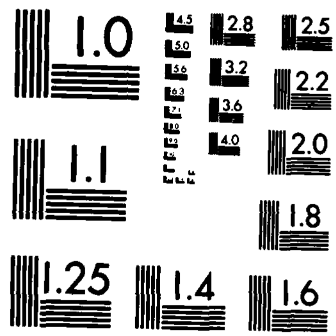
AD-A125 931 TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING(U) 1/2
ILLINOIS UNIV AT URBANA COORDINATED SCIENCE LAB
K S VASTOLA SEP 82 R-965 N00014-79-C-0424

UNCLASSIFIED F/G 12/1 NL



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

REPORT R-965 SEPTEMBER, 1982 UILU-ENG 82-2231

COORDINATED SCIENCE LABORATORY

TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING

KENNETH STEVEN VASTOLA

APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A125931	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER R-965 UILU-ENG 82-2231
7. AUTHOR(s) Kenneth Steven Vastola		8. CONTRACT OR GRANT NUMBER(s) JSEP N00014-79-C-0424 ONR N00014-81-K-0014 NSF ECS 79-16453
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory University of Illinois 1101 W. Springfield, Urbana, IL 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Electronics Program Office of Naval Research National Science Foundation		12. REPORT DATE September 1982
		13. NUMBER OF PAGES 108
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robust signal processing, Wiener-Kolmogorov theory, Spectral uncertainty, Choquet capacities

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This dissertation addresses several problems in robust signal processing. The term "robust" in this context implies insensitivity to small deviations from the assumed statistical description of the signal and/or noise.

The first part of this thesis considers the problem of linear minimum-mean-square-error estimation of a stationary signal observed in additive stationary noise when knowledge of the signal spectrum and noise spectrum is inexact. First, the performance of robust continuous-time (Wiener) noncausal

20. filters (designed using a method developed elsewhere) is examined. It is shown in a variety of situations that when spectral uncertainty exists the performance of the traditional Wiener filter degrades badly while the robust filter's insensitivity to such deviations makes it an effective alternative. Next, this design approach is developed for the general problem of robust discrete-time (Wiener-Kolmogorov) causal signal estimation, and a simple characterization of solutions to this problem is given. The method of design is then illustrated by a thorough development of the special case of one-step noiseless prediction and numerical examples which illustrate the effectiveness of the general design are given for the problem of robust causal filtering of an uncertain signal in white noise.

In the second part of this dissertation, a previously developed cohesive theory of robust hypothesis testing in which uncertainty is modeled via 2-alternating Choquet capacity classes is considered in light of recent applications of this theory to problems in robust signal processing and communication theory. In particular, a generalization of capacities is given which allows several of the most common uncertainty classes to be considered under a less restrictive compactness assumption. Results are given which generalize this robust hypothesis testing theory and which are of direct consequence for the applications. For example, it is shown how these results allow the problem of robust linear smoothing of an uncertain continuous-time signal in white noise to be fit within a general framework developed previously for robust (minimax) linear smoothing. Finally, some properties of the band model and p-point model (uncertainty classes which are especially appropriate for many applications) are developed within the context of 2-alternating capacities.

TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING

BY

KENNETH STEVEN VASTOLA

B.A., Rutgers University, 1976
M.S., University of Illinois, 1979

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1982

Urbana, Illinois

TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING

Kenneth Steven Vastola, Ph.D.
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 1982
(H. V. Poor)

This dissertation addresses several problems in robust signal processing. The term "robust" in this context implies insensitivity to small deviations from the assumed statistical description of the signal and/or noise.

The first part of this thesis considers the problem of linear minimum-mean-square-error estimation of a stationary signal observed in additive stationary noise when knowledge of the signal spectrum and noise spectrum is inexact. First, the performance of robust continuous-time (Wiener) noncausal filters (designed using a method developed elsewhere) is examined. It is shown in a variety of situations that when spectral uncertainty exists the performance of the traditional Wiener filter degrades badly while the robust filter's insensitivity to such deviations makes it an effective alternative. Next, this design approach is developed for the general problem of robust discrete-time (Wiener-Kolmogorov) causal signal estimation, and a simple characterization of solutions to this problem is given. The method of design is then illustrated by a thorough development of the special case of one-step noiseless prediction and numerical examples which illustrate the effectiveness of the general design are given for the problem of robust causal filtering of an uncertain signal in white noise.

In the second part of this dissertation, a previously developed cohesive theory of robust hypothesis testing in which uncertainty is modeled via 2-alternating Choquet capacity classes is considered in light of recent applications of this theory to problems in robust signal processing and communication theory. In particular, a generalization of capacities is given which allows several of the most common uncertainty classes to be considered under a less restrictive compactness assumption. Results are given which generalize this robust hypothesis testing theory and which are of direct consequence for the applications. For example, it is shown how these results allow the problem of robust linear smoothing of an uncertain continuous-time signal in white noise to be fit within a general framework developed previously for robust (minimax) linear smoothing. Finally, some properties of the band model and p-point model (uncertainty classes which are especially appropriate for many applications) are developed within the context of 2-alternating capacities.

To my wife, Deborah Ann Vastola.

Without her constant support and encouragement

this would not have been possible.

ACKNOWLEDGMENT

To his advisor, Professor H. Vincent Poor, the author expresses his sincerest gratitude. Professor Poor's valuable advice, unceasing support, and patient guidance have been essential to the graduate education of the author and to the preparation of this dissertation. Working with him will always be one of the highlights of the author's professional career.

The author also thanks Professors M. B. Pursley and S. L. Portnoy for their contributions to his graduate education, and he thanks S. Verdú for helpful discussions on Chapter III. Further, the author thanks C. Jewell, C. Cassells and P. Young for typing portions of this work.

Finally, the arthur (sic) would like to express his appreciation to his parents, Salvatore Vastola and Barbara Engelberg Vastola, for their support and encouragement throughout his entire education.

TABLE OF CONTENTS

Contents

I. INTRODUCTION

In the design and analysis of statistical signal processing procedures, it is usually assumed that some underlying spectral distribution or probability distribution is known precisely. Often in practice this is an unrealistic assumption. Furthermore, as we will illustrate in Chapter II of this thesis, the belief that nearly accurate models will result in nearly optimal solutions is frequently unfounded. Thus, we would like to design procedures which are insensitive to small deviations from an assumed model. Such procedures have generally been termed robust.

In 1960, Tukey [39] brought attention to the fact that a number of statistical data-analysis procedures are undesirably sensitive to small deviations from the assumed probability distribution of the observations. During the 1960's, two basic approaches to the problem of designing robust alternatives to such procedures were developed. The first, which could be termed the "minimax" or "Huber" approach consists, basically, of first, modeling the uncertainty via a class of probability distributions and, then, finding a procedure which has the best worst-case performance over this class (see [43], [14], [27], [6] and [44]). The other approach, which was originated by Hampel [45], views robustness in terms of the continuity properties of a procedure on a space of probability distributions (see [45] and [44]).

These techniques were first applied in a statistical signal processing context by Martin and Schwartz [40] who considered the design of robust signal detection procedures. Other results in robust detection were subsequently obtained by Kassam and Thomas [46], El-Sawy and VandeLinde [5], [8], Kuznetsov [22], Poor [53] and many others (for a survey, see [47]).

Robust parameter estimation has also been considered in a signal processing context by Martin [50], Papantoni-Kazakos [51], Price and VandeLinde [52] and others (see [48] for a survey). Further, results have been developed for robust nonlinear filtering (see Martin [49] for a survey of this area). Recently, Kassam and Lim [1], Cimini and Kassam [25] and Poor [2], [7] have developed a method of designing Wiener filters which are robust with respect to spectral uncertainty.

This thesis considers several top: in the general area of robust signal processing. In particular, in C ter II of this thesis, we present results from a numerical study of the p mance of the robust Wiener filters designed via the methods of [1], [2], [25]. We begin Chapter II by examining the effects of spectral uncertainty on traditional Wiener filters. We show that in many cases a clear need for robust Wiener filtering exists and that in many of these cases the robust Wiener filters developed in [1], [2], [25] are an effective alternative to traditional Wiener filters.

In Chapter III, using a general formulation analogous to that developed in [2] for robust linear continuous-time (Wiener) noncausal filtering, we develop a method of designing robust linear discrete-time (Wiener-Kolmogorov) causal signal estimators (e.g., robust n-step predictors, robust causal filters and robust n-lag smoothers). The specific problem of robust one-step noiseless prediction is developed in detail and numerical results are given for the particular problem of robust filtering in white noise.

In Chapter IV, we present a generalization of the results of Huber and Strassen [6]. In [6], a cohesive theory of (minimax) robust hypothesis testing was developed for the quite general situation in which uncertainty is modeled via classes of probability distributions dominated by 2-alternating

Choquet capacities. These results have been applied by Poor [7], [32] to problems in communication theory in which spectral uncertainty is modeled via capacity classes of spectral distributions. In this chapter we extend the theory of [6] to certain situations which are appropriate in this new context. For example, we show in Chapter IV how the problem of robust continuous-parameter smoothing of an uncertain signal in white noise now fits within the general framework developed in [7]. Furthermore, the results given in Chapter IV partially extend the usefulness of the results of [6] to noncompact measure spaces. Finally, the band model, an uncertainty class which is appropriate for many applications, is shown to be a 2-alternating capacity class and is used to illustrate certain results of this chapter.

In Chapter V, a commonly used model of uncertainty known as the p-point class is examined. It is shown that, while the p-point class is not a capacity class, it is contained in a capacity class which we call an extended p-point class. In many instances the results of [6], [7], [32] which, of course, hold for this extended p-point class are shown to hold also for the corresponding p-point class.

II. AN ANALYSIS OF THE EFFECTS OF SPECTRAL UNCERTAINTY ON WIENER FILTERING

1. Introduction

The solution to the traditional stationary linear (i.e., Wiener) filtering problem requires exact knowledge of the signal and noise spectra. Often in practice it is unrealistic to assume such knowledge. Despite this, Wiener filters are widely used for steady-state filtering. In this chapter we consider the performance of Wiener filtering when the signal and noise spectra differ to a small degree from those assumed in the design process. In Chapter III, we will consider the related problem of discrete-time (Wiener-Kolmogorov) signal estimation when spectral uncertainty exists.

In Section 2 of this chapter we consider the Wiener filter for a particular signal and noise spectral pair which would be natural to assume is the true spectral pair. We then look again at our circumstances and model the uncertainty we might have about our choice of spectra. In so doing we find that the potential exists for totally unacceptable performance degradation in the presence of even small degrees of uncertainty.

In Section 3 we consider filters termed "robust". These filters are designed to have the best "worst-case" performance over uncertainty classes of spectra. The method of design is due to Poor [2] and was based on the work of Kassam and Lim [1]. As we will see, the advantage of these robust filters is that they are least sensitive in the sense that they have the smallest possible maximum deviation from optimality within the constraints imposed by our uncertainty.

Of course there is a trade-off involved in robust filtering. While the robust filter has better worst-case performance, we cannot expect it to have optimal performance should our original choice of spectra be the true ones In Section 3 we will consider this trade-off as well.

2. The Sensitivity of the Wiener Filter to Spectral Uncertainty

The mean-square-error (MSE) for linear filtering of a signal in uncorrelated additive noise, where both signal and noise are modeled as zero-mean, second-order, wide-sense stationary random processes, is given by

$$e(\sigma,\nu;H) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[\sigma(\omega)|1-H(\omega)|^2 + \nu(\omega)|H(\omega)|^2]\, d\omega\ , \qquad (2.1)$$

where H is the transfer function of the filter and σ and ν are the power spectral densities (PSD's) of the signal and noise, respectively. For a fixed signal and noise spectral pair, (σ,ν), $e(\sigma,\nu;H)$ is minimized by the Wiener filter

$$H^*(\omega) = \frac{\sigma(\omega)}{\sigma(\omega) + \nu(\omega)} \qquad (2.2)$$

and the minimum MSE is

$$e^*(\sigma,\nu) \triangleq e(\sigma,\nu;H^*) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H^*(\omega)\ \nu(\omega)\ d\omega\ . \qquad (2.3)$$

Unfortunately, as we discussed in Section 1, it is often the case in practice that our knowledge of the signal and/or noise PSD's is inexact. If the σ and ν we choose for designing H^* are not the true spectra, then our filter will generally have less than optimal performance. To illustrate the degree of performance degradation that can result from such mis-modeling, we consider the following examples. The numerical results presented here and in the following section comprise a representative selection from an extensive numerical study.

The p-point class. For a number of applications it is natural to assume that we have a narrow-band first-order Markov signal in wide-band first-order Markov noise, i.e. that

$$\sigma_0(\omega) = \frac{2\alpha_S v_S^2}{\alpha_S^2 + \omega^2}$$

and

$$\nu_0(\omega) = \frac{2\alpha_N v_N^2}{\alpha_N^2 + \omega^2} \tag{2.4}$$

where $\alpha_S \ll \alpha_N$ are the 3 dB bandwidths and v_S^2 and v_N^2 are the powers of the signal and noise, respectively. For Fig. 1 we have $\alpha_N = 10$ and $\alpha_S = 1$.

In the figures of this chapter we have used a measure of performance which we refer to simply as output signal-to-noise ratio (SNR). The purpose of Wiener filtering is to minimize the MSE, $E\{[\hat{S}(t) - S(t)]^2\}$, between our estimate $\hat{S}(t)$ (i.e. the output of the filter) and the actual signal $S(t)$. Since the output of the filter can be written as $S(t) + (\hat{S}(t) - S(t))$, we use the signal power divided by the MSE as an output SNR. For the purpose of our graphs we translate this to dB. The horizontal axis is $10 \log_{10}(v_S^2/v_N^2)$, the input SNR in dB.

The top line in Fig. 1 gives the performance of the Wiener filter H_0^*, designed using σ_0 and ν_0 of (2.4) in equation (2.2), when σ_0 and υ_0 are, in fact, the signal and noise spectra which occur. For this case it is straightforward, via equation (2.3), to show that

$$e^*(\sigma_0, \nu_0)/v_S^2 = \alpha_S\alpha_N/\sqrt{(\alpha_S\alpha_N^2 r + \alpha_S^2\alpha_N)(\alpha_S r + \alpha_N)}$$

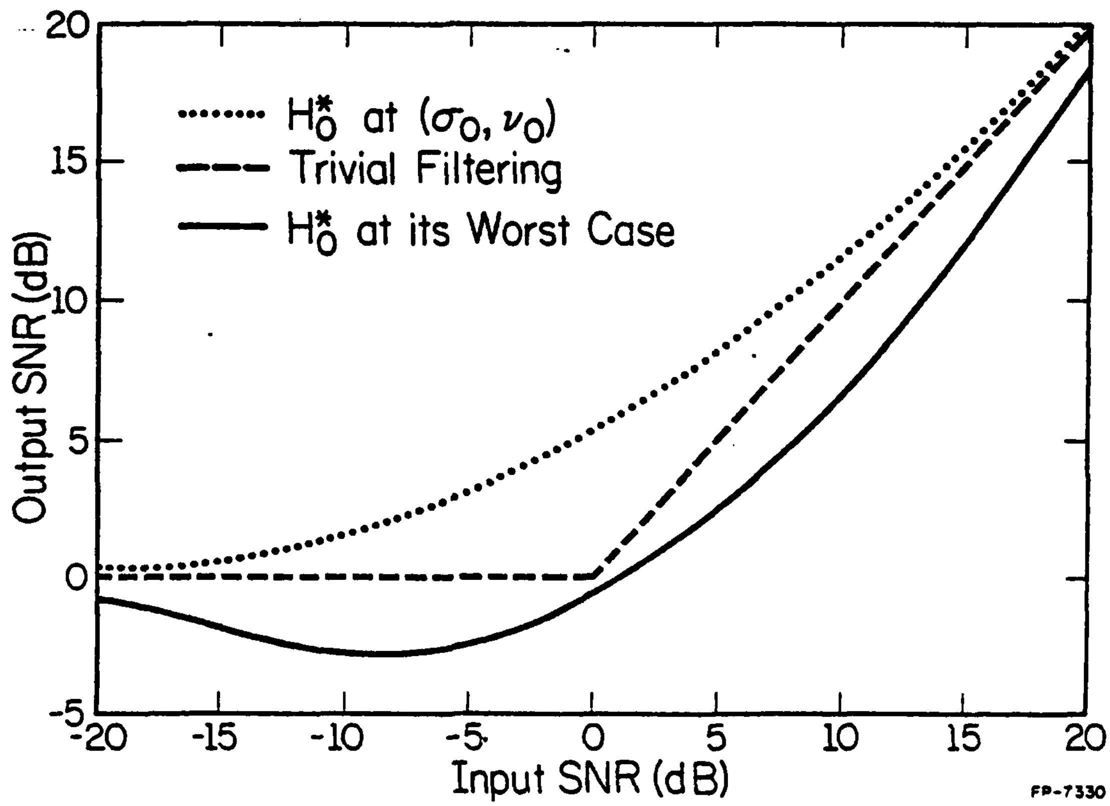


FIG. 1. p-point example. (From top to bottom) H_0^* at (σ_0, ν_0); "trivial filtering"; H_0^* at its worst case.

where

$$r \triangleq v_S^2/v_N^2 .$$

Now, suppose that the only information about which we are certain is the powers of the signal v_S^2 and the noise v_N^2 and that we have estimated with sufficient accuracy the fractional power of each on the set $S = \{\omega \text{ real} \mid |\omega| \leq 1\}$. We denote the signal and noise fractional powers by p_S and p_N, respectively (e.g. $(2\pi)^{-1}\int_S \sigma(\omega)d\omega = p_S v_S^2$). In particular, for the example considered above, we have $p_S = 0.5$ and $p_N = 0.063$. If these total powers and fractional powers are all we can really be certain of, we would like to know how badly the performance of H_0^* can deteriorate. The bottom line in Fig. 1 gives the worst-case performance of H_0^*. The middle line represents what we can do trivially for any pair of spectra by using an all-pass filter ($H \equiv 1$) when the input SNR is positive and by using a no-pass filter ($H \equiv 0$) when the input SNR is negative. Thus we see that if the spectra are actually first-order Markov then our filter does well, but if not we can do <u>significantly worse than trivial filtering</u>.

Finally we note that uncertainty classes of spectra given by assuming exact knowledge only of the total and fractional powers are called p-point classes and have been studied as models of spectral uncertainty by Cimini and Kassam [25]. An analogous uncertainty class for probabilities used in robust hypothesis testing and robust detection has been examined by El-Sawy and VandeLinde [5], [8]. These classes will be considered in greater detail in Chapter V.

<u>The ε-contamination class</u>. Suppose that we again have a particular spectral pair (σ_0, ν_0) which we believe to be the true spectra, but that we

also have a general sense of uncertainty about our choice which we model by an ε-contaminated class; i.e., we assume we know that the true spectra satisfy $(\sigma,\nu) \in \mathscr{S}_\varepsilon \times \mathscr{N}_\varepsilon$ where $0 \leq \varepsilon \leq 1$,

$$\mathscr{S}_\varepsilon = \{\sigma \mid \sigma(\omega) = (1-\varepsilon)\sigma_0(\omega) + \varepsilon\sigma'(\omega) \quad \omega \in \mathbb{R}, \int_{-\infty}^{\infty} \sigma'(\omega)\, d\omega = \int_{-\infty}^{\infty} \sigma_0(\omega)\, d\omega\}$$

and (2.5)

$$\mathscr{N}_\varepsilon = \{\nu \mid \nu(\omega) = (1-\varepsilon)\nu_0(\omega) + \varepsilon\nu'(\omega) \quad \omega \in \mathbb{R}, \int_{-\infty}^{\infty} \nu'(\omega)\, d\omega = \int_{-\infty}^{\infty} \nu_0(\omega)\, d\omega\} \quad .$$

Classes of this form have been used extensively as general models of uncertainty [39], [14], [27], [40], [1], [23].

Fig. 2 gives the performance of the Wiener filter H_0^* designed via equation (2.2) assuming a narrow-band ($\alpha_S = 1$) first-order Markov signal in wide-band ($\alpha_N = 1000$) first-order Markov noise. The upper line gives the performance of this filter when these are the true signal and noise. The lower line is the worst case of this filter over the uncertainty classes in (2.5) with σ_0 and ν_0 given by the above choices and with $\varepsilon = 0.1$. We see that, for values of input SNR near zero, the worst case is better than trivial filtering but still much worse than optimal (about 8.5 dB); for values of input SNR greater in absolute value than 60 the performance in both the nominal and worst cases is the same as trivial filtering; and for all other values the worst case is worse than trivial filtering.

An ε-contaminated signal in white noise. Fig. 3 shows the nominal and worst case performance of the nominal Wiener filter for the signal uncertainty class $\mathscr{S}_\varepsilon$ in (2.5) with $\varepsilon = 0.1$ and σ_0 first-order Markov with $\alpha_S = 1$. The noise is white noise with no uncertainty and the horizontal

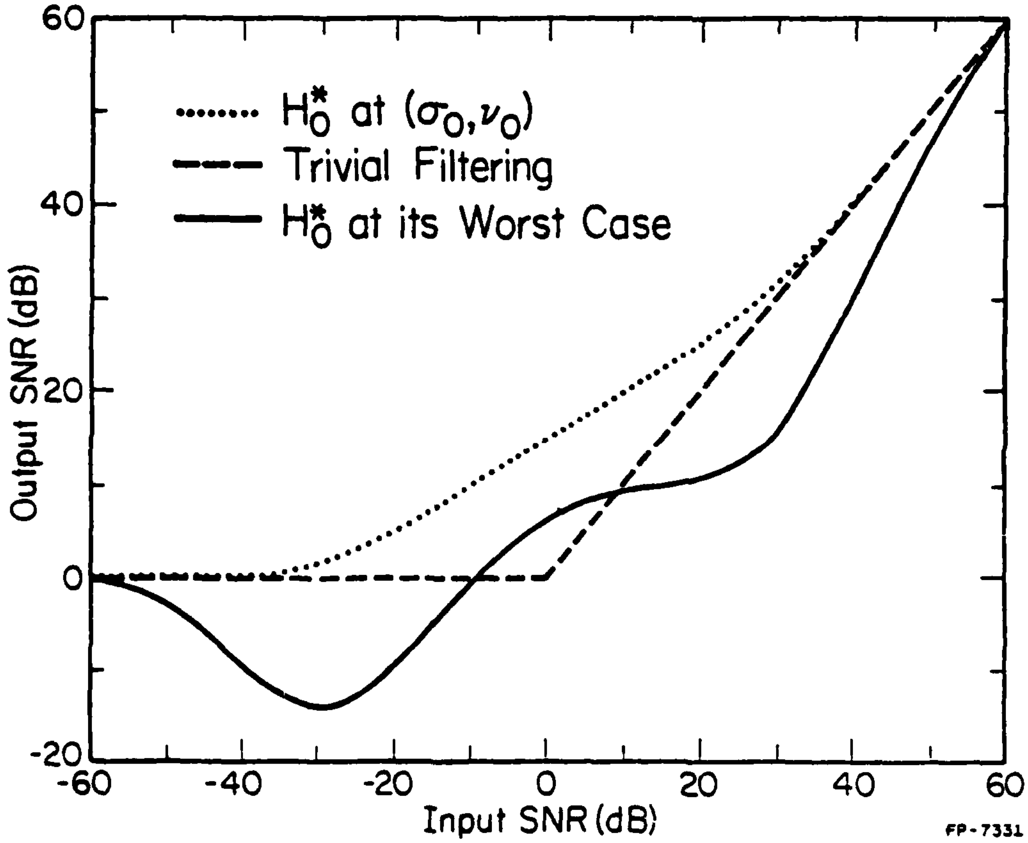


FIG. 2. ε-contaminated example. H_0^* at (σ_0,ν_0) and at its worst case.

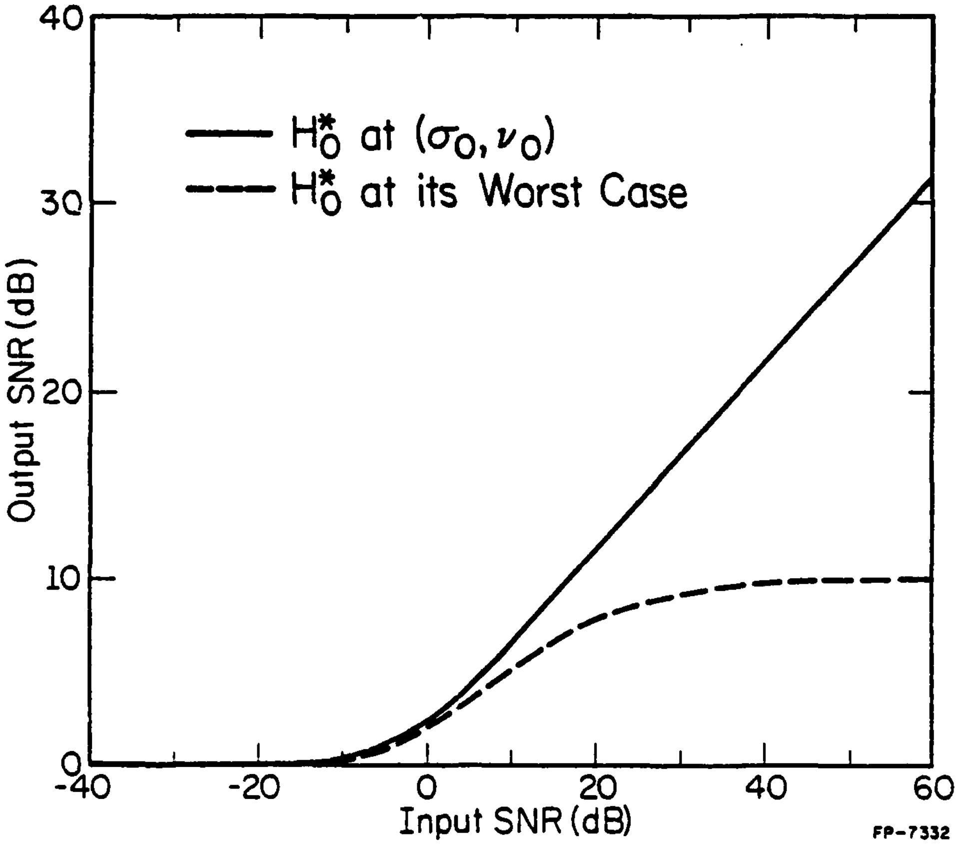


FIG. 3. ε-contaminated signal in white noise. H_0^* at (σ_0, ν_0) at its worst case.

axis is actually the ratio of signal power v_S^2 to the noise level $N_0/2$. Note that the worst case is bounded above by 10; in fact, for any choice of ε, it is bounded above by $-10 \log(\varepsilon)$.

As noted above, the optimal and worst-case performance of Wiener filtering under various conditions has been examined extensively for several uncertainty models and for a variety of signal and noise parameters (such as bandwidth and power). The above examples are representative of the sensitivity of Wiener filtering to deviations from spectral assumptions which were found in virtually every case. Further examples are pictured in the appendix.

3. Robust Wiener Filters

To remedy the problems of Wiener filtering sensitivity discussed in the preceding section, we consider the following robust filter design which was developed by Poor [2] based on the work of Kassam and Lim [1].

A most-robust Wiener filter [2] is a solution H_R to the game

$$\min_{H} \sup_{(\sigma,\nu) \in \mathcal{S} \times \mathcal{N}} e(\sigma,\nu;H) \tag{2.6}$$

where $\mathcal{S}$ and $\mathcal{N}$ are classes of spectra representing uncertainty in the signal and noise, respectively, and where $e(\sigma,\nu;H)$ is given in (2.1). Note that since the supremum in (2.6) gives the least upper bound on the error, H_R^* is a filter with the smallest possible such upper bound. In other words H_R^* is least sensitive to worst case uncertainty.

