Evaluation of the Constrained Bayesian Methodology for Signal Detection

By

RICHARD STANLEY POULSEN
B.S. (Massachusetts Institute of Technology) 1972
B.S. (Massachusetts Institute of Technology) 1972
M.S. (University of California, Davis) 1973

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved: Wm. A. Gardner

Committee in Charge

Deposited in the University Library ______ Date ______ Librarian

ABSTRACT

The constrained Bayesian methodology (CBM) is a new approach to the design of structurally constrained statistical inference and decision rules. The methodology is based on constrained minimum mean-squared error estimation of posterior probabilities. The solution for the estimates is specified by a set of linear equations in terms of only the prior probabilities, and moments and conditional moments of prescribed functionals of the observations. The CBM is developed and its applications to signal detection and estimation partially investigated by Gardner in a series of papers. This thesis broadens the development and application of the CBM for signal detection. The general results obtained include the following:

1) The CBM is shown to be equivalent to the constrained maximum generalized signal-to-noise ratio design methodology, which links the CBM to other maximum SNR approaches.
2) The CBM is shown to be a useful tool for parameter estimation, and this is exploited to compare decision rules based on parameter estimates with the rule that chooses the largest estimated posterior probability. The comparisons afford insight into estimation-based decision rules in general, and offer alternatives to degenerate decision rules arising from inadequate posterior estimates.
3) The solution to a general detection problem, that of detecting signals with separable moments in additive white Gaussian noise, is partially characterized, and it is shown that many

of the problems analyzed in this thesis are special cases of this general problem.

These results are given in Chapters II, III, and VI, respectively. In addition to these general results, performance and structural analyses of many detection problems are presented. Specifically, the linear, quadratic, and zero-memory nonlinearity-correlator structures are analyzed in detail with evaluations of probability of error, in Chapters IV, V, and VII, respectively. The analyses show that the CBM is a viable design tool for a wide variety of detection problems, being particularly useful for non-Gaussian noise for which there is no general theory for optimum receiver design. The analyses also provide insight into conventional structures such as matched-filter tapped-delay line receivers for high-speed data transmission, and suggest novel structures such as a modified quadrature correlator receiver for noncoherent reception.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to Professor William A. Gardner for suggesting the topic of this thesis as well as his guidance and encouragement throughout this work.

I would also like to thank Dr. V. Ralph Algazi and Dr. Frank J. Samaniego for serving on my thesis committee and for their time in reading this manuscript.

Finally, my greatest debt and appreciation is to my wife, María de la Luz Poulsen, for her support and understanding during this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1. Motivation and Purpose

Modern communication theory is based largely on Bayesian statistical inference and minimum-risk decision [1]. In particular, the theory of optimum signal detection has grown in the last twenty years around Bayesian and related methods [2]. However, often the Bayesian methodology is not directly applicable because the underlying probabilistic models are incomplete or overly complex. This difficulty is sometimes avoided by compromising the model or by imposing simplifying assumptions which then allow the methodology to be applied. When this is not desirable, an alternative approach would be useful. The value of any such alternative, however, should be measured by criteria such as:

1) generality of application,
2) simplicity of analysis and implementation, and
3) performance comparable to that attainable with conventional techniques.

One such approach is proposed and developed considerably by Gardner [3-8]. This approach, referred to as the constrained Bayesian methodology (CBM), is shown to satisfy the first two criteria quite well and preliminary results obtained by Gardner for some basic examples indicate promise for satisfaction of the third.

It is the purpose of this thesis to evaluate the constrained Bayesian methodology for a wide range of signal detection problems with the following goals with respect to the methodology:

1) to demonstrate and further establish its applicability, pointing out its special strengths and weaknesses,

2) to demonstrate and further establish its simplicity for analysis and implementation,

3) to carry out specific receiver designs and performance analyses, and

4) to seek new interpretations and provide more insight into its characteristics.

2. Classical Bayesian Decision

If two quantities are statistically dependent, then inferences can be made about one, given information about the other. In order to make minimum-risk decisions about one quantity, say x, given observations of the other, say y, it is sufficient to know the set of posterior probabilities $\left\{P_{x/y}[X_i/Y]\right\}_{i=1}^{M}$, where x is assumed to have a discrete distribution with M distinct samples $\{X_i\}$. For example, the minimum probability-of-error (MPE) rule for deciding which of the mutually exclusive events $\{X_i\}$ occurred, given the sample Y, is to decide x is X_i if and only if

$$P_{x/y}[X_i/Y] \geq P_{x/y}[X_j/Y] \; \forall j. \qquad 1.1$$

The posteriors, however, are seldom directly available but are obtained via Bayes' Rule from probability density functions of the observations y (which are assumed to have a continuous distribution):

$$P_{x/y}[X_i/Y] = P_i \, f_{y/x}(Y/X_i)/f_y(Y), \qquad 1.2$$

where

$$f_y(Y) = \sum_{j=1}^{M} P_j \; f_{y/x}(Y/X_j) \qquad 1.3$$

and

$$P_i \triangleq P_x[X_i]. \qquad 1.4$$

Thus, knowing the priors, $\left\{P_i\right\}_{i=1}^{M}$ and the conditional density functions (CDF's) $\left\{f_{y/x}(Y/X_i)\right\}_{i=1}^{M}$ is equivalent to knowing the posteriors. So substituting the equivalent expression for the posteriors from 1.2 into the rule 1.1 yields the equivalent rule: decide x is X_i if and only if

$$P_i \; f_{y/x}(Y/X_i) \geq P_j \; f_{y/x}(Y/X_j) \;\; \forall j. \qquad 1.5$$

3. Alternative Approaches

In practice, it often happens that the priors are unavailable or undefined, in which case it is common to use one of the following strategies [1]:

1) assume equal priors - this corresponds to maximum likelihood,
2) estimate the priors from repeated observations of y, or
3) choose the set of priors which yields the worst case (mini-max).

In other applications, the priors may be known and the CDF's unknown, in which case it is common to adopt one of the following strategies:

1) estimate the CDF's with repeated observations of y (cf., [9,10]), or

2) assume a convenient form for the CDF's, e.g., Gaussian (cf. Van Trees [11]).

Then the test is performed using the estimated or assumed information as if it were true.

On the other hand, alternative approaches are frequently used which attempt to avoid these problems. One common example is the maximum signal-to-noise ratio (SNR) approach. This approach, however, suffers from serious drawbacks which prevent it from being generally useful. For example, the SNR definition does not appear to generalize in a natural way to multiple signal detection. In fact, it seems that other than the CBM, no approach for structurally constrained receiver design has been developed which is as simple as the SNR approach yet general in application.*

4. The Constrained Bayesian Approach

The essence of the new approach is to estimate the posterior probabilities directly. To make this a meaningful estimation problem, a constraint on the form of the estimate is necessary--otherwise the "estimates" would be exactly the true posteriors, since these are non-random functions of the observables. The constraint space, L, is chosen to be linear and is generated by all linear functionals of the images of a set of prescribed transformations of the observables y.

---

*Actually, it is shown in Chapter III that the CBM is equivalent to the maximum <u>generalized SNR</u> design methodology.

The estimation criterion is minimum mean-squared error (MMSE) subject to this constraint. By using the Hilbert space orthogonal projection theorem, necessary and sufficient conditions can be obtained for the estimate, namely that the estimation error must be orthogonal to every vector in the constraint space. This results in a set of linear equations involving only the prior probabilities and moments and conditional moments of the prescribed transformations of y.

At the outset, the strengths and weaknesses of the methodology appear to be:

Strengths

1) the intimate relationship to the Bayesian methodology,
2) estimates are determined by only linear equations,
3) estimates require knowledge of only priors and moments and conditional moments of prescribed transformations of the observables,
4) without knowledge of the priors, the methodology can be used with any of the strategies used with the Bayes methodology,
5) the constraint space can be expanded to make the estimated posteriors approach the exact posteriors, and
6) the designer can choose the receiver structure and can use the posterior estimates to compute estimates of parameters for alternative parameter tests;

Weaknesses

1) the designer must have some insight into the appropriateness of candidate structures since poor performance can result if the constraint is not appropriate.

The major goal of this research is to extend these lists. One of the strengths established in this thesis merits special attention; viz., the equivalence of the CBM with the maximum generalized SNR (GSNR) design methodology. It was found in the course of this work that the receivers derived from the two methodologies are identical except for final gains and biases (i.e., threshold levels). This equivalence allows the CBM to be closely related to SNR approaches which are widely used in practice and gives support to both approaches.

5. Related Previous Work

As mentioned in Section 3, a common alternative to the Bayesian approach is based on approximations to the CDF's $\left\{f_{y/x}(Y/X_i)\right\}_{i=1}^{M}$. This might appear to be similar to the new approach, but it is in fact considerably different in concept and in procedure. The conceptual difference is that the function being estimated is not a probability distribution (density) function. The argument of the function being estimated is the conditioning event, not the event whose probability is being indicated. The procedural difference is that the estimates of the CDF's are based on repeated observations, i.e., many samples of y, whereas posterior estimates use only a single sample. The new methodology is based on the generalization of the concept of the "MMSE linear discriminant" that is employed in empirical procedures for partitioning feature spaces for pattern classification [12,13]. The basic ideas behind the constrained MMSE estimation of posteriors arose during the last decade in research work in the field of pattern recognition [14-16]. However, these basic ideas appear to have been employed only for the development and justification of certain empirical procedures [17], and were evidently not extended or generalized as needed for application to signal detection except for Gardner's work [3-7]. This is evidently related to the vague nature of pattern recognition models, whereas those in statistical communications problems are relatively specific.

There are two methodologies which bear some resemblance to the CBM. First, there is Jaynes' approach to the estimation of prior probabilities [18, see also 19] which is discussed in Section II.7 with application to posterior probability estimation. While appearing similar, Jaynes' approach does not apply to the same types of problems as the CBM. The second approach is Brick and Zames' canonical expansion of likelihood functions [9,10]. Appearing to be the only approach that is actually similar to the CBM, this is discussed further in Section II.7 and a more detailed comparison is made in Section VI.3. The series expansion approach of Schwartz [20] appears to be related to Brick and Zames approach, but is only presented for single random variable observations.

6. Preview of Results

In Chapter II, the methodology is defined and developed with a general solution for the estimates and some illustrative examples. In addition, the equivalence between the CBM and the maximum-GSNR methodology is explored and discussed. In Chapter III, the relative superiority of certain estimation-based decision rules is investigated, particularly for linear constraints. In Chapter IV, the class of linear constraints is developed in general with specific applications including passband equalization, with performance analyses and comparisons with well known results. In Chapter V, the class of quadratic constraints is investigated and developed with examples of the bandpass type. In Chapter VI, the class of N^{th}-order Volterra polynomial constraints for random signals in additive white Gaussian noise (WGN) is investigated and shown to include examples in Chapters IV and V as special cases. In Chapter VII, the zero-memory nonlinearity (ZNL)/correlator structure is analyzed in detail including comparisons of performance with that obtainable using several related design criteria, viz., the deflection (D) and the complementary deflection ($\bar{D}$). Finally, Chapter VIII summarizes and discusses obtained results with suggestions for further research.

7. General Remarks

Since the scope of this thesis is the application of the CBM to signal detection problems, much of the generality of the methodology is not made evident. Therefore, it should be noted that the CBM is potentially very broad in application. Two of the most notable restrictions of this work are that:

1) only discrete events, and therefore only discrete posterior probability distributions, are considered--the extension to continuous posterior distributions corresponds to signal or parameter estimation; and
2) only probability of error (PE) is used as a risk, although this restriction is not essential.

For a more thorough treatment and unrestricted view of the methodology, the reader is referred to the papers by Gardner [3-8], especially [3].

Most of Chapters II, IV and parts of Chapters III and V consist of Gardner's development of the CBM, and do not represent work on my part. Chapters VI and VII, and the remainder of Chapters II, III, and V do, however, consist of my original contributions to the development and application of the CBM. More specifically, my contributions in Chapter II are Sections 2, 3, 5.2, 5.3, and 7; in Chapter III are Sections 2.2, 2.3, 3, and 4; in Chapter IV are Sections 3-5; in Chapter V are part of Section 2, and Sections 3-5. More specifically, Gardner's work is referenced in the text where appropriate.

8. Glossary

In this section is a list of abbreviations and notation that should facilitate the reading of this thesis.

<u>Abbreviations</u>

ASK	amplitude-shift keyed
APK	amplitude- and phase-shift keyed
CBM	constrained Bayesian methodology
CDF	conditional density function
ET	estimation theorist's (rule)
FOB	fourth-order Butterworth (noise density)
GL	Gaussian-like (random variable or process)
GSNR	generalized signal-to-noise ratio
HT	hypothesis tester's (rule)

Abbreviations (Continued)

ISI	intersymbol interference
LC	linearly constrained (rule)
L-MMSE	L-constrained MMSE
LMS	least mean-square
M-ary	having M values or realizations (pertains to a signal)
MMSE	minimum mean-squared error
MPE	minimum probability of error
MS	mean-square
MSE	mean-squared error
OG	optimum-for-Gaussian noise
OI	orthogonal isonormal
OP	orthogonal projection
PAM	pulse amplitude modulation
PDF	probability density function
PE	probability of error
PSK	phase-shift keyed
QAM	quadrature amplitude modulation
QC	quadratically constrained (rule)
RS	regular simplex
SNR	signal-to-noise ratio
WGN	white Gaussian noise
ANL	zero-memory nonlinearity

Symbols

D	deflection
$\overline{D}$	complementary deflection

Symbols (Continued)

$E\{\cdot\}$	expected value
H	Hilbert space generated by the finite mean-square images of all functionals of the observables, y
H_i	the hypothesis that the i^{th} signal is present
K	kurtosis (ratio of the fourth centered moment to the square of the variance of a random variable)
$M_x^{(n)}$	the n^{th} moment of x (e.g., $M_x^{(2)}(t_1,t_2) = E\{x(t_1)x(t_2)\}$) - for n=1, the superscript is omitted
$M_{x/H_i}^{(n)}$	the n^{th} conditional moment of x, conditioned on H_i
P_i	the prior probability of H_i, i.e., $P_i \triangleq Pr[H_i]$
$P_{i/Y}$	the posterior probability of H_i, given Y, i.e., $P_{i/Y} \triangleq Pr[H_i/Y]$
$\bar{x}$	the centered version of x, i.e., $\bar{x} \triangleq x - E\{x\}$
$\tilde{x}$	the orthogonal projection of x onto the Hilbert space H, also equivalent to the MMSE estimate of x, given Y
$\hat{x}$	the orthogonal projection of x onto a Hilbert subspace L, of H
σ_x^2	the variance of x, i.e., $\sigma_x^2 \triangleq E\{\bar{x}^2\}$

CHAPTER II

DEFINITION AND GENERAL DEVELOPMENT OF THE METHODOLOGY FOR SIGNAL DETECTION

1. Definition

The general situation to which the methodology directly applies is the following. Given the prior probabilities, $\{P_i\}$, and some moments of the conditional distributions $\{F(\cdot/X_i)\}$, and of the distributions $\{F(\cdot)\}$, (e.g., means and covariances), find a decision rule that employs these to approximate the Bayes minimum risk rule. The methodology, to be referred to as the "constrained Bayesian methodology" (CBM), is based on constrained MMSE estimation of posterior probabilities. The posterior estimates are used as true posteriors in the decision rule: "decide X_i occurred if and only if

$$\hat{P}[X_i/Y] \geq \hat{P}[X_j/Y] \ \forall j." \tag{2.1}$$

The estimation criterion used in this methodology is minimum mean-squared error (MSE)

$$MSE \triangleq E\{(\hat{P}[X_i/y] - P[X_i/y])^2\} \tag{2.2}$$

over all random variables contained in the Hilbert space L, generated by all linear functionals of images of prescribed transformations of the observables y.

