59]. For survey papers, a bibliography and details of nonparametric detection schemes we refer the reader to [4.60] and [4.61].

In conclusion we point out a recent Russian survey paper by Krasnenker, available in English translation [4.62] which is on the subject of robust detection. Another similar survey by Ershov [4.63] is concerned with robust estimation, although it also covers some detection and hypothesis testing results. Vandelinde [4.64] gives a short survey of robust techniques in communications which covers

robust detection. Finally, Poor has given a more recent survey, emphasizing the mathematical details, in [4.65].

5. NONLINEAR METHODS FOR ROBUST ESTIMATION

In Sections 2 and 3 we discussed robust *linear* methods for estimation and detection within uncertain second-order models. In Section 4 we saw that, when the uncertainty is described in terms of the distributional model rather than the second-order model, nonlinear methods are called for to provide robustness in signal detection. Similar considerations arise in problems of estimation in uncertain distributional models, and in this section we discuss some of the main issues arising therein. Since many of these issues have been surveyed and unified elsewhere (see, e.g., the book by Huber [5.1], and the surveys by Martin [5.2], Ershov [5.3], and Poljak and Tsypkin [5.4]), we touch only briefly on the essential ideas of this area.

The large majority of work in this area has been concerned with robust nonlinear parameter estimation rather than with robust nonlinear filtering. In subsection 5.1 we outline basic methods for (minimax) robust parameter estimation by considering the important special case of robust estimation of signal amplitude. In subsection 5.2 we consider robustness in other estimation contexts including system identification and Kalman filtering. We also give brief mention to problems of robustness in other methods of time series analysis.

5.1. Robust Estimation of Signal Amplitude

Many signal processing applications arising in practice fall within the category of parameter estimation. A common example is the estimation of the amplitude of a signal embedded in additive noise. A standard model for this particular situation is that we have observations $\underline{X} = (X_1, \dots, X_n)$ given by

$$X_k = N_k + \theta s_k, \quad k=1, \dots, n, \quad (5.1)$$

where (N_1, \dots, N_n) is an i.i.d. noise sequence with symmetric marginal pdf f , (s_1, \dots, s_k) is a known signal waveform, and the signal amplitude θ is unknown and is to be estimated.

The particular situation of (5.1) in which the signal is constant ($s_1 = s_2 = \dots = s_n = 1$) is the location estimation problem of statistical inference, and it was this problem that was studied in the

seminal 1959 paper of Tukey [5.5] that demonstrated the lack of robustness of the classical estimators for this parameter, and in the pioneering 1964 paper of Huber [5.6] that established the importance of minimax methodology for robust estimation. In the following discussion of the basic results in this area, we will consider this constant signal case unless otherwise noted. Modifications for time-varying signals will be mentioned where appropriate.

Three classical estimators for the location parameters of (5.1) with constant s_i are the sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, the sample median, $\text{med} \{x_1, \dots, x_n\}$, and the maximum-likelihood estimate (MLE), $\hat{\theta}_{ML}(\underline{x}) = \arg \{ \max_{\theta} \sum_{i=1}^n \log f(x_i - \theta) \}$. The sample mean and the MLE coincide for the case in which f is a Gaussian density, $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$, and the sample median and the MLE coincide when f is a Laplacian (double-exponential) density, $f(x) = \frac{\alpha}{2} e^{-\alpha|x|}$. All three of these estimators are examples of a more general class of estimators, known as M -estimators, proposed by Huber in [5.6]. As discussed in Section 4, this class of estimators consist of those estimators of the form

$$\hat{\theta}_L(\underline{x}) = \arg \{ \min_{\theta} \sum_{i=1}^n L(x_i - \theta) \} \quad (5.2)$$

where L is a function determining the estimator. The sample mean corresponds to the choice $L(x) = x^2$, the sample median corresponds to $L(x) = |x|$, and the MLE corresponds to $L(x) = -\log f(x)$. Note that $\hat{\theta}_L$ of (5.2) is the estimate of θ that best "fits" the data when errors, $x_i - \hat{\theta}_L(\underline{x})$, are weighted with the function L .

Assuming that L is convex, symmetric about the origin, and sufficiently regular, the M -estimate based on L is consistent and has the property that $\sqrt{n} (\hat{\theta}_L(\underline{X}) - \theta)$ is asymptotically Gaussian with zero mean and variance $V(l, f)$ where

$$V(l, f) \triangleq \frac{\int l^2 f}{\int l f}, \quad (5.4)$$

with $l = L'$ (see [5.6] for details). Thus, estimators of this type can be compared on the basis of their asymptotic variances $V(l, f)$. The function l is called the *influence curve* of $\hat{\theta}$. For fixed f , the Schwarz inequality implies that

$$V(l, f) \geq V(l_0, f) = \frac{1}{I(f)}, \quad (5.5)$$

where $l_0 = -f'/f$, and where $I(f) = \int (f')^2/f$ is Fisher's information for location. (Note that the inequality $V(l, f) \geq \frac{1}{I(f)}$ is a special case of the Cramer-Rao lower bound.) This l_0 corresponds to $\rho_0(x) = -\log f(x)$; so assuming $-\log f$ is convex, the most efficient M -estimate of location is the maximum-likelihood estimate, a well-known result.

If, as in the signal-detection problems discussed in Section 4, f is not known precisely but rather is known only to be in some class \mathcal{F} of bounded symmetric pdf's, then it is possible that the performance of an improperly designed location estimator can be quite poor. For example, suppose we consider the ϵ -contaminated Gaussian class

$$\mathcal{F} = \{f \mid f = (1-\epsilon)\phi + \epsilon h, h \in \mathcal{H}\} \quad (5.6)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, \mathcal{H} is the class of all bounded symmetric pdf's, and $\epsilon \in [0,1]$. The optimum estimator based on the nominal model ϕ is the sample mean, which corresponds to $l_0(x) = x$ and so has asymptotic variance

$$V(l_0, f) = \int_{-\infty}^{\infty} x^2 f(x) dx = (1-\epsilon) + \epsilon \int_{-\infty}^{\infty} x^2 h(x) dx. \quad (5.7)$$

For $\epsilon > 0$ the asymptotic variance (5.7) can be arbitrarily large since h is arbitrary, and so the sample mean is a very nonrobust estimator of location for this type of uncertainty.

The basic problem with the sample mean is that it has an unbounded influence curve so that too many large observations (or *outliers*) can destroy its efficiency. This could be corrected by employing the sample median, which has the influence curve $l_m(x) = \text{sgn}(x)$. For the sample median the

asymptotic variance (assuming f is continuous at the origin) is

$$V(l_m, f) = 1/4 f''^2(0) \leq \frac{1}{4(1-\epsilon)\phi^2(0)} = \frac{\pi}{2(1-\epsilon)}. \quad (5.8)$$

So, the sample median is certainly more robust than the sample mean in this case; however, its efficiency relative to the sample mean is only 62 % at the nominal model. Ideally, we would like an estimator that has near the efficiency of the sample mean at the nominal and has the robustness of the sample median away from the nominal. It turns out that these goals can be achieved as we see below.

To correct the possible nonrobustness of classical estimators of location, Huber proposed the design of location estimators using a minimax formulation, viz

$$\min_l \max_{f \in \mathcal{F}} V(l, f). \quad (5.9)$$

Within some assumptions on \mathcal{F} , the solution to (5.9) is given by the influence curve $l_R = -f'_R/f_R$ where f_R is a *least-favorable* density for location estimation defined by

$$f_R = \arg \left\{ \min_{f \in \mathcal{F}} I(f) \right\}. \quad (5.10)$$

The pair (l_R, f_R) is a saddlepoint solution for (5.9) provided $-\log f_R$ is convex; i.e., under the convexity of $-\log f_R$, (l_R, f_R) satisfies

$$V(l_R, f) \leq V(l_R, f_R) \leq V(l, f_R), \quad (5.11)$$

for all $f \in \mathcal{F}$ and all l . Note that (5.11) implies that, not only does l_R have minimax asymptotic variance but also its variance is upperbounded over \mathcal{F} by its variance when f_L is the true density. This means that the performance of the estimator will never be worse than $V(l_R, f_R)$ over \mathcal{F} . Also note that, since $I(f)$ is Fisher's information for location, f_R has the interpretation of being the density for which the observations are *least informative* about θ .

For the particular case of ϵ -contaminated Gaussian data (i.e., F from (5.6)), the least favorable density turns out to be given by (4.21) with $g = \phi$ as discussed in Section 4. In this case, l_R is a soft-limiter

$$l_R(x) = \begin{cases} x, & \text{if } |x| < a \\ \text{sgn}(x), & \text{if } |x| \geq a, \end{cases} \quad (5.12)$$

where a is from (4.22)-(4.23), and so $l_R^2(x) \leq a^2$, and $l_R'(x) = 1$ for $|x| < a$ and $l_R'(x) = 0$ for $|x| > a$. Thus the numerator of $V(l_R, f)$ is

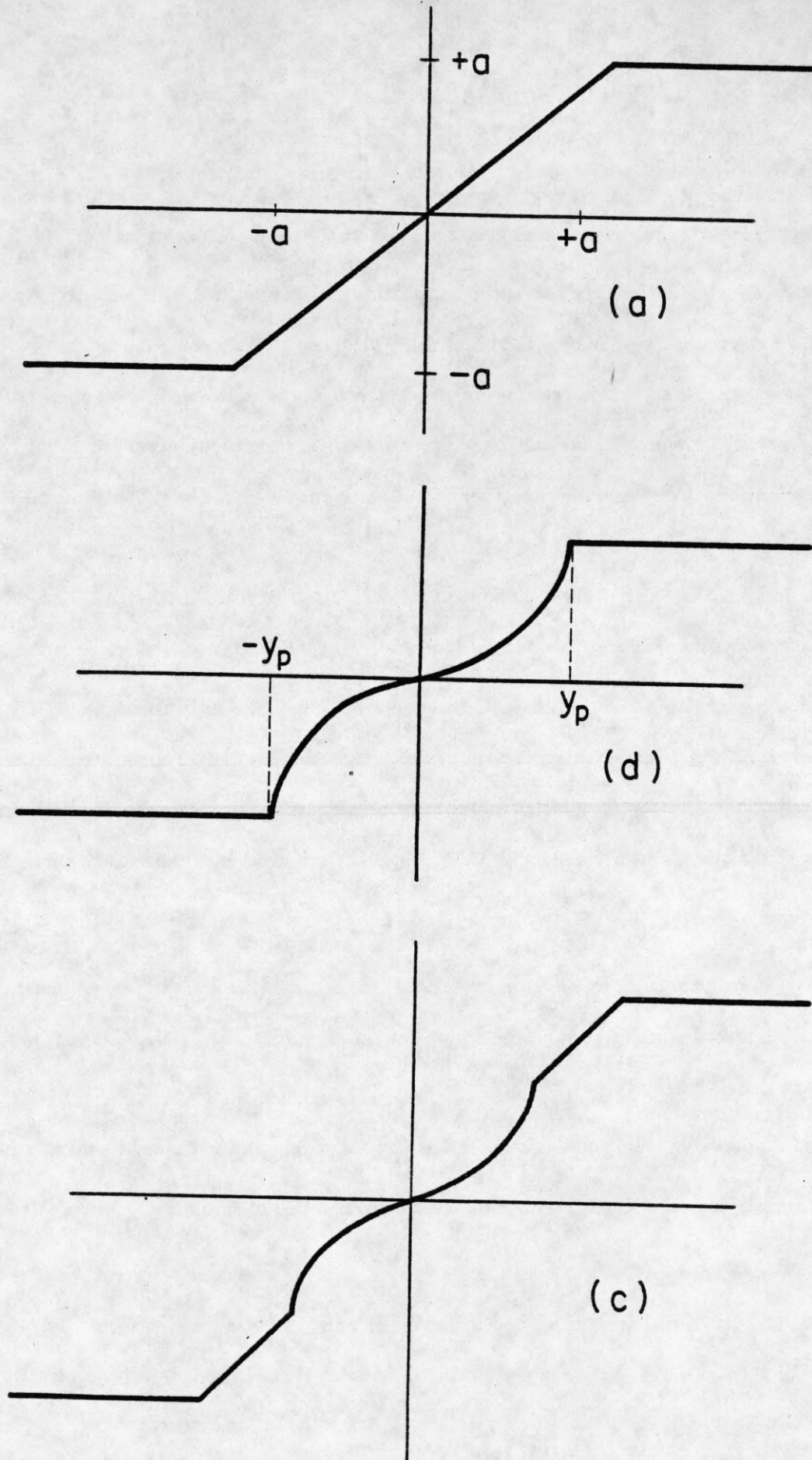
$$\begin{aligned} \int_{-\infty}^{\infty} l_R^2(x) f(x) dx &= (1-\epsilon) \int_{-\infty}^{\infty} l_R^2(x) \phi(x) dx + \epsilon \int_{-\infty}^{\infty} l_R^2(x) h(x) dx \\ &\leq (1-\epsilon) \int_{-\infty}^{\infty} l_R^2(x) \phi(x) dx + \epsilon a^2 = \int_{-\infty}^{\infty} l_R^2 f_R(x) dx, \end{aligned} \quad (5.13)$$

and the denominator is

$$\begin{aligned} \int_{-\infty}^{\infty} l_R'(x) f(x) dx &= \int_{-a}^a f(x) dx = (1-\epsilon) \int_{-a}^a \phi(x) dx + \epsilon \int_{-a}^a h(x) dx \\ &\geq (1-\epsilon) \int_{-a}^a \phi(x) dx = \int_{-\infty}^{\infty} l_R'(x) f_R(x) dx. \end{aligned} \quad (5.14)$$

From (5.13) and (5.14) the left-hand inequality of the saddlepoint condition (5.11) is readily seen to hold in this case. (The right-hand inequality of (5.10) is just a restatement of (5.5).) Similarly, for the ϵ -contamination model with ϕ replaced by any density g (with $-\log g$ convex) a saddlepoint solution to (5.9) is given by $(-f_R'/f_R, f_R)$ with f_R from (4.21).