A pair of spectra (σ_L,ν_L) is least favorable for Wiener filtering for the spectral uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ [2] if

$$e(\sigma,\nu;H_L^*) \leq e(\sigma_L,\nu_L;H_L^*) \tag{2.7}$$

for all $\sigma \in \mathcal{J}$, $\nu \in \mathcal{N}$ where H_L^* is the Wiener filter for the pair (σ_L, ν_L) as in (2.2).

It is straightforward to see that if $(\sigma_L, \nu_L) \in \mathcal{J} \times \mathcal{N}$ is least favorable for Wiener filtering for $\mathcal{J}$ and $\mathcal{N}$ then the pair $((\sigma_L, \nu_L), H_L^*)$ is a saddle-point solution to the minimax game (2.6). That is,

$$\sup_{(\sigma,\nu) \in \mathcal{J} \times \mathcal{N}} e(\sigma,\nu;H_L^*) = e(\sigma_L,\nu_L;H_L^*) = \min_H e(\sigma_L,\nu_L;H). \tag{2.8}$$

We see from this that if (σ_L, ν_L) is least favorable then H_L^* is a most-robust Wiener filter.

Thus we see that if we can find a least-favorable pair then we can design a most-robust Wiener filter. One of the methods developed by Poor [2] for finding least favorable pairs of spectra (and hence most-robust filters) involves an analogous concept in hypothesis testing: least-favorable probability density functions (PDF's) for testing one set of PDF's against another. Least-favorable PDF's have been found for a variety of classes of PDF's (see [14], [27], [20], [21], and Chapter V). If every signal spectrum in $\mathcal{J}$ has the same finite power v_S^2 and every noise spectrum in $\mathcal{N}$ has the same finite power v_N^2 then we can define classes of PDF's

$$\mathcal{P}_S = \{f_S | f_S(\omega) = \sigma(\omega)/2\pi v_S^2,\ \sigma \in \mathcal{J}\}$$

and

$$\mathcal{P}_N = \{f_N | f_N(\omega) = \nu(\omega)/2\pi v_N^2,\ \nu \in \mathcal{N}\}$$

and possibly apply the following ([2], Corollary 1).

<u>Theorem 2.1</u>: If $\mathcal{J}$ and $\mathcal{N}$ are convex and have constant powers v_S^2 and v_N^2, respectively, and $q_S \in \mathcal{P}_S$ and $q_N \in \mathcal{P}_N$ are least-favorable PDF's for $\mathcal{P}_S$ versus $\mathcal{P}_N$

then $\sigma_L \triangleq 2\pi v_S^2 q_S$ and $\nu_L \triangleq 2\pi v_N^2 q_N$ are least favorable spectra for Wiener filtering for $\mathcal{S}$ and $\mathcal{N}$.

This theorem allows us to construct most-robust Wiener filters for the first two examples considered in Section 2.

<u>The p-point class</u>. It can be seen from [25] that

$$H_R^*(\omega) = \begin{cases} \dfrac{p_S v_S^2}{p_S v_S^2 + p_N v_N^2} & \text{for } \omega \in S \\[2ex] \dfrac{(1-p_S)v_S^2}{(1-p_S)v_S^2 + (1-p_N)v_N^2} & \text{for } \omega \in S^c \end{cases}$$

and, hence,

$$e(\sigma,\nu;H_R^*) = \frac{p_S p_N}{p_S r + p_N} + \frac{(1-p_S)(1-p_N)}{(1-p_S)r + (1-p_N)} \text{ for all } (\sigma,\nu) \in \mathcal{S} \times \mathcal{N},$$

where $r \triangleq v_S^2/v_N^2$, the input SNR. In Fig. 4 we have superimposed onto Fig. 1 the performance of H_R^* (the middle line). It is clear from Fig. 4 that, unless we are extremely certain about our choice of σ and ν, H_R^* is preferable to H_0^*.

<u>The ε-contaminated class</u>. For the classes in (2.5) it can be easily seen from the above theorem and [14] that

$$H_R^*(\omega) = \begin{cases} k' \triangleq c'r/(c'r + 1) & \text{for } H_0^*(\omega) \le k' \\ H_0^*(\omega) & \text{for } k' < H_0^*(\omega) < k'' \\ k'' \triangleq c''r/(c''r + 1) & \text{for } H_0^*(\omega) \ge k'' , \end{cases}$$

where $0 \le c' < c'' \le \infty$ are constants given by Huber [14]. It is interesting to note that the robust filter H_R^* has this same form for several other

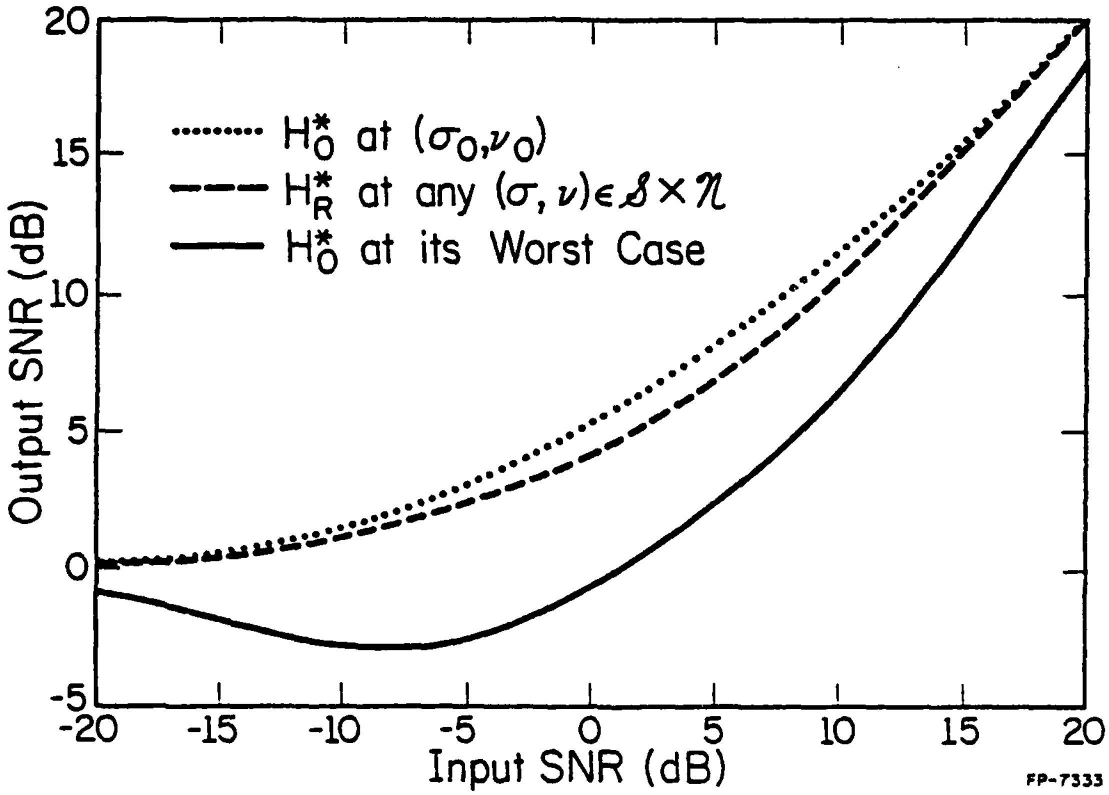


FIG. 4. p-point example. (From top to bottom) H_0^* at (σ_0,ν_0); H_R^* at any (σ,ν) ε $\mathcal{S} \times \mathcal{N}$; H_0^* at its worst case.

uncertainty models (see [2], [27]). Also note that this H_R^* will not have constant MSE over $\mathcal{S} \times \mathcal{N}$ as in the previous example. In Fig. 5 we have superimposed onto Fig. 2 the performance of H_R^* when the true spectral pair is (σ_0, ν_0) (the second line from the top) and when the true spectral pair is (σ_L, ν_L) (the third line from the top). Recall from the definition of (σ_L, ν_L) that the latter is the worst-case performance of H_R^*. For this example $c' = 1/c'' = 0.125$.

Unlike the preceding example, the preferability of the most-robust filter is not so clear-cut. If one were relatively certain about (σ_0, ν_0) being correct then H_0^* would be the better choice; however, if not, and if the guaranteed level of performance over $\mathcal{S} \times \mathcal{N}$ (given by the third line down) were adequate, we would likely choose H_R^*.

An ε-contaminated signal in white noise. Clearly the above theorem cannot be applied to find a robust filter in this case since the noise has infinite power; however a more direct approach proves fruitful here. First, we may restrict our search to $H \in L_2(d\omega)$, the mean-square integrable functions on $\mathbb{R}$, the real line, since all others have infinite MSE regardless of what σ is (cf. equation (2.1)). Second, we have, for all $H \in L_2(d\omega)$,

$$\sup_{\sigma \in \mathcal{S}_\varepsilon} e(\sigma, \nu_0; H) = \sup_{\sigma \in \mathcal{S}_\varepsilon} \frac{1}{2\pi} \int_{\mathbb{R}} [|1-H(\omega)|^2 ((1-\varepsilon)\sigma_0(\omega) + \varepsilon\sigma'(\omega)) + |H(\omega)|^2 \frac{N_0}{2}]\, d\omega$$

$$= e((1-\varepsilon)\sigma_0, \nu_0; H) + \varepsilon \sup_{\sigma'} \frac{1}{2\pi} \int_{\mathbb{R}} |1-H(\omega)|^2 \sigma'(\omega)\, d\omega \tag{2.9}$$

$$\leq e((1-\varepsilon)\sigma_0, \nu_0; H) + \varepsilon v_S^2.$$

The last step is true because $H \in L_2(d\omega)$ and it is assumed that $\int_{\mathbb{R}} \sigma'(\omega)d\omega = 2\pi v_s^2$.

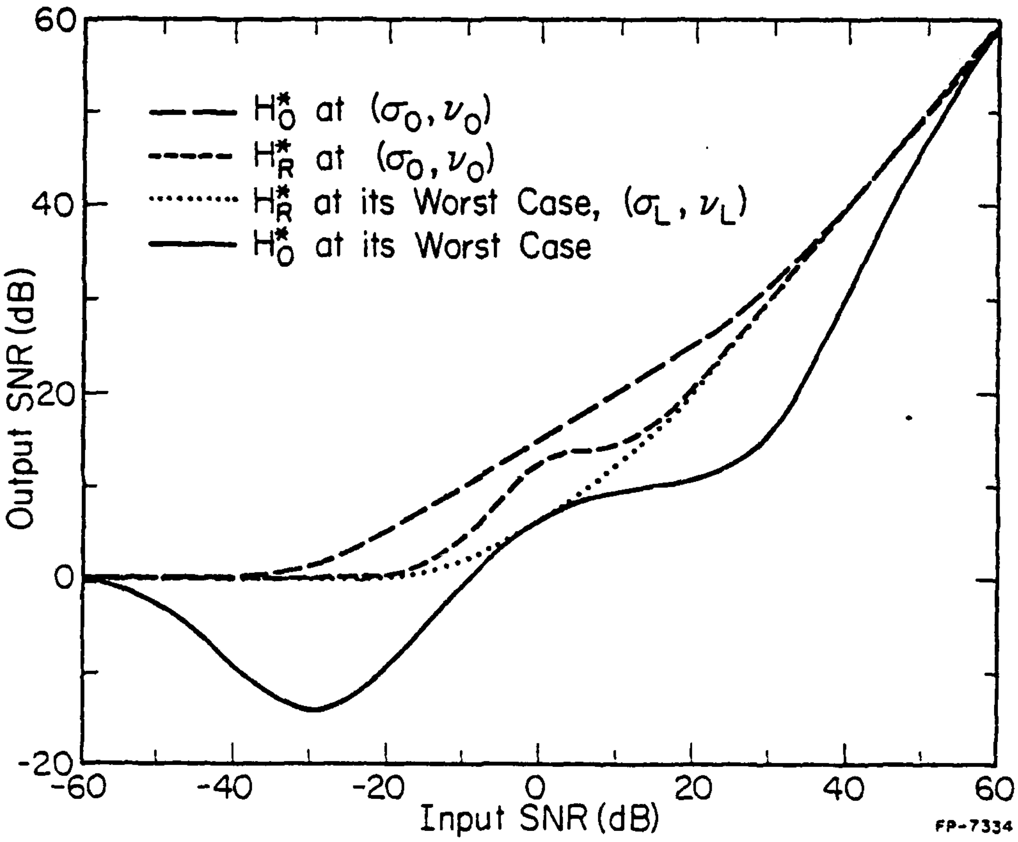


FIG. 5. ε-contaminated example. (From top to bottom) H_0^* at (σ_0,ν_0); H_R^* at (σ_0,ν_0); H_R^* at (σ_L,ν_L) (H_R^*'s worst case); H_0^* at its worst case.

Clearly equation (2.2) and equation (2.6) (the definition of H_R^*) imply that

$$\frac{(1-\varepsilon)\sigma_0(\omega)}{(1-\varepsilon)\sigma_0(\omega) + N_0/2} \tag{2.10}$$

minimizes (over H) the last expression in (2.9). But, for the value of H given in (2.10), we have equality in (2.9). Thus, H_R^* for this problem is given by (2.10).

Recall that Fig. 3 showed the performance of H_0^* in this situation in its nominal and worst cases. If we superimposed the nominal and worst cases of H_R^* onto Fig. 3, as we have done for the other examples, we would find no change; i.e., up to the accuracy of the graph the nominal cases and worst cases of H_0^* and H_R^* are the same. In fact, they differ by no more than 0.01. It should be noted that this is a singular example and the unusual performance is due to the infinite power of the white noise, not to the "very wide bandedness" which white noise is generally used to model.

4. Discussion and Conclusions

As we have discussed above, the results presented in this chapter (with the one exception of the white noise example) are representative of our findings in a wide variety of cases. For example, although it is a much harder case to solve, we have developed numerical results for causal Wiener filtering of an ε-contaminated first-order Markov signal in first-order Markov noise. The theory of the causal case has not been developed in the same generality as the noncausal case; however, this specific example can be treated using the results of Poor [2] and Yao [16]. In Fig. 6 we have presented the results for this causal filtering example

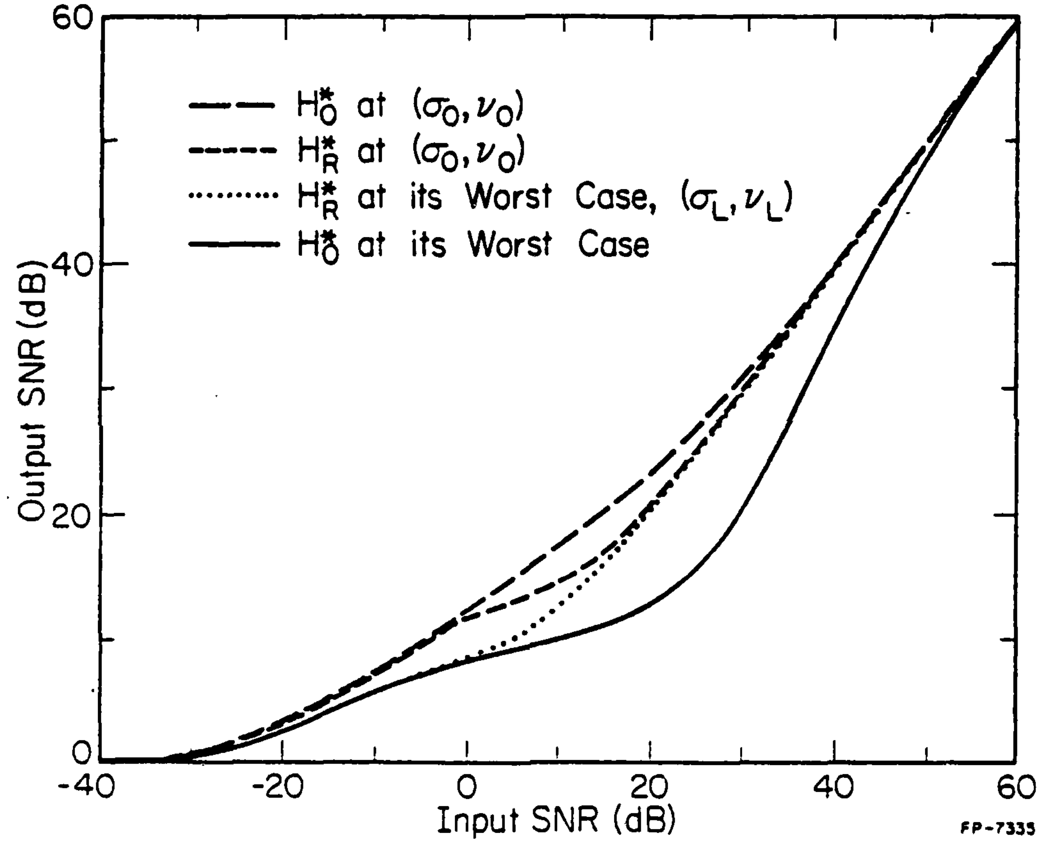


FIG. 6. Causal example. (From top to bottom) H_0^* at (σ_0, ν_0); H_R^* at (σ_0, ν_0); H_R^* at (σ_L, ν_L) (H_R^*'s worst case); H_0^* at its worst case.

with $\varepsilon = 0.1$, $\alpha_S = 1$ and $\alpha_N = 1000$. For comparison we have also included Fig. 7 which gives the results for the corresponding noncausal case. Note the similarity between the two figures. Again, this is indicative of our findings over a wide range of signal and noise bandwidths and ε's (see the appendix).

Other situations we examined in the noncausal case include ones with σ and/or ν as second-order Markov (i.e. having the form $4\alpha^3 v^2/(\alpha^2 + \omega^2)^2$) or using bandlimited white noise. The results for all these cases were similar to those already presented (e.g. Fig. 5). (Again, see the appendix.) Of particular interest is the case of an ε-contaminated first-order Markov signal in ε-contaminated bandlimited white noise. Even when the bandwidth of the noise was extremely large (e.g. 10^6) the results were similar to the other cases and unlike those involving nonbandlimited white noise (cf. the remarks at the end of Section 3 and Figure 22 in the appendix).

In summary, the Wiener filter can be undesirably sensitive to small deviations from assumed spectral models. Furthermore, while there are enough specific cases to the contrary to make caution advisable, we have found for a wide variety of situations that, when spectral uncertainty exists, the robust Wiener filter is generally preferable to the traditional Wiener filter.

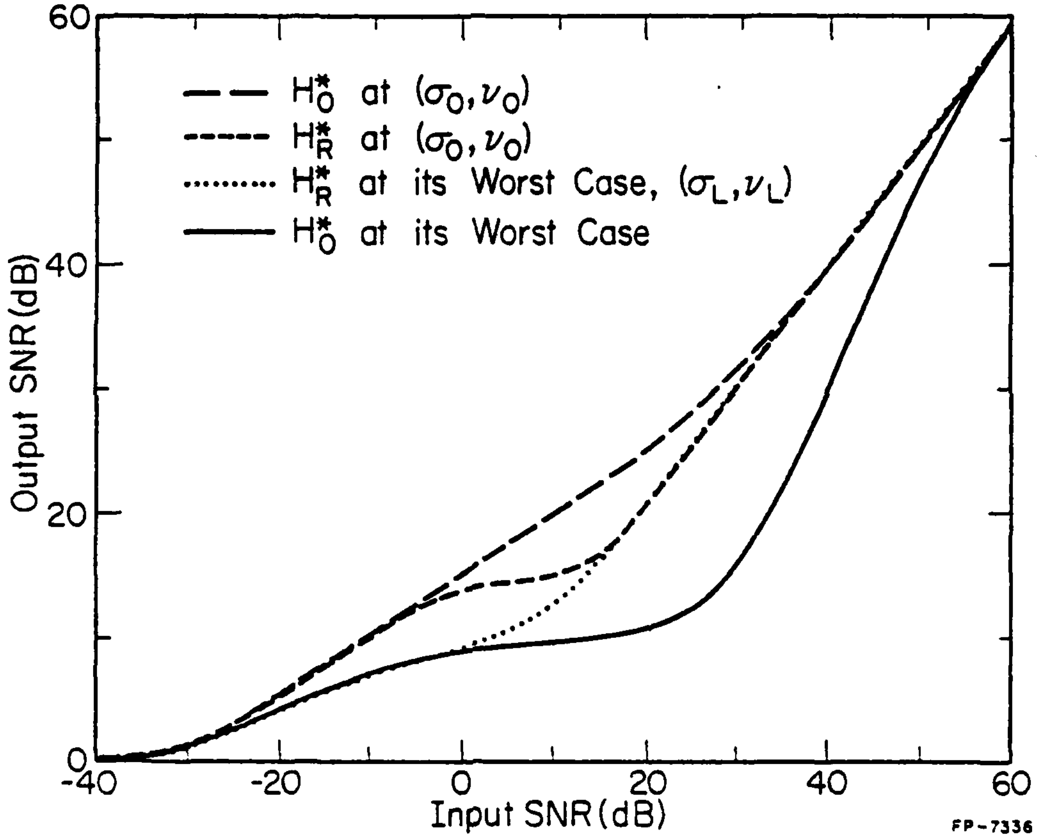


FIG. 7. Noncausal version of FIG. 6 (all parameters the same).

III. ROBUST WIENER-KOLMOGOROV THEORY

1. Introduction

We saw in Chapter II that optimal linear continuous-time (Wiener) noncausal filters can be undesirably sensitive to spectral uncertainty and that often the robust Wiener filters developed in [2] are preferable because of their insensitivity to such uncertainty. In this chapter we consider a general formulation of robust linear discrete-time causal estimation of a linear function of a wide-sense stationary signal. This formulation is analogous to the robust Wiener noncausal filtering development in [2] which is summarized in Chapter II, Section 3. Recall that the essential steps of this formulation consist of, first, choosing two classes of spectra which model the signal and noise uncertainty and, second, finding the signal estimator which minimizes the maximum error over these spectral uncertainty classes. Our main results yield, under mild conditions on these spectral uncertainty classes, a method of designing robust n-step predictors, robust causal filters and robust finite-lag smoothers and guarantee their existence. In order to illustrate this method of design, the special case of robust one-step noiseless prediction is developed in detail. Also, numerical examples are given for robust causal filtering of an uncertain signal in white noise.

In Section 2 we briefly present the traditional discrete-time (Wiener-Kolmogorov) signal estimation problem [28] and discuss those aspects which will be relevant in the sequel. In Section 3, we present the robust version of this problem and state and prove the main theorems.

Section 3 includes a discussion of commonly used models of uncertainty. In Section 4, we apply the results of Section 3 to the problem of robust one-step noiseless prediction. For this case, explicit expressions are developed using an analogy with robust hypothesis testing originally developed in [2]. In Section 5, examples are given and discussed. In Section 6 we consider the development of Section 3 in the more general situation when the signal and noise are represented by spectral distributions rather than spectral densities as in Section 3. Results of a somewhat different nature are obtained. Finally Section 7 contains conclusions and general discussion.

2. Background and Preliminaries

Throughout this chapter we assume that we observe a portion of a realization $\{y(k) | k \in \mathbf{Z}, k \leq k_0\}$ of a random process $\{Y(k) | k \in \mathbf{Z}\}$, where $\mathbf{Z}$ denotes the set of integers, and we assume that $Y(k) = S(k) + N(k)$, $\forall\, k \in \mathbf{Z}$, where $\{S(k) | k \in \mathbf{Z}\}$ and $\{N(k) | k \in \mathbf{Z}\}$ are second-order, wide-sense stationary random processes which are uncorrelated with each other. We can also assume that $\{N(k) | k \in \mathbf{Z}\}$ is zero-mean. The processes $\{S(k)\}$ and $\{N(k)\}$ represent signal and noise, respectively.

Our purpose is to form a linear causal estimate of a linear function of $\{s(k)\}$ from the observation $\{y(k)\}$. That is, we are given some function of the signal having the form

$$\ell_d(k) = \sum_{n=-\infty}^{\infty} d(k-n)s(n) \tag{3.1}$$

and we wish to find the "best" estimate of $\{\ell_d(k)\}$ among all estimates having the form

$$\hat{\ell}_h(k) = \sum_{n=-\infty}^{k} h(k-n)y(n) \tag{3.2}$$

where h satisfies $h(n) = 0$ for $n < 0$ (causality). The usual criterion of optimality is to find such an h minimizing the mean-square error (MSE).

For given functions d and h the well-known formula for the MSE is

$$E\{[\ell_d(k)-\hat{\ell}_h(k)]^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}[|D(\theta)-H(\theta)|^2 f_S(\theta)+|H(\theta)|^2 f_N(\theta)]d\lambda(\theta)$$

$$\triangleq e_D(f_S,f_N;H) \tag{3.3}$$

where D and H are the transfer functions (i.e., Fourier transforms) of the transformations d and h, respectively, and f_S and f_N are the power spectral densities (PSD's) of $\{S(k)\}$ and $\{N(k)\}$, the signal and noise, with respect to the finite measure λ. In general, we will take λ to be Lebesgue measure on $[-\pi,\pi]$ (so that f_S and f_N are just the usual PSD's) but it also might be convenient, for example, to allow λ to include some point masses in order to represent pure sinusoids in a mathematically rigorous fashion. (More will be said about this in Section 6.)

We refer to the transformation d (or its transfer function D) as the desired operation. The cases of greatest interest occur when $d(n) = 1$ for $n = n_0$ and $d(n) = 0$ for $n \neq n_0$ (correspondingly $D(\theta) = e^{in_0\theta}$). If $n_0 = 0$ our problem is causal filtering, if $n_0 < 0$ it is prediction n_0 steps ahead, and if $n_0 > 0$ we have smoothing with fixed lag n_0.

If the PSD's f_S and f_N are known and the desired operation $D(\theta)$ is given then we would like to minimize $e_D(f_S,f_N;H)$ over all causal transfer functions. We denote by $H^\dagger$ the solution to this minimization problem and refer to $H^\dagger$ as the <u>optimal causal transfer</u> function for (f_S,f_N). Note that $H^\dagger$ is unique a.e. $(f_S + f_N)d\lambda$ (i.e., if H' also minimizes $e_D(f_S,f_N;H)$ over all causal transfer functions then $\int_{\{H' \neq H^\dagger\}} [f_S(\theta) + f_N(\theta)]d\lambda(\theta) = 0$, see [3]).

For the remainder of this chapter we will assume that the PSD's f_S and f_N are bounded a.e. $[\lambda]$. Further, we will only consider filters H which are mean-square integrable on $[-\pi,\pi]$ with respect to λ, i.e. we will assume that $H \in L^2 \triangleq L^2(d\lambda(\theta))$. Of course, we only want to consider those $H \in L^2$ which are causal transfer functions. The set of such H is denoted by $H_+^2 \triangleq H_+^2(d\lambda(\theta))$ and can be defined as the (closed) subspace of the Hilbert space L^2 which is spanned by $\{e^{in\theta} \mid n = 0,1,2,\ldots\}$. H_+^2 is called a Hardy space (see [3],[4], or [18]). With these definitions the minimum-MSE problem for a specific pair of PSD's (f_S,f_N) and desired operation $D(\theta)$ can be formulated as

$$e_D^\dagger(f_S,f_N) \triangleq \min_{H \in H_+^2} e_D(f_S,f_N,H) \quad . \tag{3.4}$$

3. <u>Robust Linear Estimation of a Signal in Noise</u>

Throughout this section we will assume that the desired operation $D(\theta)$ is fixed and bounded. It is clear from Section 2 that the solution, $H^\dagger(\theta)$, to the minimum-MSE estimation problem (3.4) depends entirely on the signal and noise PSD's, f_S and f_N. As we discussed in Section 1, the spectra we choose for designing H may differ somewhat from the true signal and noise spectra. We model this spectral uncertainty by choosing appropriate classes of PSD's, $\mathcal{S}$ and $\mathcal{N}$, and assuming that $f_S \in \mathcal{S}$ and $f_N \in \mathcal{N}$. In other words, we know that f_S belongs to $\mathcal{S}$ and that f_N belongs to $\mathcal{N}$, but we do not know which elements of $\mathcal{S}$ and $\mathcal{N}$ represent the true signal and noise spectra.