2. Clarifying Example

To illustrate the procedure of the methodology, consider the following simple example. Assume that the observable, y, is a continuously distributed random variable which is related to a discretely distributed random variable x. Also assume that the constraint on the form of the estimate is an N^{th}-order polynomial

$$\hat{P}_{i/Y} \triangleq \hat{P}[X_i/Y] = \sum_{n=0}^{N} \phi_n^i \overline{Y}^n \quad , \qquad 2.3$$

where

$$\overline{Y} \triangleq Y - E\{Y\} \ . \qquad 2.4$$

Minimizing MSE (2.2) yields the following set of necessary-and-sufficient linear equations

$$\sum_{m=0}^{N} M_{n+m} \phi_m^i = P_i M_{n/X_i} \quad , \qquad 2.5$$

$$M_n \triangleq E\{\overline{Y}^n\} \ , \qquad 2.6$$

$$M_{n/X_i} = E\{\overline{Y}^n/X_i\} \ . \qquad 2.7$$

Then $\underline{\phi}^i \triangleq \{\phi_m^i\}_{m=0}^N$ can be represented as

$$\underline{\phi}^i = P_i Q^{-1} \underline{M}_i \qquad 2.8$$

where the ij^{th} element of Q is

$$Q_{ij} \triangleq M_{i+j} \quad , \qquad 2.9$$

and the j^{th} element of $\underline{M}_i$ is

$$(\underline{M}_i)_j = M_{j/X_i} \quad . \tag{2.10}$$

To illustrate the performance attainable with the simplest estimate (linear, i.e., N=1) which requires knowledge of only the variance and conditional mean of y, consider the following specific example. Let y be related to x by

$$y = x + n \tag{2.11}$$

where x is a binary random variable with samples X_1, X_2 $(X_1 < X_2)$ and prior distribution P_1, P_2, and where n is a statistically independent zero-mean random variable with Laplacian (double-sided exponential) PDF

$$f_n(N) = (\sigma_n\sqrt{2})^{-1} \exp(-|N|\sqrt{2}/\sigma_n) \quad . \tag{2.12}$$

From 2.3-2.10, the estimate is obtained:

$$\hat{P}_{i/Y} = \phi_0^i + \phi_1^i \,\overline{Y} \quad , \tag{2.13}$$

$$\phi_0^i = P_i \quad , \tag{2.14}$$

$$\phi_1^i = (-1)^i \eta d/(\sigma_x^2 + \sigma_n^2) \quad , \tag{2.15}$$

where

$$\eta \triangleq P_1P_2 \quad , \tag{2.16}$$

$$d \triangleq X_2 - X_1 \quad , \tag{2.17}$$

$$\sigma_x^2 \triangleq \eta d^2 \quad . \tag{2.18}$$

The true posterior is

$$P_{i/Y} \triangleq P[X_i/Y] = P_i\, f_{y/H_i}(Y/H_i)/f_y(Y)$$

$$= f_{y,H_i}(Y,H_i)/[f_{y,H_1}(Y,H_1) + f_{y,H_2}(Y,H_2)]$$

$$= \{1 + (P_i^{-1} - 1)\ \exp[-\sqrt{2}\ (|Y + (-1)^i d - X_i|$$

$$-|Y - X_i|)/\sigma_n]\}^{-1} \quad . \qquad 2.19$$

The normalized MSE of estimation, ε_1,

$$\varepsilon_1 \triangleq 10 \log_{10} \left\{E\{(\hat{P}_{1/Y} - P_{1/Y})^2\}/E\{P_{1/Y}^2\}\right\} \qquad 2.20$$

is plotted in Figure 2.1 for priors $P_1 = 0.1,\ 0.5,\ 0.9$ as a function of SNR, ρ_{dB},

$$\rho_{dB} \triangleq 10 \log_{10}(\sigma_x^2/\sigma_n^2) \quad . \qquad 2.21$$

The mean-squared error goes to zero as ρ_{dB} approaches either $-\infty$ or $+\infty$. This is due to the fact that the posteriors and the estimates both approach the priors as ρ_{dB} approaches $-\infty$ and that both approach either 0 or 1 as ρ_{dB} approaches $+\infty$.

3. The Relation Between Estimation Error and Decision Error

For binary testing, the rules 1.1, 2.1 are simply

$$P_{1/Y} \underset{H_2}{\overset{H_1}{\gtrless}} P_{2/Y} \qquad 2.22$$

and

$$\hat{P}_{1/Y} \mathop{\gtrless}_{H_2}^{H_1} P_{2/Y} \quad ; \qquad 2.23$$

however, using the fact that the posteriors and their estimates sum to one (see Section 4, Property 5), 2.22 and 2.23 are equivalent to

$$P_{1/Y} \mathop{\gtrless}_{H_2}^{H_1} 1/2 \quad , \qquad 2.24$$

$$\hat{P}_{1/Y} \mathop{\gtrless}_{H_2}^{H_1} 1/2 \quad . \qquad 2.25$$

The linearly constrained (LC) rule (2.25) can be simplified further for this simple case to a threshold test on Y:

$$Y \mathop{\gtrless}_{H_2}^{H_1} \gamma_L \triangleq \frac{1}{2}(X_1 + X_2) + \frac{1}{2}\sigma_n^2(P_2 - P_1)/\eta \quad . \qquad 2.26$$

This rule is similar to the optimum rule for Gaussian n (i.e., the OG rule), viz.,

$$Y \mathop{\gtrless}_{H_2}^{H_1} \gamma_G \triangleq \frac{1}{2}(X_1 + X_2) + \sigma_n^2 \ln(P_2/P_1) \quad , \qquad 2.27$$

which can be used to compare the performance of the LC rule. To illustrate the relationship between estimation error and decision error, $\hat{P}_{1/Y}$, $P_{1/Y}$, f_y, and f_{y,H_i}, $i=1,2$ ($f_{y,H_i} = P_i\, f_{y/H_i}$) are plotted in Figure 2.2 for $P_1 = 0.1$, and $\rho_{dB} = 0, 10$. As can be seen graphically, and from comparing 2.25 with 2.26, γ_L is just that value of Y for which $\hat{P}_{1/Y} = \frac{1}{2}$. Similarly for the unconstrained rule, the optimum threshold, denoted γ_o, is that value of Y for which $P_{1/Y} = \frac{1}{2}$. For a given γ, the probability of error (PE) for the rule using γ as a threshold is the area under the curve f_{y,H_2} to the left of γ, representing errors under H_2, plus the area under f_{y,H_1} to the right of γ, representing errors under H_1, i.e.,

$$PE(\gamma) = \int_{-\infty}^{\gamma} f_{y,H_2}(\sigma)\, d\sigma + \int_{\gamma}^{\infty} f_{y,H_1}(\sigma)\, d\sigma \qquad 2.28$$

$$= P_2 \int_{-\infty}^{\gamma} f_{y/H_2}(\sigma/H_2)\, d\sigma + P_1 \int_{\gamma}^{\infty} f_{y/H_1}(\sigma/H_1)\, d\sigma$$

$$= P_2 \int_{-\infty}^{\gamma} f_n(\sigma - X_2)\, d\sigma + P_1 \int_{\gamma}^{\infty} f_n(\sigma - X_1)\, d\sigma$$

$$= P_2\, g_n(\gamma - X_2) + P_1\, g_n(X_1 - \gamma),$$

where

$$g_n(x) = \int_{-\infty} f_n(\sigma)\, d\sigma \quad . \qquad 2.29$$

Both γ_L and γ_o appear naturally in Figure 2.2 and γ_G is included for comparison. The observations that result in different decisions from the two rules 2.24, 2.25, lie between γ_L and γ_o. The resultant difference in PE is the shaded area. Note that for $\rho_{dB} = 0$, γ_L and γ_G straddle the optimum value, γ_o, γ_G actually being closer (evaluation of PE shows that 2.27 is superior to 2.26). For $\rho_{dB} = 10$, however, γ_G is further away from γ_o on the same side as γ_L, and clearly 2.26 is superior to 2.27. The probability of error for both rules is plotted in Figure 2.3 for three cases where the random variable n has the PDF's:

1) Gaussian

$$f_n(N) = (\sigma_n\sqrt{2\pi})^{-1}\exp[-(N/\sigma_n)^2/2] \quad ; \qquad 2.30$$

2) Laplacian (2.12); or

3) fourth-order Butterworth (FOB)

$$f_n(N) = (\sqrt{2}/\pi\sigma_n)\,[1 + (N/\sigma_n)^4]^{-1} \quad . \qquad 2.31$$

In all three cases, the difference in PE for the two rules is negligible for all values of ρ_{dB}, and for all values of priors that are within an order of magnitude of each other. There is a distinct tendency, however, for the LC receiver to outperform the OG receiver at high SNR (except, of course, for Gaussian n).

It is not always true, however, that a more accurate (in terms of MSE) estimate will yield a lower decision error. An example of this anomaly is given here for the quadratic-plus-linear estimate:

$$P_{i/Y} = \phi^i_{0Q} + \phi^i_{1Q}\,\overline{Y} + \phi^i_{2Q}\,\overline{Y}^2 \qquad 2.32$$

where

$$\phi^i_{0Q} = P_i - \phi^i_{2Q}(\eta d^2 + \sigma^2_n) \quad , \qquad 2.33$$

$$\phi^i_{1Q} = C_i[(K - 1)\sigma^2_n + 4\eta d^2] \quad , \qquad 2.34$$

$$\phi^i_{2Q} = -C_i d(P_2 - P_1) \quad , \qquad 2.35$$

$$C_i \triangleq (-1)^i\, d[d^4 + (K+3)d^2\,\sigma^2_n + (K-1)\sigma^4_n/\eta]^{-1} \quad , \qquad 2.36$$

$$K = E\{n^4\}/\sigma_n^4 \quad . \tag{2.37}$$

To demonstrate this anomalous behavior, the quadratic (QC) estimate (2.32) is shown in Figure 2.4 along with the other curves from Figure 2.2, for Laplacian n, and in Figure 2.5 for Gaussian n. Comparison of γ_L, γ_Q, and γ_o reveals that for $\rho_{dB} = 0$, the QC rule is superior to the LC rule whereas for $\rho_{dB} = 10$, the opposite is true. It is also interesting to note that for the Laplacian case, the OG rule is superior to the QC rule for both values of ρ_{dB}. The probability of error for the two rules is plotted in Figure 2.6 for Gaussian and Laplacian PDF's. (The FOB PDF has an infinite kurtosis, K, so the QC rule is identical to the LC rule.) These plots display the relative superiority of the LC rule for higher SNR. Hence, the added accuracy of estimation only degrades performance in this case.

To further illustrate this anomaly, PE is shown in Table 2.1, for the following rules:

- LC - the linearly constrained rule (2.13),
- QC - the quadratically constrained rule (2.32),
- CC - the cubically constrained rule,
- CNQ - the cubically constrained rule, but with no quadratic term,
- MIN - the minimum PE rule,
- MAX - the maximum PE* rule, decide H_i if $P_i > \frac{1}{2}$,

---

*The maximum PE rule makes no use of the observations, simply always choosing the hypothesis with the largest prior probability.