Note that the influence curve (5.12) of the robust estimator for ϵ -contaminated Gaussian noise combines features of the influence curves for the sample mean and the sample median (see Fig. 5.1a). It has the boundedness of l_m but it has the linear shape of l_0 for $|x| \leq a$. (This eliminates the sensitivity of l_m to the value of f at zero.) For the case $\epsilon = 0.1$, the limiting point $a \cong 1.1$ and the efficiency of the robust estimator at the nominal model relative to the sample mean is 92%. Moreover,



FP-8386

Fig. 5.1: Influence curves for robust location estimates. (a) ϵ -contaminated Gaussian class, (b) p -point class, and (c) $\sup_{x \in R} |F(x) - \Phi(x)| \leq \epsilon$ class.

the worst-case variance of l_R over \mathcal{F} is $V(l_R, f_R) \cong 1.5$ compared with $\sup_{f \in \mathcal{F}} V(l_m, f) \cong 1.9$. Thus, we see that this estimator does indeed achieve the goals set forth for robust estimation.

Robust M -estimators can be found for uncertainty classes other than the ϵ -contaminated class by minimizing Fisher's information over the uncertainty class and then choosing the estimator accordingly. It is not necessary that the uncertainty class contain only distributions with density functions, although the least favorable will always correspond to a continuous distribution. A variational method for minimizing Fisher's information is discussed in Chapter 5 of [5.1]. Examples of useful uncertainty classes for which solutions are known are p -point classes, which consist of the set of all noise distributions that place a fixed amount of probability on a given interval [5.7] and the class of noise distribution functions F that satisfy $\sup_{-\infty < x < \infty} |F(x) - \Phi(x)| \leq \epsilon$, where Φ is the unit Gaussian distribution function [5.6]. Influence curves for robust estimation in these models are shown in Figs. 5.1b and 5.1c, respectively. Appropriate modifications of the influence curve for robust estimation in dependent noise and for asymmetric noise have been considered by Portnoy [5.8, 5.9] and Collins [5.10], respectively. Also, the estimate of (5.2) is straightforwardly modified to account for the time varying signal s_1, \dots, s_n , via $\hat{\theta}_L(x) = \arg \{ \min_{\theta} \sum_{i=1}^n L(x_i - \theta s_i) \}$. Assuming $C \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n s_i^2 < \infty$, this estimate has the same properties as in the constant-signal case except that the asymptotic variance is $V(l, f)/C$. Thus, the minimax robustness problem for the time varying case is identical to that for the constant signal case [5.11].

From a practical viewpoint, the robust M -estimator of signal amplitude has the disadvantage of being a "batch" procedure; i.e., all data must be stored in order to minimize $\sum_{i=1}^n L(x_i - \theta)$ iteratively. This disadvantage can be overcome by considering a class of recursive estimators of the stochastic approximation type. Robustness theory for this type of estimator was first considered by Martin [5.12], and these ideas were developed further by Martin and Masreliez [5.7] and by Price and Vandelinde [5.13].

To discuss these results, we first consider a generalization of the minimax problem of (5.9):

$$\min_{\hat{\theta} \in \Theta} \sup_{f \in \mathcal{F}} v(\hat{\theta}, f) \quad (5.15)$$

where Θ is the class of all asymptotically unbiased estimators of θ and where $v(\hat{\theta}, f)$ is the asymptotic variance of $\hat{\theta}$. A sufficient condition for $\hat{\theta}_R$ to solve (5.15) is that it, together with some $f_R \in \mathcal{F}$, satisfy the saddlepoint condition

$$v(\hat{\theta}_R, f) \leq v(\hat{\theta}_R, f_R) \leq v(\hat{\theta}, f_R) \quad (5.16)$$

for all $\hat{\theta} \in \Theta$ and $f \in \mathcal{F}$. It is possible that (5.9) and (5.15) have different solutions since, in (5.9), consideration is restricted to M -estimates. However, since the Cramer-Rao bound implies $v(\hat{\theta}, f_R) \geq 1/I(f_R) \equiv V(l_R, f_R)$ for all $\hat{\theta} \in \Theta$, we see that the minimax robust M -estimator also satisfies (5.16) and so is in fact a minimax robust estimator among all asymptotically unbiased estimators. However, there is also another saddlepoint $(\hat{\theta}_R, f_R)$ for (5.15) in which $\hat{\theta}_R$ is not the MLE for f_R but rather is a recursive estimator.

In particular, suppose \mathcal{F} , f_R and l_R are as in the robust M -estimation formulation and let $\hat{\theta}_R^{SA}(x) = \hat{\theta}_n$ be defined for each n by

$$\hat{\theta}_i = \hat{\theta}_{i-1} + l_R(x_i - \hat{\theta}_{i-1})/iI(f_R), \quad i=1, \dots, n, \quad (5.17)$$

with $\hat{\theta}_0$ arbitrary. Note that $\hat{\theta}_R^{SA}$ is computed recursively (i.e., at the i^{th} sampling instant $\hat{\theta}_i$ is computed from $\hat{\theta}_{i-1}$ and x_i). Then under regularity conditions the asymptotic variance of $\hat{\theta}_R^{SA}$ for a given noise density f by [5.13]

$$v(\hat{\theta}_R^{SA}, f) = \frac{\int l_R^2 f}{\int l_R^2 f_R [2 \int l_R' f - \int l_R' f_R]} \quad (5.18)$$

Note that $v(\hat{\theta}_R^{SA}, f_R) = 1/I(f_R)$ so that $\hat{\theta}_R^{SA}$ is asymptotically optimum for f_R . Moreover, it can also be shown that

$$v(\hat{\theta}_R^{SA}, f) \leq v(\hat{\theta}_R^{SA}, f_R) \quad (5.19)$$

for all $f \in \mathcal{F}$, so that $(\hat{\theta}_R^{SA}, f_R)$ is also a saddlepoint solution to (5.15). Thus we have the interesting conclusion that the recursive estimator of (5.17) is minimax robust over \mathcal{F} just as is the M -estimator based on l_R . However, although both estimators have the same worst-case performance, $1/I(f_R)$, one can see that their performances ($V(l_R, f_R)$ and $v(\hat{\theta}_R^{SA}, f_R)$) generally differ for given $f \neq f_R$. For the ϵ -contaminated Gaussian case it is easily shown that $v(\hat{\theta}_R^{SA}, f) > V(l_R, f)$ whenever $\int_{-a}^a h(x) dx > 0$; thus, for the computational advantage of the recursive structure one pays the penalty of some lost efficiency when worst-case conditions are not present. However, for example at $\epsilon = 0.1$ this lost efficiency is negligible ($< 1\%$) at the nominal density, ϕ .

The appropriate modifications of the algorithm of (5.17) for the time-varying case is

$$\hat{\theta}_i = \hat{\theta}_{i-1} + s_i l_R(x_i - \hat{\theta}_{i-1} s_i) / I(f_R) \sum_{k=1}^i s_k^2, \quad i=1, \dots, n. \quad (5.20)$$

This estimator has not been studied analytically; however, simulations for the case of a sinusoidal signal in ϵ -contaminated Gaussian noise indicate that the performance of (5.20) is comparable to that for the constant-signal case (Martin [5.12]). Heuristically, in either this or the constant-signal case robustness against contamination is achieved by inserting the soft-limiter l_R into the feedback loop that incorporates the residual $x_i - \hat{\theta}_{i-1} s_i$ into the updates of the amplitude estimate (see Fig. 5.2). If it were not for this limiter, then when $\hat{\theta}_i$ gets close to the time θ "too many" large noise values would tend to drive $\hat{\theta}$ away from θ .

The robustness properties of types of estimators other than the M -estimators and stochastic approximation estimators can be studied in a general framework. In particular, for the constant-signal case, many estimators of practical interest can be written in the form

$$\hat{\theta}_n = T(F_n) \quad (5.21)$$

where F_n is the sample cumulative distribution function (cdf),

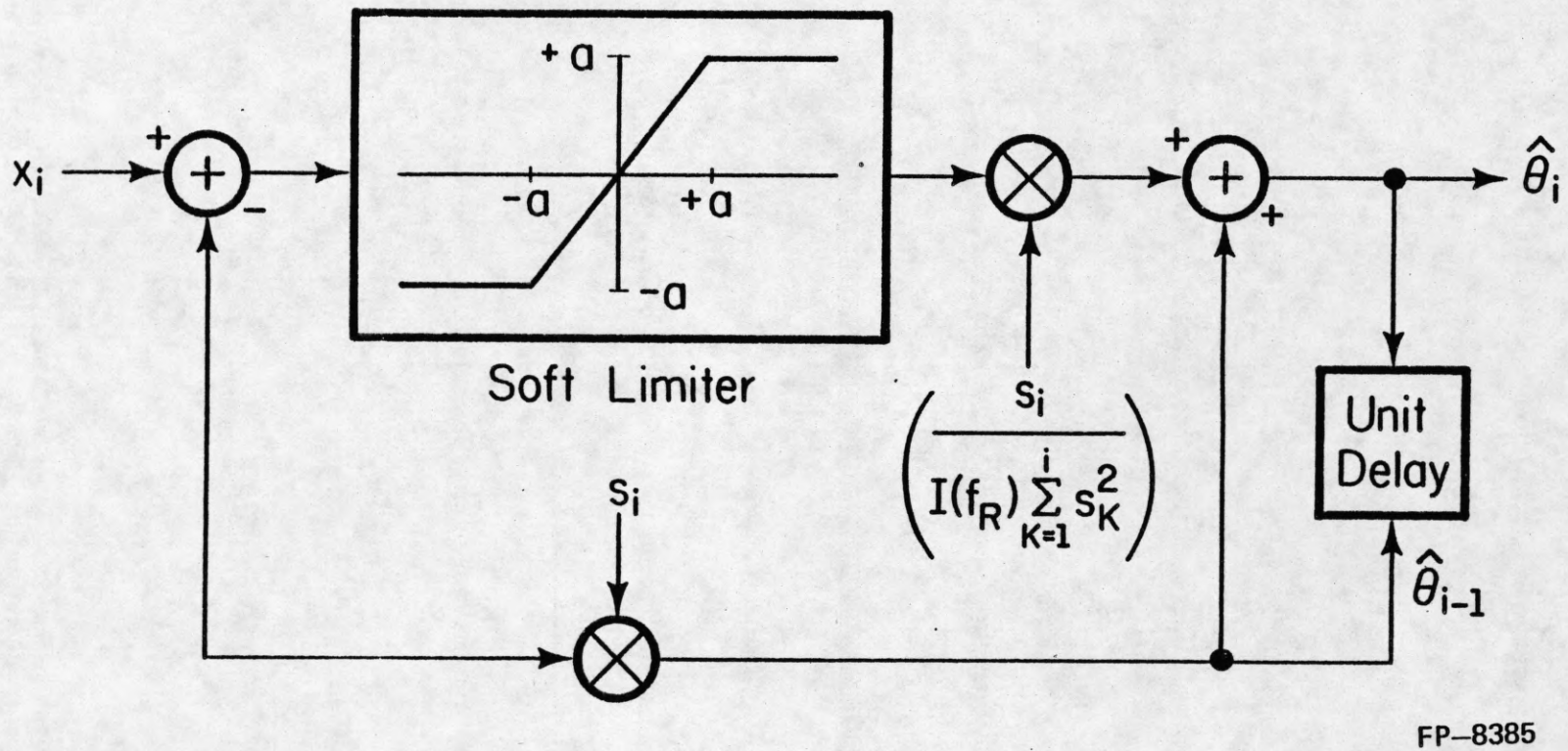


Fig. 5.2: Robust recursive amplitude estimator for ϵ -contaminated Gaussian errors.

FP-8385

$$F_n(x) \triangleq \frac{1}{n} (\#x_i \leq x), \quad -\infty < x < \infty, \quad (5.22)$$

and where T is some functional mapping the set of cdf's into \mathcal{R} . For example if we define

$$T(F) = \arg \left\{ \min_{\theta} \int_{-\infty}^{\infty} L(x-\theta) dF(x) \right\}, \quad (5.23)$$

then the M -estimate $\hat{\theta}_L$ is $T(F_n)$ since $\int_{-\infty}^{\infty} L(x-\theta) dF_n(x) = \frac{1}{n} \sum_{i=1}^n L(x_i - \theta)$. If we assume that L is convex and symmetric about zero and if F_{θ} is the common marginal cdf of $X_i = N_i + \theta$, then for (5.23) we have $T(F_{\theta}) = \theta$.

For a wide class of estimates of the form $T(F_n)$ it turns out that, when F is the true data distributions, $\sqrt{n} (T(F_n) - T(F))$ is asymptotically normal with zero mean and variance

$$A(F, T) = \int_{-\infty}^{\infty} [IC(x, F, T)]^2 dF(x) \quad (5.24)$$

where IC is the *influence curve* of the estimate (the reason for the earlier use of this term will become obvious below) defined by

$$IC(x, F, T) = \lim_{\epsilon \rightarrow 0} \frac{T((1-\epsilon)F + \epsilon F_x) - T(F)}{\epsilon}, \quad (5.25)$$

where F_x is the cdf of a random variable taking on only the value x .¹ The influence curve for an M -estimate $\hat{\theta}_L$ is

$$IC(x, F_{\theta}, T) = \frac{l(x-\theta)}{-\int l(x-\theta) dF_{\theta}(x)}, \quad (5.26)$$

which gives the asymptotic variance $V(l, f)$ of (5.4).

¹Note that, in addition to its role in the asymptotic variance formula (5.24), the influence curve characterizes the sensitivity of $T(F_n)$ to the incorporation of a datum x into the estimate; i.e., it quantifies the influence of such a datum on the estimate. Hence, the term influence curve.