Clearly any transfer function H_R for which there exists a finite upper bound on the MSE, $e_D(f_S,f_N;H_R)$ over all f_S in $\mathcal{S}$ and f_N in $\mathcal{N}$, could be termed robust since a certain level of performance can be guaranteed by using H_R. (That level being: $\sup_{(f_S,f_N)\in \mathcal{S}\times\mathcal{N}} e_D(f_S,f_N;H_R)$.) Ideally we would like a (causal) transfer function with the smallest possible such upper bound, i.e., we would like to find a solution $H_R^\dagger$ to the game

$$\inf_{H\in H_+^2} \sup_{(f_S,f_N)\in \mathcal{S}\times\mathcal{N}} e_D(f_S,f_N;H). \tag{3.5}$$

As in Chapter II, we refer to $H_R^\dagger$ as a most-robust causal transfer function for the spectral uncertainty classes $\mathcal{S}$ and $\mathcal{N}$.

We now give some specific forms for the uncertainty classes $\mathcal{S}$ and $\mathcal{N}$. These forms have been widely used to model uncertinty in both the engineering and statistics literature. We will exhibit these forms for the class $\mathcal{S}$ but, of course, they could just as easily be used to model noise spectral uncertainty.

Certainly the most commonly used uncertainty class ([6],[7],[14],[20],[23],[32],[37],[39],[40],[43],[44],[46]-[48]) is the ε-contaminated model (also called the ε-mixture or gross-error model). It has the form

$$S_\varepsilon \triangleq \{f_s | f_s(\theta) = (1-\varepsilon)f_s^0(\theta) + f_s'(\theta) \forall\theta\varepsilon[-\pi,\pi],$$

$$\int_{-\pi}^{\pi} f_s'(\theta)d\lambda(\theta) = \int_{-\pi}^{\pi} f_s^0(\theta)d\lambda(\theta)\} \tag{3.6}$$

where f_s^0 is a nominal PSD and ε $(0 \le \varepsilon \le 1)$ is the contamination parameter. This class is probably the most popular for representing uncertainty because it models the idea that we have ε of completely general uncertainty about our choice of the PSD f_s^0.

Another common model ([2],[6],[14],[20],[44]) is the <u>total variation</u> model which has the form

$$S_{TV} \triangleq \{f_s | \frac{1}{2}\int_{-\pi}^{\pi} |f_s^0(\theta) - f_s(\theta)| d\lambda(\theta) \le \varepsilon \quad ,$$

$$\int_{-\pi}^{\pi} f_s(\theta)d\lambda(\theta) = \int_{-\pi}^{\pi} f_s^0(\theta)d\lambda(\theta)\} \tag{3.7}$$

where, again, f_s^0 is a nominal PSD and ε an uncertainty parameter.

A third model is the <u>band</u> model ([1],[21],[22]) which has the form

$$S_B \triangleq \{f_s | f_s^L(\theta) \le f_s(\theta) \le f_s^U(\theta) \ \forall\theta, \ \int_{-\pi}^{\pi} f_s(\theta)d\lambda(\theta) = 2\pi w\} \tag{3.8}$$

where $\int_{-\pi}^{\pi} f_s^L(\theta)d\lambda(\theta) \le 2\pi w \le \int_{-\pi}^{\pi} f_s^U(\theta)d\lambda(\theta)$ and w is the (known) power of the signal. The name band model comes from the idea that f_s^L and f_s^U are the lower and upper bounds of a confidence band around a spectral estimate.

The last model is the <u>p-point</u> (or Sakrison's class b) model ([5],[8], [25],[41],[42]) which has the form

$$S_p \triangleq \{f_s \mid \int_{A_i} f_s(\theta)d\lambda(\theta) = 2\pi w_i \ , \ i=1,\ldots,n\} \tag{3.9}$$

where the A_i's are a partition of $[-\pi,\pi]$ and $\sum_{i=1}^{n} w_i = w$, the power of the signal. A p-point class is an appropriate model of uncertainty in situations where, for example, we can accurately measure the power p_i in each interval $A_i \triangleq [\theta_{i-1},\theta_i]$ (where $-\pi = \theta_0 < \theta_1 < \ldots < \theta_n = \pi$) using a nested bank of low pass filters or a bank of band pass filters. Note that unless n is quite large we are allowing a considerable amount of uncertainty when we use a p-point class. It might be more reasonable to obtain also some other form of spectral estimate and use a band model in addition (i.e., let our class be the intersection of a p-point model with a band model). We call this a banded p-point model.

We note that for each of these classes we have assumed that the power is known. Often it is a reasonable assumption that the power can be accurately estimated even though the shape of the PSD is uncertain. Furthermore, in all the specific cases in which we have found most-robust filters (see Section 5) it turns out that they do not depend on the specific signal and noise powers (w_S and w_N, respectively) but only on the input (to the filter) signal-to-noise ratio w_S/w_N.

Returning now to the definition of a most-robust filter (3.5) for general $\mathcal{S}$ and $\mathcal{N}$, we note that, while the minimum-MSE transfer function is known to exist for each $(f_S,f_N) \in \mathcal{S} \times \mathcal{N}$, this is no guarantee that the infimum in (3.5) is achieved. In fact, it need not even be finite. Of course, even if it were not achieved there would still exist transfer functions whose worst-case MSE over $\mathcal{S} \times \mathcal{N}$ would be arbitrarily close to this infimum.

In the following theorem we give some mild sufficient conditions for the infimum in (3.5) to be achieved and finite.

Theorem 3.1. If the spectral uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ are such that the following two conditions hold

i) $\sup_{\mathcal{S}} \frac{1}{2\pi} \int_{-\pi}^{\pi} f_S(\theta) d\lambda(\theta) = w_S < \infty$

ii) Either (a) or (b) holds for some $\varepsilon > 0$:

a) There is an $\overline{f}_N \in \mathcal{N}$ such that $\overline{f}_N(\theta) \geq \varepsilon > 0$ a.e.[λ],

b) There is an $\overline{f}_S \in \mathcal{S}$ such that $\overline{f}_S(\theta) \geq \varepsilon > 0$ a.e.[λ];

then there exists H_R^{+} achieving the infimum in (3.5). Furthermore this infimum is finite.

Proof. Let $H_0(\theta) = 0, \forall\theta$. (Note that $H_0 \in H_+^2$.) We have

$$
\begin{aligned}
\inf_{H_+^2} \sup_{\mathcal{S}\times\mathcal{N}} e_D(f_S; f_N; H) &\leq \sup_{\mathcal{S}\times\mathcal{N}} e_D(f_S, f_N, H_0) \\
&= \sup_{f_S \in \mathcal{S}} \int_{-\pi}^{\pi} |D(\theta)|^2 f_S(\theta) d\lambda(\theta) \\
&= B_D \sup_{f_S \in \mathcal{S}} \int_{-\pi}^{\pi} f_S(\theta) d\lambda(\theta) \\
&= B_D\, w_S < \infty
\end{aligned}
$$

where B_D is the essential supremum of $|D|^2$. Hence the infimum in (3.5) is finite and for any fixed M satisfying $B_D w_S < M < \infty$ we may exclude from consideration any $H \in H_+^2$ which satisfies $e_D(f_S,f_N;H) > M$ for some $(f_S,f_N) \in \mathcal{S} \times \mathcal{N}$. We will now translate this to a bound on $\|H\| \triangleq \left[\int_{-\pi}^{\pi} |H(\theta)|^2 d\lambda(\theta)\right]^{1/2}$ which is the norm of H in the Hilbert space H_+^2.

If condition (a) of (ii) in the statement of the theorem holds then for any $f_S \in \mathcal{S}$ and any $H \in H_+^2$ we have

$$
\begin{aligned}
e_D(f_S,\overline{f}_N;H) &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\theta)|^2 \overline{f}_N(\theta) d\lambda(\theta) \\
&\geq \frac{\varepsilon}{2\pi} \|H\|^2 .
\end{aligned}
$$

Hence in this case we may exclude any H satisfying $\|H\|^2 > 2\pi M/\varepsilon$. If, on the other hand, condition (b) of (ii) holds then for any $f_N \in \mathcal{N}$ and any $H \in H_+^2$ we have

$$
\begin{aligned}
e_D(\overline{f}_S,f_N;H) &\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} |D(\theta)-H(\theta)|^2 \overline{f}_S(\theta) d\lambda(\theta) \\
&\geq (\varepsilon/2\pi) \|D-H\|^2 \\
&\geq (\varepsilon/2\pi)[\,\|D\| - \|H\|\,]^2 .
\end{aligned}
$$

Hence we may exclude any H satisfying $\|H\| > \|D\| + \sqrt{2\pi M/\varepsilon}$.

Thus we have shown that if $B > \|D\| + \sqrt{2\pi B_D w_S/\varepsilon}$ then we have

$$
\inf_{H_+^2} \sup_{\mathcal{S} \times \mathcal{N}} e_D(f_S,f_N;H) = \inf_{H_+^2(B)} \sup_{\mathcal{S} \times \mathcal{N}} e_D(f_S,f_N;H) \tag{3.10}
$$

where $H_+^2(B) \triangleq [H \in H_+^2 |\ \|H\| \leq B]$. We will now show that the right hand side of (3.10) has a solution.

Step 1: $e_D(f_S, f_N; \cdot)$ is continuous on H_+^2 for each $(f_S, f_N) \in \mathcal{S} \times \mathcal{N}$.

We have

$$
\begin{aligned}
&2\pi |e_D(f_S, f_N; H_n) - e_D(f_S, f_N; H)| \\
&\leq \left| \int |D-H_n|^2 f_S d\lambda - \int |D-H_n|^2 f_S d\lambda \right| + \left| \int |H_n|^2 f_N d\lambda - \int |H|^2 f_N d\lambda \right| \\
&\leq B_{f_S} \left| \|D-H_n\|^2 - \|D-H\|^2 \right| + B_{f_N} \left| \|H_n\|^2 - \|H\|^2 \right|
\end{aligned}
$$

where B_{f_S} and B_{f_N} are the essential supremums of f_S and f_N, respectively. Now if $\|H_n - H\| \to 0$ then $\|(D-H_n)-(D-H)\| \to 0$. Hence from the inequality $\|H_1 - H_2\| \geq | \|H_1\| - \|H_2\| |$ we have that $\|H_n\| \to \|H\|$ and $\|D-H_n\| \to \|D-H\|$. Hence $e_D(f_S, f_N; \cdot)$ is continuous.

Step 2: $\sup_{\mathcal{S} \times \mathcal{N}} e_D(f_S, f_N; \cdot)$ is lower semicontinuous (l.s.c.) on H_+^2.

This is a straightforward consequence of Step 1 and Corollary 1.1, p. 77, in [55] which states that the supremum over a family of l.s.c. functions is l.s.c.

Step 3: $\sup_{\mathcal{S} \times \mathcal{N}} e_D(f_S, f_N; \cdot)$ is convex on H_+^2. This is a straightforward computation.

Step 4: Apply Theorem 1.2, p. 79, in [55], to $\sup_{\mathcal{S} \times \mathcal{N}} e_D(f_S, f_N; \cdot)$ on $H_+^2(B)$.

This Theorem 1.2 states in our case that because H_+^2 is a reflexive Banach space (see [4]); because $\sup e_D(f_S, f_N; \cdot)$ is l.s.c. (Step 2), convex (Step 3), and proper (this is trivial since $e(f_S, f_N; H) \geq 0 > -\infty$) on H_+^2 : and because

$H_+^2(B)$ is convex, closed, and bounded (these can be deduced from the fact that $H_+^2(B)$ is just a multiple of the closed unit ball of H_+^2, see [10]); we have that the infimum in (3.5) is achieved by some $H_R^\dagger$. QED.

As we noted above, the conditions of Theorem 3.1 are mild. Certainly any problem in which there was no upper bound on the signal power would be unreasonable. Further, the condition (iia) [resp. (iib)] is satisfied if $\mathcal{N}$ [resp. $\mathcal{S}$] has any of the forms discussed earlier (i.e., ε-contaminated, etc.). The only possible nontrivial exception to this would be in the case of the p-point class if some $w_i=0$ while $\lambda([\theta_{i-1},\theta_i]) > 0$.

We now turn our attention to the problem of finding H_R^+, the most robust transfer function. We begin with a definition.

Definition 3.1. A pair of PSD's (f_S^L, f_N^L) is least favorable for causal estimation[1] for the uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ if

$$e_D(f_S^L, f_N^L) = \max_{\mathcal{S} \times \mathcal{N}} e_D^+(f_S, f_N) \tag{3.11}$$

where $e_D^+(f_S, f_N)$, the minimum-MSE for (f_S, f_N), is defined in (3.4).

Note that (3.11) means that (f_S^L, f_N^L) solves the maximin game

$$\max_{\mathcal{S} \times \mathcal{N}} \; \min_{H_+^2} \; e_D(f_S, f_N, H) \quad . \tag{3.12}$$

Hence, if the minimax equality holds here (i.e., if (3.12) equals (3.5))

[1] This definition differs from the ones given in [1] and [2] (see Chapter 2) but is consistent with earlier notions of least favorability (see, for example, [26], p. 34). As was pointed out by Verdu ([54], p. 72), this discrepancy seems to have had its origins in a somewhat confusing discussion in [14].

then (f_S^L, f_N^L) is a least favorable pair if and only if (f_S^L, f_N^L) and its optimal causal transfer function H_L^+ form a saddle point solution to the game (3.5) (or, equivalently, the game (3.12)); that is (f_S^L, f_N^L) and H_L^+ satisfy

$$e_D(f_S, f_N; H_L^\dagger) \le e_D(f_S^L, f_N^L; H_L^\dagger) \le e_D(f_S^L, f_N^L; H) \tag{3.13}$$

(For a clear and thorough discussion of this point see [55], especially Section 2.3.1.) Clearly, if (3.13) holds then H_L^+ is a most robust causal transfer function.

Our next theorem gives some conditions under which the optimal transfer function for a least favorable pair is most robust. This is useful because it is often easier to solve the maximization problem (3.11) than it is to solve the minimax game (3.5).

Theorem 3.2. If the spectral uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ are such that the following three conditions hold

i) $\sup \frac{1}{2\pi} \int_{-\pi}^{\pi} f_S(\theta) d\lambda(\theta) = w_S < \infty$.

ii) $\mathcal{S}$ and $\mathcal{N}$ are convex.

iii) At least one of (a) or (b) holds for some $\varepsilon > 0$.

a) Every $f_N \in \mathcal{N}$ satisfies $f_N(\theta) \geq \varepsilon > 0$ a.e. $[\lambda]$.

b) Every $f_S \in \mathcal{S}$ satisfies $f_S(\theta) \geq \varepsilon > 0$ a.e. $[\lambda]$.

then a pair of PSD's (f_S^L, f_N^L) in $\mathcal{S} \times \mathcal{N}$ and its optimal causal transfer function H_L^+ form a saddlepoint solution to the minimax game (3.5) if and only if (f_S^L, f_N^L) is least favorable for causal estimation, i.e., solves

(3.11) (note that in this case $H_L^\dagger$ is a most robust causal transfer function).

Proof: We will need the following lemma (which is Theorem 2 in [11])

Lemma 3.1: Let X be a compact Hausdorff topological space and Y an arbitrary set. Let F be a real-valued function on X × Y such that for every $y \in Y$ $F(x,y)$ is l.s.c. on X. If for each $y \in Y$ $F(x,y)$ is convex on X and for each $x \in X$ $F(x,y)$ is concave on Y then

$$\min_{x\in X} \sup_{y\in Y} F(x,y) = \sup_{y\in Y} \min_{x\in X} F(x,y). \tag{3.14}$$

We wish to apply Lemma 3.1 with $X = H_+^2(B)$ ($H_+^2(B)$ was defined in the proof of Theorem 3.1), $Y = \mathcal{S} \times \mathcal{N}$, and $F(x,y) = F(H,(f_S,f_N)) = e_D(f_S,f_N;H)$. Since H_+^2 is a reflexive Banach space $H_+^2(B)$ is compact in the weak topology (see [10], Chapter V, especially Theorem V.4.7, or [55], Sections 1.2.2 and 1.2.3). Furthermore we saw in the proof of Theorem 3.1 that $e_D(f_S,f_N;\cdot)$ is continuous (hence l.s.c.) in the norm topology of $H_+^2(B)$ and that $e_D(f_S,f_N;\cdot)$ is convex and proper on $H_+^2(B)$. Thus by Proposition 1.5 of [55] we have that $e_D(f_S,f_N;\cdot)$ is l.s.c. in the weak topology of $H_+^2(B)$. The final condition of Lemma 3.1 is that $e_D(\cdot,\cdot;H)$ is concave on $\mathcal{S} \times \mathcal{N}$. But this is trivial since it is even linear. Thus we have shown that for any $B > 0$ we have

$$\min_{H_+^2(B)} \sup_{\mathcal{S}\times\mathcal{N}} e_D(f_S,f_N;H) = \sup_{\mathcal{S}\times\mathcal{N}} \min_{H_+^2(B)} e_D(f_S,f_N;H) . \tag{3.15}$$

Note that since the conditions of this theorem are stronger than the conditions of Theorem 3.1 we have from the proof of Theorem 3.1 that if B is large enough then the left hand side of (3.15) is equivalent to (3.5).

We will now show that the right hand side of (3.15) is equivalent to the right hand side of (3.11).

If (say) condition (a) of (iii) holds than for any $(f_S,f_N)\in \mathcal{S}\times\mathcal{N}$ and any $H \in H_+^2$ we have that

$$\begin{aligned} e_D(f_S,f_N;H) &\geq \frac{1}{2\pi} \int |H|^2 f_N d\lambda \\ &\geq (\varepsilon/2\pi) \int |H|^2 d\lambda \\ &= (\varepsilon/2\pi) \, \|H\|^2 \quad . \end{aligned}$$

Hence for any $B > \sqrt{2\pi e_D^\dagger(f_S,f_N)/\varepsilon}$ we have

$$\min_{H_+^2} \; e_D(f_S,f_N;H) = \min_{H_+^2(B)} \; e_D(f_S,f_N;H)$$

Furthermore, if $B > \sqrt{2\pi \sup_{\mathcal{S}\times\mathcal{N}} e_D^\dagger(f_S,f_N)/\varepsilon}$ we see that the right hand side of (3.15) equals the right hand side of (3.11). We note that sup $e_D^\dagger(f_S,f_N)$ is always finite under the conditions of Theorem 3.1 (hence under those of Theorem 3.2) because it is always less than or equal to

$$\min_{H\in H_+^2} \; \sup_{\mathcal{S}\times\mathcal{N}} e_D^\dagger(f_S,f_N).$$

Thus we have shown that under condition (iiia) B can be chosen so that (3.15) implies

$$\min_{H_+^2} \; \sup_{\mathcal{S}\times\mathcal{N}} e_D(f_S,f_N;H) = \sup_{\mathcal{S}\times\mathcal{N}} \; \min_{H_+^2} \; e_D(f_S,f_N;H) \quad .$$

Knowing this we have (see Section 2.3.1 of [55]) that (3.11) is equivalent to the existence of a saddlepoint solution to the game (3.5) and that, in particular, any solution (f_S^L, f_N^L) to (3.11) and its optimal causal transfer function $H_L^\dagger$ form a saddlepoint solution to (3.5). Thus $H_L^\dagger$ is a most robust transfer function . Conversely, any pair of PSD's which together with its optimal transfer function forms a saddlepoint solution to (3.5) must also solve (3.11).

Finally we note that (just as in the proof of Theorem 3.1) we may obtain these same results if condition (iiib) holds instead of (iiia) and the theorem is proved.

While the conditions of Theorem 3.2 (especially (iii)) are not as innocuous as those of Theorem 3.1, they are satisfied by any ε-contaminated (see (3.6)) whose nominal PSD, f^0, satisfies $f^0(\theta) \geq \delta$ a.e. $[\lambda]$, for some $\delta > 0$, by any band model (see (3.8)) whose lower bound f^L satisfies $f^L(\theta) \geq \delta$ a.e. $[\lambda]$ and hence by any banded p-point model whose lower bound satisfies this condition. This is not an unreasonable condition since it only need apply to one of $\mathcal{S}$ and $\mathcal{N}$, not both. Generally the noise will be wide-band with respect to the signal and this condition will be satisfied by $\mathcal{N}$. Alternatively, we might assume there is a small white component to the noise. This added component is sometimes called a "noise floor" and has been used, for example, by Van Trees to make the problem of detection in nonwhite noise analytically tractable. See [57], Section 4.3, for further justification.

The benefit of Theorem 3.2 stems from the fact that the maximization in (3.13) should be easier to solve than the original minimax problem (3.5).

In general, this will only be so if we have a closed-form expression for $e_D^{+}(f_S,f_N)$ as (f_S,f_N) ranges over $\mathcal{S} \times \mathcal{N}$. Such expressions have been found in a variety of cases by various researchers (see [3], [4], [13],[15]-[19], [33],[34]). In particular Snyders [3] has developed a general method for finding such expressions when one of f_S and f_N is fixed and rational while the other is completely arbitrary.

4. Robust Noiseless One-Step Prediction

In this section we consider in greater detail the special case of prediction one step ahead of a signal with uncertain spectrum. The signal is assumed to be received in a noiseless environment. In other words, we consider the problem (3.5) with $D(\theta) = e^{-i\theta}$ and $f_N(\theta) = 0$ for all $\theta \in [-\pi,\pi]$. This special case is the one considered previously by Hosoya [23]. The conditions on our main theorem (Theorem 3.2) when applied to this case are slightly more restrictive than those of Hosoya's analogous result (Theorem 2, p. 581 in [23]). On the other hand, in [23] only the ε-contaminated class (see (3.6)) is considered and the proofs directly depend on the specific form of the least favorable PSD for this class, whereas our treatment is valid for more general uncertainty classes.

For a known signal PSD, $f_S(\theta)$, the one-step minimum-MSE noiseless prediction problem is given by

$$e_D^{+}(f_S,0) = \min_{H_+^2} e_D(f_S,0;H) \tag{3.16}$$

where $D(\theta) = e^{-i\theta}$ and e_D^{+}, e_D and H_+^2 are defined in Section 2. Note that this is just (3.4) with $f_N(\theta) = 0 \ \forall \ \theta \in [-\pi,\pi]$. In this section we refer to the solution to (3.16) as the <u>optimal linear predictor</u> for f_S.

The Szegö-Kolmogorov-Krein Theorem [3],[4] states that (with $D(\theta) = e^{-i\theta}$) we have

$$e_D^{\dagger}(f_S,0) = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_S(\theta)d\theta\right] \tag{3.17}$$

where the right hand side is interpreted as zero if $\log f_S$ is not integrable. (Technically, the f_S in (3.17) is the density of the absolutely continuous part of $f_S(\theta)d\lambda(\theta)$ with respect to Lebesgue measure on $[-\pi,\pi]$. However, since the case of greatest interest is when λ is Lebesgue measure and since any predictor can be adjusted to perfectly predict the singular part of any signal spectrum (see [18]) we assume here that f_S is just the usual (Lebesgue) PSD).

We now consider the case where the signal spectrum is uncertain. As in Section 3 we assume that we know only that $f_S \in \mathscr{S}$, and we wish to find $H_R^{\dagger}$ solving

$$\min_{H_+^2} \sup_{\mathscr{S}} e_D(f_S,0;H) \tag{3.18}$$

where, again, $D(\theta) = e^{-i\theta}$. In this section we refer to $H_R^{\dagger}$ as the <u>most robust linear predictor</u>. Also, we define f_S^L to be <u>least favorable for one-step noiseless prediction</u> for $\mathscr{S}$ if $(f_S^L,0)$ satisfies Definition 3.1 with $\mathcal{N} = \{0\}$. From (3.17) we see that f_S^L is least favorable for one-step noiseless prediction for $\mathscr{S}$ if and only if

$$\int_{-\pi}^{\pi} \log f_S^L(\theta)d\theta = \max_{\mathcal{S}} \int_{-\pi}^{\pi} \log f_S(\theta)d\theta \; . \tag{3.19}$$

Furthermore, under the conditions of Theorem 3.2 we have that f_S^L solves (3.19) if and only if f_S^L and $H_L^\dagger$ form a saddlepoint solution to (3.18), where $H_L^\dagger$ is the optimal linear predictor for f_S^L. Thus, if we can solve (3.19) then we can find the most robust linear predictor for $\mathcal{S}$.

We will now demonstrate how to obtain the exact form of a least favorable spectrum for each of the uncertainty classes discussed in Section 3. The method we use involves an analogy to robust hypothesis testing. The advantage of this is that it allows us to make use of the considerable effort already expended in finding least favorable probability densities for that problem (defined below). This analogy was the underlying basis for the solutions given in [1] and [22] and was developed explicitly in Section III of [2] (see Chapter II).

As in each of the classes of Section 3, we will assume that satisfies a power constraint, i.e., that $\int f_S(\theta)d\theta = 2\pi w_S$, where $0 < w_S < \infty$ is the power in the signal process $\{S(k)\}$. We can now define a class of probability densities on $[-\pi,\pi]$ by

$$\mathcal{P}_S = \{p_S(\theta) \,|\, p_S(\theta) = f_S(\theta)/2\pi w_S, f_S \in \mathcal{S}\} \tag{3.20}$$

and consider the following pair of statistical hypotheses concerning a random variable X on $[-\pi,\pi]$ and its Borel σ-field

$$H_0 : X \sim p(\theta) = \frac{1}{2\pi} \; \forall \; \theta \in [-\pi,\pi]$$

versus (3.21)

$$H_1 : X \sim p \in \mathscr{P}_S.$$

For any test ϕ of H_1 versus H_0 define $R_j(\phi;P)$ to be the conditional risk of using ϕ when $X \sim P$ under $H_j (j=0,1)$ (see [26]). Consider the following (see, for example, [26]).

<u>Definition 2</u>: For the hypothesis pair in (3.21), $f_S \in \mathscr{P}_S$ is <u>least favorable in terms of risk</u> for $\mathscr{P}_S$ versus $\{1/2\pi\}$ if

$$R_1(\phi',p_S) \leq R_1(\phi',q_S) \qquad p_S \in \mathscr{P}_S$$

for every probability ratio test ϕ' between q_S and $1/2\pi$.

Least favorable densities play an important role in the design of robust hypothesis tests (see [14], [6]). They have been found for the ε-contaminated and total-variation models ((3.6) and (3.7), respectively) b. Huber [14], [27]; for the band model (3.8) by Kassam [21]; and for the p-point model (3.9) (see Chapter V). Thus the importance of the following which is related to Lemma 1 of [2], Theorem 2 of [21] and results in [29] and [30].