Table 2.1. Probability of Error ($10 \log_{10}(PE)$) for Various Binary Decision Rules

P_1 \ ρ_{dB}	0		5		10	
0.1	MIN	-10.06	MIN	-15.37	MIN	-67.74
	QC	-10.05	QC	-14.50	CC	-66.54
	LC	-10.00	LC	-13.54	LC	-65.45
	CC	-10.00	CNQ	-13.41	CNQ	-42.31
	CNQ	-10.00	MAX	-10.00	QC	-39.90
	MAX	-10.00	CC	- 0.92	MAX	-10.00
0.3	MIN	- 5.97	MIN	-12.93	MIN	-65.82
	QC	- 5.96	LC	-12.92	LC	-65.81
	LC	- 5.96	QC	-12.30	CNQ	-50.84
	CNQ	- 5.51	CNQ	- 9.57	QC	-49.20
	MAX	- 5.23	MAX	- 5.23	CC	-14.87
	CC	- 4.94	CC	- 4.94	MAX	- 5.23
0.49	MIN	- 5.11	MIN	-12.45	MIN	-65.43
	LC	- 5.11	LC	-12.45	LC	-65.43
	QC	- 5.11	QC	-12.45	CNQ	-65.37
	CNQ	- 4.91	CNQ	- 9.48	QC	-65.28
	CC	- 4.91	CC	- 9.46	CC	-61.70
	MAX	- 3.10	MAX	- 3.10	MAX	- 3.10

Gaussian noise

P_1 \ ρ_{dB}	0		5		10	
0.1	MIN	-10.00	MIN	-14.92	MIN	-35.94
	MAX	-10.00	QC	-14.29	LC	-35.47
	CC	-10.00	CNQ	-14.01	CNQ	-33.41
	LC	- 9.99	LC	-13.41	QC	-31.96
	CNQ	- 9.99	MAX	-10.00	MAX	-10.00
	QC	- 9.99	CC	- 0.61	CC	- 0.03
0.3	MIN	- 6.46	MIN	-13.10	MIN	-34.10
	CC	- 6.18	LC	-13.10	LC	-33.92
	CNQ	- 6.14	QC	-12.85	QC	-32.10
	QC	- 5.76	CNQ	-11.91	CNQ	-31.94
	LC	- 5.76	CC	- 6.68	CC	-17.79
	MAX	- 5.23	MAX	- 5.23	MAX	- 5.23
0.49	MIN	- 6.08	MIN	-12.72	MIN	-33.72
	LC	- 6.08	LC	-12.72	LC	-33.72
	QC	- 6.08	QC	-12.72	CNQ	-33.72
	CNQ	- 6.04	CNQ	-12.08	QC	-33.71
	CC	- 6.04	CC	-12.08	CC	-33.42
	MAX	- 3.10	MAX	- 3.10	MAX	- 3.10

Laplacian noise

for Gaussian and Laplacian n, $P_1 = 0.1, 0.3, 0.49$, and for $\rho_{dB} = 0$, 5, 10. Note that the term "cubically constrained," for example, implies that all order terms up to the third order are included in the constraint, unless otherwise noted. Hence, the QNC rule computes terms of third and first orders. The first four rules listed above were the most robust of all possible polynomial forms up to third degree. The CC rule, while performing poorly in general, is included to emphasize the extreme degradation in performance that is possible with added estimation accuracy. This section illustrates the seemingly strange fact that a more accurate estimate may yield a larger probability of decision error. The conclusion to be drawn from this is that the type of estimation constraint is highly important to the performance of the estimate. Of course, in the limit as the number of terms in the constraint approaches infinity, i.e., as the estimation error goes to zero, the decision error must approach the minimum possible value.

4. Equivalences and Properties of Posterior Estimates

A part of the theoretical background for the methodology is a number of equivalences and properties for posterior probabilities and their estimates developed by Gardner [3,6,7]. For brevity, they will be summarized here without proof. To do this, it is first necessary to introduce the following definitions:*

1) a random function called a random indicator:

$$\delta(X) = \begin{cases} 1 \text{ if } X \text{ occurs} \\ 0 \text{ otherwise,} \end{cases} \qquad 2.38$$

*The use of the symbols ^ and ~ has been interchanged from the usage in [3,6,7].

2) H, the Hilbert space generated by the finite mean-square (MS) images of all functionals of the observables, y,

3) L, any Hilbert subspace of H,

4) $\tilde{\delta}(X)/y$, the orthogonal projection (OP) of $\delta(X)$ onto H, and

5) $\hat{\delta}(X)/y$, the OP of $\delta(X)$ onto L.

With these definitions and the basic properties of probabilities and OP's, the equivalences presented in Subsection 4.1 can be derived.

4.1 Equivalences

1) $\hat{\delta}(X)/y$ is MS equivalent to the OP of $\tilde{\delta}(X)/y$ onto L. This follows from the smoothing property of OP's [21,22].

2) The OP, $\tilde{\delta}(X)/y$, is MS equivalent to the mean of $\delta(X)$, conditioned on y. This follows from the smoothing property of expectations [22,23] and the orthogonality condition.

3) The conditional mean, $E\{\delta(X)/Y\}$, is the posterior probability $P[X/Y]$. This is true by definition [23].

4) The OP of $\delta(X)$ onto H, $\tilde{\delta}(X)/y$, is MS equivalent to the random posterior probability $P[X/y]$. This follows from equivalences 2 and 3.

5) The OP of $\delta(X)$ onto L is MS equivalent to the OP of $P[X/y]$ onto L, denoted $\hat{P}[X/y]$. This follows from equivalences 1-4.

6) $P[X/y]$ is MS equivalent to the MMSE estimator of $\delta(X)$, given y [24]. This follows from equivalence 4 and the OP theorem [21].

7) $\hat{P}[X/y]$ is MS equivalent to the L-constrained MMSE (L-MMSE) estimator of $P[X/y]$, and of $\delta(X)$, given y. This follows from equivalence 5 and the OP theorem.

4.2 Properties

Throughout this thesis, it is assumed that the subspace L contains zero-variance random variables (i.e., non-random variables), so that L-MMSE estimates exhibit zero-mean L-constrained minimum variance error. Let ϕ denote the null event and Ω the certain event. From the equivalences in Section 4.1 and the linearity of OP operators [21], the following properties of L-MMSE estimators for posterior probabilities can be verified:

1) L-MMSE estimates of posteriors are non-negative on the average:

$$E\{\hat{P}[X/y]\} = P[x] \geq 0 \quad . \qquad 2.39$$

2) L-MMSE estimates of the posterior probability of the certain event equal one (1):

$$\hat{P}[\Omega/Y] = 1 \; . \qquad 2.40$$

3) L-MMSE estimates of the posterior probability of the union of two mutually exclusive events equal the sum of L-MMSE estimates of the posterior probabilities of the individual events:

$$\hat{P}[X \cup Z/Y] = \hat{P}[X/Y] + \hat{P}[Z/Y], \quad X \cap Z = \phi \quad . \qquad 2.41$$

4) L-MMSE estimates for posterior probabilities possess the following higher level properties, which follow from the basic properties 1-3:

$$\hat{P}[\phi/Y] = 0 \ , \tag{2.42}$$

$$\hat{P}[X \cap Z/Y] = \hat{P}[X/Y] - \hat{P}[X \cap Z^c/Y] \ , \tag{2.43}$$

$$\hat{P}[X^c/Y] = 1 - \hat{P}[X/Y] \ , \tag{2.44}$$

$$\hat{P}[X \cup Z/Y] = \hat{P}[X/Y] + \hat{P}[Z/Y] - \hat{P}[X \cap Z/Y], \tag{2.45}$$

$$E\{\hat{P}[X/Y]\} \leq E\{\hat{P}[Z/Y]\} \quad \text{if} \quad X \subset Z \ . \tag{2.46}$$

The above properties parallel those of true probabilities although 2.39 and 2.46 are weaker than the corresponding properties (for which $E\{\cdot\}$ is deleted) for true probabilities.

5) L-MMSE estimates of joint posteriors sum to L-MMSE estimates of corresponding marginal posteriors:

$$\sum_i \hat{P}[X \cup Z_i/Y] = \hat{P}[X/Y] \ , \tag{2.47}$$

$$\sum_i \hat{P}[Z_i/Y] = 1. \tag{2.48}$$

6) L-MMSE estimates of posterior probabilities decompose into the product of the prior probability and the L-MMSE estimate of the ratio of PDF's $f_{y/x}/f_y$, i.e.,

$$\hat{P}[X/Y] = P[X] \ [f_{y/x}(Y/X)/f_y(Y)]^{\wedge} \ , \tag{2.49}$$

where $[\cdot]^{\wedge}$ denotes the L-MMSE estimate of the random variable within the brackets. Let x be a discrete random variable with realizations $\{X_i\}$, and let $\hat{E}\{\cdot/Y\}$ denote wide sense conditional expectation [25].

7) Wide sense, relative to L, conditional moments are MS equivalent to the moments of the L-MMSE estimated posterior distribution:

$$\hat{E}\{x^n/Y\} = \sum_i X_i^n \hat{P}_{i/y} \quad . \tag{2.50}$$

It should be mentioned that most of the results in this section extend to continuously distributed posteriors, which can be used, for example, in a study of signal and parameter estimation.

5. Equivalence of the Constrained Bayesian Methodology to the Generalized Signal-to-Noise Ratio Methodology

5.1 Definition of GSNR and Its Relation to Other SNR Definitions (for Binary Hypotheses)

Gardner [8] has proposed a generalized measure of signal-to-noise ratio, denoted by GSNR, which includes several different definitions of SNR as special cases. The definition applies to the binary test of a statistic τ which is a functional of the observations:

$$\tau = \phi(Y) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma, \tag{2.51}$$

where γ is a threshold level and H_i, $i=1,2$ are mutually exclusive and exhaustive hypotheses being tested. For this situation, the definition of GSNR is

$$\text{GSNR} \triangleq \frac{[E\{\tau/H_1\} - E\{\tau/H_2\}]^2}{\alpha \operatorname{Var}\{\tau/H_1\} + (1-\alpha) \operatorname{Var}\{\tau/H_2\}}, \quad 0 \leq \alpha \leq 1. \qquad 2.52$$

Depending on the choice of α, this measure includes deflection (D) ($\alpha = 0$, H_2 corresponds to no signal present) [26-33] and complementary deflection ($\overline{D}$) ($\alpha = 1$, H_2 corresponds to no signal present) [34,35]. In addition, for situations in which $\operatorname{Var}\{\tau/H_1\} = \operatorname{Var}\{\tau/H_2\}$, such as sure signals in additive noise, GSNR is equivalent to both D and $\overline{D}$ for any choice of α. However, when τ is to be used in the test 2.51, a more natural choice would seem to be $\alpha = P_1$, the prior probability of H_1.

5.2 The Unconstrained and the Constrained GSNR Methodologies

In [8] it is proved that the decision function which maximizes GSNR (using $\alpha = P_1$) subject to a zero-bias constraint is identical to the decision function which minimizes PE. When structural constraints are imposed on the decision function ϕ, a modified general equivalence exists as shown in [8]. This result is given here as a theorem and proved in Appendix I.* To facilitate the comparison of the L-constrained PE-type approach (CBM) with the L-constrained GSNR approach, the probability-estimation decision rule (2.1) from the CBM is reexpressed, using the sum-to-one property of posterior estimates (Property 5, 2.48), by

*My contributions to this result are explained in the acknowledgements in [8].

$$\phi_{PE}(Y) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma_{PE} , \tag{2.53}$$

where

$$\gamma_{PE} \triangleq (P_2/P_1 - 1)/2 \tag{2.54}$$

and

$$\Phi_{PE}(Y) = \hat{P}_{1/Y}/P_1 - 1 , \tag{2.55}$$

and the maximum-GSNR decision rule is expressed (with $\alpha = P_1$) by

$$\Phi_{GSNR}(Y) \underset{H_2}{\overset{H_1}{\gtrless}} \gamma_{GSNR} . \tag{2.56}$$

Then, the result for L-constraints is given by the following theorem.

Theorem: If the minimum attainable PE is nonzero and $\alpha = P_1$ in GSNR, then for any L-constraint the L-constrained maximum-GSNR decision function, subject to the zero-bias constraint

$$E\{\Phi_{GSNR}(y)\} = 0 , \tag{2.57}$$

is identical (except for an arbitrary scale factor β) to the L-constrained minimum-MSE probability-estimation decision function:

$$\Phi_{GSNR}(\cdot) = \beta\Phi_{PE}(\cdot) . \tag{2.58}$$

These equivalences have a two-way effect. First, the second equivalence (constrained) provides considerable justification for the use of the CBM since maximum-SNR techniques are well established, widely used, and reasonably well understood. Second, since the CBM is linked to the optimum Bayes theory by formulation, this gives much insight into and justification for the maximum-GSNR approach.

The fact that the two methodologies are equivalent can be understood both intuitively and analytically from their relationship in terms of quadratic performance measures. The CBM performs minimum mean-squared error estimation of posterior probabilities, and SNR inherently involves quadratic operations: the square of the difference of the conditional means divided by the variances of the statistic. It is nevertheless surprising that the particular quadratic forms that arise in the two approaches are identical.

5.3 Extension to Multiple Signal Detection

There does not appear to be a natural extension of the single SNR performance measure for binary hypotheses to the multiple hypothesis testing problem. However, a definition has been found for a set of individual SNR measures which reduces to the binary case properly and which is equivalent, in the sense described in Section 5.2, to the constrained and unconstrained Bayesian methodologies. The definition applies to the testing of the hypotheses

$$H_i: \ Y \text{ is a sample from } y_i, \quad i=1,\ldots,M \qquad 2.59$$

with the test statistics $\tau_i = \Phi_i(Y)$ in the decision rule: "decide H_i if and only if

$$a_i\tau_i + b_i \geq a_j\tau_j + b_j \qquad \forall j \text{ ,"} \qquad 2.60$$

where the gains and biases $\{a_i, b_i\}$ are to be specified. The GSNR definition for multiple hypotheses is then

$$\mathrm{GSNR}(i) = \frac{[E\{\tau_i/H_i\} - E\{\tau_i/\overline{H}_i\}]^2}{P_i \ \mathrm{Var}\{\tau_i/H_i\} + (1-P_i)\ \mathrm{Var}\{\tau_i/\overline{H}_i\}} \tag{2.61}$$

subject to $E\{\tau_i\} = 0$, where $\overline{H}_i$ denotes the complement of the hypothesis H_i, i.e., not H_i. Now, taking a single hypothesis pair, e.g., $(H_i, \overline{H}_i)$, the results for binary hypotheses apply directly. Thus, the equivalence between the maximum-GSNR methodology and the CBM holds for every i and therefore in general.

6. Solution for L-MMSE Estimates of Posterior Probabilities

6.1 General Solution

For applications to signal detection problems, it will be assumed that the subspace L corresponds to linear functionals of images of prescribed nonlinear transformations of the observables, y, i.e.,

$$\begin{aligned}\hat{P}[X/Y] &= \ell[\{g(y,v),\ v \in V\}] \\ &= \ell[\{g(\{y(t),\ t \in T\},\ v),\ v \in V\}]\end{aligned} \tag{2.62}$$

where ℓ is a linear functional to be optimized, T is an indexing set which is a subset of the reals, and the indexing set V is a Cartesian product of subsets of the reals, and $\{g(\cdot,v),\ v \in V\}$ is a prescribed set of functionals. As mentioned in Section 4.2, it is assumed that there exists a v_o in V such that $g(Y,v_o) = 1$. For an illustrative example of the quantities g and V, see the specific example at the end of the next subsection.