Other useful classes of estimates of the form $T(F_n)$ are the so-called L -estimates, which estimate θ by linear combinations of the order statistics (i.e., the observation sequence put in numerically increasing order), and R -estimates which are based on the ranks of the data. Explicit representations of these type estimators in the form $T(F)$ and the corresponding influence curves can be found in [5.1]. For any fixed noise density, an asymptotic variance equal to $1/I(f)$ can be achieved within each of these classes. Of particular interest in the class of L -estimates is the so-called α -trimmed mean, which estimates θ by first removing the $[\alpha n]$ largest and $[\alpha n]$ smallest samples ($0 \leq \alpha \leq 1/2$), and then computing the sample mean of the remaining sample. The α -trimmed mean can be written as $T_\alpha(F_n)$ where

$$T_\alpha(F) = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} F^{-1}(x) dx, \quad (5.27)$$

where F^{-1} is defined by $F^{-1}(x) = \inf \{y \mid F(y) \geq x\}$. For a symmetric noise distribution F_N the influence curve of the α -trimmed mean is given by

$$IC(x; F_{\theta T_\alpha}) = \frac{l_{F_N, \alpha}(x - \theta)}{1-2\alpha} \quad (5.28)$$

where

$$l_{F_N, \alpha}(x) = \begin{cases} F_N^{-1}(\alpha), & \text{if } x < F_N^{-1}(\alpha) \\ x, & \text{if } F_N^{-1}(\alpha) \leq x \leq F_N^{-1}(1-\alpha) \\ F_N^{-1}(1-\alpha), & \text{if } x > F_N^{-1}(1-\alpha). \end{cases} \quad (5.29)$$

Note the similarity of this influence curve to that of the robust M -estimator for ϵ -contaminated Gaussian noise. In fact, by choosing $\alpha = \int_a^\infty f_R(x) dx$ where f_R is the least-favorable density from (4.21)

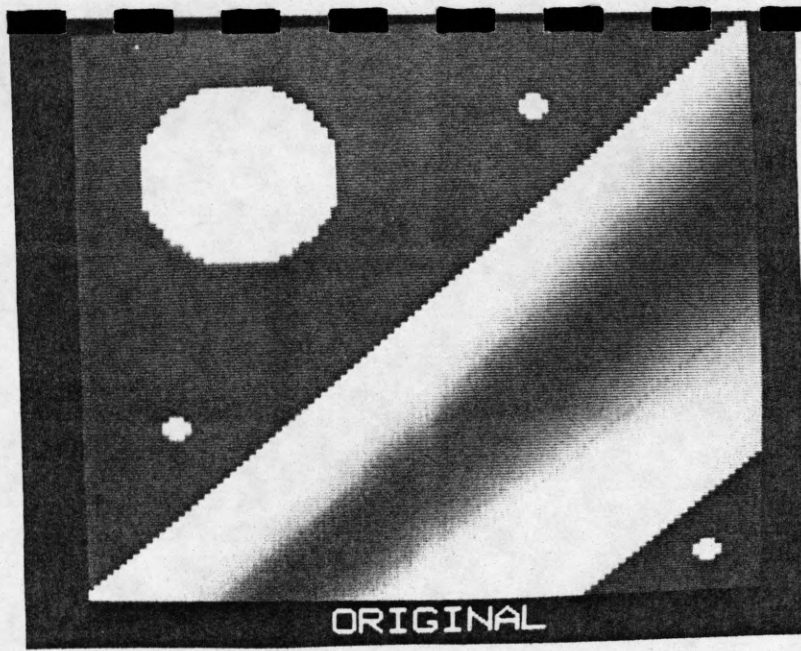
and a is the limiter breakpoint from (4.22)-(4.23), we see that $|IC(x, F_{\theta T_\alpha})| \leq |IC(x, F_{R, \theta T_\alpha})|$

where $F_{R, \theta}(x) = \int_{-\infty}^x f_R(y - \theta) dy$. It follows that this α -trimmed mean is also a minimax robust location estimate for ϵ -contaminated Gaussian noise. Unfortunately, there is no known time-varying

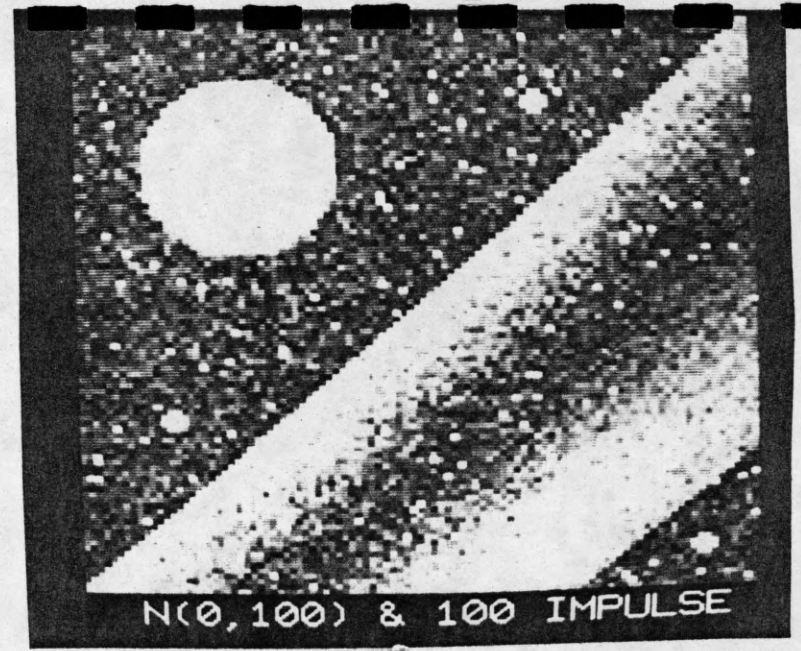
analog to the α -trimmed mean that carries its minimax property.

A different, intuitive, notion of robustness of an estimator is that small changes in the data set x_1, \dots, x_n should not change the estimate much. (This notion is sometimes called *resistance*.) This property is assured for an estimate $T(F_n)$ if the functional T is continuous. Thus one way of defining robustness of an estimator of the type $T(F_n)$ at a given nominal F_0 is in terms of the continuity of T at F_0 . A related robustness notion arises if we view an estimator $\hat{\theta}(x)$ as a mapping from the marginal distribution of the observations X_1, \dots, X_n to the distribution of the estimate $\hat{\theta}(X)$. Robustness can be formulated in this context by requiring this mapping to be continuous in some way; i.e., a small change in the data distribution should cause only a small change in the distribution of the estimate. It turns out that, for estimates of the form $T(F_n)$, these two continuity definitions of robustness are equivalent (within the proper definitions of continuity). The notion of robustness as a continuity property was introduced in the context of parameter estimation by Hampel [5.14] in 1968, although similar ideas were set forth in the context of signal detection in Root's earlier study of stability in detection [5.15]. Robustness of this continuity type is usually known as *qualitative robustness*.

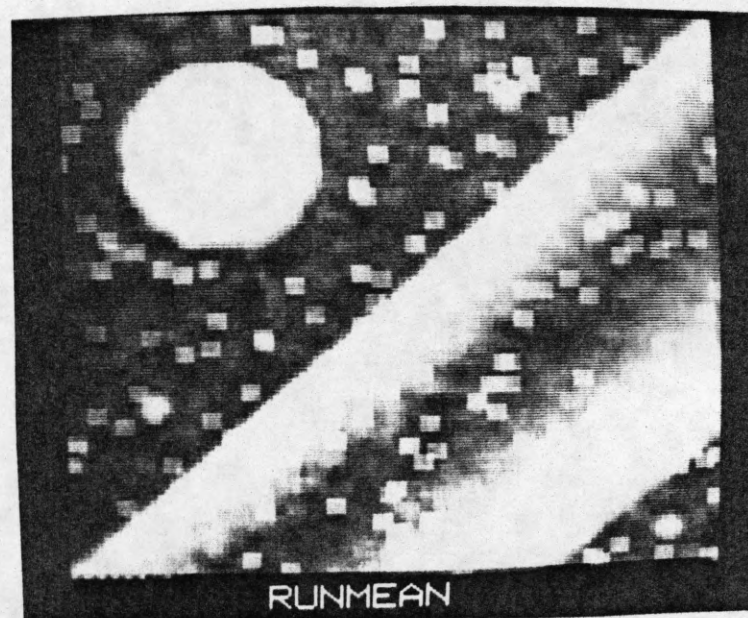
The robust estimators discussed above can be extended to the estimation of signal parameters other than amplitude (see, e.g., Huber [5.1], Kelly [5.16], and El-Sawy [5.17]). Robust estimators have found numerous applications in statistics and signal processing. For example, location M -estimates, α -trimmed means, and modifications thereof have been applied successfully to the problem of image enhancement Bovik, Huang and Munson [5.18] and by Lee and Kassam [5.19-5.20]. This is a natural application for robust methods since image noise typically consists of a Gaussian-like background with occasional impulsive or "spiky" components. For example the image of Fig. 5.3a is shown corrupted by a combination of additive Gaussian and impulsive noises in Fig. 5.3b. The effects of smoothing this image with a running mean (analogous to the sample mean) and with an M -estimate are shown in Figs. 5.3c and 5.3d, respectively. The beneficial effects of robustifying the estimate are quite dramatic in this case.



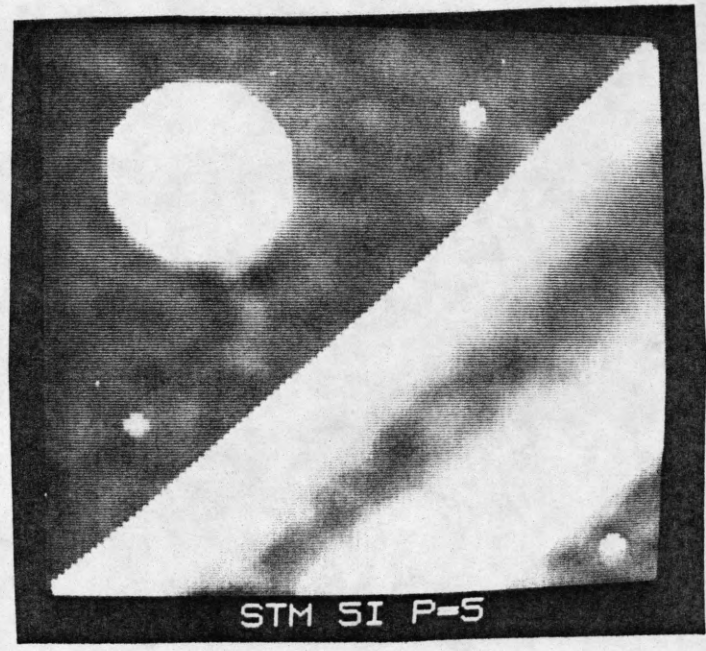
(a)



(b)



(c)



(d)

Fig. 5.3: Application of robust amplitude estimation to image enhancement. (a) Original image, (b) noisy image (Gaussian background noise with impulsive outliers), (c) running-mean processing, and (d) M -estimate processing.

5.2. Robust Nonlinear Recursive Filtering and Identification

One of the most commonly used signal processing algorithms is the Kalman filter, which is based on the linear observation model

$$Y_n = H_n X_n + V_n, \quad n=0,1,2, \dots \quad (5.30)$$

where, for each n , Y_n is an $r \times 1$ observation vector, H_n is the observation matrix, X_n is the $n \times 1$ state vector, and V_n is the observation noise; and on the state model

$$X_{n+1} = F_n X_n + W_n, \quad n=0,1, \dots \quad (5.31)$$

where F_n is the state transition matrix and W_n is the state noise. Assuming that X_0 , $\{W_n\}_{n=0}^{\infty}$, and $\{V_n\}_{n=0}^{\infty}$ are all Gaussian, are independent of one another and that $\{W_n\}_{n=0}^{\infty}$ and $\{V_n\}_{n=0}^{\infty}$ are independent sequences, the optimum (MMSE) estimators of X_n and X_{n+1} given Y_0, \dots, Y_n are given recursively by the well-known Kalman filtering algorithm

$$\hat{X}_{n|n} = \hat{X}_{n|n-1} + M_n H_n^T (H_n M_n H_n^T + R_n)^{-1} (Y_n - H_n \hat{X}_{n|n-1}) \quad (5.32)$$

and

$$\hat{X}_{n+1|n} = F_n \hat{X}_{n|n}, \quad (5.33)$$

where $R_n = \text{cov}(W_n)$, $M_n = F_{n-1} P_{n-1} F_{n-1}^T + Q_{n-1}$, $Q_n = \text{cov}(V_n)$, and $P_n = \text{cov}(\hat{X}_{n|n} - X_n)$. P_n is found recursively from a standard formula [5.21]. The relations (5.32) and (5.33) are called the *measurement update* and *time update*, respectively.

Of course the linearity of (5.32) and (5.33) follows from the assumptions of Gaussian statistics for the state and the observation noise. If either of these quantities has a distribution that deviates from this nominal assumption in that it allows an unexpected number of large observations, then the Kalman estimator may perform poorly. This may be seen from (5.32). In particular, the prediction residual $(Y_n - H_n \hat{X}_{n|n-1})$ will contain outliers if either of the Gaussian assumptions is violated towards a heavier tail behavior. Since this residual is fed directly into the estimate, a distorted value can cause

severe degradation of the estimation performance; and, because of the dynamics, such errors will propagate. However, there is an additional dimension to this problem in that outliers in the state are of interest and should be tracked rather than limited, and so additional considerations arise here. A method for robustification of the Kalman filter against modeling uncertainty has been proposed by Masreliez and Martin in [5.22].

In view of the ideas described in the preceding subsection, a natural way to try to protect against the detrimental effects noted in the above paragraph is to somehow place a limiter or similar nonlinearity in the feedback loop of the filter (5.32)-(5.33). This would limit the effects that any outlier could have on performance, and this in fact is the way robustness is achieved in this model if all uncertainty is in the observation noise distribution. However, because of the vector nature of the Kalman filtering problem, some scaling and transformation of the residual is necessary both in modeling the uncertainty and in processing the residuals for robust estimation. In particular, for p -point uncertainty, the robust version of (5.32)-(5.33) when the prediction errors $X_n - \hat{X}_{n|n-1}$ are Gaussian but the innovations $(Y_n - H_n \hat{X}_{n|n-1})$ have uncertain distribution (this case corresponds to uncertainty in the observation noise) is of the form

$$\hat{X}_{n|n} = \hat{X}_{n|n-1} + M_n H_n^T T_n^T \Psi(\nu_n) \quad (5.34)$$

$$\hat{X}_{n+1|n} = F_n \hat{X}_{n|n} \quad (5.35)$$

with

$$\nu_n = T_n (y_n - H_n \hat{X}_{n|n-1}), \quad (5.36)$$

where T_n is a scaling transformation described in [5.22] and where $\Psi(\nu_n)$ is the vector whose j^{th} component is the j^{th} component of ν_n , $(\nu_{n,j})$, replaced by $\psi_p(\nu_{n,j})$, ψ_p being the influence curve for robust location estimation in the p -point model. The error covariance of this estimator for p -point uncertainty is always less than its value when the components of the transformed residual vector are i.i.d. with the least-favorable marginal distribution. Moreover, this worst-case covariance is the optimum covariance for the case in which the transformed residuals have this least-favorable property. However, this estimate does not provide a saddlepoint because of the constraints of the linear model as

discussed in [5.22].