<u>Proposition 3.1</u>: Let φ be any differentiable concave function on $(0,\infty)$. Let $\mathscr{P}_S$ be any convex set of probability densities on $[-\pi,\pi]$. If $q_S \in \mathscr{P}_S$ is least favorable in terms of risk for $\mathscr{P}_S$ versus the uniform distribution on $[-\pi,\pi]$ then, for all $p_S \in \mathscr{P}_S$, we have

$$\int\varphi(q_S(\theta))d\theta \geq \int\varphi(p_S(\theta))d\theta . \qquad (3.22)$$

__Proof__: We wish to show that $\int \varphi(p_S(\theta))d\theta$ achieves a maximum over $\mathcal{P}_S$ at q_S. Since $\int \varphi(\cdot)$ is concave we only need to show that

$$\frac{\partial}{\partial\varepsilon}\left[\int \varphi[(1-\varepsilon)q_S(\theta) + \varepsilon p_S(\theta)]d\theta\right]_{\varepsilon=0} \leq 0 \quad , \tag{3.23}$$

$\forall\ p_S \in \mathcal{P}_S$. The left hand side of (3.23) is

$$\lim_{\varepsilon\downarrow 0} \int \frac{\varphi[(1-\varepsilon)q_S(\theta) + \varepsilon p_S(\theta)] - \varphi(q_S(\theta))}{\varepsilon} d\theta \tag{3.24}$$

But, by the concavity of φ, as $\varepsilon\downarrow 0$ the function inside the integral in (3.24) converges pointwise up to $\varphi'[q_S(\theta)][p_S(\theta) - q_S(\theta)]$. So the left hand side of (3.23) is less than or equal to $\int \varphi'[q_S(\theta)][p_S(\theta) - q_S(\theta)]d\theta$. This latter term is nonpositive since (by concavity) φ' is nonincreasing and q_s is the distribution making q_S stochastically smallest over $\mathcal{P}_S$ (see [14], [27]) therefore implying

$$\int \varphi'(q_S(\theta))p_S(\theta)d\theta \leq \int \varphi'(q_S(\theta))q_S(\theta)d\theta \ .$$

This concludes the proof of Proposition 3.1.

From this proposition and Section 3 in [14] we have that if $\mathcal{J}$ is an ε-contaminated model (i.e. has the form (3.6)) then

$$f_S^L(\theta) = \begin{cases} (1-\varepsilon)f_S^0(\theta) & \text{for } f_S^0(\theta) > c' \\ \\ c'(1-\varepsilon)2\pi w_S & \text{for } f_S^0(\theta) \leq c' \end{cases} \tag{3.25}$$

where f_S^0 is the nominal PSD in (3.6) and the constant c' can be determined so that

$$\int f_S^L(\theta)d\theta = \int f_S^0(\theta)d\theta \ .$$

Note that this agrees with the result given in [23]. Similarly f_S^L may be found for the other uncertainty classes discussed in Section 3.

We also note that $-[\int_{-\pi}^{\pi} \varphi(p(\theta))d\theta]$, as in Proposition 3.1, may be thought of as a "measure of the distance" between p_S and Lebesgue measure on $[-\pi,\pi]$. $-\int\varphi(\cdot)$ is related to a class of divergences sometimes referred to as Ali-Silvey distances (see [29]-[31], [21]). Thus, Proposition 3.1 says that f_S is an element of $\mathcal{S}$ which is "closest" to Lebesgue measure. This makes f_S^L of (3.25) intuitively appealing since it is the "flattest" element of $\mathcal{S}$.

Another interesting fact about f_S^L is that it is a least favorable spectral density for the problem of calculating the rate-distortion function of the class of discrete-parameter homogeneous Gaussian sources whose spectra are contained in an ε-contaminated class (see [32]).

5. <u>Robust Filtering in White Noise: Numerical Results</u>

Abstractly, the significance of <u>robust</u> signal estimation is clear: to be able to put the tightest upper bound on the error when the possibility of deviation from the assumed spectra exists is clearly desirable. However, as we discussed in Chapter II, in most situations we must also expect that the robust estimator will not perform as well as the assumed (or nominal) estimator if the true spectra are the nominal spectra. So there is a trade-off. Thus the questions that naturally arise are how much is gained by the robust estimator in <u>its</u> worst case (at (f_S^L, f_N^L)) as compared to the nominal estimator at <u>its</u> worst case and how much is lost in using the robust estimator should the true spectra be the nominal ones. Clearly a blanket statement of the superiority of one estimator over the other in all cases is not likely to prove correct. Thus, we consider these questions for two numerical examples.

Specifically, we consider robust filtering of a first-order Markov signal in white noise. It has been shown [33], [34] that , if f_N^W is the spectrum of white noise (i.e., $f_N^W(\theta) \equiv N_0$ for some positive constant N_0) and if $D(\theta)$ represents filtering (i.e., $D(\theta) \equiv 1$), then

$$e_1^{\dagger}(f_S, f_N^W) = N_0\{1 - \exp \frac{-1}{2\pi} [\int \log(1+N_0^{-1} f_S(\theta))d\theta] \} \tag{3.26}$$

for any signal PSD f_S. In view of Eq. (3.26), the results of Section 3 and Proposition 3.1 in Section 4 can be applied to any convex signal uncertainty class $\mathcal{J}$. In particular, for normalized classes, results from robust hypothesis testing can be used to obtain least favorable signal spectra and hence most-robust filters.

For a wide variety of applications it is appropriate [35] to assume that the noise is white and that the signal is first-order Markov, i.e., has PSD f_S^0 where

$$f_S^0(\theta) = \frac{(1 - r^2)w_S}{1 - 2r\cos\theta + r^2} \quad , \tag{3.27}$$

for some $r \in [-1,1]$. A process with this spectrum has power w_S and, for $r \geq 3 - 2\sqrt{3} \approx 0.172$, it has 3 dB power bandwidth

$$\cos^{-1} \left[\frac{r^2 - 4r + 1}{-2r} \right] \quad . \tag{3.28}$$

Substituting the expression for f_S^0 into (3.26) we obtain $e_1^{\dagger}(f_S^0, f_N^W)$.

Alternatively, since $f_S(\theta)$ is rational, we can determine that the optimal causal transfer function for the nominal pair (f_S^0, f_N^W) is given by

$$H_0^{\dagger}(\theta) = \frac{K}{1 - \alpha e^{-i\theta}} \quad \theta \in [-\pi,\pi] \tag{3.29}$$

where

$$\alpha = \frac{b - \sqrt{b^2 - 4r^2}}{2r} \quad ,$$

$$K = \frac{\alpha w_S (1-r^2)}{N_0(1-r\alpha)} \quad ,$$

and

$$b = \frac{w_S(1-r^2)}{N_0} + (1+r^2) \quad .$$

We can then substitute f_S^0, f_N^W, and (3.29) into (3.3) to obtain $e_1^{\dagger}(f_S^0, f_N^W)$.

As in Chapter II, we use a measure of performance in the figures which we refer to simply as output signal-to-noise ratio (SNR). The purpose of Wiener-Kolmogorov filtering is to minimize the MSE, $E\{|\hat{S}(t) - S(t)|^2\}$, between our estimate $\hat{S}(t)$ (i.e., the output of the filter) and the actual signal $S(t)$. Since the output of the filter can be written as $S(t) + (\hat{S}(t) - S(t))$, we use the signal power divided by the MSE as an output SNR. For the purpose of our graphs we translate this to dB. The horizontal axis is $10 \log_{10}(w_S/N_0)$, the input SNR in dB.

The top line in Fig. 8 gives the performance of the filter H_0, given in (3.29), (which is optimal for the pair (f_S^0, f_N^W)) when f_S^0 and f_N^W are, in fact, the signal and noise spectra which occur. In Fig. 8, the signal bandwidth is 0.105.

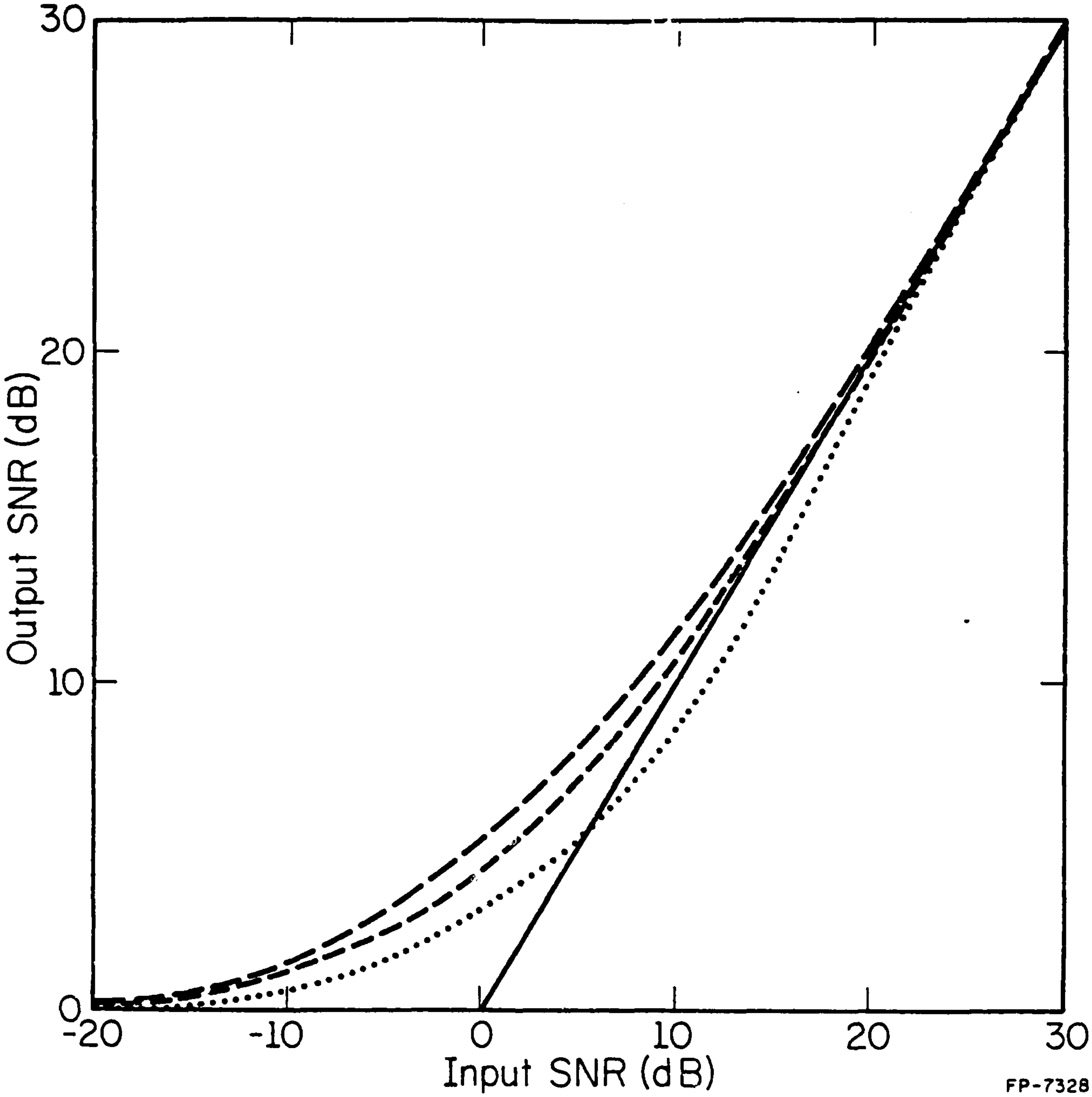


FIG. 8. p-point example. (From top to bottom) $H_0^{\div}$ at (f_S^0, f_N^W); $H_R^{\div}$ at (f_S, f_N^W) for $f_S \in \mathcal{P}$; $H_0^{\div}$ at its worst case (straight line is performance of trivial all-pass filter).

Suppose, now, that we are not completely certain about our choice of signal spectrum. In particular, assume we know only the total signal power w_S and that the fractional power w_1 on the set $A_1 \triangleq \{\theta \mid |\theta| < 0.125\}$ is given by $w_1 = 0.555\ w_S$. In other words, we are modeling our uncertainty about the signal by defining $\mathcal{S}$ as the p-point class (3.9) with A_1 and w_1 as above and $A_2 = A_1^c$ and $w_2 = 0.445\ w_S$.

If all we know is that $f_S \in \mathcal{S}$ then we would like to know how badly the performance of H_0^+ can deteriorate. The bottom curved line in Fig. 8 gives the worst case performance of H_0^+. The straight line in Fig. 8 was included to show how bad this deterioration is. It gives the performance of an all-pass filter (i.e. $H \equiv 1$).

So we see that the performance of the optimal or nominal filter H_0^+ can deteriorate by as much as 3 dB; and, for input SNR above 6 dB, its performance can be significantly worse than no filtering at all. Thus there is a clear need for robust filtering.

Applying Proposition 3.1 with $\varphi(x) = \log(1 + x/N_0)$ (see (3.26)), it can be shown (see Chapter V) that f_S^L is least favorable for causal filtering for the uncertainty classes $\mathcal{S}$ and $\mathcal{N} \triangleq \{f_N^W\}$ if f_S^L is given by

$$f_S^L(\theta) = \begin{cases} 2\pi w_1/0.25 & \text{for } \theta \in A_1 \\ 2\pi w_2/(2\pi - 0.25) & \text{for } \theta \in A_2 \end{cases} \tag{3.30}$$

It is clear that the hypotheses of Theorem 3.2 (specifically i, ii, and iiia) are satisfied for this $\mathcal{S}$ and $\mathcal{N}$. Thus, the most robust filter H_R^+ is the optimal filter for (f_S^L, f_N^W). It can be seen from [16] and (3.29) that $|1 - H_R^+(\theta)|^2$ is constant on A_1 and on A_2. Since every element of $\mathcal{S}$

has the same power on A_1 and on A_2, the error $e_1(f_S,f_N^W;H_R)$ is constant over all $f_S \in \mathcal{S}$; in particular, $e_1^\dagger(f_S^L,f_N^W) = e_1(f_S,f_N^W;H_R^\dagger)$, for all $f_S \in \mathcal{S}$, and we can calculate $e_1^\dagger(f_S^L,f_N^W)$ using (3.26). The constant (over $\mathcal{S}$) performance of $H_R^\dagger$ is given by the second line from the top in Fig. 8. It is clear from Fig. 8 that, unless we are very confident about our original choice of signal spectrum, f_S^0, the robust filter $H_R^\dagger$ is preferable to $H_0^\dagger$.

For our second example, we assume that we have again chosen f_S^0 (first-order Markov signal) and f_N^W (white noise) as our signal and noise spectra; but instead of assuming we know the fractional powers, as in the previous example, we assume, now that we have a general sense of uncertainty about our choice of signal spectrum, i.e., we assume that $\mathcal{N} = \{f_N^W\}$ and that $\mathcal{S}$ is an ε-contaminated class (see (3.6)) with nominal spectrum f_S^0 given by (3.27).

Of course, $e_1(f_S,f_N^W;H_0^\dagger)$ is the same as in the previous example; in particular, $e_1^\dagger(f_S^0,f_N^W)$ is the same. From the expression for $e_1(f_S,f_N^W;H_0^\dagger)$ we can calculate $e_1(f_S^{WC}, f_N^W;H_0^\dagger) = \sup_{f_S \in \mathcal{S}} e_1(f_S,f_N^W;H_0^\dagger)$. It is easy to see that f_S^{WC} (WC stands for worst case) is given by $(1-\varepsilon)f_S^0 + \varepsilon f_S'$, for any f_S' satisfying $f_S'(\theta) = 2\pi w_1\delta(\theta-\pi)+2\pi w_2\delta(\theta+\pi)$ where $w_1 + w_2 = w_S$, since $|1 - H_0^\dagger(-\pi)|^2 = |1 - H_0^\dagger(\pi)|^2 = \sup_{\theta\in[-\pi,\pi]} |1 - H_0^\dagger(\theta)|^2$. (Of course, these f_S' are not actual PSD's but since we can get arbitrarily close to their value in $e_1(f_S,f_N^W;H_0^\dagger)$ using the usual limit arguments for dirac delta functions we will use this notation.) The performances of $H_0^\dagger$ at (f_S^0,f_N^W) and (f_S^{WC},f_N^W) are given by the top and bottom curved lines, respectively, in Fig. 9. For Fig. 9, $\varepsilon = 0.1$ and the signal bandwidth is 0.001. As in the previous example, the straight line gives the performance of an all-pass filter; and, for input SNR between 10 and 60 dB, the performance of the "optimal" filter $H_0^{\dagger}$ can be much worse than that of the all-pass filter. Once again, a clear need for robust filtering exists.

In order to find $f_S^L \in \mathcal{S}$, we can apply Proposition 3.1 with the same φ function as in the first example. Thus we see that the least favorable spectrum f_S^L is given by (3.25) if f_S^0 is the first-order Markov spectral density given in (3.27).

We can now calculate $e_1^{\dagger}(f_S^L, f_N^W)$ by substituting (3.25) into (3.26). On the other hand, $e_1(f_S^0, f_N^W; H_R^{\dagger})$ is more difficult to calculate. Fortunately, Yao has developed an expression (see equation (36') and ff. in [16]) which can be used to find $|1 - H_R^{\dagger}(\theta)|^2$. And we can write $e_1(f_S^0, f_N^W; H_R^{\dagger})$ as

$$e_1^{\dagger}(f_S^L, f_N^W) - \frac{1}{2\pi}\int_{-\pi}^{\pi} |1 - H_R^{\dagger}(\theta)|^2 (f_S^L(\theta) - f_S^0(\theta))d\theta \ . \tag{3.31}$$

The second and third lines from the top in Fig. 9 give the performance of $H_R^{\dagger}$ at (f_S^0, f_N^W) and (f_S^L, f_N^W), respectively. For input SNR below 0 dB or above 60 dB there is essentially no difference between $H_0^{\dagger}$ and $H_R^{\dagger}$; between 0 and 30 dB the insensitivity of $H_R^{\dagger}$ makes it preferable to $H_0^{\dagger}$ unless we are fairly certain about our choice of f_S; and between 30 and 60 dB $H_R^{\dagger}$ is clearly preferable to $H_0^{\dagger}$. We also note that above 20 dB the performance of $H_R^{\dagger}$ is the same as the all-pass filter. Hence, for high input SNR, we are better off doing no filtering at all than using $H_R^{\dagger}$ or $H_0^{\dagger}$.

The examples given in this section are in close agreement with those for the continuous-time case given in Chapter II and the appendix. Based on this experience we can conclude that the robust filter design developed

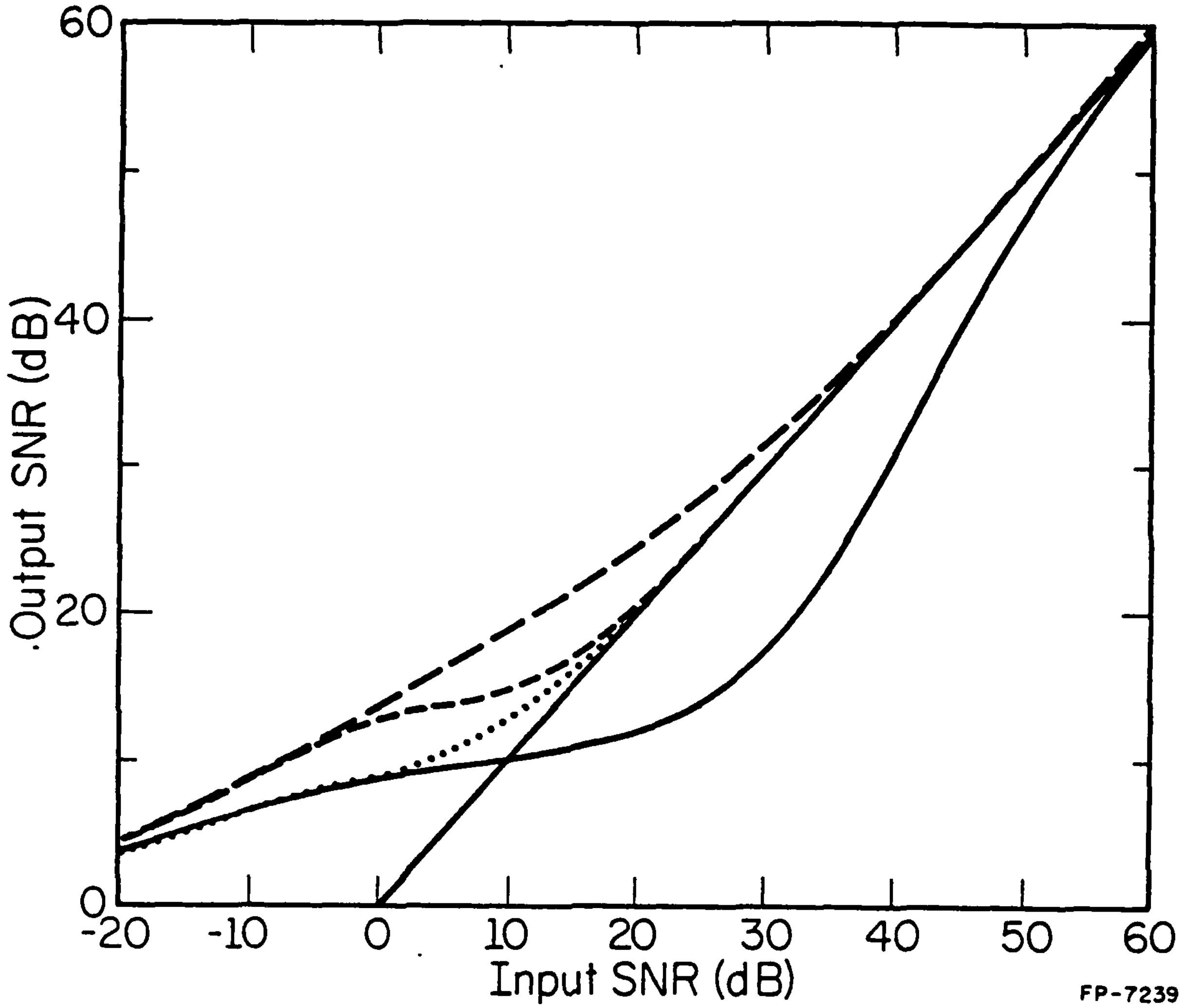


FIG. 9. ε-Contaminated Example. (From top to bottom) $H_0^{\dagger}$ at (f_S^0, f_N^W); $H_R^{\dagger}$ at (f_S^0, f_N^W); $H_R^{\dagger}$ at (f_S^L, f_N^W) ($H_R^{\dagger}$'s worst case; $H_0^{\dagger}$ at (f_S^{WC}, f_N^W) ($H_0^{\dagger}$'s worst case)(straight line is performance of trivial all-pass filter).

here and in [2] is always worth considering and is often preferable to the nominal filter in situations where spectral uncertainty might exist.

6. General Spectral Uncertainty

In the preceeding sections we have modeled spectral uncertainty via classes of bounded spectral densities. While this formulation is quite general and allows for accurate modeling of uncertainty in most situations, there are still a number of practical circumstances in which greater generality is needed. For example, in many situations there can be a contamination of an assumed noise PSD by small amounts of noise generated by rotating machinery located in close proximity to the receiver. This noise is often best modeled by sinusoids of random phase at frequencies which are imprecisely known. Similarly in, for example, an active sonar system such contamination can be present in the signal spectrum due to engine noise generated by the target or due to jamming (see [56]).

Thus, it is clear that a completely general model of spectral uncertainty should include pure tones. This is modeled mathematically using spectral distributions or, equivalently, spectral measures. This means, for example, that if the signal covariance function is $R_S(k)$ then the spectral measure m_S of the signal is a Borel measure on $[-\pi,\pi]$ satisfying

$$R(k) = \frac{1}{2\pi}\int e^{-ik\theta}dm_S(\theta)$$

(See [36]. Chapter X, Theorem 3.2). The spectral measure m_N of the noise is similarly defined.

In this setting (3.3), the MSE, becomes

$$e_D(m_S m_N;H) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} |D(\theta)-H(\theta)|^2 dm_S(\theta) + \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(\theta)|^2 dm_N(\theta) \qquad (3.32)$$

where D and H are the transfer functions of the desired and performed operations, respectively, and now must be mean-square integrable with respect to m_S and $m_S + m_N$, respectively. Furthermore we have that such a transfer function H is causal if and only if $H \in H_+^2(d(m_S+ m_N)(\theta))$, the Hardy subspace of $L^2\ (d(m_S+m_N)(\theta))$ (see Section 2). We also note that the minimum MSE, equation (3.4), becomes

$$e_D^{\dagger}(m_S,m_N) \triangleq \min_{H \in H_+^2(d(m_S+m_N)(\theta))} e_D(m_S,m_N;H) \quad . \qquad (3.33)$$

We now consider, as in Section 3, the situation where we only know that for some classes $\mathcal{S}$ and $\mathcal{N}$ of spectral measures the true signal spectrum satisfies $m_S \in \mathcal{S}$ and the true noise spectrum satisfies $m_N \in \mathcal{N}$. In this setting a most robust causal transfer function $H_R^{\dagger}$ is a solution to the game

$$\inf_{H \in \mathcal{K}} \ \sup_{(m_S,m_N) \in \mathcal{S} \times \mathcal{N}} e_D(m_S,m_N;H) \qquad (3.34)$$

where $\mathcal{K} = \bigcap_{(m_S,m_N)\in \mathcal{S}\times\mathcal{N}} H_+^2\ (d(m_S+m_N)(\theta))$. We do not need to consider any H not in $\mathcal{K}$ because if $H \in \mathcal{K}$ then either H is not causal or for some $(m_S,m_N) \in \mathcal{S}\times\mathcal{N}$ we have that $e_D(m_S,m_N;H) = \infty$. Similarly to definition 3.1 we define $(m_S^L,\ m_N^L)$ to be a <u>least favorable pair of spectral measures for the uncertainty classes $\mathcal{S}$ and $\mathcal{N}$</u> if

$$e_D^{\dagger}(m_S^L, m_N^L) = \max_{\mathcal{S} \times \mathcal{N}} e_D^{\dagger}(m_S, m_N) \tag{3.35}$$

where $e_D^{\dagger}(m_S, m_N)$ is defined in (3.33).

Of course, our objective at this point is to prove results like Theorems 3.1 and 3.2 for this more general setting; i.e., to show that a most-robust transfer function always exists and that if a least favorable pair exists then the optimal transfer function for that pair is a most robust transfer function. We have been unable to show that this is the case here. However, the following result (Theorem 3.3) is in some sense symmetric to Theorems 3.1 and 3.2. That is, this result states (under a mild condition on the uncertainity classes $\mathcal{S}$ and $\mathcal{N}$) that a least favorable pair always exists and that a most robust transfer function exists if and only if the optimal transfer function for the least favorable pair is also a most robust transfer function.

In order to formulate the hypothesis of Theorem 3.3, we consider $\mathcal{S}$ and $\mathcal{N}$ as subsets of the Banach space β of Borel measures on $[-\pi,\pi]$ (see [4]). We consider β to be endowed with the topology induced by $\mathcal{C}$, the Banach space of continuous functions on $[-\pi,\pi]$; i.e., a sequence $\{m_n\} \subseteq \beta$ converges to $m \in \beta$ in this topology if and only if, for all $f \in \mathcal{C}$,

$$\int_{-\pi}^{\pi} f(\theta)dm_n(\theta) \rightarrow \int_{-\pi}^{\pi} f(\theta)dm(\theta). \tag{3.36}$$

Note that β is the dual space of $\mathcal{C}$ and that, in this setting, this topology is called the <u>$\mathcal{C}$-topology</u> or <u>weak*topology</u> on β (see [24] or [10]).

In probability theory a sequence $\{m_n\}$ of probability measures is said to be <u>weakly convergent</u> if (3.36) holds for some $m \in \beta$ (see [9]).

We are now ready to present the main theorem of this section.