For constraints of the above type, the necessary-and-sufficient condition obtained from the Hilbert space OP theorem [6,25],

$$E\{\hat{P}[X/Y]z\} = E\{P[X/Y]z\}$$

$$= P[X]\ E\{z/X\} \quad \forall z \in L \qquad 2.63$$

can be reduced to a set of explicit linear equations. Summarizing the results obtained by Gardner [6] using the Riesz representation [25],

$$\hat{P}[X/Y] = \langle g(Y),\phi\rangle \ , \qquad 2.64$$

where ϕ is the Riesz representor for the optimum linear functional ℓ^o, these equations are

$$\langle E\{g(y,v)g(y,\cdot)\},\phi(\cdot)\rangle = P[X]\ E\{g(y,v)/X\} \quad \forall v \in V \ . \qquad 2.65$$

Since a constant term is included in the constraint, the solution for $\hat{P}_{i/y}$ is of the form

$$\hat{P}_{i/y} = P_i[1 + \overline{\tau}_i(Y)] \ , \qquad 2.66$$

where the functional τ_i includes only terms that depend on Y.

6.2 General Example

The following is an example of the application of 2.65, to which all of the problems in this thesis will conform (including analogous discrete-time problems). Consider an observation Y that is a sample of a continuous parameter process $\{y(t);\ t \in T\}$, dependent probabilistically on the signal $s(t)$ which has M realizations $\{S_i(t);\ t \in T\}_{i=1}^{M}$. By H_i denote the hypothesis that s is S_i. Let the estimation constraint be a generalized N^{th}-order Volterra polynomial [7,36] in $\overline{Y}$. Then the estimated posterior takes the form

$$\hat{P}_{i/Y} = \sum_{n=0}^{N} \int_{T} \cdots \int_{T} \phi_n^i(\tau_1,\ldots,\tau_n)\overline{Y}(\tau_1)\ldots\overline{Y}(\tau_n)d\tau_1\ldots d\tau_n \qquad 2.67$$

and the solution for $\{\phi_n^i\}_{n=0}^N$ is expressed in terms of the N+1 simultaneous linear equations from 2.65:

$$\sum_{n=0}^{N} \int_{T} \cdots \int_{T} \phi_n^i(\tau_1,\ldots,\tau_n)M_{\overline{y}}^{(n+k)}(\tau_1,\ldots,\tau_n,\ t_1,\ldots,t_k)d\tau_1\ldots d\tau_n$$

$$= P_i\ M_{\overline{y}/H_i}^{(k)}(t_1,\ldots,t_k)\ ,\quad \forall t_j \in T,\ j=1,\ldots,k;\quad k=1,\ldots,N \qquad 2.68$$

where

$$M_{\overline{y}}^{(k)}(t_1,\ldots,t_k) \triangleq E\{\overline{Y}(t_1)\ldots\overline{Y}(t_k)\}\ , \qquad 2.69$$

$$M_{\overline{y}/H_i}^{(k)}(t_1,\ldots,t_k) \triangleq E\{\overline{Y}(t_1)\ldots\overline{Y}(t_k)/H_i\}\ . \qquad 2.70$$

Note that 2.67-2.70 are just the continuous-time counterpart of 2.3-2.6. Gardner has shown that explicit solutions to these integral equations can be obtained for a number of interesting signal detection (and estimation) problems. Chapters IV through VII describe these preliminary results and extend them to a wider range of signal detection problems. Chapter VI investigates constraints of this form in more depth for additive white Gaussian noise (WGN) and signals whose autocorrelation functions are separable. Chapters IV and V examine the classes of linear (N=1) and quadratic (N=2) constraints, respectively, and Chapter VII investigates the zero-memory nonlinearity (ZNL)/correlator structure. In all cases, receiver structures and/or performance are analyzed. The example 2.67 has the natural discrete-time counterpart, which will be used in Chapters III, V, and VII:

$$\hat{P}_{i/Y} = \sum_{n=0}^{N} \left[\sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} \phi_n^i(i_1,\ldots,i_n)\overline{Y}_{i_1}\ldots\overline{Y}_{i_n} \right] . \qquad 2.71$$

A specific (quadratic) example of the above is given below for $N=2$:

$$\hat{P}_{i/Y} = \phi_0^i + \sum_{j=1}^{q} \phi_1^i(j)\, \overline{Y}_j + \sum_{j=1}^{q} \sum_{k=1}^{q} \phi_2^i(j,k)\, \overline{Y}_j\overline{Y}_k \qquad 2.72$$

where $\underline{\phi}^i \triangleq \{\phi_0^i, \phi_1^i, \phi_2^i\}$ is the solution to the following set of $1+q+q^2$ linear algebraic equations

$$\phi_0^i M_{\overline{y}}^{(n)}(i_1,\ldots,i_n) + \sum_{j=1}^{q} \phi_1^i(j) M_{\overline{y}}^{(n+1)}(i_1,\ldots,i_n,j)$$

$$+ \sum_{j,k=1}^{q} \phi_2^i(j,k) M_{\overline{y}}^{(n+2)}(i_1,\ldots,i_n,j,k)$$

$$= P_i M_{\overline{y}/H_i}^{(n)}(i_1,\ldots,i_n) \; ; \quad n=0,1,2; \quad \forall j,k=1,\ldots,q \qquad 2.73$$

where $M_{\overline{y}}^{(n)}$ and $M_{\overline{y}/H_i}^{(n)}$ are as defined in 2.69 and 2.70. Note that for this example, g and V of the general formulation are given by

$$\{g(Y,v),\ v \in V\} = \{1,\ Y(t),\ Y(\tau)Y(\sigma);\ t \in T,\ (\tau,\sigma) \in TxT\}, \qquad 2.74$$

$$V = \{\phi,\ T,\ TxT\} . \qquad 2.75$$

7. Comparison with Prior Methodologies

As mentioned in Section I.5, the CBM is quite different from approaches based on the estimation of conditional density functions, although there are two such approaches that appear somewhat similar to the CBM.

The first is Jaynes' approach to estimating prior probabilities [18, see also 19] using a maximum entropy criterion:

$$\text{Max } H \triangleq -\sum_{i=1}^{M} P_i \log(P_i) \ , \tag{2.76}$$

with the constraints that the estimates must be positive and sum to one. The estimates are obtained from linear equations that are fully specified by (prior) moments of the distribution to be estimated. While seeming similar to the CBM, this approach is not applicable to the same types of problems. If Jaynes' approach is applied to the estimation of posterior probabilities, then specification of posterior moments is necessary, and these moments are all that one is really after in applications to signal estimation (and some applications to signal detection). For example, the posterior mean $E\{S/Y\}$ of a random signal parameter is the MMSE estimate of that parameter, and to require its specification is to require the solution of the "signal estimation" problem. In contrast, the CBM requires specification of conditional moments of the observations (e.g., $E\{Y/S\}$)--quantities that can in many applications be specified (or estimated) without first solving the "signal estimation" problem.

The approach which appears to be most similar to the CBM is Brick and Zames' approach based on Wiener canonical expansions [9,10], described in more detail in Chapter VI. This approach is based on truncated series expansions of likelihood functions, and like the CBM, requires knowledge of conditional means of prescribed functionals of the observations, but unlike the CBM, does not require the solution of linear equations. Thus, while the CBM is probably superior for low-order constraints since it yields optimum approximations with specification of only a few low-order moments required, Brick and Zames' approach may be preferable for applications where many moments can be computed or estimated (e.g., for ergodic models and adaptive procedures). The truncated series is not optimum in any sense, however, as are the functionals obtained via the CBM, and there appears to be no way to gain insight into appropriate choices of bases for the expansions. Thus, Brick and Zames' approach is more ad hoc than the CBM.

Another approach that appears to be related to the two approaches described above is the series expansion of conditional densities of Schwartz [20]. However, it appears that only the single random variable observation problem is studied, and extension to multiple random variables might considerably complicate the procedure. Extension to waveform observations is not straightforward.

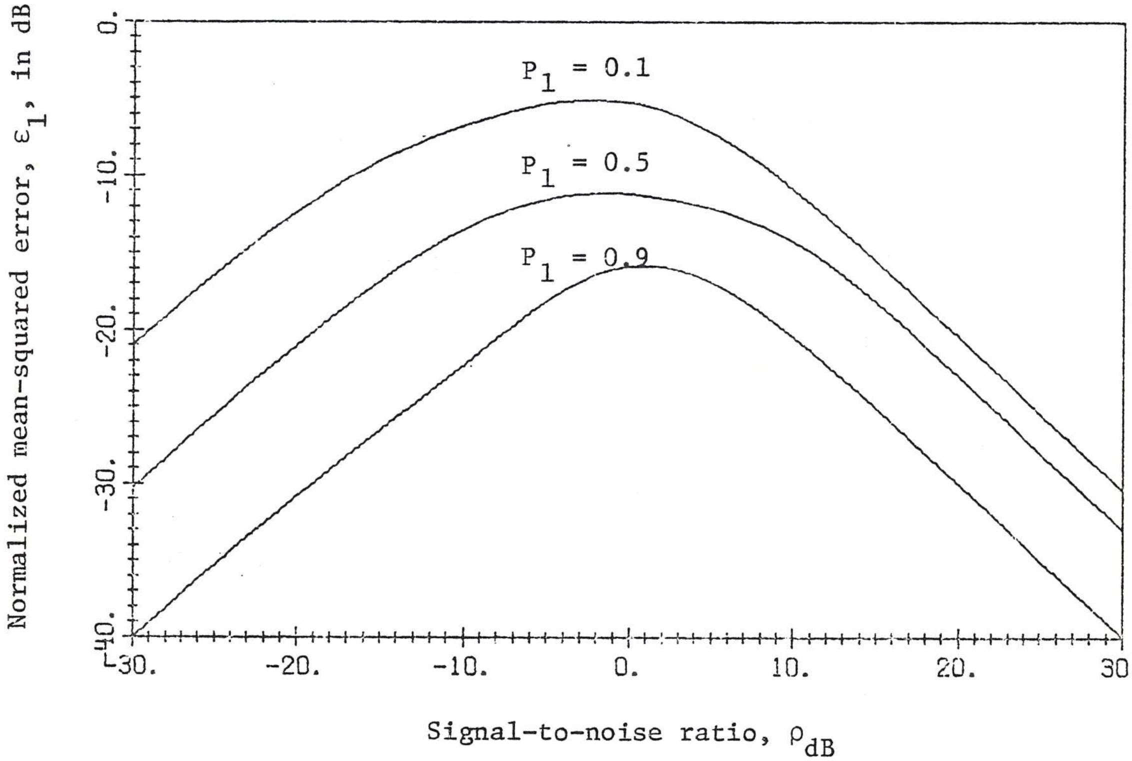


Figure 2.1. Normalized mean-squared error in estimating a posterior probability.

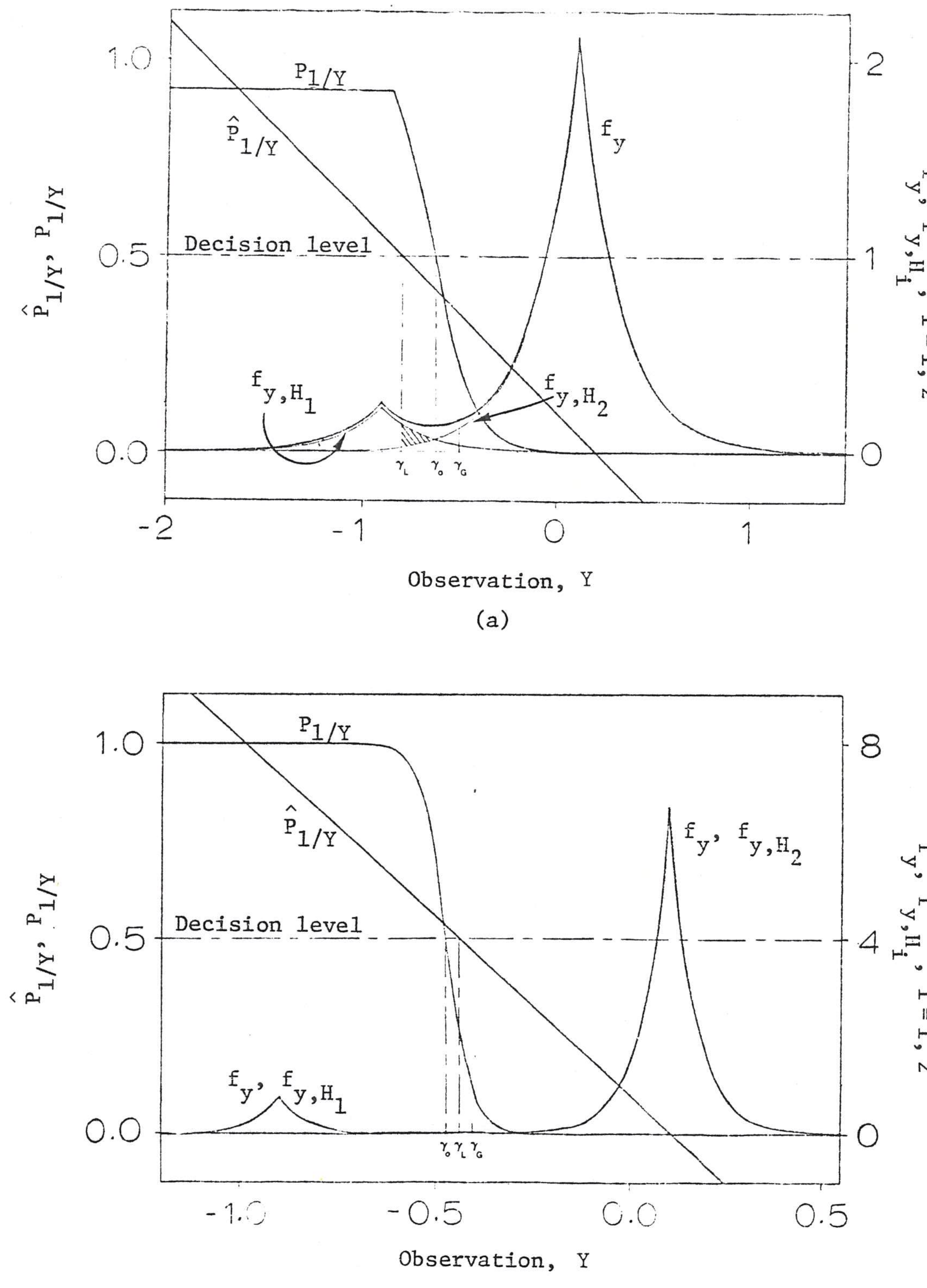
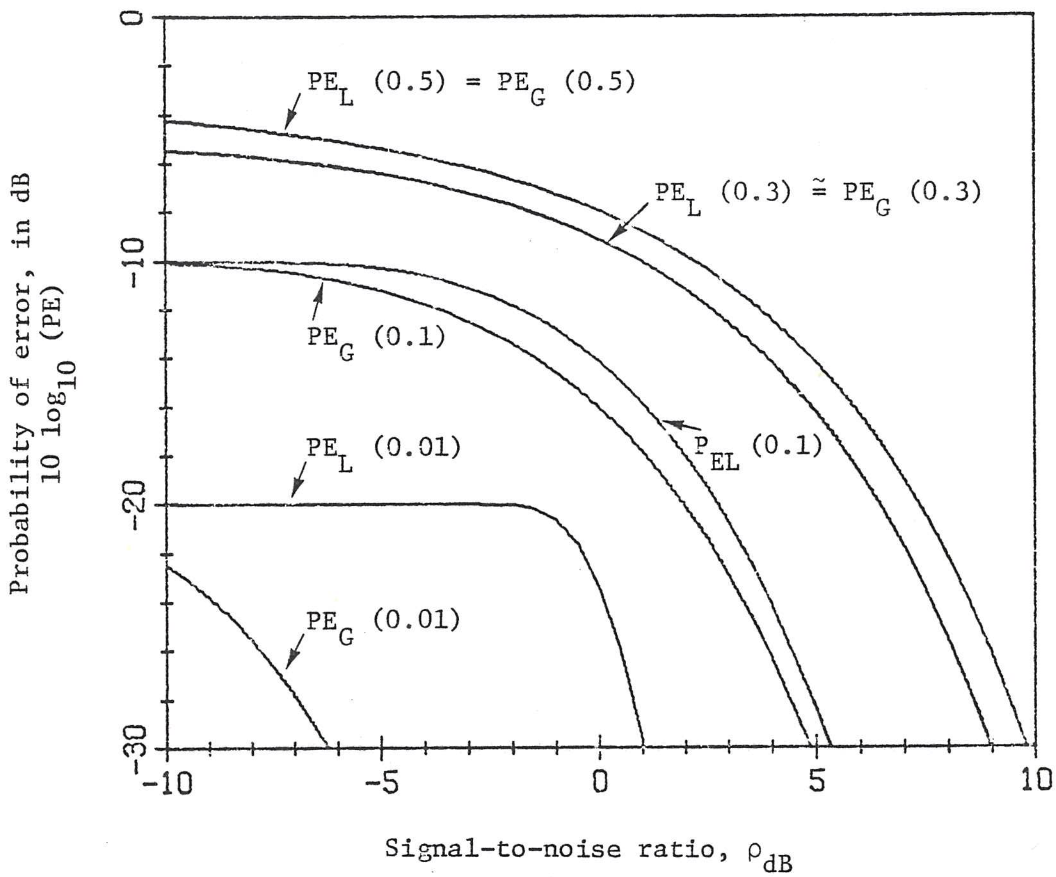