For the case in which the observation noise is Gaussian but the residual distribution is uncertain due to uncertainty in the state distribution, the Kalman filter is robustified differently than for the case uncertain observation noise. In particular, without going into detail, the measurement update (5.32) still incorporates Y_n linearly into the estimate of $\hat{X}_n |_{n-1}$ but the transformed residual ν_n is limited. The linear incorporation of Y_n is done because an outlier in Y_n indicates an outlier in X_n , which should be incorporated into the estimate of X_n . On the other hand, an outlier in ν_n does not necessarily indicate a bad prediction ($\hat{X}_n |_{n-1}$) and so should not be treated as such.

For further details and simulation results for robust Kalman filtering, the interested reader is referred to [5.22].

A problem related to Kalman filtering is that of system parameter identification. In this problem we have a set of scalar observations Y_1, \dots, Y_n and a set of system input vectors X_1, \dots, X_n that are related through the equation

$$Y_i = s(c, X_i) + N_i, \quad i=1, \dots, n \quad (5.37)$$

where c is a vector parameter and N_1, \dots, N_n is an i.i.d. noise sequence with marginal pdf f . The system identification problem is to estimate c from observation of Y_1, \dots, Y_n and X_1, \dots, X_n . Note that (5.37) is a generalization of the amplitude estimation problem of (5.1) in which X_i , θ , and s_i in (5.1) correspond respectively to Y_i , c , and X_i in (5.37), and $s(c, X_i) = cX_i$.

As in the estimation of signal amplitude, the conventional estimators of c in (5.37) are of the form of M -estimators; namely,

$$\hat{c} = \arg \left\{ \min_c \sum_{i=1}^n L(y_i - s(c, x_i)) \right\}, \quad (5.38)$$

with L an error-weighting function. We get least-squares estimation with $L(x) = x^2$, least-moduli estimation with $L(x) = |x|$, and maximum-likelihood estimation with $L(x) = -\log f(x)$. If the X_i 's are i.i.d. and L is convex, then within mild regularity conditions the estimate \hat{c} of (5.38) is

strongly consistent. Moreover if $s(c, X_i)$ is linear (i.e., $s(c, X_i) = c^T X_i$), then $\sqrt{n}(\hat{c} - c)$ is asymptotically Gaussian with zero mean and covariance matrix

$$\beta^{-1}V(l, f), \quad (5.39)$$

where $\beta = E\{X_i X_i^T\}$ and $V(l, f)$ is from (5.4). Thus, just as in the location estimation case the minimax robust estimate of c in (5.33) for a given noise uncertainty class \mathcal{F} is the optimum M -estimate ($l_R = -f'_R/f_R$) for the least favorable $f_R = \arg\{\min_{f \in \mathcal{F}} I(f)\}$.

As noted in the preceding subsection, since the M -estimator is a batch process, we are often more interested in recursive estimators for on-line applications. In this case (with linear $s(c, X_i)$) the appropriate robust recursive identifier for a class \mathcal{F} is given by

$$\hat{c}_{n+1} = \hat{c}_n - \beta^{-1} \hat{c}_n l_R(y_n - s(\hat{c}_n, x_n)) / nI(f_R), \quad (5.40)$$

which, together with f_R , also gives a saddlepoint for the minimax robustness problem. For further details on this and related topics the reader is referred to the survey by Poljak and Tsympkin [5.4].

The problem of system identification is closely related to the problem of estimating the parameters of a linear time series. An excellent survey of methodology for robust estimation in this context is found in [5.2]. To illustrate the type of results that can be obtained, we mention briefly the problem of estimating the parameters of an autoregression. Specifically, suppose we have observations

$$Y_i = \theta + X_i, \quad i=1, 2, \dots, n, \quad (5.41)$$

where θ is a location parameter and $\{X_i\}$ is a p^{th} -order autoregression

$$X_i = \phi_1 X_{i-1} + \phi_2 X_{i-2} + \dots + \phi_p X_{i-p} + \epsilon_i, \quad (5.42)$$

where $\phi = (\phi_1, \dots, \phi_p)^T$ is a vector of constants, and $\{\epsilon_i\}$ is an i.i.d. innovations sequence with marginal pdf f . M -estimates of ϕ and θ can be computed by finding

$$(\hat{\gamma}, \hat{\phi}) = \arg \left\{ \min_{\gamma, \phi} \sum_{i=p+1}^n L \left(\frac{Y_i - \gamma - \phi_1 Y_{i-1} - \dots - \phi_p Y_{i-p}}{\hat{s}} \right) \right\}$$

where $\gamma \triangleq \theta / (1 - \phi_1 - \dots - \phi_p)$ and \hat{s} is an estimate of the innovations scale parameter (see [5.2]), and then setting $\hat{\theta} = \hat{\gamma}(1 - \hat{\phi}_1 - \dots - \hat{\phi}_p)$. These estimates of θ and ϕ are consistent and are asymptotically Gaussian and independent with

$$n \text{ Var}(\hat{\theta}) \sim V(l, f) / (1 - \phi_1 - \dots - \phi_p)^2 \quad (5.43)$$

and

$$n \text{ cov}(\hat{\phi}) \sim V(l, f) \hat{C}^{-1} / \sigma_\epsilon^2, \quad (5.44)$$

where $V(l, f)$ is the asymptotic variance expression for M -estimation of location from (5.4), σ_ϵ^2 is the variance of the innovations, and \hat{C} is the $p \times p$ covariance matrix of the $\{X_i\}$ process when $\sigma_\epsilon^2 = 1$.

Note from (5.43)-(5.44) that $l_R = -f'_R / f_R$ and $f_R = \arg \{ \min_{f \in \mathcal{F}} I(f) \}$ gives a saddlepoint for the minimax robust estimation of θ when the innovations distribution is uncertain. However, this is not necessarily so for robust estimation of ϕ because of the σ_ϵ^2 term in (5.44). It is interesting to note from (5.44) that a heavy-tailed innovations distribution may actually be beneficial in estimating ϕ , since $\text{cov}(\hat{\phi})$ depends inversely on σ_ϵ^2 . This is not surprising if we note that outliers in the innovations should actually aid in the identification of ϕ just as the insertion of an impulse into a system allows one identifying its impulse response. Thus, we have here a situation in which impulsive phenomena are beneficial to inferences.

Extensions of the above ideas to problems of estimation in ARMA models, nonlinear models, models with additive observation noise (in which outliers are again detrimental), and models of unknown order are discussed in [5.2]. Similar ideas can also be applied to other problems of time series analysis such as forecasting [5.2] and spectrum estimation [5.23]. Connections between the problems of robust Kalman filtering and robust time-series regression have been noted in a recent paper by Boncelet and Dickinson [5.24].

6. ROBUST DATA QUANTIZATION

Data quantization is a necessary function of systems which digitally process or transmit signals. The optimum design of quantizers based on minimum-distortion criteria has been considered extensively over the past three decades and a number of references on this subject can be found in surveys by Morris [6.1] and Gersho [6.2] and a recent special issue of the *IEEE Transactions on Information Theory* (Gray [6.3]). More recently, designs for quantizers which are optimum for signal detection or estimation purposes have also been developed by [6.4]-[6.7].

Optimum quantizer design is based primarily on statistical definitions of optimality, such as minimum mean-distortion [6.8] or maximum divergence [6.6]. Thus, the optimum designs usually require an accurate statistical model for the data to be quantized. Since such models are rarely exact, the study of quantizer design for inaccurate models is of practical interest. One approach to this problem is that of adaptive quantization (see, for example, [6.9-6.11]), which is appropriate for very inaccurate models or for situations in which data statistics are changing significantly over moderate time periods. An alternate approach which is primarily of interest when there are relatively small inaccuracies in the statistical model or when a fixed structure is preferred, is a game theoretic one in which a quantizer with best worst-case performance is sought. As demonstrated in the preceding sections, this general approach has also been applied successfully to many other problems of signal processing with inaccurate statistical models, and the resulting designs are usually robust. In this section, we consider the problem of minimax design of quantizers for imprecisely modeled data. In particular, we survey several recent results pertaining to the problem of minimax distortion quantization. These results include both asymptotic (as the number of quantization levels becomes infinite) and nonasymptotic treatments of this problem.

6.1. Robust Quantization for a Small Number of Levels

An M -level quantizer Q can be represented by a set of M output levels q_1, q_2, \dots, q_M and a set of $(M+1)$ input breakpoints t_0, t_1, \dots, t_M , satisfying $-\infty = t_0 < t_1 < \dots < t_{M-1} < t_M = +\infty$, where the quantized value of a real input x is given by

$$Q(x) = q_k, \text{ if } x \in (t_{k-1}, t_k], k = 1, \dots, M. \quad (6.1)$$

(See Fig. 6.1.) Thus, the design of an M -level quantizer is an optimization problem on \mathbb{R}^{2M-1} . The most common quantizer design criterion for a random input X is to choose the quantizer parameters to minimize a mean-distortion quantity,

$$E \{D[X, Q(X)]\}, \quad (6.2)$$

where $D[;]$ is some appropriate measure of distortion. Usually $D[;]$ is a difference distortion measure (i.e., it depends only on the difference, $|X - Q(X)|$), and the most useful of these are the p^{th} -difference distortion measures given by $D[a, b] = |a - b|^p$.

The study of minimum-distortion quantization is exemplified by the classical work of Max [6.7], in which design conditions for the t_k 's and q_k 's of minimum-distortion quantizers are derived. In particular, for a wide class of difference distortion measures (including the p^{th} -difference ones), Max shows that the breakpoints should be chosen to satisfy

$$t_k = \frac{q_k + q_{k+1}}{2}, k = 1, \dots, M-1, \quad (6.3)$$

and he also gives a second set of necessary conditions which, together with (6.3), give a set of nonlinear equations to be solved for the optimum quantizer parameters. For example, if we consider mean-square distortion ($D[a, b] = |a - b|^2$) and if X has a probability density function f , then the necessary conditions are given by

$$q_k = \frac{\int_{t_{k-1}}^{t_k} x f(x) dx}{\int_{t_{k-1}}^{t_k} f(x) dx}, k = 1, \dots, M; \quad (6.4)$$

i.e., q_k is the centroid of $(t_{k-1}, t_k]$ weighted with f . (An equivalent interpretation is that $q_k = E\{X | X \in (t_{k-1}, t_k]\}$). Many important refinements and generalizations of this theory as well as

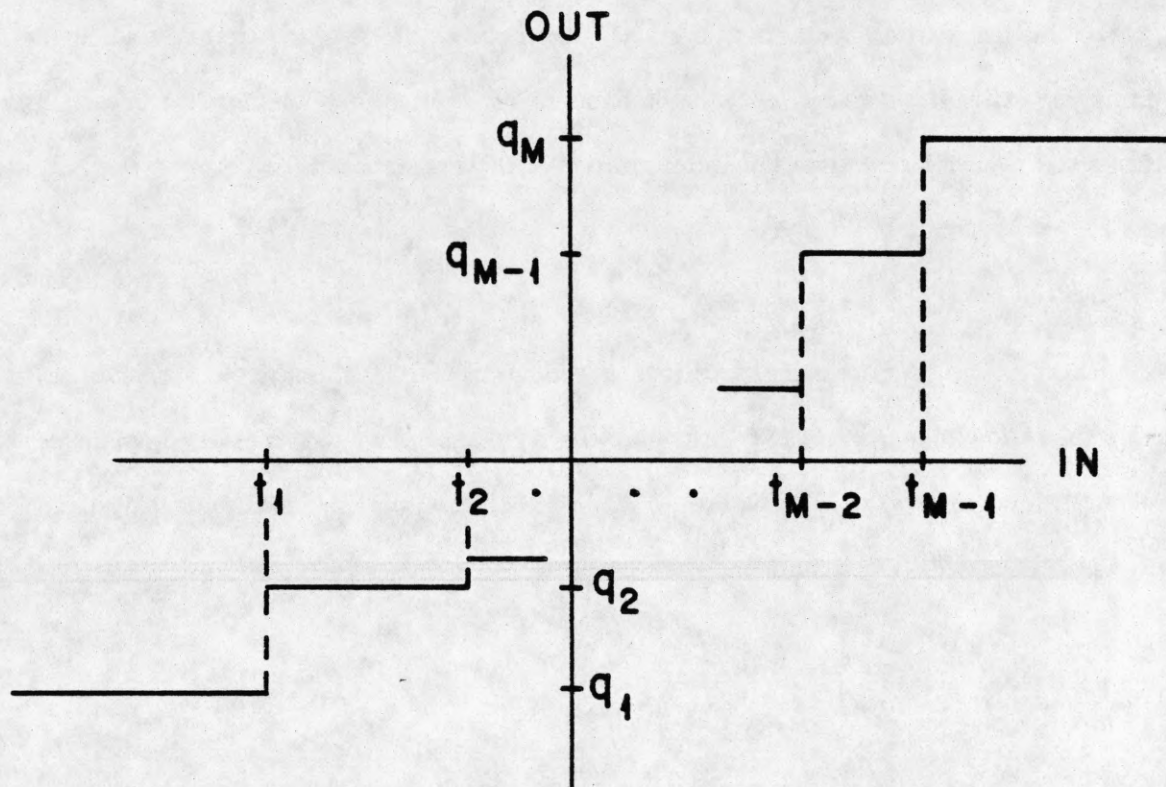


Fig. 6.1: Input/output characteristics of an M -level quantizer.

studies of performance, approximately optimal design, etc., can be found in the literature, and again the reader is referred to [6.1], [6.2], and [6.3] for further discussion.

Of course, designs based on minimizing the quantity of (6.2) will depend on knowledge of the probability distribution of X , as can be seen, for example, in (6.4). Thus, as discussed in the introduction, when this distribution is not known exactly it is necessary to seek an alternative design strategy to minimizing (6.2). If we assume that the probability distribution of X lies in some uncertainty class \mathcal{F} , then a useful design objective is the minimization of the alternate quantity

$$\sup_{\mathcal{F}} E \{D[X, Q(X)]\}. \quad (6.5)$$

Several recent studies have considered various aspects of the minimization of (6.5) and these are discussed in the following paragraphs. For the purposes of discussion, we will consider only the particular case of mean-square distortion, although all of the cited studies consider more general distortion measures as well. Thus, we will be considering the problem

$$\min_{Q \in \mathcal{Q}_M} \sup_{f \in \mathcal{F}} \int_{-\infty}^{\infty} |x - Q(x)|^2 dF(x), \quad (6.6)$$

where \mathcal{Q}_M denotes the class of all M -level quantizers.