<u>Theorem 3.3</u>: Assume the transfer function of the desired operation $D(\theta)$ is continuous on $[-\pi,\pi]$. If the spectral uncertainty classes $\mathcal{S}$ and $\mathcal{N}$ are convex and weak*compact then there exists a least favorable pair $(m_S^L, m_N^L) \in \mathcal{S} \times \mathcal{N}$; i.e. (m_S^L, m_N^L) satisfies (3.35). Furthermore, a most robust transfer function exists if and only if the optimal transfer function for (m_S^L, m_N^L), $H_L^\dagger$, is a most robust transfer function; that is, if and only if

$$e_D(m_S^L, m_N^L; H_L^\dagger) = \max_{\mathcal{S} \times \mathcal{N}} e_D(m_S, m_N; H_L^\dagger) \tag{3.37}$$

holds.

<u>Proof</u>. The proof of this theorem is quite similar to that of Theorem 3.2. In fact, we begin by applying Lemma 3.1. Let $X = \mathcal{S} \times \mathcal{N}$ (endowed with the weak*product topology), $Y = \mathcal{H} \cap \mathcal{C}$ and $F(x,y) = -e_D(m_S, m_N; H)$. $F(x,y)$ is clearly convex on X and concave on Y, and X is compact by hypothesis. We have only to show that $F(.,y)$ is lower semicontinuous on X. We will actually show that it is continuous. Let $(m_S^k, m_N^k) \in \mathcal{S} \times \mathcal{N}$ converge to $(m_S^0, m_N^0) \in \mathcal{S} \times \mathcal{N}$ in the (weak*product) topology of $\mathcal{S} \times \mathcal{N}$. Since $H \in \mathcal{H} \cap \mathcal{C}$ and D is continuous by hypothesis, we have that $|D(\theta)-H(\theta)|^2 \in \mathcal{C}$ and $|H(\theta)|^2 \in \mathcal{C}$. Hence, from equation (3.36) we have that

$$\int_{-\pi}^{\pi} |D(\theta) - H(\theta)|^2 dm_s^k(\theta) \to \int_{-\pi}^{\pi} |D(\theta) - H(\theta)|^2 dm_s^0(\theta) \text{ and}$$

and $\int_{-\pi}^{\pi} |H(\theta)|^2 dm_N^k(\theta) \to \int_{-\pi}^{\pi} |H(\theta)|^2 dm_N^0(\theta)$. Hence $e_D(m_S^k, m_N^k; H) \to e_D(m_S^0, m_N^0; H)$, i.e., $e_D(.,.;H)$ is continuous on $\mathcal{S} \times \mathcal{N}$. We can now apply Lemma 3.1 to yield

$$\max_{\mathcal{S}\times\mathcal{N}} \inf_{\mathcal{K}\cap\mathcal{C}} e_D\,(m_S,m_N;H) = \inf_{\mathcal{K}\cap\mathcal{C}} \max_{\mathcal{S}\times\mathcal{N}} e_D(m_S,m_N;H) \quad . \tag{3.38}$$

Now, for each pair $(m_S,m_N)\in\mathcal{S}\times\mathcal{N}$, we have that $\mathcal{C}_+ \triangleq \{H \in \mathcal{C}\,|\,H \text{ is causal}\}$ (i.e. $\mathcal{C}_+$ is the subspace of $\mathcal{C}$ spanned by $\{e^{in\theta}|\ n=0,1,2,\ldots\}$) is dense in $H_+^{\ 2}(d(m_S+m_N(\theta))$. We have that $\mathcal{C}_+\subseteq H_+^{\ 2}(d(m_S+m_N)(\theta))$ since H continuous on $[-\pi,\pi]$ implies H is bounded and, hence,

$$\int |H(\theta)|^2\ d(m_S+m_N)(\theta) \leq [\sup_\theta |H(\theta)|^2]\ (m_S+m_N)\ ([\pi,\pi]) < \infty.$$

That $\mathcal{C}_+$ is dense in $H_+^{\ 2}(d(m_S+m_N)(\theta))$ now follows from the fact that they are both spanned by $\{e^{in\theta}|\ n=0,1,2,\ldots\}$. This implies that for each $(m_S,m_N)\in\mathcal{S}\times\mathcal{N}$ we have $\mathcal{C}_+ = \mathcal{K}\cap\mathcal{C} = H_+^2(d(m_S+m_N)(\theta))\cap\mathcal{C}$ and we have $\inf_{\mathcal{K}\cap\mathcal{C}} e_D(m_S,m_N;H) = \min_{H_+^2(d(m_S+m_N)(\theta))} e_D(m_S,m_N;H)$. This and (3.38) imply

$$\inf_{\mathcal{K}\cap\mathcal{C}} \max_{\mathcal{S}\times\mathcal{N}} e_D(m_S,m_N;H) = \max_{\mathcal{S}\times\mathcal{N}} \min_{H_+^2(d(m_S+m_N)(\theta))} e_D(m_S,m_N;H). \tag{3.39}$$

But for any minimax problem we have sup inf $\leq$ inf sup, hence the right hand side of (3.39) is less than or equal to

$$\inf_{H_+^2(d(m_S+m_N)(\theta))} \max_{\mathcal{S}\times\mathcal{N}} e_D(m_S,m_N;H) \leq \inf_{\mathcal{K}\cap\mathcal{C}} \max_{\mathcal{S}\times\mathcal{N}} e_D(m_S,m_N;H) \quad . \tag{3.40}$$

This last inequality holds because $\mathcal{K} \cap \mathcal{C} \subseteq H^2_+(d(m_S+m_N)(\theta))$. Since the right hand side of (3.40) equals the left hand side of (3.39) we have

$$\inf_{\mathcal{K}} \max_{\mathcal{S}\times\mathcal{N}} e_D(m_S,m_N;H) = \max_{\mathcal{S}\times\mathcal{N}} \min_{H^2_+(d(m_S+m_N)(\theta))} e_D(m_S,m_N;H)$$

$$= \max_{\mathcal{S}\times\mathcal{N}} e_D^{\dagger}(m_S,m_N). \tag{3.41}$$

In particular, there exists a pair $(m_S{}^L, m_N{}^L) \in \mathcal{S}\times\mathcal{N}$ solving the right hand side of (3.41), i.e. $(m_S{}^L, m_N{}^L)$ is a least favorable pair. Further, we see from (3.41) that a saddle point exists if and only if a most robust transfer function (i.e. a solution to the left hand side of (3.41)) exists. Finally, this implies (see [55], Section 2.3.1) that a transfer function $H_R^{\dagger}$ is most robust if and only if it satisfies

$$e_D(m_S{}^L, m_N{}^L; H_R^{\dagger}) = \min_{H^2_+(d(m_S+m_N)(\theta))} e_D(m_S^L, m_N^L; H)$$

Hence, by the uniqueness of optimal transfer functions (see [3]), if $H_R^{\dagger}$ is most robust then $H_R^{\dagger} = H_L^{\dagger}$ a.e.$[m_S+m_N]$ and $H_L^{\dagger}$ is a most robust causal transfer function. QED

For the remainder of this section we will consider a general formulation of uncertainty classes involving 2-alternating capacities (defined below). As we shall see, this formulation includes most types of classes that have been used to model uncertainty in the robustness literature and all such classes are convex and weak* compact, i.e. they satisfy the hypothesis of Theorem 3.3 regarding $\mathcal{S}$ and $\mathcal{N}$. Furthermore, least favorable pairs have been found for many of these classes (in particular, the five discussed below) for the analogous problem of robust hypothesis testing, and we have shown in Sections 4 and 5 how these least favorable pairs for robust hypothesis testing can often be used to find pairs which are least favorable for robust signal estimation. Finally, we note that 2-alternating capacities are central to efforts being made to develop unifying theories of robust statistics and of robust statistical communication (see [37], also see [6], [7], [32] and Chapters IV and V).

We define $\mathcal{B}_+ \triangleq \{m \in \mathcal{B} \mid m \text{ is nonnegative}\}$ and we let $\mathcal{A}$ denote the Borel σ-algebra on $[-\pi,\pi]$. Suppose $\mathcal{M} \subseteq \mathcal{B}_+$. We will be thinking of $\mathcal{M}$ as a possible uncertainty class of signal or noise spectral measures. Thus, it is not unreasonable to assume as we did in Section 3 that, while we are uncertain about the spectrum, we still are able to make an accurate estimate of the power of the process. Thus we assume $m([-\pi,\pi]) = 2\pi w$, $\forall m \in \mathcal{M}$, where w is this constant power.

If $\mathcal{M} \subseteq \mathcal{B}_+$ then we can define the <u>upper measure</u>, v, of $\mathcal{M}$ as

$$v(A) = \sup\{m(A) \mid m \in \mathcal{M}\}$$

for each $A \in \mathcal{A}$. Clearly, if $m([-\pi,\pi]) = 2\pi w$ $\forall m \in \mathcal{M}$, for some $w < \infty$, then v satisfies

$$v(\phi) = 0, \; v([-\pi,\pi]) < \infty \tag{3.42}$$

$$A \subseteq B \Rightarrow v(A) \leq v(B), \tag{3.43}$$

$$A_n \uparrow A \Rightarrow v(A_n) \uparrow v(A). \tag{3.44}$$

Of course we actually have $v([-\pi,\pi]) = 2\pi w$. If $\mathcal{M}$ is also weak* compact then v satisfies ([6], Lemma 2.3)

$$F_n \downarrow F, \; F_n \text{ closed} \Rightarrow v(F_n) \downarrow v(F). \tag{3.45}$$

Any set function v on $\mathcal{A}$ satisfying (3.42)-(3.45) is called a <u>(Choquet) capacity</u> [12]. If v further satisfies

$$v(A \cup B) + v(A \cap B) \leq v(A) + v(B) \tag{3.46}$$

then v is called a 2-alternating capacity.

For a 2-alternating capacity v we define

$$\mathcal{M}_v = \{m \in \mathcal{B}_+ | \, m(A) \leq v(A), \; \forall A \in \mathcal{A}, m([-\pi,\pi]) = v([-\pi,\pi])\}. \tag{3.47}$$

Robust noncausal (infinite-lag) smoothers have been developed for classes of this form in [7]; and, in [32], a method of finding the rate-distortion function for classes of discrete-parameter homogeneous Gaussian sources whose spectra belong to a class of this form is developed. Also, classes of probability measures having the form (3.47) were considered in detail by Huber and Strassen [6] as uncertainty models for robust hypothesis testing. Most importantly for our purposes, it was shown in [6] that $\mathcal{M}_v$ is weak* compact (note that $\mathcal{M}_v$ is also clearly convex), that the upper measure of $\mathcal{M}_v$ is v, and that three commonly used uncertainty models have the form (3.47).

The first of these models is the ε-contaminated model

$$\mathcal{M}_\varepsilon \triangleq \{m\in\mathcal{B}_+ \,|\, m(A) = (1-\varepsilon)m^0(A) + m'(A), \forall A\in\mathcal{A}; m^0([-\pi,\pi]) = m'([-\pi,\pi]); m'\in\mathcal{B}_+\} \tag{3.48}$$

where $\varepsilon \in [0,1]$ and m^0 is a nominal measure. This has the form (3.47) with $v(A) = (1-\varepsilon)m^0(A) + \varepsilon m^0([-\pi,\pi])$ for $A \neq \phi$. Let λ be a finite Borel measure on $[-\pi,\pi]$ such that $m^0 << \lambda$ and $dm^0/d\lambda$ is bounded a.e. $[\lambda]$, and consider the following set of spectral densities $\{f = dm/d\lambda |$ where $m\in\mathcal{M}_\varepsilon$ such that $m<<\lambda$ and f is bounded a.e. $[\lambda]\}$. Clearly this is nothing more than the ε-contaminated class $\mathcal{S}_\varepsilon$ of PSD's defined in (3.6) with $f_S^0 = dm^0/d\lambda$. Similarly classes of measures can be defined to correspond to the total variation class given in (3.7), the band model given (3.8), and the p-point model in (3.9). It was shown in [6] that the total variation class of measures can be generated by a 2-alternating capacity having the form

$$v(A) = \min\{m^0(A) + \varepsilon m^0([-\pi,\pi]),\ m^0([-\pi,\pi])\}$$

for $A \neq \phi$.

It will be shown in Chapter IV that the band model can be generated by a 2-alternating capacity. In fact the band model is a rather nice special case. If we let $m^U(A) = \int_A f^U(\theta)d\lambda(\theta)$ and $m^L(A) = \int_A f^L(\theta)d\lambda(\theta)$ for all $A\in\mathcal{A}$ where F^U and F^L are the upper and lower PSD's of a band model as in (3.8) then the corresponding band model of measures is given by

$$\mathcal{M}_B \triangleq \{m\in\mathcal{B}_+ \,|\, m^L(A) \leq m(A) \leq m^U(A), \forall A\in\mathcal{A};\ m([-\pi,\pi]) = 2\pi w\} \tag{3.49}$$

where w is the constant power and m^L and m^U are lower and upper measures, respectively, satisfying $m^L([-\pi,\pi]) \leq 2\pi w \leq m^U([-\pi,\pi])$. So we see that for the band model there is a one-to-one correspondence between the density version and the measure version. Furthermore, if $f^U(\theta)$ does not satisfy condition (ii) of Theorem 3.1 we can take the densities with respect to m^U instead of λ. Once we do this we have $f^U(\theta) = 1, \forall\, \theta$, and the conditions of Theorem 3.1 will be satisfied. Hence a most robust filter exists. But since $\mathcal{M}_B$ is a 2-alternating capacity class (it is generated by $v(A) = \min\{2\pi w - m^L(A^c),\ m^U(A)\}$, see Chapter IV) we have that the conclusions of Theorem 3.3 also apply to the band model. Hence we have

Theorem 3.4. If $\mathcal{S}$ and $\mathcal{N}$ are band models as in (3.8), or equivalently, as in (3.49), then there exists a pair $(f_S^L, f_N^L) \in \mathcal{S} \times \mathcal{N}$ such that (f_S^L, f_N^L) is least favorable and $H_L^\dagger$, the optimal transfer function for (f_S^L, f_N^L), is a most robust transfer function.

Note that a singleton is, of course, a special case of a band model.

The p-point model is, unfortunately, not a capacity class and is never weak* compact, but in Chapter V we will show that the weak* closure of a p-point model is a 2-alternating capacity class and in many situations the results for the closure can also be shown to hold for the actual p-point class.

The last of the 2-alternating capacity classes considered in [6] (the first two being the ε-contaminated and total variation classes) is called the Prokhorov class. It has the form

$$\mathcal{M}_p \triangleq \{m \in \mathcal{P}_+ \mid m(A) \leq m^0(A^\delta) + \varepsilon m^0([-\pi,\pi]), \forall A \in \mathcal{A};\ m([-\pi,\pi]) = m^0([-\pi,\pi])\} \quad (3.50)$$

where $\delta \geq 0$ and $\varepsilon \in [0,1]$, m^0 is a nominal, and A^δ is the closed δ-neighborhood of the set A, i.e. $A^\delta = \{\theta \in [-\pi,\pi] | \inf_{a\in A} |\theta - a| \leq \delta\}$. The 2-alternating capacity v which gives $\mathcal{M}_P$ the form (3.47) is defined by letting

$$v(A) = \min\{m^0(A^\delta) + \varepsilon m^0([-\pi,\pi]), m^0([-\pi,\pi])\}$$

for compact $A \neq \phi$ and, then, extending v to $\mathcal{Q}$ via (3.44) and (3.45). While the Prokhorov class has no immediate intuitive appeal, it has some nice theoretical properties; for example, the set consisting of all classes having the form (3.50) (i.e. as δ varies over $[0,2\pi]$ and ε varies over [0,1]) forms a base for (i.e. generates) the weak* topology (see [9]).

From the above discussions we see that many of the classes one might use to model spectral uncertainty are 2-alternating capacity classes and hence are weak* compact and convex. Thus we see that the hypothesis of Theorem 3.3 is quite general.

7. Discussion

In this chapter, we have considered the problems of robust causal smoothing, filtering and prediction of a discrete-time signal in noise and the special case of robust noiseless one-step prediction. Our formulation is analogous to that developed in [2] for robust noncausal continuous-time filtering; however, the proof of the main theorem (Theorem 3.2) required a more abstract approach since no completely general expressions exist for the optimal transfer function or the minimum error. This same difficulty has prevented us from proving a general theorem stating that if the uncertainty classes of Theorem 3.2 each satisfy power constraints then a least favorable pair of PSD's can be found directly from the (in many cases, known) solutions to an analogous robust hypothesis testing problem. Fortunately, as we showed for the robust one-step noiseless prediction problem in Section 4 and for the filtering in white noise problem in Section 5, Proposition 3.1 can be applied to yield this result in many cases. For example, it is straightforward to see from the expression given in Theorem 1 of [3] that this approach will work for the problem of robust one-step prediction in white noise. Many other cases could be handled in this manner or by, first, proving a more general version of Proposition 3.1 to suit the other cases for which minimum error expressions have been developed [3], [15]-[19], [33], [34].

In Section 6 we saw that the notion of a 2-alternating capacity is useful as a general model of uncertainty. In the next two chapters we consider capacities and some of their properties in detail.

IV. A GENERALIZATION OF THE HUBER-STRASSEN DERIVATIVE

1. Introduction

As we discussed in Chapter III, Section 6, the formulation of uncertainty in terms of classes of measures dominated by 2-alternating Choquet capacities, first considered by Huber and Strassen [6], is quite general. It includes most classes commonly used to model spectral uncertainty or to model uncertainty in robust hypothesis testing. In [6], Huber and Strassen develop the Neyman-Pearson Lemma for classes of probability measures whose upper probabilities are 2-alternating capacities. In particular, they prove the existence of a minimax test statistic between two such classes (this statistic is actually a derivative between the capacities which dominate these classes) and the existence of a least favorable pair (Q_0, Q_1) such that for each fixed sample size the Neyman-Pearson tests between Q_0 and Q_1 constitute a minimal essentially complete class of minimax tests between these two classes. In addition to the obvious importance of these results in unifying the theory of robust statistics, they have been used to obtain several general results in robust statistical communication theory [7], [32].

The one shortcoming of Huber and Strassen's fundamental paper is that the capacities which generate the three classes of probability measures most commonly used to model uncertainty in robust hypothesis testing (i.e., those generating the ε-contaminated, total-variation and Prokhorov neighborhoods) must be restricted to a compact space in order to satisfy property (2.4) in the definition of a capacity given by Huber and Strassen [6]. Property (2.4) insists that a capacity be continuous on decreasing sequences of closed sets. In this chapter we relax this restriction by only insisting that our set function (which, for lack of a better term we call a "generalized capacity") be continuous on decreasing sequences of <u>compact</u> sets. This minor alteration allows us to consider the three aforementioned neighborhoods on noncompact spaces.

For many robust statistical communication theory results it is of interest to consider σ-finite as well as finite measures (the generalization of Huber and Strassen's results to capacities v satisfying $v(\Omega) < \infty$ rather than $v(\Omega) = 1$ being straightforward). For example, in [7] spectral uncertainty for the problem of robust linear smoothing is modeled via capacity classes of spectral distributions on $\mathbb{R}^n$. This excludes the possibility of "white noise" whose spectral measure is given by Lebesque-Borel measure on $\mathbb{R}^n$.

The purpose of this chapter is to develop Huber-Strassen type results for a 2-alternating generalized capacity class versus a σ-finite measure. In Section 2 we give the definition of a 2-alternating generalized capacity and some preliminaries. In Section 3 we construct the Huber-Strassen derivative π between a 2-alternating generalized capacity and a σ-finite measure m and we prove that if a least favorable distribution Q exists then $\pi = dQ/dm$ a.e.[m].

Our main theorem (Theorem 4.2) gives an easily verifiable necessary and sufficient condition for the main result of Huber and Strassen [6] to hold for a distribution Q which we construct from π. Corollary 4.1 states that this condition always holds if the generalized capacity is actually a capacity. Thus, for the problem of a capacity versus a σ-finite measure, we always have a least favorable distribution and a Huber-Strassen derivative.

Section 5 contains some examples and applications. In particular, we give an example of a situation in which the condition of the main theorem is not satisfied. Also, we introduce a new 2-alternating capacity which generates a widely used uncertainty class known as the band model. This class is an accurate model of uncertainty for many applications. Furthermore, the upper measure of such a class is a capacity even if the sample space is not compact.

2. Generalized Choquet Capacities

Let Ω be a complete separable metrizable space, $\mathcal{A}$ its Borel σ-algebra and $\mathfrak{M}$ the set of all nonnegative finite Borel measures on Ω.

Definition 4.1: A set function v on $(\Omega, \mathcal{A})$ is a 2-alternating generalized (Choquet) capacity if v satisfies

$$v(\phi) = 0, \; v(\Omega) < \infty, \tag{4.1}$$

$$A \subseteq B \Rightarrow v(A) \leq v(B), \tag{4.2}$$

$$A_n \uparrow A \Rightarrow v(A_n) \uparrow v(A), \tag{4.3}$$

$$F_n \downarrow F, \; F_n \text{ compact} \Rightarrow v(F_n) \downarrow v(F), \tag{4.4}$$

$$v(A \cup B) + v(A \cap B) \leq v(A) + v(B). \tag{4.5}$$

Note that this definition "generalizes" the definition of a 2-alternating capacity [6] by changing the condition "F_n closed" to "F_n compact" in property (4.4)

For any 2-alternating generalized capacity v we define

$$\mathfrak{M}_v = \{P \in \mathfrak{M} : P(A) \leq v(A), \; \forall A \in \mathcal{A} \; ; \; P(\Omega) = v(\Omega)\}. \tag{4.6}$$

It was shown by Huber and Strassen [6] (Examples 3-5) that three of the most important uncertainty classes in robust statistics (the ε-contamination, total variation and Prokhorov classes) have the form (4.6) where v is a 2-alternating capacity only if Ω is compact (see Chapter III). The motivation for the generalization in Definition 4.1 is that these three set functions are 2-alternating generalized capacities even if Ω is not compact. Moreover, the "special capacities" considered by Rieder [20] are 2-alternating generalized capacities.

For all the results of this chapter we assume that Ω is σ-compact.

The following extends Lemma 2.5 of [6] to generalized capacities on σ-compact spaces.

Lemma 4.1: Let v be a 2-alternating generalized capacity on $(\Omega,\mathcal{A})$. For each $A \in \mathcal{A}$, there is a $Q \in \mathcal{M}_v$ such that $Q(A) = v(A)$.

Proof: Since Ω is σ-compact, let $\{K_n\}$ be a sequence of compact sets such that $K_n \uparrow \Omega$. We denote the restriction of v to K_n by v^n; i.e., $v^n(A) = v(A \cap K_n)$, $\forall A \in \mathcal{A}$; $\forall n$. Clearly, for each n, v^n is a 2-alternating capacity. Thus, we may apply Lemma 2.5 of [6] to v^n.

Let $A \in \mathcal{A}$. Denote $A \cap K_n$ by A^n. From Lemma 2.5 of [6], there exists $Q_1 \in \mathcal{M}_{v^1}$ such that $Q_1(A^1) = v^1(A^1)$ and, for each $n \geq 2$, there exists $Q_n \in \mathcal{M}_{v^n}$ such that $Q_n(A^n \setminus A^{n-1}) = v^n(A^n \setminus A^{n-1})$. For all $B \in \mathcal{A}$, define

$$Q(B) = Q_1(B) + \sum_{n=2}^{\infty} \left\{ \left[\frac{v(A^n)-v(A^{n-1})}{Q_n(A^n \setminus A^{n-1})} \right] Q_n(B \cap A^n \setminus A^{n-1}) + \left[\frac{v(K_n)-v(A^n)+v(A^{n-1})-v(K_{n-1})}{Q_n[K_n \setminus (A^n \cup K_{n-1})]} \right] Q_n[B \cap K_n \setminus (A^n \cup K_{n-1})] \right\}$$

It is straightforward to verify using (4.5) that Q satisfies the conclusion of the lemma.QED.

3. <u>The Huber-Strassen Derivative Between a Generalized Capacity and a σ-finite Measure</u>

Since Ω is assumed to be σ-compact, we can fix a sequence of compact sets $K_n \uparrow \Omega$. For a 2-alternating generalized capacity v and a σ-finite nonnegative measure m on $(\Omega,\mathcal{A})$ we denote by v^n and m^n the restrictions of v and m, respectively, to the set K_n. Clearly, for all n, v^n is a 2-alternating capacity and m^n is a finite nonnegative measure on $(\Omega,\mathcal{A})$ (hence, m^n is also a 2-alternating capacity).

Thus, for each n, the theory of Huber and Strassen [6] can be applied to the pair (v^n,m^n). In particular, for each n and for each $t\in[0,\infty]$, there exists a set A_t^n minimizing, over all $A\in\mathcal{A}$, the set function $w_t^n(A) \triangleq t\, m^n(A) + v^n(A^c)$. Further, for each n, there is a function π^n (the Huber-Strassen derivative of v^n with respect to m^n) such that $A_t^n = \{\pi^n > t\}$ for every $t\in[0,\infty]$. Finally, there is a least favorable distribution $Q^n\in\mathcal{M}_{v^n}$, i.e. $Q^n(\{\pi^n \leq t\}) = v^n(\{\pi^n \leq t\}) \geq P(\{\pi^n \leq t\})$ for all $P\in\mathcal{M}_{v^n}$ and for every $t\in[0,\infty]$

Regarding the above situation we have the following lemma.

<u>Lemma 4.2</u>: The collection of sets $\{A_t^n | t\in[0,\infty], n = 1,2,\ldots\}$ may be chosen so that $A_t^{n+1} \subseteq A_t^n$ for each $t\in[0,\infty]$ and $n \geq 1$. Hence, for all n, $\pi^{n+1}(x) \leq \pi^n(x)$ for all $x\in\Omega$.

<u>Proof</u>: We first note that, for any n and t, $w_t^n(A_t^n\cup K_n^c) = w_t^n(A_t^n)$, so we may assume that $K_n^c \subseteq A_t^n$. Now fix n and t; we will show that $w_t^n(A_t^{n+1}\cup A_t^n) \leq w_t^n(A_t^n)$ i.e. that

$$t\,m[(A_t^{n+1} \cup A_t^n) \cap K_n] + v[(A_t^{n+1} \cup A_t^n)^c] \le t\,m(A_t^n \cap K_n) + v(A_t^{n^c}). \tag{4.7}$$

By adding $t\,m(K_{n+1} \setminus K_n)$ to both sides of (4.7) and using the additivity of m and the fact that $K_n^c \subseteq A_t^n$ we obtain the following equivalent expression

$$t\,m[(A_t^{n+1} \cup A_t^n) \cap K_{n+1}] + v[(A_t^{n+1} \cup A_t^n)^c] \le t\;m(A_t^n \cap K_{n+1}) + v(A_t^{n^c}). \tag{4.8}$$

We prove (4.8)(and hence the lemma) by first noting that the right hand side of (4.8) is $w_t^{n+1}(A_t^n)$, then showing (using (4.5) and some straightforward manipulations) that the left hand side is less than or equal to $w_t^{n+1}(A_t^n)$ plus $w_t^{n+1}(A_t^{n+1}) - w_t^{n+1}(A_t^{n+1} \cap A_t^n)$ and finally noting that this additional term is nonpositive.

Lemma 4.2 allows us to define $\pi(x) = \lim_{n\to\infty} \pi^n(x)$ for each $x \in \Omega$. Further, it implies that the definition of $\pi(x)$ does not depend on the choice of the sequence $\{K_n\}$, since for any alternative sequence $\{K'_\ell\}$ the ordered (by set inclusion) union of the two sequences must produce the same limit as each of the original two.