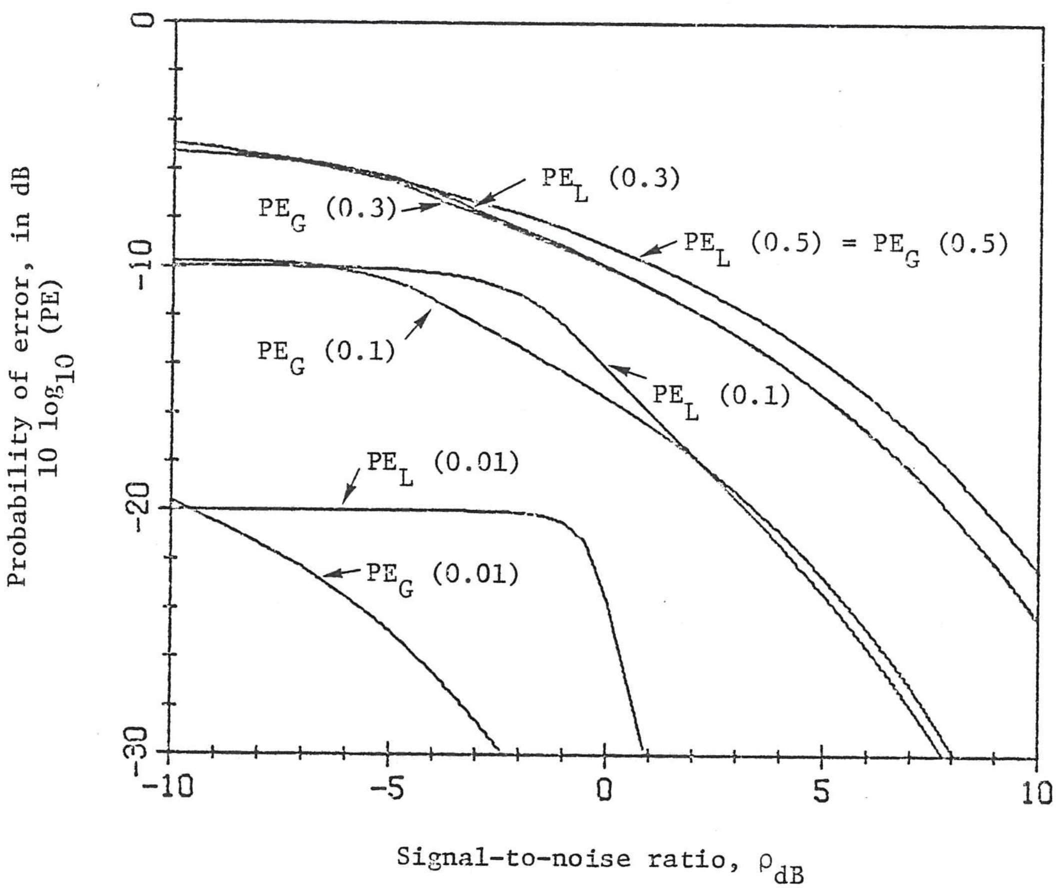
Figure 2.2. Estimated and true posterior probabilities with the densities of the observations for Laplacian noise with $P_1 = 0.1$, and $\rho_{dB} = 0$ (a), 10 (b).

Figure 2.3. Probability of error for the linearly constrained binary detector ($PE_L(P_1)$) and for the optimum-for-Gaussian noise binary detector ($PE_G(P_1)$) with (a) Gaussian (2.30), (b) Laplacian (2.12), and (c) FOB (2.31) noise distributions.

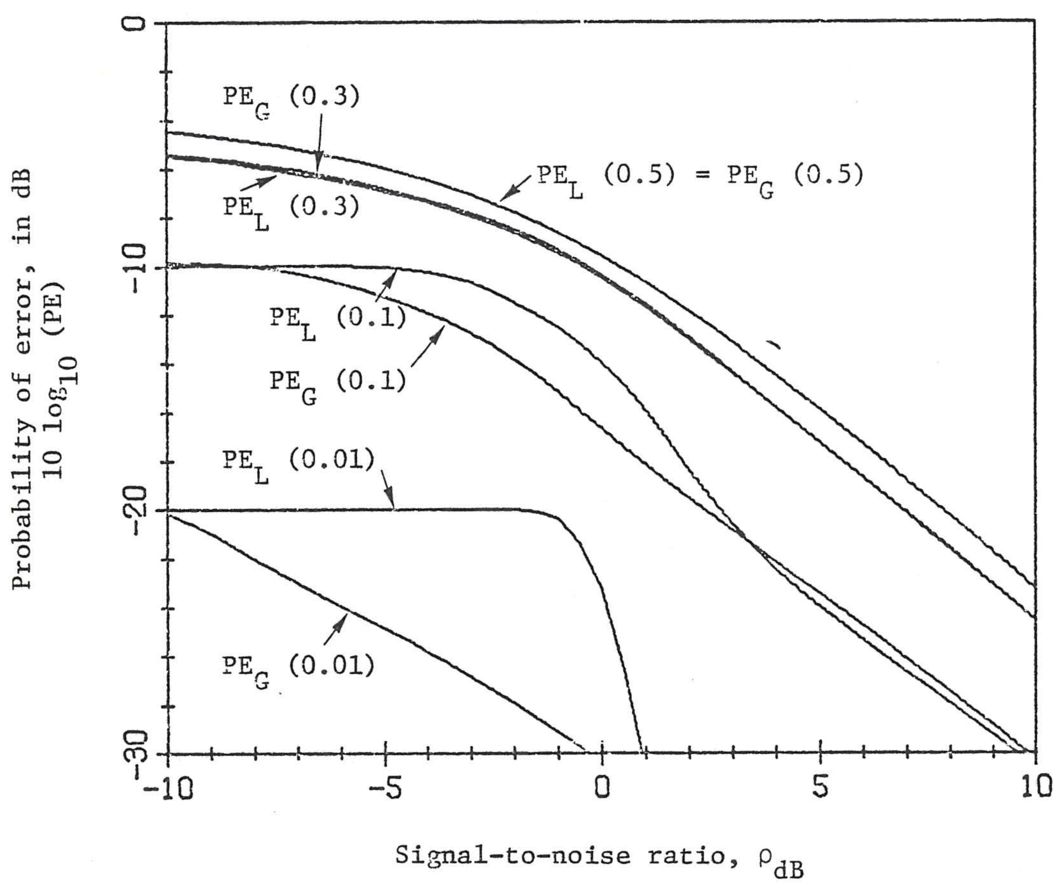


PE_L (0.5) = PE_G (0.5)
PE_L (0.3) ≅ PE_G (0.3)
PE_G (0.1)
P_EL (0.1)
PE_L (0.01)
PE_G (0.01)
0
-10
-20
-30
-10
-5
0
5
10
Probability of error, in dB
10 log10 (PE)
Signal-to-noise ratio, ρdB

(a)


0
-10
-20
-30
Probability of error, in dB
10 $\log_{10}$ (PE)
PE_L (0.3)
PE_G (0.3)
PE_L (0.5) = PE_G (0.5)
PE_G (0.1)
PE_L (0.1)
PE_L (0.01)
PE_G (0.01)
-10
-5
0
5
10
Signal-to-noise ratio, ρ_{dB}

(b)

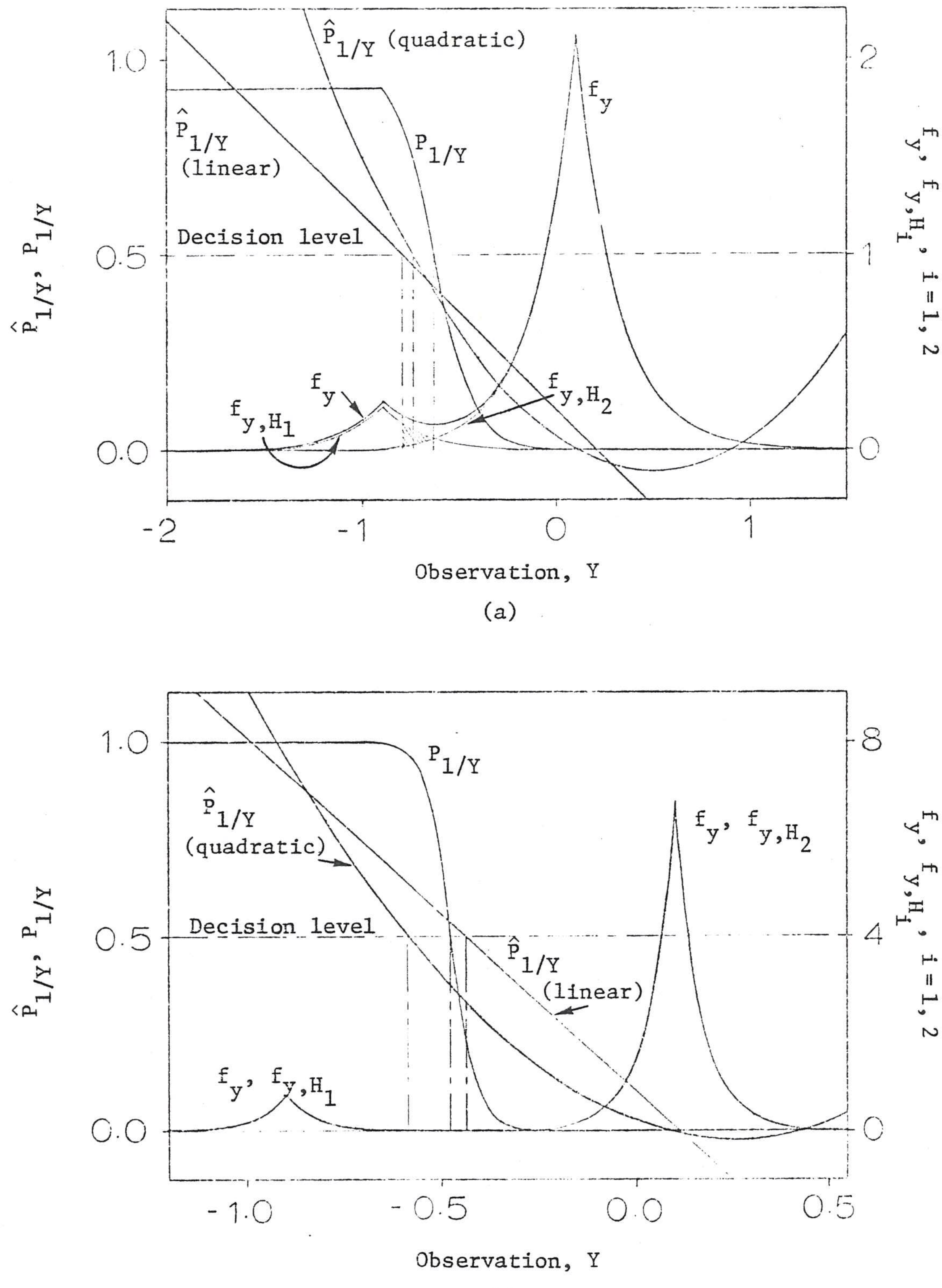


0
-10
-20
-30
-10
-5
0
5
10
PE_G (0.3)
PE_L (0.3)
PE_L (0.5) = PE_G (0.5)
PE_L (0.1)
PE_G (0.1)
PE_L (0.01)
PE_G (0.01)
Probability of error, in dB
$10 \log_{10}$ (PE)
Signal-to-noise ratio, ρ_{dB}

(c)

Figure 2.4. Linear and quadratic estimated posterior probabilities with the true posteriors and the densities of the observations for Laplacian noise, with $P_1 = 0.1$ and $\rho_{dB} = 0$ (a), 10 (b).