The first study to consider (6.6) in the context of minimax robustness is that of Morris and Vandelinde [6.12] in which \mathcal{F} is taken to be the class of all possible probability distributions on a fixed interval $[-V, V]$. In this case, it is shown in [6.12] that the minimax quantization problem (6.6) is solved by the M -level uniform quantizer on $[-V, V]$, which is given by

$$\begin{aligned} t_k &= -V + 2kV/M, & k &= 1, \dots, M-1 \\ q_k &= t_k - V/M, & k &= 1, \dots, M-1 \end{aligned} \quad (6.7)$$

and

$$q_M = V - V/M.$$

The solution to (6.6) was considered next by Bath and VandeLinde in [6.13], in which the class \mathcal{F} is taken to be a unimodal generalized moment constrained (UGMC) set; i.e., \mathcal{F} is assumed to consist of the distribution functions which are unimodal (with mode zero) and which satisfy the generalized moment constraint

$$\int_{-\infty}^{\infty} \rho(x) dF(x) \leq c, \quad (6.8)$$

where the constraint function ρ is symmetric, continuous, is strictly increasing on $(0, \infty)$, and satisfies $\rho(0) = 0$ and $\rho(x) \rightarrow \infty$ as $x \rightarrow \infty$. In particular, it is shown in [6.13] via the Lagrange duality theorem that the minimax quantizer for this problem is given by the solution to

$$\min_{Q \in \mathcal{Q}_M} \left\{ \begin{array}{l} \min \\ \lambda_1, \lambda_2 \leq 0 \\ H(x) \leq 0 \end{array} \quad \{ \lambda_1 + \lambda_2 c \} \right\} \quad (6.9)$$

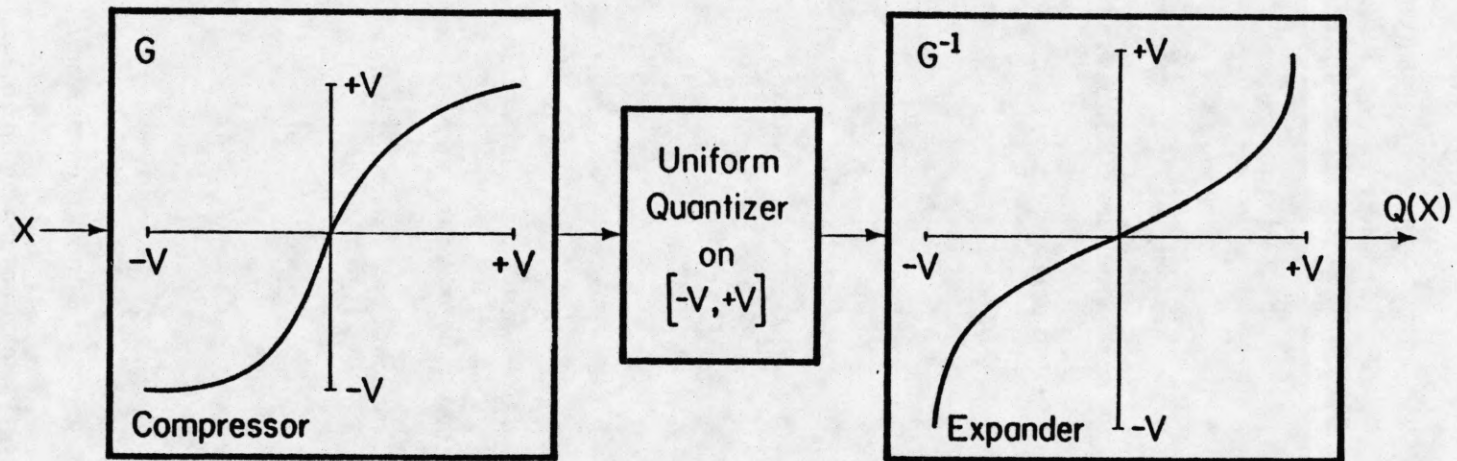
where c is from (6.8) and

$$H(x) = \int_0^x [|y - Q(y)|^2 - \lambda_1 - \lambda_2 \rho(y)] dy. \quad (6.10)$$

An $O(M)$ algorithm for solving (6.9) is also given in [6.13], and it is demonstrated numerically that the resulting minimax quantizer can perform much better in the worst-case over \mathcal{F} than both the uniform quantizer and the quantizer which is optimum for Gaussian data.

6.2. Asymptotic Robust Quantization

Further work on the minimax-distortion quantization problem considers the asymptotic case as the number, M , of quantization levels becomes infinite. Assuming the data is confined to an interval $[-V, V]$, it is convenient to represent (and to implement) a quantizer as an increasing invertible function $G : [-V, V] \rightarrow [-V, V]$, followed by a uniform quantizer, which is followed in turn by the inverse of G . (See Fig. 6.2.) The function G is termed a *compressor* and its inverse an *expander*, so



FP-8400

Fig. 6.2: Configuration for companding quantization.

that the whole scheme is termed *companding*. In general, the compressor G can be any invertible function; however, it is usually assumed to have a continuous derivative, g . It is also common to assume, for simplicity, that the data distribution F has a density f . Under these and further mild regularity conditions it can be shown that the mean-squared error associated with the companding scheme is asymptotically of the form $D(f, g) \cdot M^{-2}$, where the functional $D(f, g)$ is given by

$$D(f, g) = \frac{V^2}{3} \int_{-V}^V f(x) [g(x)]^{-2} dx. \quad (6.11)$$

The function g describes the relative density of the quantization intervals within the range of the data.

By way of (6.11), an asymptotically optimum compressor curve G can be chosen by minimizing over g . Straightforward minimization (see, e.g., Gersho [6.2]) yields that the minimizing compressor is given by

$$G_0(x) = 2V \left[\frac{\int_{-V}^x f^{1/3}(y) dy}{\int_{-V}^V f^{1/3}(y) dy - \frac{1}{2}} \right], \quad (6.12)$$

which yields a value of (6.11) of

$$D(f, g_0) = \frac{1}{12} \left(\int_{-V}^V f^{1/3}(x) dx \right)^3, \quad (6.13)$$

where $g_0 = G_0'$. Since it is known that (6.13) gives the limiting distortion of the minimum-mean-square-error quantizer (see, for example, Bucklew and Wise [6.14]), it follows that the companding structure causes no performance loss for optimum quantization in the asymptotic case.

Note that, as one would expect, the optimum compressor characteristic of (6.12) depends on an exact knowledge of the probability distribution of the data. Thus if, as above, F is known only to belong to some class \mathcal{F} of possible data distributions, then it is reasonable to replace the problem of

minimizing (6.11) over g with that of minimizing over g the worst-case over \mathcal{F} of (6.11). That is, it is of interest to consider the problem

$$\min_{g \in S} \{ \sup_{f \in \mathcal{P}} D(f, g) \}, \quad (6.14)$$

where $\mathcal{P} = \{ f \mid f = F', f \in \mathcal{F} \}$ and where S is a set of admissible compressor curve derivatives.

The problem in (6.14) was first considered by Bath and VandeLinde in [6.15], where the case in which \mathcal{F} is a UGMC set as in (6.8) is treated. It follows from [6.15] that the minimax compressor for this case is given by

$$G_1^*(x) = \int_{-V}^x g_1^*(y) dy - V, \quad (6.15)$$

where

$$g_1^*(x) = (V/3^{1/2})[\lambda_1^* + \lambda^{2*} \rho(x)]^{-1/2} \quad (6.16)$$

with λ_1^* and λ^{2*} chosen to solve

$$\min_{\lambda_1, \lambda_2 \geq 0} \{ \lambda_1 + \lambda_2 c \}$$

and

$$\int_{-V}^V g_1^*(x) dx = 2V. \quad (6.17)$$

Here, the constant c and the function ρ are from (6.8). It is noted in [6.15] that the case $\rho(x) = x^2$, the solution to (6.17) involves only finding the root of a simple transcendental equation. It is also shown in [6.15] that the worst-case performance over the UGMC set \mathcal{F} of the minimax compander is much better than that of the corresponding optimum companders for the Laplace, uniform, and Gaussian distributions. However, surprisingly, it is found that "robustified" versions of the classical A-law and μ -law companders (see, for example, Cattermole [6.16]) perform nearly as well in their worst case as does the minimax compander.

A slightly different approach to solving the robust companding problem is proposed by Kazakos in [6.17] and [6.18]. In particular, [6.17] and [6.18] consider uncertainty classes \mathcal{F} for which saddlepoint solutions to (6.14) can easily be demonstrated. Note that $D(f, g)$ is linear (and hence concave) in f and is convex in g . In [6.17] the following uncertainty class is considered (here V is taken to be equal to 1):

$$\begin{aligned} \mathcal{F}_2 = \{ & F \mid F(x_{k+1}) - F(x_k) = p_k, k = 0, 1, \dots, N-1, \\ & F(x) + F(-x) = 1, 0 = x_0 < x_1 < \dots < x_N = 1 \}, \end{aligned} \quad (6.18)$$

where N , the x_k 's, and the p_k 's are fixed and known with $\sum_{k=0}^{N-1} p_k = 1$. This class is an example of the p -point class discussed in previous sections. Kazakos shows that the member of (6.18) which has the uniform density on each of the intervals $(x_{k-1}, x_k]$ and its corresponding optimum compressor from (6.12) form a saddlepoint for (6.14) in this case. That is, with

$$f_2^*(x) = p_k / (x_{k+1} - x_k), x \in (x_k, x_{k+1}], k = 0, 1, \dots, N-1, \quad (6.19)$$

and g_2^* from (6.12) with $f = f_2^*$, we have

$$D(f, g_2^*) \leq D(f_2^*, g_2^*) \leq D(f_2^*, g), \quad (6.20)$$

for all $g \in \mathcal{G}$. Note that the compressor curve G_2^* corresponding to this g_2^* will be piecewise linear. Also note that, of course, the existence of (f_2^*, g_2^*) satisfying (6.20) is equivalent to the condition

$$\min_{f \in \mathcal{P}_2} \max_{g \in \mathcal{G}} D(f, g) = \max_{g \in \mathcal{G}} \min_{f \in \mathcal{P}_2} D(f, g). \quad (6.21)$$

In [6.18], the saddlepoint properties of (6.14) are considered in a more general setting. In particular, it is noted in [6.18] that within regularity conditions on \mathcal{P} and \mathcal{G} , the equality of (6.21) holds for general convex classes \mathcal{P} . Thus, in view of (6.13), saddlepoint solutions to (6.14) can be sought by looking for solutions to the alternate problem

$$\max_{f \in \mathcal{P}} J(f), \quad (6.22)$$

where

$$J(f) = \int_{-V}^V f^{1/3}(x) dx. \quad (6.23)$$

That is, under mild conditions, a solution f^* to (6.22) and its corresponding optimum compressor characteristic g^* from (6.12) form a saddle point for (6.14). As was noted in [6.19], since $J(f)$ is of the form $\int_{-V}^V C(f(x))dx$ with C concave, solutions to (5.22) are those "closest" to the uniform density on $[-V, V]$. Thus, solutions for several uncertainty classes are straightforward. For example, with

$$\mathcal{P}_3 = \{f \mid f = (1-\epsilon)f_0 + \epsilon h\}, \quad (6.24)$$

where f_0 and ϵ are fixed and h is arbitrary, the density solving (6.22) is given by

$$f_3^*(x) = \max\{(1-\epsilon)f_0(x), m\}, \quad (6.25)$$

where m is a constant chosen so that f_3^* integrates to unity.

Unfortunately, as pointed out in [6.20] the asymptotic formulation of [6.17]-[6.18] is flawed by the fact that, for some models (e.g., ϵ -contaminated data), the minimum of the maximum asymptotic error is different from the minimum of asymptotic maximum error due to the discrete nature of the finite- M problem. (This problem does not occur for the UGMC model of Bath and VandeLinde.) This problem is corrected in [6.20], and the solution for ϵ -contaminated data is still of the form of (6.25) but with a higher degree of limiting. A typical solution is shown in Fig. 6.3. Note that, for this case, the levels of the robust quantizer are distributed more or less as those of the nominal quantizer near the center of the range but they are spaced uniformly (and closer together than for the nominal quantizer) near the ends of the range. Thus, the robust quantizer is a mixture of the nominally optimum quantizer and the uniform quantizer (which is universally minimax). Numerical results illustrating the effectiveness of the robust quantizer are given in [6.20].