The following theorem justifies our defining π as the Huber-Strassen derivative of v with respect to m.

Theorem 4.1: Let v be a 2-alternating generalized capacity and m a σ-finite nonnegative measure on $(\Omega, \mathcal{A})$, and let π be defined as above. If there exists a measure $Q \in \mathcal{M}_v$ such that Q is least favorable for $\mathcal{M}_v$ versus m, i.e. $\forall\ t \in [0,\infty]$

$$Q(\{\tfrac{dQ}{dm} \le t\}) = v(\{\tfrac{dQ}{dm} \le t\}), \tag{4.9}$$

then

$$\pi = \frac{dQ}{dm} \text{ a.e. } [m].$$

Proof: For every $t \in (0,\infty]$, there exists $t_k \uparrow t$. Since $t_k \uparrow t$ implies $\{dQ/dm \leq t_k\} \uparrow \{dQ/dm < t\}$, (4.9) implies that $\forall t \in (0,\infty]$

$$Q(\{\frac{dQ}{dm} < t\}) = v(\{\frac{dQ}{dm} < t\}). \tag{4.10}$$

Also $t_k \uparrow t$ implies that, $\forall n$, $\{\pi^n \leq t_k\} \uparrow \{\pi^n < t\}$. This fact, property (4.3) and the definition of $A_t^n = \{\pi^n > t\}$ imply that, $\forall t \in (0,\infty]$ and $\forall n$,

$$t\, m^n(\{\pi^n \geq t\}) + v^n(\{\pi^n < t\}) \leq t\, m^n(\{\frac{dQ}{dm} \geq t\}) + v^n(\{\frac{dQ}{dm} < t\}). \tag{4.11}$$

Since the right hand side of (4.11) is clearly less than or equal to $t\, m(\{dQ/dm \geq t\}) + v(\{dQ/dm < t\})$ and $n \to \infty$ implies $\{\pi^n < t\} \uparrow \{\pi < t\}$, (4.11) and (4.3) imply

$$t\, m(\{\pi \geq t\}) + v(\{\pi < t\}) \leq t\, m(\{\frac{dQ}{dm} \geq t\}) + v(\{\frac{dQ}{dm} < t\}). \tag{4.12}$$

By (4.10) the right hand side of (4.12) is equal to $t \times m(\{dQ/dm \geq t\}) + Q(\{dQ/dm < t\})$. Since $Q \in \mathcal{M}_v$, the lefthand side of (4.12) is greater than or equal to $t\, m(\{\pi \geq t\}) + Q(\{\pi < t\})$. Thus we have, $\forall t \in (0,\infty]$

$$t\, m(\{\pi \geq t\}) + Q(\{\pi < t\}) \leq t\, m(\{\frac{dQ}{dm} \geq t\}) + Q(\{\frac{dQ}{dm} < t\}). \tag{4.13}$$

Since, for any $t \in [0,\infty)$, $t_k \downarrow t$ implies $\{\pi \geq t_k\} \uparrow \{\pi > t\}$ and $\{dQ/dm \geq t_k\} \uparrow \{dQ/dm > t\}$, we have from (4.13) and the continuity of measures that

$$t\, m(\{\pi > t\}) + Q(\{\pi \leq t\}) \leq t\, m(\{\frac{dQ}{dm} > t\}) + Q(\{\frac{dQ}{dm} \leq t\}), \tag{4.14}$$

$\forall t \in [0,\infty]$. By the uniqueness of the Radon-Nikodym derivative between two measures (see, for example, Royden [38]) we have that $\pi = dQ/dm$ a.e. [m] (cf. the remarks near the end of Section 3 in [6]). QED.

4. The Least Favorable Distribution

By Fatou's Lemma (see, for example, Royden [38], Proposition 17, p. 231) we have, for each $A \in \mathcal{A}$, that

$$\int_A \pi dm \leq \liminf_{n\to\infty} \int_A \pi^n dm^n \leq \liminf_{n\to\infty} Q^n(A), \tag{4.15}$$

where π, π^n, and Q^n are defined at the beginning of Section 3. By Lemma 4.1 there is a distribution $\tilde{Q} \in \mathcal{M}_v$ such that $\tilde{Q}(\{\pi < \infty\}) = v(\{\pi < \infty\})$. We define the distribution Q on $(\Omega, \mathcal{A})$ by

$$Q(A) = \int_A \pi dm + \tilde{Q}(A \cap \{\pi = \infty\}), \tag{4.16}$$

for each $A \in \mathcal{A}$. If there exists a least favorable distribution then Theorem 4.1 and the Fundamental Theorem of Calculus imply that the Q given in (4.16) is also least favorable. Furthermore, we have

Theorem 4.2: Assume v is a 2-alternating generalized capacity and m is a σ-finite nonnegative measure on $(\Omega, \mathcal{A})$. Let π be defined as in Section 3 and let Q be defined by (4.16). If $Q(\{\pi < \infty\}) = v(\{\pi < \infty\})$ (or equivalently if $Q(\Omega) = v(\Omega)$) then Q is a least favorable distribution for $\mathcal{M}_v$ versus m; i.e., for every $t \in [0,\infty]$

$$Q(\{\pi \leq t\}) = v(\{\pi \leq t\}) \tag{4.17}$$

and π is a version of dQ/dm.

Note that the hypothesis is also trivially necessary. Also, note that if $m(\Omega) < \infty$ and the conclusion of Theorem 4.2 holds then, by [20, Proposition 2.1], Q is least favorable for any minimax testing problem for $\mathcal{M}_v$ versus $\{m\}$.

Proof: From the definition of Q^n, the fact that $t_k \uparrow \infty$ implies $\{\pi \leq t_k\} \uparrow \{\pi < \infty\}$, and property (4.3) we have that $Q^n(\{\pi^n < \infty\}) = v^n(\{\pi^n < \infty\}) \uparrow v(\{\pi < \infty\})$. Thus, the hypothesis of the theorem implies that $Q^n(\{\pi^n < \infty\}) \uparrow Q(\{\pi < \infty\})$; i.e.

$$\lim_{n\to\infty} \int_\Omega \pi^n \, dm^n = \int_\Omega \pi dm.$$

For any set $A \in \mathcal{A}$, we apply the Generalized Dominated Convergence Theorem (Royden [38], Proposition 18, p. 232) to the sequence $\{\pi^n|_A \mid n=1,2,\ldots\}$ ($\pi^n|_A$ is the restriction of π^n to A) to obtain $\lim_{n\to\infty} \int_A \pi^n \, dm^n = \int_A \pi dm$; i.e.,

$$\lim Q^n(A) = Q(A), \ \forall A \in \mathcal{A} . \tag{4.18}$$

If we set $A = \{\pi < t\} = \bigcup_{k=1}^{\infty} \{\pi^k < t\}$ in (4.18), we have $Q(\{\pi < t\}) = \lim_{n\to\infty} Q^n(\bigcup_{k=1}^{\infty} \{\pi^k < t\}) \geq \lim_{n\to\infty} Q^n(\{\pi^n < t\}) = \lim_{n\to\infty} v^n(\{\pi^n < t\}) = v(\{\pi < t\})$. Thus, we have $Q(\{\pi < t\}) \geq P(\{\pi < t\})$, for all $P \in \mathcal{M}_v$ and all $t \in [0,\infty]$. From property (4.3) we have $Q(\{\pi \leq t\}) \geq P(\{\pi \leq t\})$, for all $P \in \mathcal{M}_v$ and all $t \in [0,\infty]$, since $t_k \downarrow t$ implies $\{\pi < t_k\} \downarrow \{\pi \leq t\}$. But, by Lemma 4.1, for each $t \in [0,\infty]$ there is some $P_t \in \mathcal{M}_v$ such that $P_t(\{\pi \leq t\}) = v(\{\pi \leq t\})$ so we must have $Q(\{\pi \leq t\}) = v(\{\pi \leq t\}), \forall t \in [0,\infty]$. QED.

Corollary 4.1: If v is a 2-alternating capacity and m, π and Q are as above then $Q(\{\pi < \infty\}) = v(\{\pi < \infty\})$; hence, the conclusions of Theorem 4.2 hold.

Proof: We wish to show that $\int_\Omega \pi^n dm^n \to \int_\Omega \pi dm$. Let $\tilde{\pi}^n = \pi^n$ on K_n and 0 on K_n^c; we will show $\int_\Omega \tilde{\pi}^n dm \to \int_\Omega \pi dm$. By the Vitali convergence theorem (see Dunford and Schwartz [10], pp. 150, 173) this happens if for any sequence

$\{E_k\} \subseteq \mathcal{A}$ such that $E_k \downarrow \phi$ we have $\lim_{k\to\infty} \int_{E_k} \tilde{\pi}^n dm = 0$ uniformly in n. To prove this we use the fact that a capacity class is tight (Huber and Strassen [6], Lemma 2.2) to pick (given $\varepsilon > 0$) K_N such that $\int_{K_N^c} \tilde{\pi}^n dm \leq Q^n(K_N^c) <$ $\varepsilon/2$ for all n. Then since $\{\tilde{\pi}^n | n=N,N+1,\ldots\}$ is decreasing on K_N we have that $\int_{E_k \cap K_N} \tilde{\pi}^n dm \leq \int_{E_k \cap K_N} \tilde{\pi}^N dm \leq Q^N(E_k \cap K_N)$. Now for each $n=1,\ldots,N$ there is an $L_{n,\varepsilon}$ such that, $\forall\, k \geq L_{n,\varepsilon}$, $Q^n(E_k) < \varepsilon/2$. Now let $L_\varepsilon = \max\{L_{1,\varepsilon},\ldots,L_{N,\varepsilon}\}$, we have that, for all $k \geq L_\varepsilon$, $\int_{E_k} \tilde{\pi}^n dm \leq \int_{K_N^c} \tilde{\pi}^n dm + \int_{E_k \cap K_N} \tilde{\pi}^n dm < \varepsilon/2 + \varepsilon/2$, for every n.

5. Examples and Applications

Our first example shows a situation in which the hypothesis of Theorem 4.2 fails to hold.

Let $\Omega = \mathbf{R}$ and let v be the 2-alternating generalized capacity which setwise dominates an ε-contaminated neighborhood of probability distributions, i.e., let $\varepsilon \in (0,1)$ and let $v(A) = (1-\varepsilon)P(A) + \varepsilon$, for $A \neq \phi$, where P is a probability measure on $(\mathbf{R},\mathcal{A})$. Also, let m be Lebesgue-Borel measure on $(\mathbf{R},\mathcal{A})$. Assume that P has a density p with respect to m. Choose $K_n \uparrow \mathbf{R}$ with K_n compact $\forall n$. From [14], Section 3, we have that $\pi^n(x) = \max\{c_n,(1-\varepsilon)p(x)\}$, $\forall x \in K_n$, where $c_n \geq 0$ can be chosen so that $\int_{K_n} \pi^n dm = v(K_n)$. It is straightforward to show that $\lim_{n\to\infty} c_n = 0$. Thus, $\pi(x) = (1-\varepsilon)p(x)$, $\forall x \in \mathbf{R}$, and we have that $\{\pi < \infty\} = \Omega$ and $Q(\{\pi < \infty\}) = \int_{\mathbf{R}} \pi dm = 1-\varepsilon < v(\{\pi < \infty\})$.

Intuitively each least favorable Q^n tries to be as much as possible like Lebesgue-Borel measure by flattening the tails of π^n with the ε of contamination; but, because $\mathbf{R}$ is not compact, this ε of contamination is allowed to slip off the ends of the real line as $n \to \infty$.

On the other hand, as we mentioned earlier, all the neighborhoods of probability distributions commonly used to model inexact knowledge in robust hypothesis testing are 2-alternating generalized capacities even if the sample space Ω is not compact. Since least favorable pairs have been found for all these classes [14], [27], [20], Theorem 4.1 implies that, for any of these specific but important examples, we have succeeded in extending the main result of Huber and Strassen [6] to a generalized capacity class versus a finite measure on a σ-compact space. Further, Corollary 4.1 implies that this is true for any 2-alternating capacity class versus any σ-finite measure on a σ-compact space. To illustrate this result we introduce the following example of a 2-alternating capacity on a noncompact space.

Let m_L and m_U be finite nonnegative measures on $(\Omega,\mathcal{A})$ such that

$$m_L(A) \le m_U(A), \ \forall A \in \mathcal{A} . \tag{4.19}$$

Define

$$v(A) = \min\{w-m_L(A^c), m_U(A)\}, \tag{4.20}$$

for $A \in \mathcal{A}$, where w is a positive constant such that $m_L(\Omega) \le w \le m_U(\Omega)$. For this v, $\mathcal{M}_v$ is given by

$$\{m \in \mathcal{M} \mid m_L(A) \le m(A) \le m_U(A), \ \forall A \in \mathcal{A} ; \ m(\Omega) = w\}. \tag{4.21}$$

This class is called the band model. Note that every element of $\mathcal{M}_v$ is absolutely continuous with respect to m_U, so if $w=1$, $\mathcal{M}_v$ is equivalent to the class of pdf's considered by Kassam [21]. Most importantly, least favorable pairs are given for these classes by Kassam [21].

Classes of the form (4.21) arise naturally as a confidence band around an estimate of a pdf (where, of course, $w=1$) or around an estimate of a spectral density (where $w/2\pi$ is the known power of the process). Also, as has been noted by Kassam [21], if we define $\varepsilon = 1-m_L(\Omega)/w$ then the band model(4.21) contains those elements of an ε-contaminated class which are bounded above by m_U. Thus, the ε-contaminated class may be thought of as a limit of classes which are band models.

Proposition 4.1: The set function v in (4.20), which defines the band model (21), is a 2-alternating capacity even if Ω is not compact.

Proof: The 2-alternating property (given by (4.5)) is the only defining property which is not straightforward. We handle (4.5) case by case. If A and B are such that $v(A) = m_U(A)$ and $v(B) = m_U(B)$ then, by (4.20), $v(A \cup B) + v(A \cap B) \leq m_U(A \cup B) + m_U(A \cap B) = m_U(A) + m_U(B) = v(A) + v(B)$. Similarly, if $V(A) = w-m_L(A^c) = m_L(A) + (w-m_L(\Omega))$ and $v(B) = m_L(B) + (w-m_L(\Omega))$ then(4.5)holds. If, say, $v(A) = m_L(A) + (w-m_L(\Omega))$ and $v(B) = m_U(B)$ (A and B are interchangable) then we must have $v(A \cap B) = m_U(A \cap B)$ and $v(A \cup B) = m_L(A \cup B) + (w-m_L(\Omega))$ because if $m_L(A \cap B) + (w-m_L(\Omega)) < m_U(A \cap B)$ then, since $m_L(B) + (w-m_L(\Omega)) \geq m_U(B)$, we have (by subtraction) that $m_L(B\setminus A) > m_U(B\setminus A)$ which contradicts(4.19). (The possibility of $m_U(A \cup B) < m_L(A \cup B) + (w-m_L(\Omega))$ is similarly disallowed.) The one possible case is $v(A \cup B) = m_L(A \cup B) + (w-m_L(\Omega))$ and $v(A \cap B) = m_U(A \cap B)$. In this case(4.5) is equivalent to $m_L(B\setminus A) \leq m_U(B\setminus A)$ which always holds by (4.19). Thus v of (4.20) which gives rise to the band model (4.21) is a 2-alternating capacity. QED.

Theorem 2.2 of [7] states that if v_S and v_N are 2-alternating capacities on $\mathbb{R}^n$ and π is a version of the Huber-Strassen derivative between v_S and v_N then $h^* = \pi/(1+\pi)$ is a minimax linear smoother for

$\mathcal{M}_{v_S}$ and $\mathcal{M}_{v_N}$, where $\mathcal{M}_{v_S}$ and $\mathcal{M}_{v_N}$ are uncertainty classes of signal and noise spectral measures, respectively. A careful examination of the proof of this result shows that it holds for any two set functions v_S and v_N for which the conclusions of Theorem 4.1 of Huber and Strassen [6] hold. Thus we see from Corollary 4.1 that Theorem 2.2 of [7] holds for v_S and m_N where v_S is a 2-alternating capacity and m_N is a nonnegative σ-finite Borel measure on $\mathbb{R}^n$. Unquestionably, the most important examples to which this extension can be applied are those where m_N is Lesbesgue-Borel measure on $\mathbb{R}^n$, which is the spectral measure of continuous-parameter white noise.

V. ON THE P-POINT UNCERTAINTY CLASS

1. Introduction

As we have discussed in the preceding chapters, there are many applications of statistical communication theory for which it is inappropriate to assume that we have exact knowledge of some underlying spectral distribution or probability distribution. A common approach to such situations (and the one we have used in this thesis) involves choosing classes of distributions which accurately model this uncertainty. We have also discussed the fact that a number of results have been obtained for situations where uncertainty is modeled via classes whose upper measures are 2-alternating Choquet capacities (see [6], [7], [32] and Section 6 of Chapter III).

The significance of these results is twofold. First, in each case, general existence results are given which unify the theory involved (especially considering the generality of modeling uncertainty via 2-alternating capacity classes; a topic we will discuss below). Second, least favorable pairs for the robust hypothesis testing problem [6] have already been found for each of the classes which have been shown to be 2-alternating capacity classes and the results given in [32] and [7] allow us to solve the problems considered therein directly from such least favorable pairs.

Among the classes which have been used to model uncertainty (and for which least favorable pairs have been found) are the ε-contaminated, total-variation, Prokhorov and band models. The first three are shown to be 2-alternating capacity classes (when restricted to have compact support) in [6]. The band model was shown to be a capacity class in Chapter IV. Furthermore, least favorable pairs are known for these classes (see [14] and [27] and [21] for the band model).

One class which has been used to model uncertainty in detection problems [5], [8], rate-distortion problems [41] and robust smoothing problems [25] which is not a 2-alternating capacity class is given by

$$\mathcal{P}_{pp} \triangleq \{P \in \mathcal{M} \mid P(A_j) = w_j,\ j=1,\dots,n\} \tag{5.1}$$

where $\mathcal{M}$ is the set of all finite nonnegative Borel measures on a compact Polish space Ω, the w_j's are positive real constants, and $\{A_j \mid j=1,\dots,n\}$ is a fixed partition of Ω such that each A_j is a Borel set with nonempty interior. If Ω is a compact subset of the real line, if each of the A_j's is the union of a symmetric pair of intervals and if the P's are <u>spectral</u> distributions then $\mathcal{P}_{pp}$ is essentially the form of the class referred to by Sakrison [41] as "class b" and by Cimini and Kassam [25] as a "p-point class." The term p-point class is also used by El-Sawy and VandeLinde [5], [8] for $\mathcal{P}_{pp}$ when the P's are probability distributions and, of course, $\sum_{j=1}^{n} w_j = 1$. The class given in (5.1) is an appropriate model of spectral uncertainty when, for example, we are able to use power measurements from a bank of low-pass filters [41], and they are appropriate models for probabilistic uncertainty since $P([-a,a])$ "is one of the most easily measured parameters of a distribution" [5, p. 725].

In this chapter we consider uncertainty classes of the form $\mathcal{P}_{pp}$ (which we henceforth refer to as p-point classes). In particular, we show how, in many cases, the results of [6], [7], [32] can be applied to these p-point classes by embedding them in slightly larger 2-alternating capacity classes which we term extended p-point classes (see below).

2. Development

Unfortunately, the p-point class $\mathcal{P}_{pp}$ is not weak* compact (see Chapter III for a definition of the weak* topology) and, as we mentioned in Section 1, the set function v_{pp} defined, for each A, by

$$v_{pp}(A) = \sup_{P \in \mathcal{P}_{pp}} P(A) \tag{5.2}$$

is not a capacity. (For a class $\mathcal{P} \subseteq \mathcal{M}$, the set function v defined as the setwise supremum over $P \in \mathcal{P}$ is called the upper measure of $\mathcal{P}$ and Huber and Strassen have shown [6, Lemma 2.5] that if v is a 2-alternating capacity then v is the upper measure of the set $\mathcal{P}_v \triangleq \{P \in \mathcal{M} \mid P(A) \leq v(A), \forall A,\ P(\Omega) = v(\Omega)\}$.) Basically, the reason for this is that $\mathcal{P}_{pp}$ is not weak* closed. To illustrate this, suppose $\Omega = [0,2]$, $n = 2$, $A_1 = [0,1]$, $A_2 = (1,2]$ and $w_1 = w_2 = \frac{1}{2}$ then for each $k \geq 1$, the probability distribution P_k defined by $P_k(\{1\}) = \frac{1}{2}$ and $P_k(\{1+1/k\}) = \frac{1}{2}$ is an element of $\mathcal{P}_{pp}$. But $\{P_k\}$ converges to P_0 weak* where $P_0(\{1\}) = 1$. Clearly $P_0 \notin \mathcal{P}_{pp}$.

We now consider a new uncertainty class which we term the extended p-point class. It is given by

$$\bar{\mathcal{P}}_{pp} \triangleq \{P \in \mathcal{M} \mid P(A_j^o) \leq w_j,\ P(\bar{A}_j) \geq w_j,\ j=1,\ldots,n,\ P(\Omega) = \sum_{j=1}^{n} w_j\} \tag{5.3}$$

where A_j^o is the (nonempty) interior of A_j and $\bar{A}_j$ is the closure. We use the notation $\bar{\mathcal{P}}_{pp}$ because this class is actually the weak* closure of $\mathcal{P}_{pp}$.

The upper measure of $\bar{\mathcal{P}}_{pp}$ is given by

$$v_{epp}(A) \triangleq \begin{cases} 0 & \text{if } A \neq \phi; \\ w_j & \text{if } A \subseteq A_j^o,\ j=1,\ldots,n; \\ \sum\limits_{\substack{k=1 \\ k\neq j}}^{n} w_k & \text{if } A \subseteq \bar{A}_j^c \text{ and } A \not\subseteq A_k^o \text{ for } k=1,\ldots,n,\ j=1,\ldots,n; \\ \sum\limits_{j=1}^{n} w_j & \text{otherwise;} \end{cases} \tag{5.4}$$

which is a 2-alternating capacity (recall that Ω is compact here). Thus, the importance of this extended p-point class is that the results of [6], [7], [32] can be applied to them. Furthermore, as we will show below, the least favorable pairs for two nonintersecting extended p-point classes are also contained in the corresponding p-point classes and, hence, are also least favorable for the p-point classes.

We now consider the problem of a p-point class versus a single distribution. This is relevant for tests between a composite hypothesis and a simple alternative and robust smoothing of an uncertain signal in noise (e.g. band limited white noise). Moreover, Poor [32] has shown that if $\mathcal{P}_v$ is a 2-alternating capacity class of spectral distributions then the spectral distribution $Q \in \mathcal{P}_v$ which is least favorable for $\mathcal{P}_v$ versus Lebesgue-Borel measure on $[-\pi,\pi]$ (least favorable in the sense discussed in Chapter IV) has rate-distortion function equal to the rate-distortion function over $\mathcal{P}_v$.

Let $\bar{\mathcal{P}}_{pp}$ be an extended p-point class as in (5.3) and let $P_0 \in \mathcal{M}$ be such that $P_0 \notin \bar{\mathcal{P}}_{pp}$. For ease of exposition we assume that Ω is a compact subset of $\mathbb{R}^n$ and that P_0 has a density p_0 with respect to Lebesgue measure. It is fairly straightforward to see from the definition of dv_1/dv_0 given in [6] that dv_{epp}/dP_0 (where v_{epp} has the form (5.4) and is the capacity dominating $\bar{\mathcal{P}}_{pp}$) must be constant a.e.$[P_0]$ on each of the A_j's. Thus, a least favorable $Q \in \bar{\mathcal{P}}_{pp}$ can be given in terms of its density with respect to Lebesgue measure:

$$q(x) = \frac{w_j}{P_0(A_j)} p_0(x) \;\; \forall\; x \in A_j,\; j=1,\ldots,n . \tag{5.5}$$

Thus we have that

$$Q(\{\frac{dQ_1}{dP_0} \leq t\}) \geq P(\{\frac{dQ_1}{dP_0} \leq t\}) \qquad (5.6)$$

for all $P \in \bar{\mathcal{P}}_{pp}$. Hence, (5.6) holds for all $P \in \mathcal{P}_{pp}$ (where $\mathcal{P}_{pp}$ is given in (5.1)) and, since $Q \in \mathcal{P}_{pp}$, we have that Q (with density given in (5.5) is least favorable for the p-point class $\mathcal{P}_{pp}$ versus the single measure P_0. Note that(5.5) agrees with the form given by Sakrison [41, equation (37)] for the spectrum achieving the maximum rate over his class b.

So we have seen that by considering the extended p-point class the results of [6],[7],[32] can be applied to the usual p-point class versus a single distribution. In many cases, the problem of one p-point class versus another can be handled in a similar manner. We illustrate this possibility in a simplified case.

Let A and B be Borel subsets of Ω such that $\bar{A} \subseteq B^o$. Let $\mathcal{P}_1$ and $\mathcal{P}_0$ be p-point classes as in (5.1) with n=2 and based on A and B, respectively. That is, let

$$\mathcal{P}_1 = \{P \in \mathcal{M} \mid P(A) = w_1^1,\ P(A^c) = w_2^1\}$$

and

$$\mathcal{P}_0 = \{P \in \mathcal{M} \mid P(B) = w_1^0,\ P(B^c) = w_2^0\} \qquad (5.7)$$

We can further assume that $w_1^1 > w_1^0$ or that $w_2^1 < w_2^0$, because otherwise we would have $\mathcal{P}_0 \cap \mathcal{P}_1 \neq \phi$ and the problem would be trivial. Again, it is not difficult to show from [6], that if v_1 and v_0 are the 2-alternating capacities determining the extended p-point classes corresponding to $\mathcal{P}_1$ and $\mathcal{P}_0$ of (5.7) then dv_1/dv_0 is constant on A and on B^c (note that $A \cap B^c = \phi$) and that any pair $(Q_0, Q_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ satisfying $Q_0(A') = Q_1(A')\, w_1^1/w_1^0$, $\forall A' \subseteq A$, and $Q_1(B') = Q_0(B')\, w_2^1/w_2^0$, $\forall B' \subseteq B^c$, is least favorable for $\mathcal{P}_1$ versus $\mathcal{P}_0$.

At the beginning of this example we assumed that $\bar{A} \subseteq B^o$. One reason for this is that if there was a point x which was contained in the boundary of A and the boundary of B and if, say, $w_1^0 + w_2^0 = w_1^1 + w_2^1 = 1$ then the $Q \in \mathcal{M}$ which satisfies $Q(\{x\}) = Q(\{\Omega\}) = 1$ also satisfies $Q \in \bar{\mathcal{P}}_0 \cap \bar{\mathcal{P}}_1$, where, for i=1,2, $\bar{\mathcal{P}}_i$ is the extended p-point class corresponding to $\mathcal{P}_i$. In this case, $(Q,Q) \in \bar{\mathcal{P}}_0 \times \bar{\mathcal{P}}_1$ is least favorable for $\bar{\mathcal{P}}_1$ versus $\bar{\mathcal{P}}_0$ but $(Q,Q) \notin \mathcal{P}_0 \times \mathcal{P}_1$. Thus the approach used above will not work in this case, and we must note that this case is important. For example, for the robust linear smoothing application, if we used power measurements from a bank of low pass filters (as suggested in [41]) to determine p-point classes to model our uncertainty about the signal and noise spectra then the boundary points would be the same for both classes and the corresponding extended p-point classes would overlap. Of course, this case can be handled more directly as in [25].