Figure 2.5. Same as Figure 2.4 for Gaussian noise.



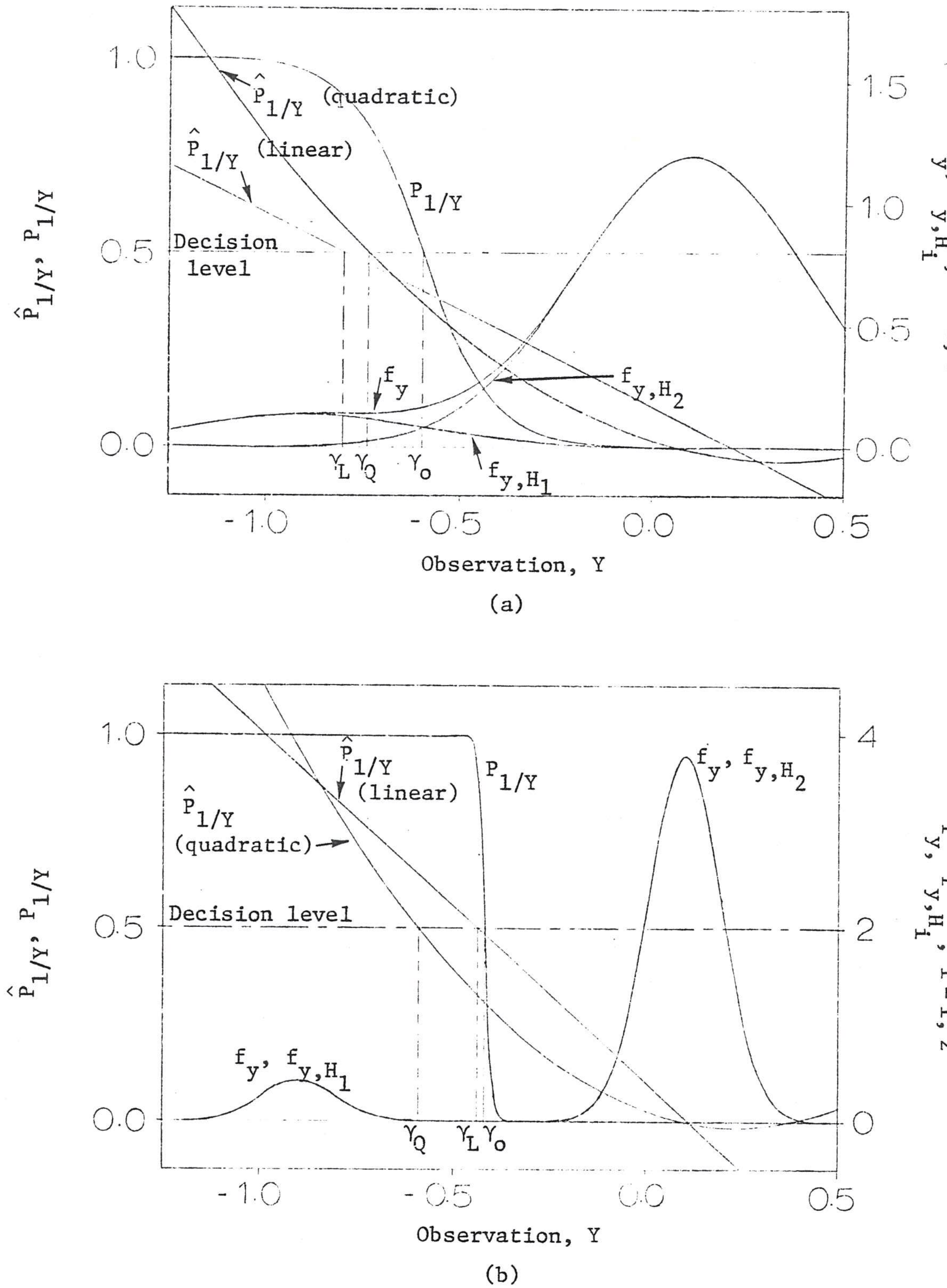
$\hat{P}_{1/Y}$ (quadratic)
$\hat{P}_{1/Y}$ (linear)
$P_{1/Y}$
f_y
Decision level
f_{y,H_1}
f_{y,H_2}
$\hat{P}_{1/Y}, P_{1/Y}$
$f_y, f_{y,H_i}, i=1,2$
1.0
0.5
0.0
2
1
0
-2
-1
0
1
Observation, Y

(a)

$P_{1/Y}$
$\hat{P}_{1/Y}$ (quadratic)
f_y, f_{y,H_2}
Decision level
$\hat{P}_{1/Y}$ (linear)
f_y, f_{y,H_1}
$\hat{P}_{1/Y}, P_{1/Y}$
$f_y, f_{y,H_i}, i=1,2$
1.0
0.5
0.0
8
4
0
-1.0
-0.5
0.0
0.5
Observation, Y

(b)

Figure 2.4.



1.0
0.5
0.0
1.5
1.0
0.5
0.0
$\hat{P}_{1/Y}$ (quadratic)
$\hat{P}_{1/Y}$ (linear)
$P_{1/Y}$
Decision level
f_y
f_{y,H_2}
f_{y,H_1}
γ_L γ_Q γ_o
-1.0
-0.5
0.0
0.5
Observation, Y
$\hat{P}_{1/Y}$, $P_{1/Y}$
f_y, f_{y,H_i}, i = 1, 2
(a)
1.0
0.5
0.0
4
2
0
$\hat{P}_{1/Y}$ (linear)
$P_{1/Y}$
$\hat{P}_{1/Y}$ (quadratic)
f_y, f_{y,H_2}
Decision level
f_y, f_{y,H_1}
γ_Q γ_L γ_o
-1.0
-0.5
0.0
0.5
Observation, Y
$\hat{P}_{1/Y}$, $P_{1/Y}$
f_y, f_{y,H_i}, i = 1, 2
(b)

Figure 2.5.

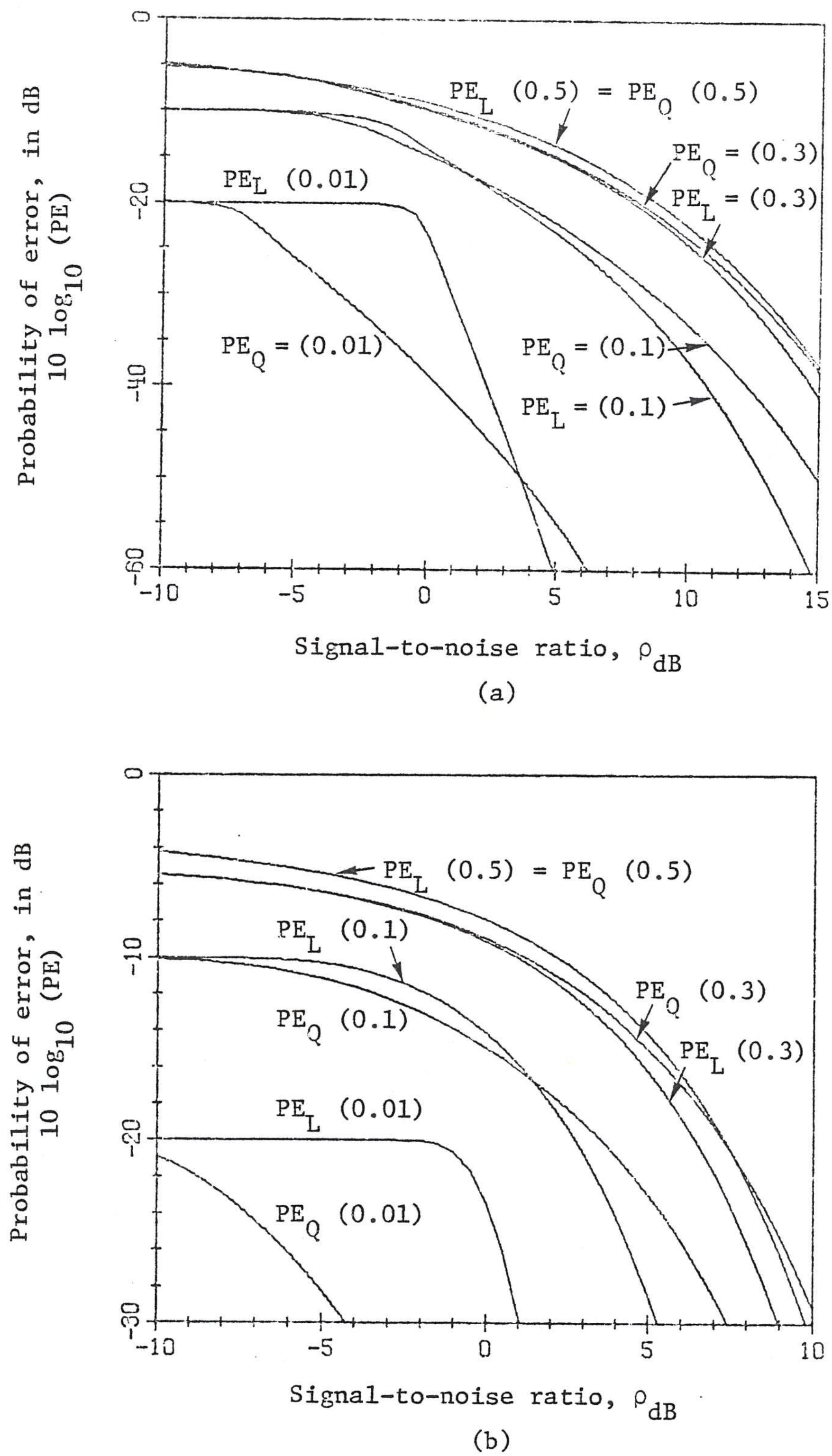


Figure 2.6. Probability of error for linearly ($PE_L(P_1)$) and quadratically ($PE_Q(P_1)$) constrained binary detectors with (a) Gaussian (2.30) and (b) Laplacian (2.12) noise distributions.

CHAPTER III

COMPARISON OF ESTIMATION-BASED DECISION RULES

1. Introduction

The decision rule that chooses the mode (maximum) of the estimated posterior distribution, i.e.,

$$\max_i \hat{P}_{i/Y} \quad , \qquad 3.1$$

referred to by Gardner [3,7] as the hypothesis tester's* (HT) rule, serves as the basis for the CBM. Assuming that the observations Y depend probabilistically on a discretely distributed random q-vector $\underline{x}$, with realization $\underline{X}_i$, given H_i, an alternative decision rule to the HT rule can be based on the L-MMSE estimate of $\underline{x}$, $(\hat{\underline{X}}/Y)$ instead of the L-MMSE estimate of $P_{i/y}$. Specifically, the decision rule that chooses the closest value in the range $\left\{\underline{X}_i\right\}_{i=1}^{M}$ to the estimate, i.e.,

$$\min_i \|\hat{\underline{X}}/Y - \underline{X}_i\| \quad , \qquad 3.2$$

where $\|\cdot\|$ is an appropriate norm for the parameter space, is such a rule. Following the terminology of Ziv and Zakai [37], Gardner [3,7] calls 3.2 the estimation theorist's (ET) rule.

Now, when constructing a decision rule, such as 3.2, based on a structurally constrained estimate of some quantity $\underline{x}$, the most obvious question to ask is "what quantity should be estimated?" The answer

---

*This term is derived from the original minimum PE rule which comes from the Bayes decision theoretic approach to hypothesis testing.

to this question is essential to the comparison of the CBM to other L-MMSE-estimation-based methodologies. The answer is partially given by property 7 in Chapter II, subsection 4.2, namely that the L-MMSE estimate of any quantity, referred to as a parameter (e.g., of a signal), can be obtained from the L-MMSE estimates of posterior probabilities:

$$\hat{\underline{X}}/Y = \sum_{i=1}^{M} \underline{X}_i \, \hat{P}_{i/Y} \,, \qquad 3.3$$

i.e., that the L-MMSE estimate of $\underline{x}$ is the mean of the estimated posterior distribution. Using this fact, the posterior estimates can be used to compute the L-MMSE estimate of a more preferable parameter. This reveals that for the particular L-constraint being used, the posterior estimates form a set of "sufficient statistics," i.e., they contain all the relevant information there is for L-MMSE estimation. The relation 3.3 is analogous to, and in fact, reduces to the well-known result for no constraint that the MMSE estimate of $\underline{x}$, given Y, is the conditional mean of $\underline{x}$, given Y, i.e.,

$$\hat{\underline{X}}_{MMSE} = \sum_{i=1}^{M} \underline{X}_i \, P_{i/Y} \,. \qquad 3.4$$

The comparison of the HT and ET rules is facilitated by the re-expression of 3.2. Expanding the norm into an inner product and changing the minimization into the maximization of the negative of the inner product yields, after a little manipulation,

$$\mathrm{Max}(\bar{\hat{\underline{X}}},\bar{\underline{X}}_i) - \frac{1}{2} X_{ii} \,, \qquad 3.5$$

which is equivalent to

$$\operatorname*{Max}_{i} \sum_{j=1}^{M} X_{ii} \hat{P}_{j/y} - \frac{1}{2} X_{ii} , \qquad 3.6$$

where

$$X_{ij} \triangleq \overline{\underline{X}}_i^T \, \overline{\underline{X}}_j , \qquad 3.7$$

and, e.g., $\overline{\underline{X}}_i$ is the centered version of $\underline{X}_i$ (2.4). Note that this is identical in form to the optimum rule for detecting M signals in additive WGN, except for the addition of the term $\rho^{-1}\ln(P_i)$ [11, pg. 259], where ρ is a suitably defined SNR.