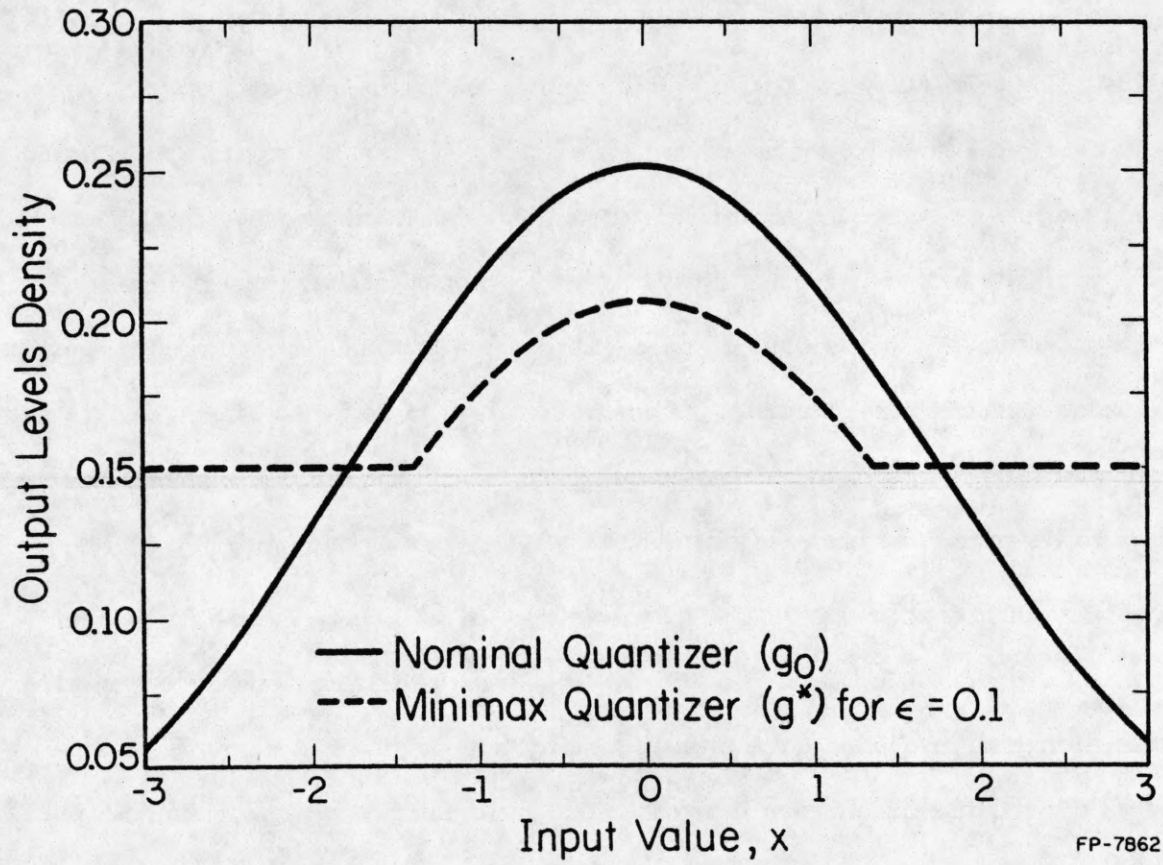


Fig. 6.3: Nominal and robust level densities for ϵ -contaminated truncated-Gaussian data ($\epsilon = 0.1$).

7. CONCLUSIONS

In this paper we have considered minimax robustness in the context of the signal processing tasks of estimation, detection, and data quantization. We have seen that these robustness formulations take two basic forms: robustness with respect to uncertain second-order statistical properties (e.g., spectral properties) of signals or noise and robustness with respect to uncertainty in the marginal distribution of the noise or signal process. In the first of these cases, the robust methods take the form of the linear procedures discussed in Sections 2 and 3, whereas in the second case nonlinear procedures are called for. In either case the typical robustification procedure has the effect of lowering the sensitivity of a nominally optimum procedure by tempering those characteristics that are accentuated by the nominal model. Thus, in a nominally Gaussian noise model with a small fraction of "outliers", enough limiting is introduced into the optimum procedure to keep these outliers from destroying the action of the procedure. The frequency domain robustness methods can also be thought of in the same way in which the gain of the appropriate filter is reduced in certain spectral regions to limit the effects of a more than expected amount of energy in those regions (i.e., "spectral outliers").

The relationship of these two types of robustness has been discussed by Franke and Poor in [7.1] in the context of estimation. In [7.1] it is noted that, if one knows only spectral or other second-order properties, then linear estimation procedures are globally minimax over all estimation schemes. It is only when distributional information is provided that nonlinearity arises in the minimax solutions. Thus, the robust filters of Section 2 are globally minimax over all filters (linear or nonlinear) and all random processes (Gaussian or otherwise) with the given spectral properties. Note that the relevant observation or noise processes in Section 5 have very special spectral characteristics (they are usually white), and only their marginal distribution is allowed to vary. Very little progress has been made in dealing with simultaneous spectral and distributional uncertainty models (an exception is a recent paper by Moustakides and Thomas [7.2]), although this certainly gives rise to an interesting class of problems from a practical viewpoint.

REFERENCES

- [1.1] H. L. Van Trees, *Detection, Estimation and Modulation Theory, I.* New York: Wiley, 1968.
- [1.2] W. L. Root, "Stability in Signal Detection Problems," *Proc. Symposia in Applied Mathematics*, vol. 16, American Math. Soc., 1964.
- [1.3] F. R. Hampel, "A General Qualitative Definition of Robustness," *Ann. Math. Stat.*, vol. 42, pp. 1887-1896, 1971.
- [1.4] Ya. Z. Tsytkin, "Adaptive Optimization Algorithms When There Is A Priori Uncertainty," *Automation and Remote Control*, vol. 40, no. 6, pp. 857-868, June 1979.
- [1.5] J. W. Tukey, "A Survey of Sampling from Contaminated Distributions," in *Contributions to Probability and Statistics*, I. Olkin, et al., Eds. Stanford University Press: Stanford, CA, 1960.
- [1.6] P. J. Huber, "Robust Estimation of a Location Parameter," *Ann. Math. Stat.*, vol. 35, pp. 73-104, 1964.
- [1.7] P. J. Huber, "A Robust Version of the Probability Ratio Test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.
- [1.8] L. A. Zadeh, "General Filters for Separation of Signals and Noise," *Proc. Symp. on Information Networks*, Poly. Inst. of Brooklyn, pp. 31-49, April 1954.
- [1.9] W. L. Root, "Communications Through Unspecified Additive Noise," *Inform. and Control*, vol. 4, pp. 15-29, March 1961.
- [1.10] W. L. Root, "Some Notes on Jamming," *MIT Lincoln Lab. Tech. Rept. 103*, 1956.
- [1.11] M. C. Yovits and J. L. Jackson, "Linear Filter Optimization with Game Theory Considerations," *IRE National Conv. Record*, pt. 4, pp. 193-199, 1955.
- [1.12] N. J. Nilsson, "An Application of the Theory of Games to Radar Reception Problems," *IRE National Conv. Record*, pt. 4, pp. 130-140, March 1959.
- [1.13] L.-H. Zetterberg, "Signal Detection Under Noise Interference in a Game Situation," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 47-57, September 1962.
- [1.14] N. M. Blachman, "Communication as a Game," *IRE Wescon Conv. Rec.*, pt. 2, pp. 61-66, August 1957.
- [1.15] R. L. Dobrushin, "Optimum Information Transmission Through a Channel with Unknown Parameters," *Radio Eng. and Electronics*, vol. 4, no. 12, pp. 1-8, 1959.
- [1.16] M. Yu. Gadzhiev, "Determination of the Optimum Variation Mode of the Useful Signal and Noise Carrier Frequencies in Detection Problems Based on the Theory of Games," *Automation and Remote Control*, vol. 22, no. 1, pp. 31-39, 1961.
- [1.17] G. E. P. Box, "Non-Normality and Tests on Variances," *Biometrika*, vol. 4D, pp. 318-335, 1953.
- [1.18] P. J. Huber, "Episodes from the Early History of Robustness," *Abstracts of Papers: 1982 IEEE International Symp. on Information Theory*, Les Arcs, France, p. 72, June 1982.
- [1.19] P. J. Huber, "Robust Statistics: A Review," *Ann. Math. Stat.*, vol. 43, p. 1041, 1972.
- [1.20] P. J. Huber, *Robust Statistical Procedures*. SIAM Press: Philadelphia, 1977.
- [1.21] F. R. Hampel, "Robust Estimation: A Condensed Partial Survey," *Z. Wahr. verw. Geb.*, vol. 27, pp. 87-104, 1973.
- [1.22] P. J. Bickel, "Another Look at Robustness: A Review of Reviews and Some New Developments," *Scand. J. Statist.*, vol. 3, pp. 145-168, 1976.
- [1.23] R. V. Hogg, "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," *J. Amer. Stat. Assoc.*, vol. 69, pp. 909-927, 1974.

- [1.24] R. V. Hogg, "Statistical Robustness: One View of its Use in Applications Today," *The American Statistician*, vol. 33, pp. 108-115, August 1979.
- [1.25] A. A. Ershov, "Stable Methods of Estimating Parameters (Survey)," *Automation and Remote Control*, vol. 39, no. 7, pt. 2, pp. 1152-2181, July 1978.
- [1.26] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location: Survey and Advances*. Princeton University Press: Princeton, NJ, 1972.
- [1.27] R. L. Launer and G. N. Wilkinson, Eds., *Robustness in Statistics*. Academic: New York, 1979.
- [1.28] P. J. Huber, *Robust Statistics*. Wiley: New York, 1981.
- [1.29] S. S. Wolff and J. L. Gastwirth, "Robust Two-Input Correlators," *J. Acoust. Soc. Am.*, vol. 41, pp. 1212-1219, May 1967.
- [1.30] R. D. Martin and S. C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50-56, 1971.
- [1.31] S. A. Kassam and H. V. Poor, "Robust Signal Processing for Communication Systems," *IEEE Commun. Magazine*, vol. 21, pp. 20-28, January 1983.
- [1.32] V. D. VandeLinde, "Robust Techniques in Communication," in *Robustness in Statistics*, R. L. Launer and G. N. Wilkinson, Eds., Academic Press: New York, 1979, pp. 177-200.
- [1.33] V. M. Krasnenker, "Stable (Robust) Detection Methods for Signals Against a Noise Background (Survey)," *Automation and Remote Control (Engl. transl.)*, vol. 41, no. 5, pt. 1, pp. 640-659, May 1980.
- [1.34] H. V. Poor, "Robustness in Signal Detection," in *Advances in Communications and Networks*, I. F. Blake and H. V. Poor, Eds. Springer-Verlag: New York, 1965 (to appear).
- [1.35] B. Kleiner, R. D. Martin and D. J. Thomson, "Robust Estimation of Power Spectra," *J. Roy. Stat. Soc., Series B*, vol. 41, no. 3, pp. 313-351, 1979.
- [1.36] R. D. Martin and D. J. Thomson, "Robust-Resistant Spectrum Estimation," *Proc. IEEE*, vol. 70, pp. 1097-1115, September 1982.
- [1.37] R. D. Martin, "Robust Methods for Time Series," in *Applied Time Series II*, D. F. Findley, Ed. Academic Press: New York, 1981.
- [1.38] B. T. Poljak and Ya. Z. Tsympkin, "Robust Identification," *Automatica*, vol. 16, p. 53-63, 1980.
- [2.1] J. B. Thomas, *An Introduction to Statistical Communication Theory*. Wiley: New York, 1968.
- [2.2] K. S. Vastola and H. V. Poor, "An Analysis of the Effects of Spectral Uncertainty on Wiener Filtering," *Automatica*, vol. 28, pp. 289-293, 1983.
- [2.3] T. L. Marzetta and S. W. Lang, "Power Spectral Density Bounds," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 117-122, January 1984.
- [2.4] S. A. Kassam and T. L. Lim, "Robust Wiener Filters," *J. Franklin Inst.*, vol. 304, pp. 171-185, 1977.
- [2.5] H. V. Poor, "On Robust Wiener Filtering," *IEEE Trans. Automatic Control*, vol. AC-25, pp. 531-536, 1980.
- [2.6] S. M. Ali and S. D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another," *J. Roy. Stat. Soc., Series B*, vol. 28, pp. 131-142, 1966.
- [2.7] I. Csiszar, "Information-Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 229-318, 1967.
- [2.8] P. J. Huber, "A Robust Version of the Probability Ratio Test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.

- [2.9] H. V. Poor, "Minimax Linear Smoothing for Capacities," *Ann. Prob.*, vol. 10, pp. 504-507, 1982.
- [2.10] H. V. Poor and D. P. Looze, "Minimax State Estimation for Linear Stochastic Systems with Noise Uncertainty," *IEEE Trans. Autom. Control*, vol. AC-26, pp. 902-906, 1981.
- [2.11] L. J. Cimini and S. A. Kassam, "Robust and Quantized Wiener Filters for p-point Spectral Classes," *Proc. 1980 Conf. on Inform. Sciences and Systems*, Princeton University, pp. 314-318, March 1980.
- [2.12] K. S. Vastola and H. V. Poor, "On Generalized Band Models in Robust Detection and Filtering," *Proc. 1980 Conf. on Inform. Sciences and Systems*, Princeton University, pp. 1-5, March 1980.
- [2.13] L. J. Cimini and S. A. Kassam, "Optimum Piecewise Constant Wiener Filters," *J. Optical Soc. Am.*, vol. 77, pp. 1162-1171, October 1981.
- [2.14] L. Brieman, "A Note on Minimax Filtering," *Ann. Prob.*, vol. 1, pp. 175-179, 1973.
- [2.15] S. A. Kassam, "The Bounded p-point Classes in Robust Hypothesis Testing and Filtering," *Proc. 20th Annual Allerton Conf. on Comm., Control and Computing*, Monticello, IL, pp. 526-534, October 6-8, 1982.
- [2.16] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov Theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316-327, March 1984.
- [2.17] E. Wong, *Stochastic Processes in Information and Dynamical Systems*. Wiley: New York, 1971.
- [2.18] J. Franke, "A General Approach to the Robust Prediction of Discrete Time Series," SFB123 "Stochestische Mathematische Modelle," Preprint No. 74, Univ. of Heidelberg.
- [2.19] K. S. Vastola, "On Robust Linear Caused Estimation of Continuous-Time Signals," *Abstracts of Papers: 1982 IEEE International Symp. on Inform. Theory*, Les Arcs, France, p. 124, June 21-25, 1982.
- [2.20] J. Franke and H. V. Poor "Minimax Robust Filtering and Finite-Length Robust Predictors," in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Hardle, and R. D. Martin, Eds. Springer-Verlag: Heidelberg, 1984.
- [2.21] K. Yao, "An Alternative Approach to the Linear Causal Least-Square Filtering Theory," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 232-240, 1971.
- [2.22] K. Yao, "On the Direct Calculation of MMSE of Linear Realizable Estimator by Toeplitz Form Method," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 95-97, 1971.
- [2.23] J. Snyders, "Error Formulae for Optimal Linear Filtering, Prediction and Interpolation of Stationary Time Series," *Ann. Math. Statist.*, vol. 43, pp. 1935-1943, 1972.
- [2.24] J. Snyders, "Error Expressions for Optimal Linear Filtering of Stationary Processes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 574-582, 1972.
- [2.25] K. Hoffman, *Banach Spaces of Analytic Functions*. Prentice-Hall: Englewood Cliffs, NJ, 1962.
- [2.26] Y. Hosoya, "Robust Linear Extrapolations of Second-Order Stationary Processes," *Ann. Prob.*, vol. 6, pp. 574-584, 1978.
- [2.27] H. V. Poor, "The Rate-Distortion Function on Classes of Sources Determined by Spectral Capacities," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 19-26, 1982.
- [2.28] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day: San Francisco, 1964.
- [2.29] A. Papoulis, "Maximum Entropy and Spectrum Estimation: A Review," *IEEE Trans. Acoust. Speech and Signal Processing*, vol. ASSP-29, pp. 1176-1186, December 1981.