3. Conclusions

In this chapter we have considered an uncertainty class which is appropriate for many applications. We have shown that by embedding this class in a slightly larger 2-alternating capacity class the results of [6], [7], [42] can be shown to hold for the original class in most cases. Actually in those cases where this approach cannot be utilized a more direct approach can be shown to work [58].

VI. SUMMARY AND CONCLUSIONS

In this thesis several problems in robust statistical signal processing have been considered. In this final chapter we briefly summarize the results obtained and propose some related topics for possible further research.

In Chapters II and III and the appendix, we presented a varied selection of numerical results which indicate that the robust Wiener and Wiener-Kolmogorov filters, developed in [1], [2], [25] and in Chapter III, are often preferable to the corresponding traditional filters in situations where deviations from assumed spectra might occur. In Chapter III, we also gave a method of obtaining robust n-step predictors and robust n-lag smoothers. Further, we illustrated in the case of robust one-step noiseless prediction how this method could be used to design robust signal estimators utilizing least-favorable pairs from an analogous robust hypothesis testing problem. One possible topic for further work is extending this result to cases other than those cases considered in Sections 4 and 5 of Chapter III; perhaps, by generalizing Proposition 3.1 and, if needed, developing error expressions for those cases not treated in [3], [15]-[19]. Another subject which needs to be considered concerns the implementation of these robust filters. We have, in this study, made the standard assumption that we have knowledge of the infinite past. For most applications this is an unrealistic assumption; thus, an examination of the effects of finite memory on robust signal estimation would be of interest.

In Chapter IV, we introduced the notion of a 2-alternating generalized capacity because several of the most important uncertainty classes must be restricted to compact spaces in order to be capacity classes, but are generalized capacity classes even if the space is not compact. We then developed a Huber-Strassen derivative between a 2-alternating generalized capacity and a σ-finite measure and defined a distribution which Theorem 4.2 guarantees to be least favorable for this problem if any least favorable distribution exists. It was also shown that, for the problem of a capacity class versus a σ-finite measure, a least favorable distribution always exists. Finally, in Chapters IV and V two uncertainty classes, the band model and the extended p-point model, were shown to be 2-alternating capacity classes. The fact that the band model is a capacity class even if the underlying space is not compact is especially significant in view of the results of Chapter IV (especially Corollary 4.1). The significance of the extended p-point class is that the p-point class, which is appropriate for many applications, is contained in it and, in many cases, results obtained for capacity classes can be applied to the p-point class directly from the corresponding extended p-point class results.

Further study regarding the topics of Chapters IV and V might be directed toward finding a Huber-Strassen derivative between two generalized capacities and a corresponding least favorable pair of distributions. Such a result would allow the full generality of the Huber-Strassen theory to be applied to a larger class of problems.

APPENDIX

The purpose of this appendix is to give a further selection from our numerical study of robust Wiener filtering (see Chapter II, Section 3). Figures 10-14 give further results regarding the examples presented in Chapter II. In particular, for the p-point example of Chapter II, Figures 10 and 11 give the performances of the nominal filter H_0^* and of the robust filter H_R^* for wider noise bandwidths α_N. In Figure 10, $\alpha_N = 100$ and, in Figure 11, $\alpha_N = 1000$ (recall that in Figures 1 and 4 we had $\alpha_N = 10$). Note that as α_N increases the performance of H_0^* at (σ_0, ν_0) improves but at its worst case H_0^*'s performance degradates further. Meanwhile H_R^*'s performance changes little. Figures 12, 13 and 14 give further results for the ε-contaminated example of Chapter II. In Figures 2 and 5, we had $\alpha_N = 1000$; in Figures 12, 13 and 14 we have $\alpha_N = 100$, $\alpha_N = 10^4$ and $\alpha_N = 10^6$, respectively. Again we see that H_R^* is relatively insensitive to changes in the noise bandwidth, but H_0^* is not.

Recall that the example which we have referred to here and in Chapter II as the ε-contaminated example involved robust filtering of an ε-contaminated first-order Markov signal in ε-contaminated first-order Markov noise. In Figures 15-22, we consider other nominal signal and noise models. In particular, Figure 15 gives the performances of H_0^* and H_R^* when the nominal signal and noise power spectral densities, σ_0 and ν_0, are second-order Markov, i.e.

$$\sigma_0(w) = \frac{4\,\alpha_S^3\,v_S^2}{(\alpha_S^2 + w^2)^2}$$

and

$$\nu_0(w) = \frac{4\,\alpha_N^3\,v_N^2}{(\alpha_N^2 + w^2)^2}$$

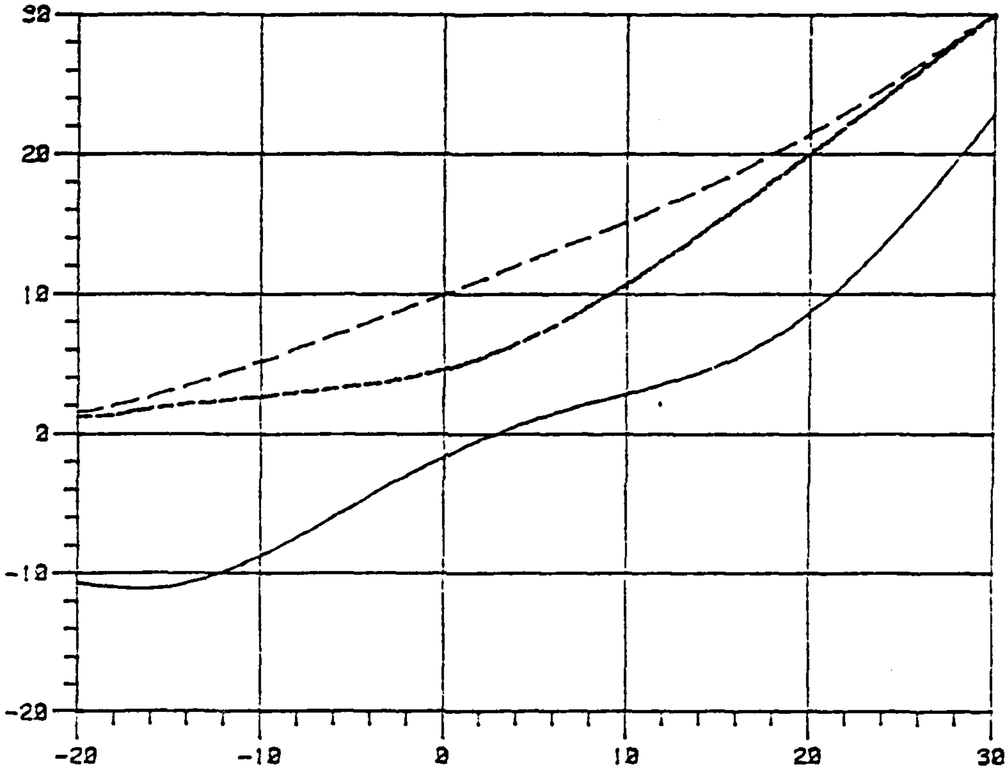


Figure 10. p-point example from Chapter II, $\alpha_s = 1$, $\alpha_N = 100$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at any $(\sigma, \nu) \in \mathcal{S} \times \mathcal{N}$, H_0^* at its worst case.

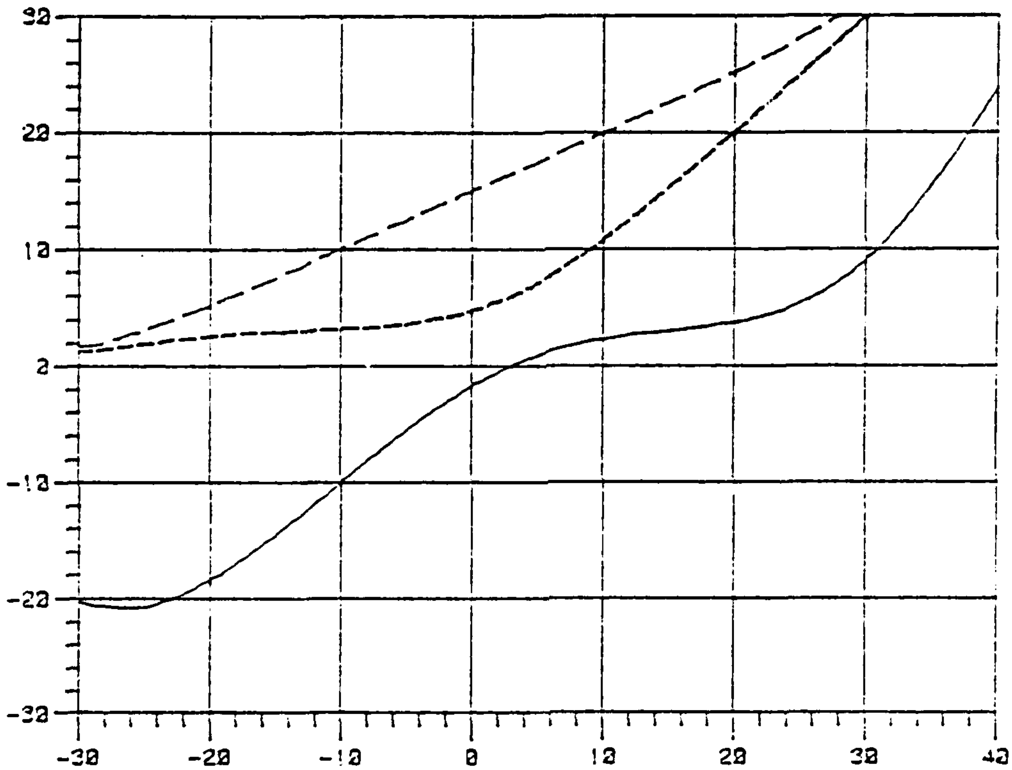


Figure 11. p-point example from Chapter II, $\sigma_S = 1$, $\sigma_N = 1000$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at any $(\sigma, \nu) \in \mathcal{S} \times \mathcal{N}$, H_0^* at its worst case.

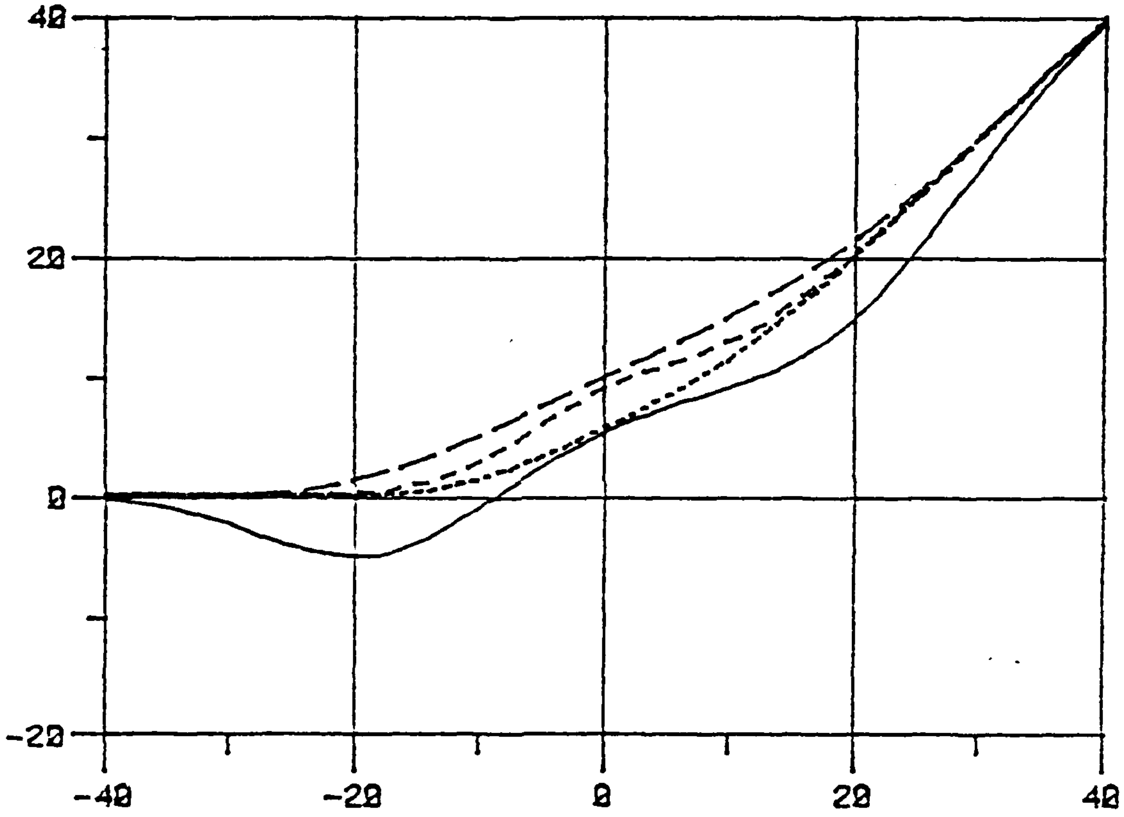


Figure 12. ε-contaminated example from Chapter II, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 100$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

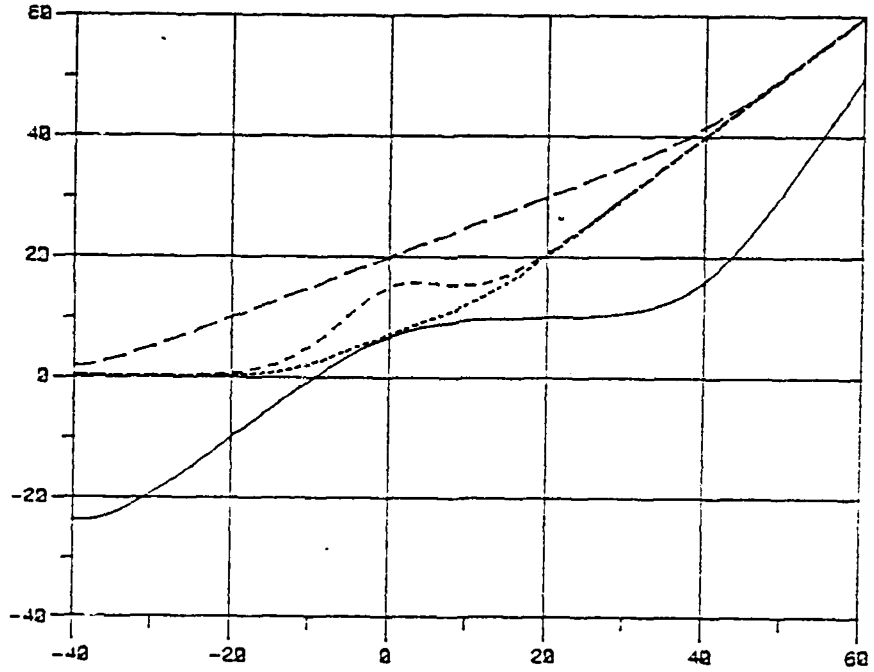


Figure 13. ε-contaminated example from Chapter II, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10^4$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

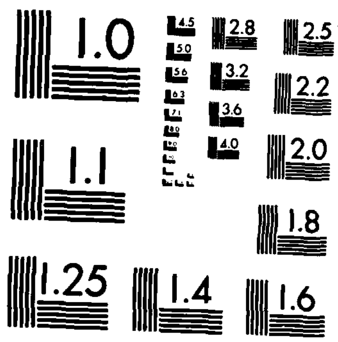
AD-A125 931 TOPICS IN ROBUST STATISTICAL SIGNAL PROCESSING(U) 2/2
ILLINOIS UNIV AT URBANA COORDINATED SCIENCE LAB
K S VASTOLA SEP 82 R-965 N00014-79-C-0424

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

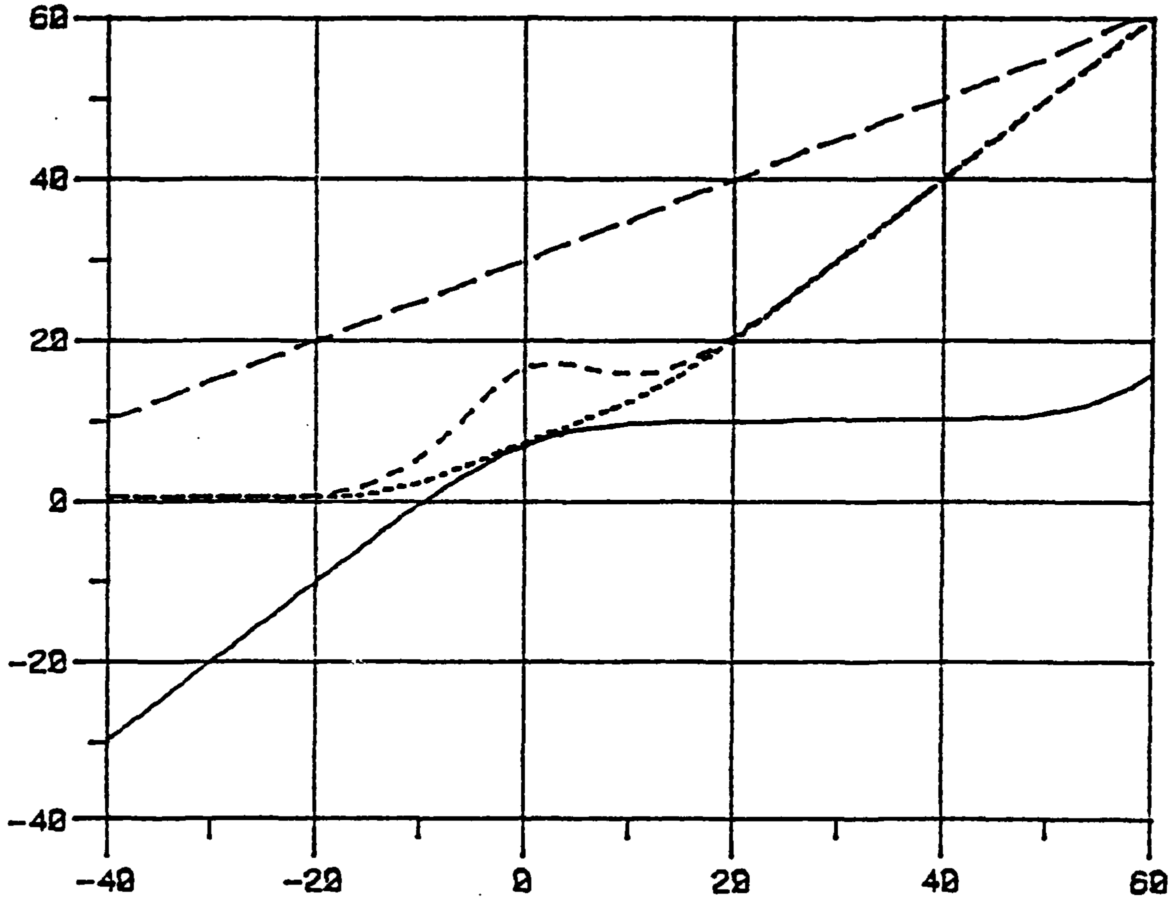


Figure 14. ε-contaminated example from Chapter II, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10^6$, (from top to bottom) H_0^* at (σ_0,ν_0), H_R^* at (σ_0,ν_0), H_R^* at (σ_L,ν_L) (H_R^*'s worst case), H_0^* at its worst case.

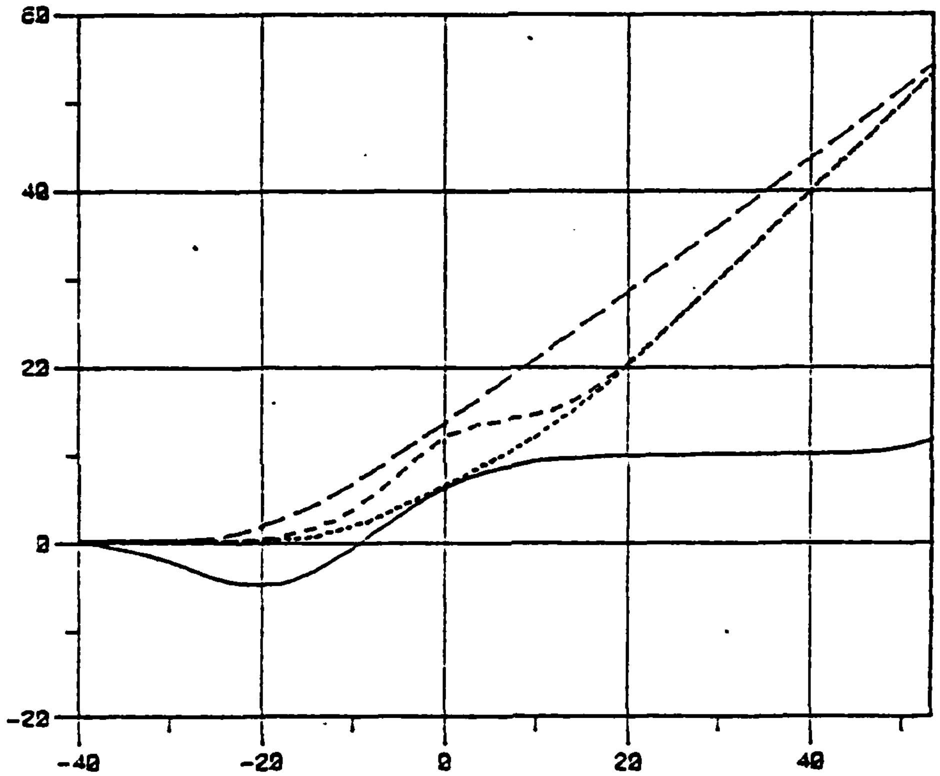


Figure 15. ε-contaminated second-order Markov signal in ε-contaminated second-order Markov noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 100$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

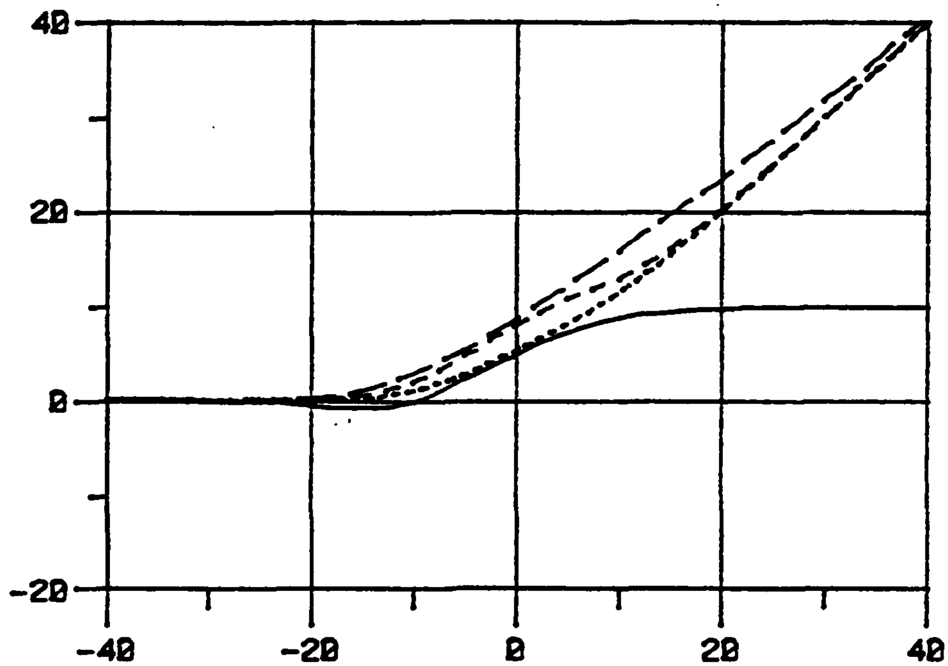


Figure 16. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

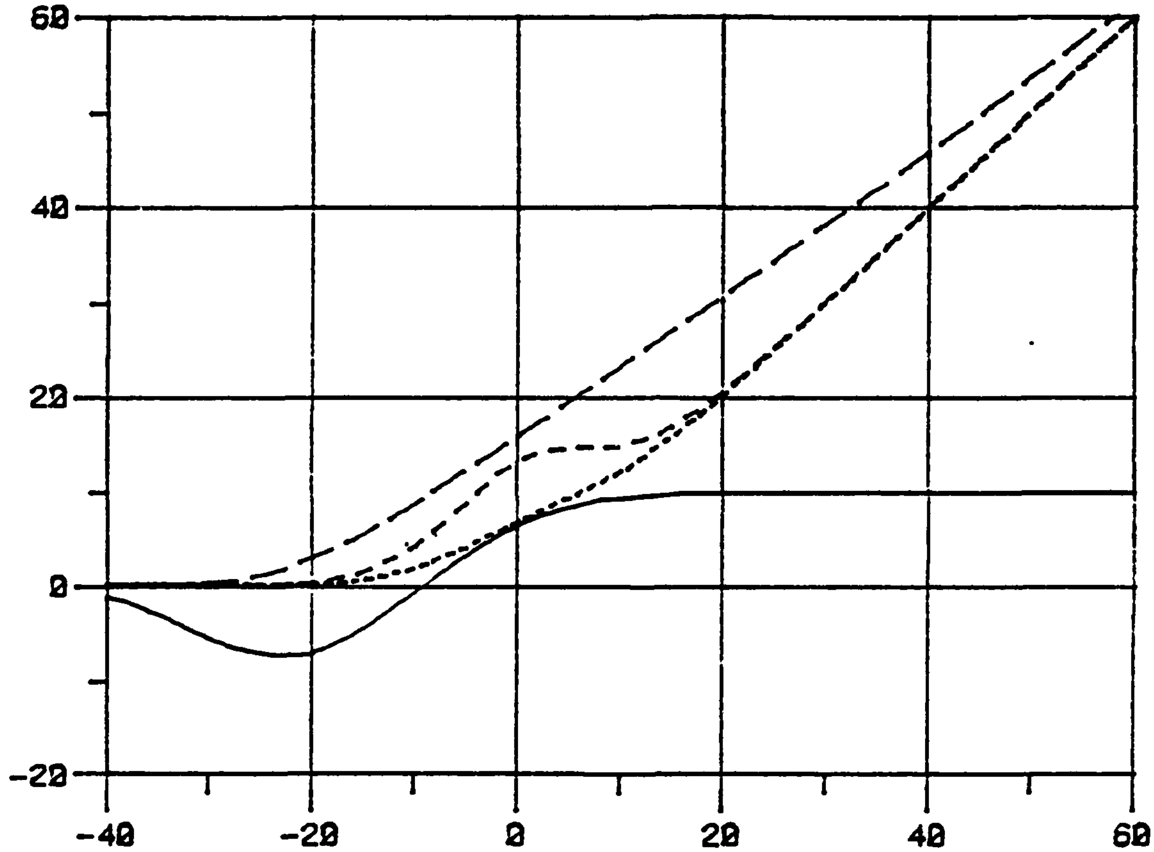


Figure 17. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 100$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