Garner [3,7] shows that the two rules 3.1 and 3.2 are equivalent for binary (M=2) signals and that the HT rule without constraints is optimum, so cannot be worse than the unconstrained ET rule. He also shows that in general for L-constraints:

1) neither rule is always superior or equivalent to the other, and

2) the rules are equivalent when the quantities $\underline{X}_i\,{}_{i=1}^{M}$ are mutually orthogonal and isonormal (OI), i.e.,

$$(\underline{X}_i, \underline{X}_j) = \delta_{ij} \, E , \quad i,j = 1,\ldots,M . \qquad 3.8$$

In light of the fact that the posterior estimates are MS equivalent to L-MMSE estimates of the random indicator functions (see Subsections II.4.1 and II.2.2), this can be interpreted as the equivalence among OI bases, i.e., decisions can be made equally well based on any set of OI vector quantities.

Another aspect to the equivalence of the posterior estimates to estimates of indicators is that the HT rule can be regarded as an ET rule. Thus, an alternative perspective of the comparison of the HT and ET rules is the comparison among various ET rules. The purpose of this chapter is to extend and generalize the above results by establishing some conditions under which one rule is equivalent or superior to the other. In terms of the above mentioned alternate perspective, this amounts to investigating what are appropriate parameters to be estimated.

2. Equivalence of the HT and ET Rules

2.1 Arbitrary Binary Signals

For binary signals, substituting 3.3 into 3.6 and using the sum-to-one property of posterior estimates (2.48) yields after a little manipulation,

$$\max_{\substack{i,j\\ i\neq j}} (\hat{P}_{i/Y} - \tfrac{1}{2})\, X_{ii} + (1 - \hat{P}_{i/Y})\, X_{ij} \; . \tag{3.9}$$

Since $i\neq j$ in the above expression, the term X_{ij} is constant, so 3.9 is equivalent to

$$\max_{\substack{i,j\\ i\neq j}} (\hat{P}_{i/Y} - \tfrac{1}{2})(X_{ii} - X_{ij}) \quad . \tag{3.10}$$

Now, since $X_{ii} - X_{ij}$ equals $P_j \; \|\underline{\overline{X}}_i - \underline{\overline{X}}_j\| > 0$, and since $\hat{P}_{i/Y} > \frac{1}{2}$ if and only if $\hat{P}_{i/Y} > \hat{P}_{j/Y}$, then 3.10 is equivalent to 3.1 and the HT and ET rules are equivalent for binary signals.

2.2 Regular Simplex M-ary Signals

The equivalence of the HT and ET rules can be extended from OI signals (3.8) [3,7] to a related* class of signals, called regular simplex (RS) signals [38-45] for which

$$X_{ij} = E\,(M\,\delta_{ij} - 1)/(M - 1)\ . \tag{3.11}$$

Substituting either 3.8 or 3.11 into 3.6 yields 3.1 after very little manipulation, and the two rules are thereby shown equivalent for RS signals. As mentioned in the introductory section, an interesting example of this equivalence is for $\underline{x} = \underline{\delta}$, where $\underline{\delta}$ is the random indicator vector

$$(\underline{\delta})_i = \delta_{ij}\ , \quad \text{given } H_j\ , \qquad i=1,\ldots,M\ . \tag{3.12}$$

For this case, 3.11 (and 3.8) is satisfied, and the parameter estimate is identical to the posterior estimate (see Subsection II.4.1, Equivalence 7). Note that the results of this section are valid for any L-constraints, for any type of observations (e.g., not only sure signals or additive noise), and for arbitrary prior probability distributions.

2.3 Linear Estimates

To obtain further results, linear constraints are focused on in this subsection. In Chapter IV, the linear estimate is shown to be

---

*An OI set with M elements lies in a M-1 dimensional space in which the centered version of the set is a RS. Thus, if the set $\{\underline{X}_i\}$ is OI, the set $\{\overline{\underline{X}}_i\}$ is a RS. The extension, however, allows for a set that is an arbitrary translation of an OI set.

$$\hat{P}_{i/Y} = P_i[1 + \overline{\tau}_i] \quad , \tag{4.5}$$

where

$$\overline{\tau}_i \triangleq \iint_{T\,T} Q(t,\tau)\, M_{\overline{y}/H_i}(t)\, \overline{Y}(\tau)\, dt\, d\tau \quad , \tag{4.6}$$

and Q is the solution to the integral equation

$$\int_T M^{(2)}_{\overline{y}}(t,\tau)\, Q(\tau,\sigma)\, d\tau = \delta(t - \sigma) \quad . \tag{4.3}$$

Let W be a factor of Q, i.e.,

$$Q(t,\sigma) = \int_T W(t,\tau)\, W(\tau,\sigma)\, d\tau \quad , \tag{3.13}$$

then the whitened version of $\overline{y}$ can be defined as

$$\overline{z}(t) \triangleq \int_T W(t,\sigma)\, \overline{y}(\sigma)\, d\sigma \quad , \tag{3.14}$$

and $\overline{z}$ is called white since, from 3.14,

$$E\{\overline{z}(t)\, \overline{z}(\tau)\} = \delta(t - \tau) \quad . \tag{3.15}$$

Now consider the linear vector space S spanned by the M-1 dimensional set of functions $\left\{M_{\overline{z}/H_i}(t)\right\}_{i=1}^{M}$. Denote an orthonormal basis for S by $\{S_i(t)\}_{i=1}^{N}$, where $N < M$. Then denote the representation for $M_{\overline{z}/H_i}(t)$, relative to this basis, by $\overline{\underline{\mu}}_i$, i.e.,

$$M_{\overline{z}/H_i}(t) = \overline{\underline{\mu}}_i^T\, \underline{s}(t) \quad , \tag{3.16}$$

$$\overline{\underline{\mu}}_i \triangleq \int_T \underline{s}(t)\, M_{\overline{z}/H_i}(t)\, dt \quad , \tag{3.17}$$

$$\int_T S_i(t)\, S_j(t)\, dt = \delta_{ij} \quad . \qquad 3.18$$

Using this representation, 3.5 can be written as

$$\hat{P}_{i/Y} = P_i[1 + \underline{\overline{\mu}}_i^T\, \underline{\overline{Z}}] \;, \qquad 3.19$$

where $\underline{\overline{Z}}$ is the representation of the projection of $\overline{Z}(t)$ onto S:

$$\underline{\overline{Z}} \triangleq \int_T \overline{Z}(t)\, \underline{x}(t)\, dt \quad . \qquad 3.20$$

Substituting 3.19 into 3.3 gives

$$\hat{\underline{X}} = \sum_{i=1}^{M} \underline{X}_i\, P_i[1 + \underline{\overline{\mu}}_i^T\, \underline{\overline{Z}}] \quad . \qquad 3.21$$

Subtracting M_x from both sides of 3.21 yields the simpler expression

$$\hat{\underline{\overline{X}}} = \sum_{i=1}^{M} \underline{\overline{X}}_i\, P_i\, \underline{\overline{\mu}}_i^T\, \underline{\overline{Z}} \quad . \qquad 3.22$$

Now defining

$$K_{\overline{\mu x}} = \sum_{i=1}^{M} P_i\, \underline{\overline{\mu}}_i\, \underline{\overline{X}}_i^T \quad , \qquad 3.23$$

3.22 and 3.6 are, after some manipulation, simply

$$\hat{\underline{\overline{X}}} = K_{\overline{\mu x}}^T\, \underline{\overline{Z}} \quad , \qquad 3.24$$

$$\underset{i}{\text{Max}}\; \underline{\overline{X}}_i^T\, K_{\overline{\mu x}}^T\, \underline{\overline{Z}} - \frac{1}{2} X_{ii} \quad . \qquad 3.25$$

Substituting 3.19 into 3.1, the HT rule for linear estimates is

$$\max_i P_i \, [1 + \overline{\underline{\mu}}_i^T \, \overline{\underline{Z}}] \quad . \tag{3.26}$$

Now, if the HT rule is not degenerate (see Section 3), i.e., if there exists a vector $\underline{Z}_i$ such that $\hat{P}_{i/Y}$ is the largest estimate, for each $i = 1,\ldots,M$, then general necessary-and-sufficient conditions can be derived under which the rules 3.25 (ET) and 3.26 (HT) are equivalent. If the HT rule is degenerate, then sufficient conditions for equivalence can be derived although the results are of little practical importance. Assuming the HT rule is not degenerate, the rules are equivalent if and only if both 3.25 and 3.26 partition the observation space into the same decision regions. This means that one rule must be at most a scaled shifted version of the other, where the scale and shift must be independent of i, i.e.,

$$P_i[1 + \overline{\underline{\mu}}_i^T \, \overline{\underline{Z}}] = a \, [\overline{\underline{X}}_i^T \, K_{\overline{\mu}\overline{x}}^T \, \overline{\underline{Z}} - \frac{1}{2} X_{ii}] + b + \underline{c}^T \overline{\underline{Z}} \quad , \tag{3.27}$$

where $a > 0$ and b are arbitrary scalars and $\underline{c}$ is an arbitrary N-vector. Since 3.27 must hold for each $\overline{\underline{Z}}$, the following two conditions must hold for equivalence:

$$P_i = b - \frac{1}{2} a \, X_{ii} \quad , \tag{3.28a}$$

$$P_i \, \overline{\underline{\mu}}_i = a \, K_{\overline{\mu}\overline{x}} \, \overline{\underline{X}}_i + \underline{c} \quad , \qquad i = 1,\ldots,M \quad . \tag{3.28b}$$

These two conditions are met for equal priors by any $\underline{x}$ such that

$$X_{ij} = d \, \delta_{ij} + e \quad , \tag{3.29}$$

where $d > 0$ and e are constants. This condition is satisfied by any RS (3.11) of M-1 dimensions ($e = -d/M$) which coincides with the general result of Section 2.2. In addition, if the priors are equal, then 3.28 is satisfied by any parameter set of the form

$$\underline{X}_i = a\, A\, \underline{\mu}_i \quad , \quad i=1,\ldots,M \quad , \qquad 3.30$$

where a is a constant, and where A is a NxN matrix satisfying

$$A^T A = K_{\overline{\mu}\overline{\mu}}^{-1} \qquad 3.31$$

where I is the NxN identity matrix and $K_{\overline{\mu}\overline{\mu}}$ is defined in 3.23 with $\underline{x}$ replaced by $\underline{\mu}$. This is verified by substituting 3.30 and 3.31 into 3.28 and noting that for equal priors, $\underline{c} = 0$ in 3.28b. Note that a solution to 3.31 for A always exists since $K_{\overline{\mu}\overline{\mu}}$ is easily shown to be positive definite. Equation 3.28a is interesting in that, for the equivalence of the two rules, the energies X_{ii} can be equal if and only if the priors are equal, and if either is unequal, a low energy signal must be compensated by a high prior probability, and conversely. Perhaps more interesting, however, is 3.28b, which says that the dimension $R_{\overline{x}}$ of the space spanned by $\{\overline{\underline{X}}_i\}$, referred to as the rank of the matrix whose i^{th} row vector is $\overline{\underline{\mu}}_i$, for the two rules to be equivalent. Since $R_{\overline{x}}$ must be less than M, it can be concluded that for equivalence,

$$R_{\overline{\mu}} \leq R_{\overline{x}} \leq M-1 \quad . \qquad 3.32$$

Parameters satisfying 3.28 can be constructed for which $R_{\overline{x}}$ ranges from $R_{\overline{\mu}}$ to M-1 for the case $R_{\overline{\mu}} = 1$, $M = 3$, which shows that such parameters exist for unequal priors at least for some cases, although this has not been proved yet in general. Also not proved, except for equal priors, is the conjecture that if a parameter set

$\{\underline{X}_i\}$ with rank $R_{\overline{x}}$ greater than $R_{\overline{\mu}}$ is found satisfying 3.28, then there must exist an equivalent parameter $\underline{x}'$ of rank $R_{\overline{\mu}}$, with the relationship

$$\underline{X}_i' = A\,\underline{X}_i + \underline{Y}_i \tag{3.33}$$

where

$$\underline{X}_i^T\,\underline{Y}_j = 0 \quad , \qquad \forall i,j = 1,\ldots,M \ . \tag{3.34}$$

Whether or not there exist parameters satisfying 7.27, a natural question is "Does a parameter $\underline{x}$ exist for which the ET_x rule is better than the HT rule (in terms of PE performance)?" This question is approached in Section 4. The importance of finding parameters $\underline{x}$ which satisfy 3.28 for the case of unequal priors is reduced by the questionable appropriateness of the ET rule for unequal priors. Perhaps a more suitable rule would take account of prior imbalances other than by weighting the more likely points more heavily in the estimate. For example, the OG rule (see Section 1) includes the term $\rho^{-1}\ln(P_i)$. This would then change condition 3.28.

As seen above, when the priors are not equal, the comparison of the HT and ET rules is not easy. The following example is offered as an indication of the kind of behavior the two rules can have for unequal priors. Shown in Figure 3.1 are plots of PE for the two rules in detecting a ternary random variable, x, in additive noise, n, where the parameter for the ET rule is the "signal," x. The observation, y, is

$$Y = X_i + N \ , \qquad i = 1,2,3 \ , \tag{3.35}$$

where $X_i = i-2$ and $P_1 = P_3$. The three plots are for Gaussian (2.30), Laplacian (2.12), and FOB (2.31) noise, respectively, each for various priors, P_2. It can be seen that the ET rule is superior to the HT rule for high SNR values in all cases. Typically, the HT rule becomes superior for SNR values below 1 to 7dB for $P_2 > \frac{1}{2}$, and for SNR values below -4 to -7dB for $P_2 < \frac{1}{2}$, whereas the rules are equivalent for $P_2 = \frac{1}{2}$. The exception to this is that for Gaussian noise, with $P_2 < \frac{1}{2}$, the ET rule is superior to the HT rule for all values of SNR, although for the other noise types, the superiority of the HT rule is only at levels of PE around $1 - \mathrm{Min}(P_1, P_2)$, which is attainable simply by choosing H_i if P_i is largest.