- [2.30] G. Moustakides and S. A. Kassam, "Minimax Robust Equalization for Random Signals through Uncertain Channels," *Proc. 20th Annual Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, pp. 945-954, October 1982.
- [2.31] G. V. Moustakides and S. A. Kassam, "Robust Wiener Filters for Random Signals in Correlated Noise," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 614-619, July 1983.
- [2.32] G. Moustakides and S. A. Kassam, "Robust Wiener Filters for Random Signals in Correlated Noise," *IEEE Trans. Inform. Theory*, vol. IT-29, 1983.
- [2.33] P. J. Huber and V. Strassen, "Minimax Tests and the Neyman-Pearson Lemma for Capacities," *Ann. Statist.*, vol. 1, pp. 251-263, 1973.
- [2.34] H. Rieder, "Least Favorable Pairs for Special Capacities," *Ann. Statist.*, vol. 5, pp. 909-921, 1977.
- [2.35] T. Bednarski, "On Solutions of Minimax Test Problems for Special Capacities," *Z. Wahr. verw. Geb.*, vol. 58, pp. 397-405, 1981.
- [2.36] K. S. Vastola and H. V. Poor, "On the p-point Uncertainty Class," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 374-376, March 1984.
- [2.37] M. Taniguchi, "Robust Regression and Interpolation for Time Series," *J. Time Series Analysis*, vol. 2, pp. 53-62, 1981.
- [2.38] S. A. Kassam, "Robust Hypothesis Testing on Robust Time Series Interpolation and Regression," *J. Time Series Analysis*, vol. 3, pp. 185-194, 1982.
- [2.39] C. T. Chen and S. A. Kassam, "Robust Wiener Filtering for Multiple Inputs with Channel Distortion," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 674-677, July 1984.
- [2.40] J. C. Darragh and D. P. Looze, "Noncausal Minimax Linear State Estimation for Systems with Uncertain Second-Order Statistics," *IEEE Trans. Autom. Control*, vol. AC-29, pp. 555-557, June 1984.
- [2.41] H. Tsaknakis and P. Papantoni-Kazakos, "Robust Linear Filtering for Multivariable Stationary Time Series," *Proc. 1984 Conf. on Inform. Sciences Syst.*, Princeton University, Princeton, NJ, March 14-16, 1984.
- [2.42] J. A. D'Appolito and C. E. Hutchinson, "A Minimax Approach to the Design of Low Sensitivity State Estimators," *Automatica*, vol. 18, pp. 599-608, 1972.
- [2.43] V. D. VandeLinde, "Robust Properties of Solution to Linear-Quadratic Estimation and Control Problems," *IEEE Trans. Autom. Control*, vol. AC-22, pp. 138-139, February 1977.
- [2.44] J. M. Morris, "The Kalman Filter: A Robust Estimator for Some Classes of Linear Quadratic Problems," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 526-534, 1976.
- [2.45] H. V. Poor and D. P. Looze, "Minimax State Estimation for Linear Stochastic Systems with Noise Uncertainty," *IEEE Trans. Autom. Control*, vol. AC-26, pp. 902-906, 1981.
- [2.46] S. Verdú and H. V. Poor, "Minimax Linear Observers and Regulators for Stochastic Systems with Uncertain Second-Order Statistics," *IEEE Trans. Autom. Control*, vol. AC-29, pp. 499-511, June 1984.
- [2.47] C. Martin and M. Mintz, "Robust Filtering and Prediction for Linear Systems with Uncertain Dynamics: A Game-Theoretic Approach," *IEEE Trans. Autom. Control*, vol. AC-28, pp. 888-896, September 1983.
- [3.1] W. L. Root, "Stability in Signal Detection Problems," *Proc. Symposia in Applied Mathematics*, vol. 16, American Math. Soc., 1964.
- [3.2] T. Kailath, "RKHS Approach to Detection and Estimation Problems - Part I: Deterministic Signals in Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 530-549, September 1971.

- [3.3] V. P. Kuznetsov, "Stable Detection When the Signal and Spectrum of Normal Noise are Inaccurately Known," *Telecomm. and Radio Engineering* (Engl. transl.), vol. 30/31, pp. 58-64, March 1976.
- [3.4] S. A. Kassam, T. L. Lim and L. J. Cimini, "Two-Dimensional Filters for Signal Processing Under Modeling Uncertainties," *IEEE Trans. Geoscience and Remote Sensing*, vol. GRS-18, pp. 331-336, October 1980.
- [3.5] W. C. Knight, R. G. Pridham, and S. M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, vol. 69, pp. 1451-1506, November 1981.
- [3.6] N. J. Nilsson, "An Application of the Theory of Games to Radar Reception Problems," *IRE National Conv. Rec.*, pt. 4, pp. 130-140, March 1959.
- [3.7] L.-H. Zetterberg, "Signal Detection Under Noise Interference in a Game Situation," *IEEE Trans. Inform. theory*, vol. IT-8, pp. 47-57, September 1962.
- [3.8] C. R. Cahn, "Performance of Digital Matched Filter Correlator with Unknown Interference," *IEEE Trans. Commun.*, vol. COM-19, pp. 1163-1172, December 1971.
- [3.9] G. L. Turin, "Minimax Strategies for Matched-Filter Detection," *IEEE Trans. Commun.*, vol. COM-23, pp. 1370-1371, November 1975.
- [3.10] C.-T. Chen, *Robust and Quantized Linear Filtering for Multiple Input Systems*. Ph.D. Dissertation, Dept. of Systems Engineering, Univ. of Pennsylvania, 1983.
- [3.11] C.-T. Chen and S. A. Kassam, "Robust Multiple-Input Matched Filtering: Frequency and Time-Domain Results," *IEEE Trans. Inform. Theory*, (submitted).
- [3.12] S. Verdú and H. V. Poor, "Minimax Robust Discrete-Time Matched Filters," *IEEE Trans. Commun.*, vol. COM-31, pp. 208-215, February 1983.
- [3.13] V. P. Kuznetsov, "Synthesis of Linear Detectors when the Signals is Inexactly Given and the Properties of the Normal Noise are Incompletely Known," *Radio Eng. and Electronic Physics*, vol. 19, no. 12, pp. 65-73, 1974.
- [3.14] M. V. Burnashev, "On the Minimax Detection of an Inaccurately Known Signal in a White Gaussian Noise Background," *Th. Prob. Appl.*, vol. 24, pp. 107-119, 1979.
- [3.15] R. Sh. Aleyner, "Synthesis of Stable Linear Detectors for an Inaccurately Known Signal," *Radio Eng. and Electronic Physics*, vol. 22, no. 1, pp. 142-145, 1977.
- [3.16] S. Verdú and H. V. Poor, "Signal Selection for Robust Matched Filtering," *IEEE Trans. Commun.*, vol. COM-31, pp. 667-670, May 1983.
- [3.17] V. P. Kuznetsov, "Minimax Linear Neyman-Pearson Detectors for an Imprecisely Known Signal and Noise," *Radio Eng. and Electronic Physics*, vol. 19, no. 12, pp. 73-81, 1974.
- [3.18] K. S. Vastola, J. S. Farnbach, and S. C. Schwartz, "Maximin Sonar System Design for Detection," *Proc. 1983 IEEE International Conf. on Acoustics, Speech and Signal Processing*, April 1983.
- [3.19] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
- [3.20] K. M. Ahmed and R. J. Evans, "Robust Signal and Array Processing," *IEE Proc.*, vol. 129, pt. F, no. 4, pp. 297-302, August 1982.
- [3.21] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-24, pp. 320-327, August 1976.
- [3.22] E. K. Al-Hussaini and S. A. Kassam, "Robust Eckart Filters for Time Delay Estimation," *IEEE Trans. Acoustic, Speech and Signal Processing*, vol. ASSP-32, October 1984.
- [3.23] H. V. Poor, "Robust Decision Design Using a Distance Criterion," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 575-587, September 1980.

- [3.24] H. V. Poor, "Robust Matched Filters," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 677-687, September 1983.
- [3.25] J. B. Thomas, *An Introduction to Statistical Communication Theory*. Wiley: New York, 1969.
- [3.26] C. W. Helstrom, *Statistical Theory of Signal Detection, 2nd Ed.* Pergamon: Oxford, 1968.
- [3.27] D. Slepian, "On Bandwidth," *Proc. IEEE*, vol. 64, pp. 292-300, March 1976.
- [3.28] H. L. Van Trees, *Detection, Estimation and Modulation Theory, I*. Wiley: New York, 1968.
- [3.29] S. Verdú and H. V. Poor, "On Minimax Robustness: A General Approach and Applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 328-340, March 1984.
- [4.1] P. J. Huber, "Robust Estimation of a Location Parameter," *Ann. Math. Stat.*, vol. 35, pp. 73-104, 1964.
- [4.2] P. J. Huber, "A Robust Version of the Probability Ratio Test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.
- [4.3] P. J. Huber and V. Strassen, "Minimax Tests and the Neyman-Pearson Lemma for Capacities," *Ann. Statist.*, vol. 1, pp. 251-263, 1973.
- [4.4] S. A. Kassam, "Robust Hypothesis Testing for Bounded Classes of Probability Densities," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 242-247, 1981.
- [4.5] V. P. Kuznetsov, "Stable Methods of Discriminating Between Two Signals," *Radio Eng. Electron. Phys.*, vol. 19, no. 10, pp. 42-49, 1974.
- [4.6] V. P. Kuznetsov, "Stable Detection of a Signal that is not Precisely Known," *Problems of Information Transmission* (Engl. transl.), vol. 12, no. 2, pp. 119-129, 1976.
- [4.7] V. P. Kuznetsov, "Stable Rules for Discrimination of Hypotheses," *Prob. Inform. Trans.*, vol. 18, no. 1, pp. 41-51, 1982.
- [4.8] S. A. Kassam, "Bounded p-point Classes in Robust Hypothesis Testing and Filtering," *Proc. 20th Annual Allerton Conf. on Comm., Control and Comp.*, pp. 526-534, 1982.
- [4.9] E. L. Price and V. D. VandeLinde, "Robust Estimation Using the Robbins-Munro Stochastic Approximation Algorithm," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 698-704, 1979.
- [4.10] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov Theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316-327, March 1984.
- [4.11] H. V. Poor, "Robust Decision Design Using a Distance Criterion," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 575-587, 1980.
- [4.12] P. M. Schultheiss and J. J. Wolcin, "Robust Sequential Probability Ratio Detectors," *EASCON '75 Convention Rec.*, pp. 36-A-36-H, 1975.
- [4.13] R. D. Martin and S. C. Schwartz, "Robust Detection of a Known Signal in Nearly Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50-56, 1971.
- [4.14] J. Capon, "On the Asymptotic Efficiency of Locally Optimum Detectors," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, April 1961.
- [4.15] J. H. Miller and J. B. Thomas, "Detectors for Discrete-Time Signals in Non-Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 241-250, March 1972.
- [4.16] S. A. Kassam and J. B. Thomas, "Asymptotically Robust Detection of a Known Signal in Contaminated Non-Gaussian Noise," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 22-26, January 1976.
- [4.17] J. H. Miller and J. B. Thomas, "Robust Detectors for Signals in Non-Gaussian Noise," *IEEE Trans. Commun.*, vol. COM-25, pp. 686-691, July 1977.
- [4.18] H. V. Poor and J. B. Thomas, "Asymptotically Robust Quantization for Detection," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 222-229, March 1978.

- [4.19] A. H. El-Sawy and V. D. VandeLinde, "Robust Detection of Known Signals," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 722-727, November 1977.
- [4.20] A. H. El-Sawy and V. D. VandeLinde, "Robust Sequential Detection of Signals in Noise," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 346-353, May 1979.
- [4.21] G. V. Trunk, "Small- and Large-Sample Behavior of Two Detectors Against Envelope-Detected Sea Clutter," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 95-99, January 1970.
- [4.22] G. V. Trunk and S. F. George, "Detection of Targets in Non-Gaussian Sea Clutter," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-6, pp. 620-628, September 1970.
- [4.23] G. V. Trunk, "Non-Rayleigh Sea Clutter: Properties and Detection of Targets," Naval Research Lab. Rept. 7986, June 1976.
- [4.24] S. S. Wolf and J. L. Gastwirth, "Robust Two-Input Correlators," *J. Acoust. Soc. Am.*, vol. 41, pp. 1212-1219, May 1967.
- [4.25] S. A. Kassam, "Locally Robust Array Detectors for Random Signals," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 309-316, May 1978.
- [4.26] M. Kanefsky, "Detection of Weak Signals with Polarity Coincidence Arrays," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 260-268, April 1966.
- [4.27] R. D. Martin and C. P. McGath, "Robust Detection of Stochastic Signals," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 537-541, July 1974.
- [4.28] H. V. Poor, M. Mami, and J. B. Thomas, "On Robust Detection of Discrete-Time Stochastic Signals," *J. Franklin Inst.*, vol. 309, pp. 29-53, January 1980.
- [4.29] H. V. Poor and J. B. Thomas, "Locally-Optimum Detection of Discrete-Time Stochastic Signals in Non-Gaussian Noise," *J. Acoust. Soc. Am.*, vol. 63, pp. 75-80, January 1978.
- [4.30] J. W. Modestino and A. Y. Ningo, "Detection of Weak Signals in Narrowband Non-Gaussian Noise," *IEEE Trans. on Inform. Theory*, vol. IT-25, pp. 592-600, September 1979.
- [4.31] S. A. Kassam, "Detection of Narrowband Signals: Asymptotic Robustness and Quantization," *Proc. 1984 Conf. on Inform. Sciences and Systems*, Princeton Univ., pp. 297-301, March 1984.
- [4.32] S. A. Kassam, "Nonparametric Detection of Narrowband Signals," *Proc. 27th Midwest Symp. on Circuits and Systems*, June 1984.
- [4.33] A. H. El-Sawy, "Detection of Signals with Unknown Phase," *Proc. 17th Ann. Allerton Conf. on Comm., Control and Computing*, pp. 152-161, 1979.
- [4.34] P. A. Kelly, "Robust Estimation and Detection of Signals with Arbitrary Parameters," *Proc. 21st Annual Allerton Conf. on Commun., Control and Computing*, Monticello, IL, pp. 602-609, October 5-7, 1984.
- [4.35] J. G. Shin and S. A. Kassam, "Robust Detector for Narrowband Signals in Non-Gaussian Noise," *J. Acoust. Soc. Amer.*, vol. 74, pp. 527-533, August 1983.
- [4.36] S. S. Wolff and F. J. M. Sullivan, "Robust many-input signal detectors," in *Signal Processing: Proc. of NATO Advanced Study Inst. on Signal Processing*, J. W. R. Griffiths, et al., Eds. Academic Press: London, 1973.
- [4.37] V. P. Kuznetsov, "Synthesis of Stable, Small Signal Detectors," *Radio Eng. and Electronic Physics*, vol. 21, no. 2, pp. 26-34, 1967.
- [4.38] V. P. Kuznetsov, "Minimax Linear Neyman-Pearson Detectors for an Imprecisely Known Signal and Noise," *Radio Eng. and Electronic Physics*, vol. 19, no. 12, pp. 73-81, 1974.
- [4.39] H. V. Poor, "Signal Detection in the Presence of Weakly Dependent Noise - Part II: Robust Detection," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 744-752, 1982.
- [4.40] S. L. Portnoy, "Robust Estimation in Dependent Situations," *Ann. Statist.*, vol. 5, pp. 22-43, 1977.