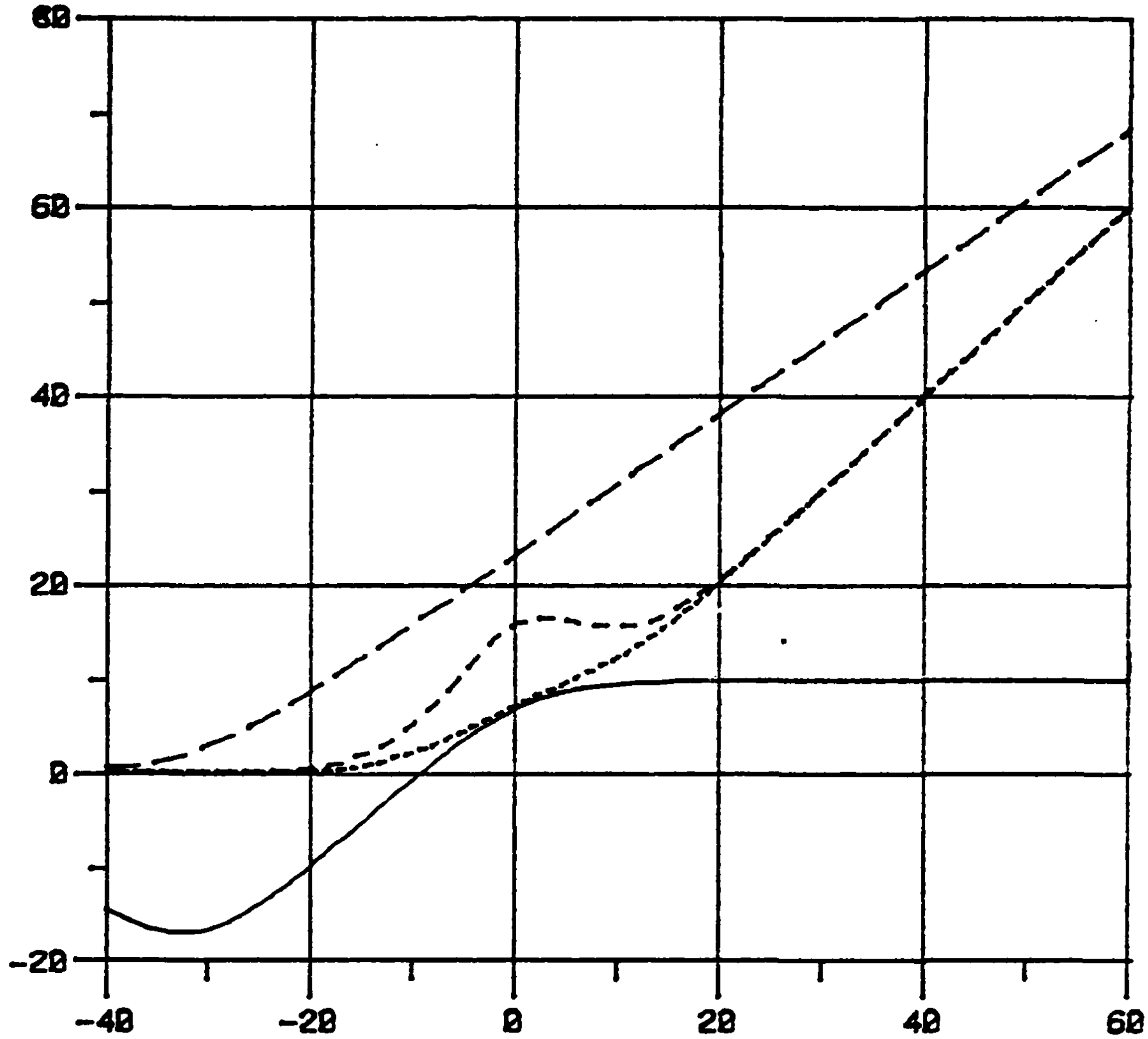


Figure 18. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10^3$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

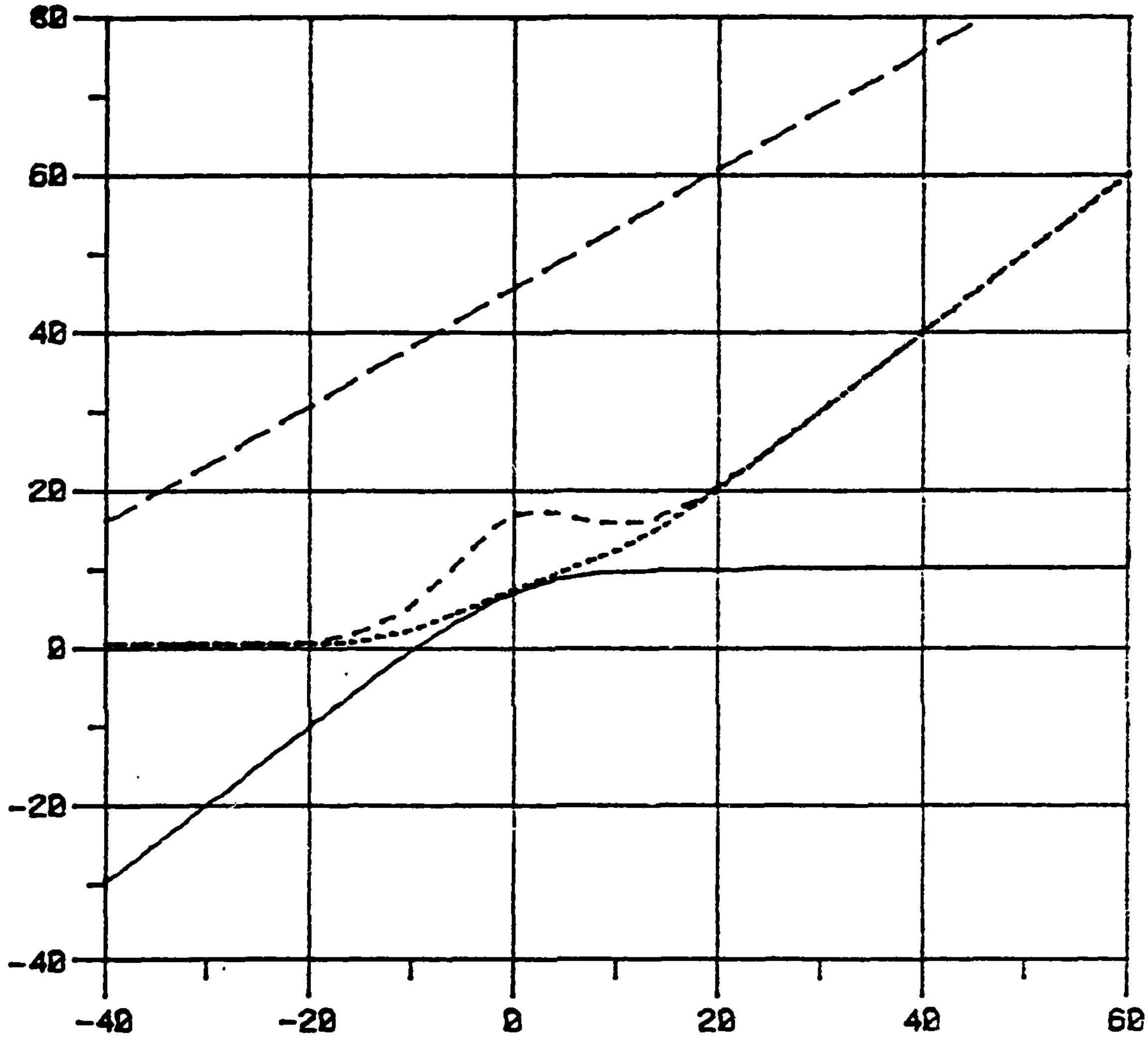


Figure 19. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10^6$, (from top to bottom) H_0^* at (σ_0,ν_0), H_R^* at (σ_0,ν_0), H_R^* at (σ_L,ν_L) (H_R^*'s worst case), H_0^* at its worst case.

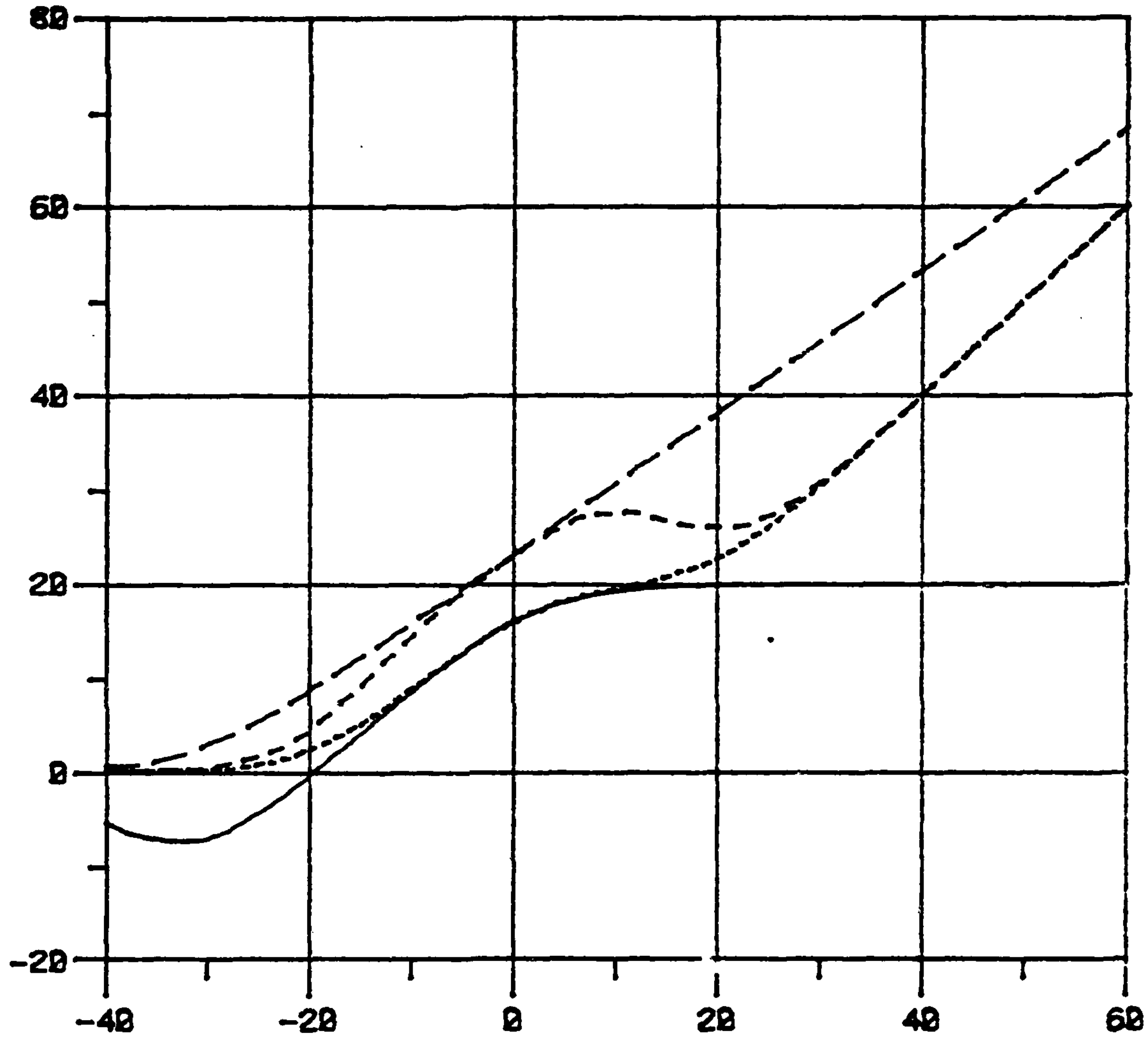


Figure 20. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .01$, $\alpha_S = 1$, $\alpha_N = 10^3$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

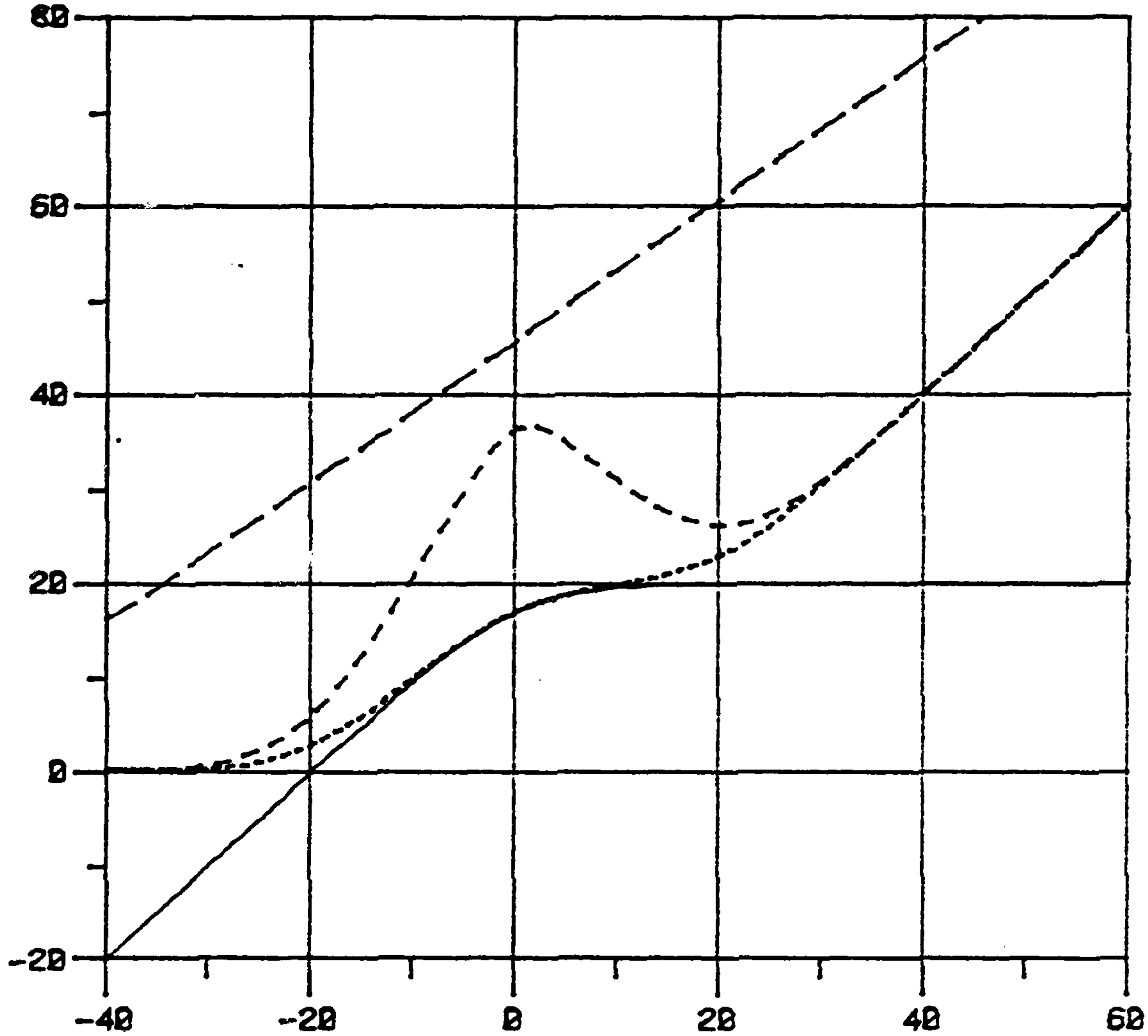


Figure 21. ε-contaminated second-order Markov signal in ε-contaminated first-order Markov noise, $\varepsilon = .01$, $\alpha_S = 1$, $\alpha_N = 10^6$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

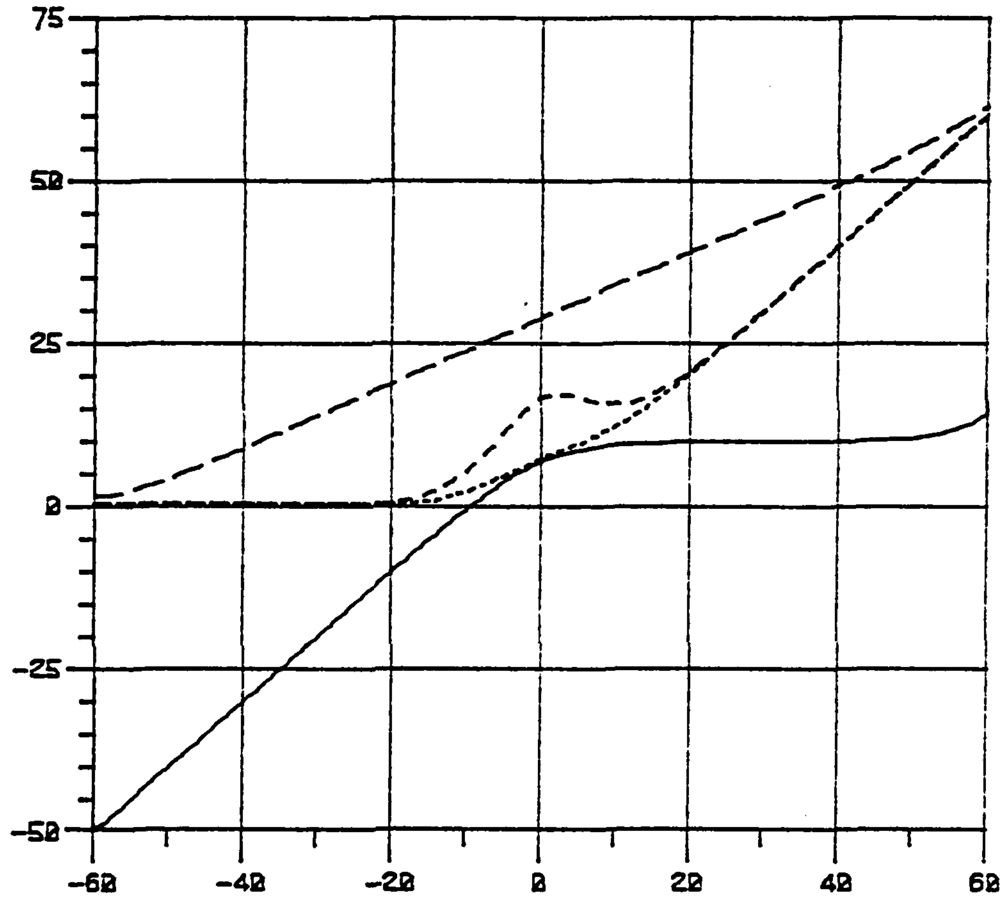


Figure 22. ε-contaminated first-order Markov signal in ε-contaminated band-limited white noise, $\varepsilon = .1$, $\alpha_S = 1$, $\alpha_N = 10^6$, (from top to bottom) H_0^* at (σ_0, ν_0), H_R^* at (σ_0, ν_0), H_R^* at (σ_L, ν_L) (H_R^*'s worst case), H_0^* at its worst case.

In Figures 16-21, σ_0 is second-order Markov and ν_0 is first-order Markov. In all these cases (Figures 15-21) both signal and noise uncertainty are modeled via ε-contaminated classes. In Figures 15-19, we have $\varepsilon = .1$ and α_N varying, and we see that the results are fairly similar to those already presented when both σ_0 and ν_0 are first-order Markov. In Figures 20 and 21, we have $\alpha_N = 10^3$ and $\alpha_N = 10^6$ as in Figures 18 and 19, respectively; however we have $\varepsilon = .01$ in Figures 20 and 21. Note the surprisingly strong similarity between 18 and 20 and between 19 and 21. Finally, Figure 22 is included to substantiate the claim made in the penultimate paragraph of Chapter II. Figure 22 gives performances for the case when σ_0 is an ε-contaminated first-order Markov spectrum and ν_0 is ε-contaminated bandlimited white noise. In Chapter II, we claimed that even when the bandwidth of the ε-contaminated bandlimited white noise is very large (in Figure 22 it is 10^6) the results for H_R^* are similar to the other cases and unlike those involving nonbandlimited white noise as in Figure 3. Compare Figure 22 with Figures 14 and 19, for example; they are virtually identical.

REFERENCES

1. S.A. Kassam and T.L. Lim, "Robust Wiener filters," J. Franklin Inst., Vol. 304, pp. 171-185, 1977.

2. H.V. Poor, "On robust Wiener filtering," IEEE Trans. Automatic Control, Vol. AC-25, pp. 531-536, 1980.

3. J. Snyders, "Error formulae for optimal linear filtering, ·diction and interpolation of stationary time series," Ann. Math. S ·ist., Vol. 43, pp. 1935-1943, 1972.

4. K. Hoffman, Banach Spaces of Analytic Functions. Englewoo ·iffs, NJ: Prentice-Hall, 1962.

5. A.H. El-Sawy and V.D. VandeLinde, "Robust detection of known signals," IEEE Trans. Inform. Theory, Vol. IT-23, pp. 722-727, 1977.

6. P.J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson Lemma for capacities," Ann. Statist., Vol. 1, pp. 251-263, 1973.

7. H.V. Poor, "Minimax linear smoothing for capacities," Ann. Prob., Vol. 10, pp. 504-507, 1982.

8. 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, 1979.

9. P. Billingsley, Convergence of Probability Measures. New York: John Wiley and Sons, 1968.

10. N. Dunford and J.T. Schwartz, Linear Operators, Part I. New York: Interscience, 1958.

11. K. Fan, "Minimax theorems," Proc. Nat. Acad. Sci., Vol. 39, pp. 42-47, 1953.

12. G. Choquet, "Theory of capacities," Ann. Inst. Fourier, Vol. 5, pp. 131-292, 1953-54.

13. S.A. Tretter, Introduction to Discrete-time Signal Processing. New York: John Wiley and Sons, 1976.

14. P.J. Huber, "A robust version of the probability ratio test," Ann. Math. Stat., Vol. 36, pp. 1753-1758, 1965.

15. J. Snyders, "Error expressions for optimal linear filtering of stationary processes," IEEE Trans. Inform. Theory, Vol. IT-18, pp. 574-582, 1972.

16. K. Yao, "An alternative approach to the linear causal least-square filtering theory," IEEE Trans. Inform. Theory, Vol. IT-17, pp. 232-240, 1971.

17. 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.

18. J. Snyders, "On optimal linear estimation of signals with general spectral distribution," IEEE Trans. Inform. Theory, Vol. IT-20, pp. 654-658, 1974.

19. J. Snyders, "On the error matrix in optimal linear filtering of stationary processes," IEEE Trans. Inform. Theory, Vol. IT-19, pp. 593-599, 1973.

20. H. Rieder, "Least favorable pairs for special capacities," Ann. Statist., Vol. 5, pp. 909-921, 1977.

21. S.A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," IEEE Trans. Inform. Theory, Vol. IT-27, pp. 242-247, 1981.

22. V.P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," Telecomm. Radio Eng. (English translation), Vol. 30/31, pp. 58-64, 1976.

23. Y. Hosoya, "Robust linear extrapolations of second-order stationary processes," Ann. Prob., Vol. 6, pp. 574-584, 1978.

24. D.G. Luenberger, Optimization by Vector Space Methods. New York: John Wiley and Sons, 1969.

25. L.J. Cimini and S.A. Kassam, "Robust and quantized Wiener filters for p-point spectral classes," Proc. 14th Conf. on Inform. Sciences and Systems, Princeton University, Princeton, NJ, pp. 314-319, 1980.

26. T.S. Ferguson, Mathematical Statistics. New York: Academic Press, 1967.

27. P.J. Huber, "Robust confidence limits," Z. Wahrscheinlichkeitstheorie verw. Geb., Vol. 10, pp. 269-678, 1968.

28. A. Kolmogorov, "Interpolation and extrapolation," Bull. Acad. Sci., USSR, Ser. Math., Vol. 5, pp. 3-14, 1941.

29. R.N. Bradt and S. Karlin, "On the design and comparison of certain dichotomous experiments," Ann. Math. Statist., Vol. 27, pp. 390-409, 1956.

30. H. Kobayashi and J.B. Thomas, "Distance measures and related criteria," Proc. 5th Allerton Conf. Circuits Syst., Monticello, IL, pp. 491-500, 1967.

31. S.M. Ali and S.D. Silvey, "A general class of coefficients of divergence of one distribution from another," J. Roy. Stat. Soc. Ser. B, Vol. 28, pp. 131-142, 1966.

32. 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.

33. D.L. Snyder, "Some useful expressions for optimal linear filtering in white noise," Proc. IEEE, Vol. 53, pp. 629-630, 1965.

34. A.J. Viterbi, "On the minimum mean square error resulting from linear filtering of stationary signals in white noise," IEEE Trans. Inform. Theory, Vol. IT-11, pp. 594-595, 1965.

35. A. Gelb, et al., Applied Optimal Estimation. Cambridge, MA: MIT Press, 1974.

36. J.L. Doob, Stochastic Processes. New York: John Wiley and Sons, 1953.

37. P.J. Huber, "The use of Choquet capacities in statistics," Bull. Internat. Statist. Inst., Proc. 39th Session, Vol. 45, No. 4, pp. 181-191, 1973.

38. H.L. Royden, Real Analysis. Toronto: Macmillian, 1968.

39. J.W. Tukey, "A survey of sampling from contaminated distributions," in Contributions to Probability and Statistics (I. Olkin, editor), pp. 448-485, Stanford: Stanford Univ. Press, 1960.

40. 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.

41. D.J. Sakrison, "The rate of a class of random processes," IEEE Trans. Inform. Theory, Vol. IT-16, pp. 10-16, 1970.

42. D.J. Sakrison, "The rate distortion function for a class of sources," Inform. and Control, Vol. 15, pp. 165-195, 1969.

43. P.J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist., Vol. 35, pp. 73-104, 1964.

44. P.J. Huber, Robust Statistics. New York: John Wiley and Sons, 1981.

45. F.R. Hampel, "A general qualitative definition of robustness," Ann. Math. Statist., Vol. 42, pp. 1887-1896, 1971.

46. 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, 1976.

47. V.M. Krasnenker, "Stable (robust) detection methods for signal against a noise background (survey)," Automation and Remote Control (English Translation), No. 5, pp. 640-659, 1980.

48. A.A. Ershov, "Robust methods of evaluating parameters (survey)," Automation and Remote Control (English Translation), No. 8, pp. 66-100, 1978.

49. R.D. Martin, "Robust methods for time series," in Applied Time Series II, (Friendly, ed.). New York: Academic Press, 1981.

50. R.D. Martin, "Robust estimation of signal amplitude," IEEE Trans. Inform. Theory, Vol. IT-18, pp. 596-606, 1972.

51. P. Papantoni-Kazakos, "Robustness in parameter estimation," IEEE Trans. Inform. Theory, Vol. IT-23, pp. 223-231, 1977.

52. E.L. Price and V.D. VandeLinde, "Minimax robust estimation of location and minimizing Fisher information," Proc. IEEE Conf. Decis. and Contr. 15th Symp. Adapt. Process., Clearwater, FL, pp. 443-444, 1976.

53. H.V. Poor, "A general approach to robust matched filtering," Proc. 22nd Midwest Symposium on Circuits and Systems, Philadelphia, PA, pp. 514-518, June 1979.

54. S. Verdú, "A General Approach to Minimax Robust Filtering," M. S. Thesis, Univ. of Illinois, 1981. [Also coordinated Sci. Lab. Tech. Report R-933].

55. V. Barbu and Th. Precupanu, Convexity and Optimization in Banach Spaces. Alphen aan de Rijr, The Netherlands: Sijthoff and Noordhoff, 1978.

56. W.C. Knight, R.G. Pridham, and S.M. Kay, "Digital signal processing for sonar," IEEE Proc. Vol. 69, pp. 1451-1506, 1981.

57. H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, New York: John Wiley and Sons, 1968.

58. K. S. Vastola and H. V. Poor, "On generalized band models in robust detection and filtering," Proc. 14th Conf. Inform. Sci. Sys., Princeton University, Princeton, NJ, pp. 1-5, 1980.

VITA

Kenneth Steven Vastola was born in Brooklyn, New York, on January 21, 1954. He received the Bachelor of Arts degree in Mathematics from Rutgers University in May 1976 and the Master of Science degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in January 1979.

From August 1976 to August 1977 he held a University of Illinois Fellowship and was a Teaching Assistant in the Department of Mathematics of the University of Illinois. From August 1977 to August 1981 he was a Research Assistant with the Coordinated Science Laboratory of the University of Illinois. Since September 1981 he has been a Research Staff Member in the Department of Electrical Engineering and Computer Science of Princeton University.