- [4.41] S. L. Portnoy, "Further Remarks on Robust Estimation in Dependent Situations," *Ann. Statist.*, vol. 7, pp. 244-251, 1979.
- [4.42] G. V. Moustakides and J. B. Thomas, "Min-Max Detection of Weak Signals in ϕ -mixing Noise," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 529-537, May 1984.
- [4.43] S. A. Kassam and J. B. Thomas, "Dead-Zone Limiter: An Application of Conditional Tests in Nonparametric Detection," *J. Acoust. Soc. Amer.*, vol. 60, pp. 857-862, 1976.
- [4.44] R. D. Martin, "Robust Estimation of Signal Parameters with Dependent Data," *Proc. 21st IEEE Conf. on Decision and Control*, pp. 433-436, December 1982.
- [4.45] S. A. Kassam, G. Moustakides and J. G. Shin, "Robust Detection of Known Signals in Asymmetric Noise," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 84-91, January 1982.
- [4.46] J. R. Collins, "Robust Estimation of a Location Parameter in the Presence of Asymmetry," *Ann. Statist.*, vol. 4, pp. 68-85, 1976.
- [4.47] R. D. Martin and S. C. Schwartz, "On Mixture, Quasi-Mixture, and Nearly Normal Random Processes," *Ann. Math. Stat.*, vol. 43, pp. 948-967, 1972.
- [4.48] S. A. Kassam, "Optimum Data Quantization for Signal Detection," in *Advances in Communications and Networks*, I. F. Blake and H. V. Poor, Eds. Springer-Verlag: New York, 1985.
- [4.49] L. Kurz, "Nonparametric Detectors Based on Partition Tests," in *Nonparametric Methods in Communications*, P. Papantoni-Kazakos and D. Kazakos, Eds. Marcel Dekker: New York, 1977.
- [4.50] R. F. Dwyer, "Robust Sequential Detection of Weak Signals in Undefined Noise Using Acoustical Arrays," *J. Acoust. Soc. Amer.*, vol. 67, pp. 833-841, 1980.
- [4.51] Y. C. Ching and L. Kurz, "Nonparametric Detectors Based on m-Internal Partitioning," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 251-257, March 1972.
- [4.52] P. Kerstun and L. Kurz, "Robustized Vector Robbins-Monro Algorithm with Applications to m-Interval Detection," *Inform. Sci.* vol. 11, pp. 121-140, 1976.
- [4.53] P. Kersten and L. Kurz, "Bivariate m-Interval Classifiers with Application to Edge Detection," *Inform. and Control*, vol. 34, pp. 152-168, 1977.
- [4.54] R. F. Dwyer, "Robust Sequential Detection of Narrowband Acoustic Signals in Noise," *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, pp. 140-143, April 1979.
- [4.55] R. F. Dwyer, "A Technique for Improving Detection and Estimation of Signals Contaminated by Under Ice Noise," *J. Acoust. Soc. Amer.*, vol. 74, pp. 124-130, July 1983.
- [4.56] J. H. Miller and J. B. Thomas, "The Detection of Signals in Impulsive Noise Modeled as a Mixture Process," *IEEE Trans. on Comm.*, vol. COM-24, pp. 559-563, May 1976.
- [4.57] J. W. Modestino, "Adaptive Detection of Signals in Impulsive Noise Environments," *IEEE Trans. on Comm.*, vol. COM-25, pp. 1022-1027, September 1977.
- [4.58] L. B. Milstein, D. D. Schilling and J. K. Wolf, "Robust Detection Using Extreme-Value Theory," *IEEE Trans. Inform. Theory*, vol. 15, pp. 370-375, 1969.
- [4.59] P. J. Huber, *Robust Statistical Procedures*. SIAM Press: Philadelphia, 1977.
- [4.60] S. A. Kassam and J. B. Thomas, Eds., *Nonparametric Detection: Theory and Applications*. Dowden, Hutchinson and Ross: Stroudsburg, PA, 1980.
- [4.61] S. A. Kassam, "A Bibliography on Nonparametric Detection," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 595-602, September 1980.
- [4.62] V. M. Krasnenker, "Stable (Robust) Detection Methods for Signals Against a Noise Background (Survey)," *Automation and Remote Control (Engl. transl.)*, vol. 41, no. 5, pt. 1, pp. 640-659, 1980.

- [4.63] A. A. Ershov, "Stable Methods of Estimating Parameters (Survey)," *Automation and Remote Control*, vol. 39, no. 7, pt. 2, pp. 1152-1181, July 1978.
- [4.64] V. D. VandeLinde, "Robust Techniques in Communication," in *Robustness in Statistics*, R. L. Launer and G. M. Wilkinson, Eds. Academic Press: New York, 1979, pp. 177-200.
- [4.65] H. V. Poor, "Some Results on Robust Signal Detection," *Proc. 22nd IEEE Conf. on Decision and Control*, San Antonio, TX, pp. 1098-1105, December 1983.
- [5.1] P. J. Huber, *Robust Statistics*. Wiley: New York, 1981.
- [5.2] R. D. Martin, "Robust Methods for Time Series," in *Proc. of Nottingham International Time Series Meeting*, O.D. Anderson, Ed. North-Holland: Amsterdam, 1980.
- [5.3] A. A. Ershov, "Robust Parameter Estimation (A Survey)," *Automation and Remote Control*, vol. 7, p. 66, 1978.
- [5.4] B. T. Poljak and Ya. Z. Tsytkin, "Robust Identification," *Automatica*, vol. 16, p. 52, 1980.
- [5.5] J. W. Tukey, "A Survey of Sampling from Contaminated Distributions," in *Contributions to Probability and Statistics*, I. Olkin, et al., Eds. Stanford University Press: Stanford, CA, 1960.
- [5.6] P. J. Huber, "Robust Estimation of a Location Parameter," *Ann. Math. Stat.*, vol. 35, pp. 73-104, 1964.
- [5.7] R. D. Martin and D. J. Masrelietz, "Robust Estimation via Stochastic Approximation," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 263-271, May 1975.
- [5.8] S. L. Portnoy, "Robust Estimation in Dependent Situations," *Ann. Statist.*, vol. 5, pp. 22-43, 1977.
- [5.9] E. L. Price and N. H. Gholson, "Robust Estimation of Random Parameters with Incompletely Defined Uncertainty," *1977 IEEE Conf. on Decision and Control*, pp. 439-443, December 1977.
- [5.10] J. R. Collins, "Robust Estimation of a Location Parameter in the Presence of Asymmetry," *Ann. Statist.*, vol. 4, pp. 68-85, 1976.
- [5.11] A. H. El-Sawy and V. D. VandeLinde, "Robust Detection of Known Signals," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 722-727, 1977.
- [5.12] R. D. Martin, "Robust Estimation of Signal Amplitude," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 596-606, 1972.
- [5.13] E. L. Price and V. D. VandeLinde, "Robust Estimation Using the Robbins-Munro Stochastic Approximation Algorithm," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 698-704, 1979.
- [5.14] F. R. Hampel, "A General Qualitative Definition of Robustness," *Ann. Math. Stat.*, vol. 42, pp. 1887-1896, 1971.
- [5.15] W. L. Root, "Stability in Signal Detection Problems," *Proc. Symposia in Applied Mathematics*, vol. 16, American Math. Soc., 1964.
- [5.16] P. A. Kelly, "Robust Estimation and Detection of Signals with Arbitrary Parameters," *Proc. 21st Annual Allerton Conf. on Comm., Control and Computing*, Monticello, IL, pp. 602-609, October 5-7, 1984.
- [5.17] A. H. El-Sawy, "Detection of Signals with Unknown Phase," *Proc. 17th Ann. Allerton Conf. on Comm., Control and Computing*, pp. 152-161, 1979.
- [5.18] A. C. Bovik, T. S. Huang and A. C. Bovik, "A Generalization of Median Filtering Using Linear Combinations of Order Statistics," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-31, pp. 1342-1350, December 1982.
- [5.19] Y. H. Lee and S. A. Kassam, "Nonlinear Edge Preserving Filtering Techniques for Image Enhancement," *Proc. 27th Midwest Symp. Circuits Syst.*, June 1984.
- [5.20] Y. H. Lee and S. A. Kassam, "Applications of Nonlinear Adaptive Filters for Image Enhancement," *Proc. 7th International Conf. Pattern Recog.*, Montreal, Canada, 1984.

- [5.21] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, NJ, 1979.
- [5.22] C. J. Masreliez and R. D. Martin, "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Trans. Automatic Control*, vol. AC-22, pp. 361-371, 1977.
- [5.23] B. Kleiner, R. D. Martin and D. J. Thompson, "Robust Estimation of Power Spectra," *J. Roy. Stat. Soc., Series B*, vol. 41, pp. 313-351, 1979.
- [5.24] C. G. Boncelet, Jr. and B. W. Dickinson, "An Approach to Robust Kalman Filtering," *Proc. 22nd IEEE Conf. Decision Control*, San Antonio, TX, pp. 304-305, December 14-16, 1983.
- [6.1] J. M. Morris, *Quantization and Source Encoding with a Fidelity Criterion: A Survey*. Naval Research Laboratory Report No. 8082, March 25, 1977.
- [6.2] A. Gersho, "Principles of Quantization," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 427-436, July 1978.
- [6.3] R. M. Gray, Ed., *Special Issue on Quantization*. *IEEE Trans. Inform. Theory*, vol. IT-28, March 1982.
- [6.4] S. A. Kassam, "Quantization for Signal Detection," in *Advances in Communications and Networks*. I. F. Blake and H. V. Poor, Eds. Springer-Verlag, New York, 1985.
- [6.5] B. Aazhang and H. V. Poor, "On Optimum and Nearly Optimum Data Quantization for Signal Detection," *IEEE Trans. Commun.*, vol. COM-32, pp. 745-751, July 1984.
- [6.6] H. V. Poor, "A Companding Approximation for the Statistical Divergence of Quantized Data," *Proc. 22nd IEEE Conf. Decision Control*, San Antonio, TX, pp. 697-702, December 14-16, 1983.
- [6.7] H. V. Poor, "The Effects of Data Quantization on Filtering of Stationary Gaussian Processes," *Proc. 23rd IEEE Conf. Decision Control*, Las Vegas, NV, December 12-14, 1984.
- [6.8] J. Max, "Quantizing for Minimum Distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7-12, March 1960.
- [6.9] A. Gersho, "Principles of Quantization," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 427-436, July 1978.
- [6.10] D. J. Goodman and A. Gersho, "Theory of an Adaptive Quantizer," *IEEE Trans. Commun.*, vol. COM-22, pp. 1037-1045, August 1974.
- [6.11] A. Gersho and D. J. Goodman, "A Training Mode Adaptive Quantizer," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 746-749, November 1974.
- [6.12] J. M. Morris and V. D. VandeLinde, "Robust Quantization of Discrete-Time Signals with Independent Samples," *IEEE Trans. Commun. Technol.*, vol. COM-22, pp. 1897-1901, December 1974.
- [6.13] W. G. Bath and V. D. VandeLinde, "Robust Memoryless Quantization for Minimum Signal Distortion," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 296-306, March 1982. [Also: W. G. Bath and V. D. VandeLinde, "Robust Quantizers for Signals with Known Moments and Modes," *Proc. 22nd Midwest Symp. Circuits Syst.*, Philadelphia, PA, pp. 509-513, June 1979.]
- [6.14] J. A. Bucklew and G. L. Wise, "Multidimensional Asymptotic Quantization Theory with r-th Power Distortion Measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239-247, March 1982.
- [6.15] W. G. Bath and V. D. VandeLinde, "Robust Quantizers Designed Using the Companding Approximation," *Proc. 18th IEEE Conf. on Decision and Control*, Ft. Lauderdale, FL, pp. 483-487, December 1979.
- [6.16] P. W. Cattermole, *Principles of Pulse Code Modulation*. Iliffe: London, 1969.

- [6.17] D. Kazakos, "On the Design of Robust Quantizers," *1981 IEEE Telecomm. Conf., Conf. Record*, New Orleans, LA, pp. F4.5.1-F4.5.4, December 1981.
- [6.18] D. Kazakos, "New Results on Robust Quantization," *IEEE Trans. Commun.*, vol. COM-31, pp. 965-974, August 1983.
- [6.19] H. V. Poor, "Some Results in Robust Data Quantization," *Proc. 21st IEEE Conf. Decision Control*, Orlando, FL, pp. 440-445, December 8-10, 1982.
- [6.20] H. V. Poor, "Robust Quantization of ϵ -Contaminated Data," *IEEE Trans. Commun.*, vol. COM-33, 1985.
- [7.1] J. Franke and H. V. Poor "Minimax Robust Filtering and Finite-Length Robust Predictors," in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Hardle and R. D. Martin, Eds. Springer-Verlag: Heidelberg, 1984.
- [7.2] G. V. Moustakides and J. B. Thomas, "Min-Max Detection of Weak Signals in ϕ -Mixing Noise," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 529-537, May 1984.