3. Degeneracies in the HT Rule

For "signals" with rank $R_{\overline{\mu}} < M-1$, it is possible for a degeneracy to occur in the HT test. The problem arises whenever some of the signals cannot be detected. An example which is easily visualized is shown in Figure 3.2a for $M=5$, $R_{\overline{\mu}} = 2$, and equal priors. For this configuration, H_5 is never selected, H_i always being selected whenever $\underline{Z}$ is in the i^{th} quadrant.

A related problem can occur whenever $M > 2$, namely that the neighborhood of a signal conditional mean, $\overline{\underline{\mu}}_i$, is not in the appropriate decision region, $\mathscr{R}_i$. This degeneracy is illustrated by the signal set in Figure 3.2b. The transformation W (3.14) attempts to correct this situation, but when the signal set differs too much from a hyperellipsoidal shape, a degeneracy can remain. The larger M is for a given $R_{\overline{\mu}}$, the more the signal set must conform to a hyperellipsoidal shape if this degeneracy is to be avoided, assuming the priors maintain sufficient "continuity" as M increases. In fact, it can be shown that for fixed $R_{\bar{\mu}}$, as $M \to \infty$, any nondegenerate signal set must approach a hypersphere if the priors are equal. For unequal priors and fixed $R_{\bar{\mu}}$, the class of limiting ($M \to \infty$) nondegenerate signal sets contains only convex (hyperellipsoidal) but not hyperspheroidal sets.

With the use of a proper parameter in an ET test, however, these HT rule degeneracies can be avoided. Alternatively, a more accurate estimate of the posterior can avoid this problem, as illustrated in the following example. Consider an amplitude-shift keyed (ASK) signal in additive WGN of spectral height N_o:

$$Y(t) = A_i\, S(t) + N(t) \ , \qquad P_i = M^{-1} \ , \qquad i = 1,\ldots,M \ , \tag{3.36}$$

$$A_i = i - \tfrac{1}{2}(M + 1) \ . \tag{3.37}$$

Then the linear estimate of the posterior is

$$P_{i/Y} = M^{-1}[1 + A_i \tau] \ , \tag{3.38}$$

where

$$\tau \triangleq [N_o + (M^2 - 1)\, E/12]^{-1} \int_T S(t)\, Y(t)\, dt \ , \tag{3.39}$$

$$E \triangleq \int_T S^2(t)\, dt \ . \tag{3.40}$$

If $\tau > 0$, then $\hat{P}_{M/Y}$ will always be the largest estimate; if $\tau < 0$, then $\hat{P}_{1/Y}$ will always be the largest, and if $\tau = 0$, then all the estimates will be equal. Hence, all other signals will never be detected. Figure 3.3 shows the minimum order constraint, N_{min}, needed for the detection of all ASK signals in the absence of noise, i.e., for high SNR. N_{min} was found by assuming $n = N_o = 0$ and considering all possible constraints of order i, for $i = 1,2,\ldots,$ until a constraint was found that did not possess a degeneracy. It turned out that for each constraint order i, the structure possessing terms of all orders up to i showed nondegenerate behavior for higher values of M than any other i^{th}-order constraint. A structure was determined degenerate by computing the elements $\{P_{jk} \triangleq \hat{P}_{j/Y}/H_k\}_{j,k=1}^{M}$ and then testing to make sure that

$$P_{kk} > P_{jk} , \quad \forall j \neq k \quad . \tag{3.41}$$

Also shown in Figure 3.3 is $N_{min}-1$ to indicate the behavior of N_{min} as the smallest integer greater than a quantity N^*_{min} which may have an analytic dependence on M. This functional dependence seems to be a power relation close to

$$N^*_{min} \cong c \ (M^{1/2} - 1) \ , \tag{3.42}$$

with $c \in [1.7179, 1.7266]$, although for $M \leq 9$, especially $M=9$, the actual values do not follow 3.42 exactly, as do the other points, being slightly lower than N^*_{min}.

The nature of the degeracies of the HT rule are related to both the structural constraint on the posterior estimates as well as the structure of the signal set. To isolate the effects of these two factors, it may be enlightening to consider the unconstrained situation. To begin, the unconstrained HT rule is optimum, being the minimum PE rule. As illustrated below, the ET rule (both constrained and unconstrained) can behave poorly at low SNR values for degenerate signal sets (i.e., those for which the constrained HT rule is degenerate), compared with nondegenerate signal sets. Consider the two signal sets shown in Figure 3.4 for equal priors. At low SNR, the estimate of $\underline{s}$ will be close to the mean of $\underline{s}$, $\underline{M}_s$. In the first set, the four signal points lie on a line and the set is degenerate. Since the signal mean is halfway between $\underline{\mu}_2$ and $\underline{\mu}_3$, the estimate will be closest to either $\underline{\mu}_2$ or $\underline{\mu}_3$, making the detection of $\underline{\mu}_1$ or $\underline{\mu}_4$ unlikely. Looking at the second set, $\underline{M}_s$ is centered between all four points, equally spaced from each. Thus, the slightest perturbation from the mean will favor one of the points in favor of the others, and each of the four points is equally detectable. As illustrated by the previous ASK example, a similar phenomenon occurs with the linear HT rule at all values of SNR for the same reason. Again, this is avoided for the second signal set and is also avoided for the first set at high SNR values by using a higher order constraint. Thus, it can be concluded that:

1) the lower the order of the constraint, the more important is the signal structure to the HT rule at all SNR values, and
2) the signal structure is very important with regard to the degeneracy of both the constrained and unconstrained ET rules at low SNR.

4. Strengths and Weaknesses of the ET Rule

This section serves to show that:

1) the ET rule is potentially superior to the HT rule, and

2) the ET rule can fail when the parameter to be estimated is inappropriately chosen.

The former is accomplished by showing that unequal "signal" energies of $M_{\overline{z}/H_i}$ do not affect the ET_μ rule as they affect the HT rule. For reliable detection using a linearly constrained ET rule, the signal means $\overline{\underline{\mu}}_i$ should fall in their respective decision regions. If very little noise is present and a "signal" mean does not fall in its decision region, the rule will not be asymptotically error free, as it should. Since the appropriateness of the ET rule is questionable for unequal priors, the case of equal priors will be focused on in this section. Whether or not the noise is additive, represent $\overline{\underline{Z}}$ for the linear receiver (see 3.14) as

$$\overline{\underline{Z}} = \overline{\underline{\mu}}_k + \underline{N}_k \ , \qquad \text{given } H_k \tag{3.43}$$

(note that $E\{\underline{N}_k/H_k\} = 0$). Then the HT rule is, from 3.26,

$$\underset{i}{\text{Max}}\ V_{ik} + N_{ik} \ , \qquad \text{given } H_k \ , \tag{3.44}$$

and the ET_x rule is, from 3.25,

$$\underset{i}{\text{Max}}\ M^{-1} \sum_{j=1}^{M} X_{ij} V_{jk} - \frac{1}{2} X_{ii} + U_{ik} \ , \qquad \text{given } H_k \ , \tag{3.45}$$

where

$$V_{ik} \triangleq \overline{\underline{\mu}}_i^T \overline{\underline{\mu}}_k = \int_T M_{\overline{z}/H_i}(t)\, M_{\overline{z}/H_k}(t)\, dt \ , \tag{3.46}$$

$$N_{ik} \triangleq \overline{\underline{\mu}}_i^T \underline{N}_k = \int_T M_{\overline{z}/H_i}(t)\, \overline{Z}(t)\, dt - V_{ik} \ , \tag{3.47}$$

$$U_{ik} \triangleq M^{-1} \sum_{j=1}^{M} X_{ij} N_{jk} \quad . \tag{3.48}$$

Comparing 3.44 with 3.45, if $\underline{x}$ has an inappropriate structure (e.g., $R_{\overline{x}}$ is too low), the HT rule can easily outperform the ET_x rule. As an example of this, consider a phase-shift keyed (PSK) signal in additive WGN:

$$Y(t) = A \cos(\omega_o t + \phi_i) + N(t) \ , \quad \text{given } H_i \ , \tag{3.49}$$

where the amplitude, A, is constant, the phases $\{\phi_i\}_{i=1}^{M}$ are equally likely and uniformly spaced over the interval $[-\pi,\pi]$, and N is a sample from a white noise process, n, of spectral height N_o. From 4.3-4.6, the posterior estimates are

$$\hat{P}_{i/Y} = M^{-1}\{1 + C[Y_c \cos \phi_i + Y_s \sin \phi_i]\} \ , \tag{3.50}$$

$$C \triangleq A/(N_o + A^2T/4) \ , \tag{3.51}$$

$$Y_c \triangleq \int_T Y(t) \cos \omega_o t \, dt \ , \tag{3.52}$$

$$Y_s \triangleq -\int_T Y(t) \sin \omega_o t \, dt \ . \tag{3.53}$$

From 3.50, it is easily seen that the HT rule is equivalent to picking the closest phase to $\tan^{-1}(Y_s/Y_c)$. However, the ET_ϕ rule, which estimates the signal phase, picks the closest phase to the estimate

$$\hat{\phi} = M^{-1} \, C \, Y_s \sum_{i=1}^{M} \phi_i \sin \phi_i \tag{3.54}$$

$$\triangleq C \, Y_s \, S_M \ . \tag{3.55}$$

Now, S_M is a constant ranging from $\pi/2$ for $M=2$, to 1 as M increases ($S_6 = \pi/3 = 1.047$), and

$$Y_s = \frac{1}{2} A\ T \sin \phi_j - N_s \ , \quad \text{given } H_j \ , \tag{3.56}$$

$$N_s = \int_T N(t) \sin \omega_o t \ dt \ . \tag{3.57}$$

Assuming there is no noise ($N \equiv 0$, $N_o = 0$), C becomes $4/AT$, Y_s/H_j becomes $\frac{1}{2}AT \sin \phi_j$, and

$$\hat{\phi} = 2\ S_M \sin \phi_j \ , \quad \text{given } H_j \tag{3.58}$$

$$\cong 2 \sin \phi_j \ . \tag{3.59}$$

Now, for any value of $M > 3$, this estimate will have an associated PE of 1 or $1 - 2M^{-1}$, so this shows that the signal phase is not an appropriate parameter to estimate. Rather, the in-phase and quadrature components of the signal, i.e., $A \cos \phi_i$ and $A \sin \phi_i$, which form a basis for the conditional means of the observations (see (3.16), would be more appropriate. It seems that an appropriate parameter needs to have the rank of $\underline{\bar{\mu}}$, although contrived counter examples can be found. One natural example of such a full-rank parameter is $\underline{x} = \underline{\mu}$. (Note that, as shown in Section 2.3, there exists a transformation A for which the $ET_{A\mu}$ rule is equivalent to the HT rule, although this may not perform as well as the ET_μ rule.) For large SNR, using 3.46 and 3.15,

$$M^{-1} \sum_{i=1}^{M} \underline{\bar{\mu}}_i \ \underline{\bar{\mu}}_i^T = E\{\underline{Z}\ \underline{Z}^T\} = I \ , \tag{3.60}$$

and replacing $\underline{x}$ with $\underline{\mu}$ in 3.45, the ET_μ rule becomes

$$\max_i V'_{ik} + N_{ik} \quad , \quad \text{given } H_k \quad , \tag{3.61}$$

where

$$V'_{ik} \triangleq V_{ik} - \frac{1}{2} V_{ii} \quad . \tag{3.62}$$

Since both 3.44 and 3.62 involve the same noise terms, the two rules can be compared using the measures

$$D_{HT} \triangleq \min_{i \neq k} V_{kk} - V_{ik} \quad , \tag{3.63}$$

$$D_{ET} \triangleq \min_{i \neq k} V'_{kk} - V'_{ik} \quad . \tag{3.64}$$

The reasoning behind the choice of these measures is that for a given H_k, the correct decision will be made only if the value of V_{kk} (using the HT rule 3.44 for example) is larger than the value of any other V_{ik}. For high SNR, the amount of PE will depend only on the term $V_{ik} + N_{ik}$ for which V_{ik} is closest to V_{kk}, given H_k. Thus, as long as both D_{HT} and D_{ET} are positive, the larger quantity will indicate the asymptotically superior rule. Substituting 3.62 into 3.64 yields

$$D_{ET} = \min_{i \neq k} \frac{1}{2} (V_{kk} + V_{ii} - 2V_{ik}) \quad . \tag{3.65}$$

Using the facts that

$$\min_x\{f(x)\} + \min_x\{g(x)\} \leq \min_x\{f(x) + g(x)\} \tag{3.66}$$

(where equality holds if and only if the minima of f and g occur for the same value of x) and that the indices in the definition of D_{HT} (3.63) can be interchanged with no effect on the value of D_{HT}, it follows that

$$\min_{i \neq k}(V_{kk} - V_{ik}) + \min_{k \neq i}(V_{ii} - V_{ki}) \leq \min_{i \neq k}(V_{kk} + V_{ii} - 2\, V_{ik}) \ , \qquad 3.67$$

from which it follows that

$$D_{HT} \leq D_{ET} \ . \qquad 3.68$$

If the energies V_{ii} are equal, 3.68 will be satisfied with equality, and the two rules are equivalent. However, if the energies are not all equal, the two rules cannot be equivalent because of condition 3.28a, yet 3.68 can still be satisfied with equality. This happens when there are many index pairs (i,k) that simultaneously minimize the expressions in 3.63 and 3.64, but when there are more such pairs for D_{HT} than for D_{ET}, in which case the ET rule is superior. It is easily shown that the reverse situation cannot occur, so whenever 3.68 is satisfied with equality for unequal energies, the ET rule is superior. This seems to be an uncommon, almost contrived situation, as illustrated by the signal set in Figure 3.5. Setting aside this anomaly, the behavior of the two rules is nicely illustrated by the following example, shown in Figure 3.6. The four points on the coordinate axes are fixed, and the points on the diagonals vary in location as a ranges from 0 to 1. Plotted in Figure 3.7 are the values of $D_{HT}(a)$ and $D_{ET}(a)$ over this range:

$$D_{HT}(a) = \mathrm{Min}(1 - a,\ 2a^2 - a) \ , \qquad 3.69$$

$$D_{ET}(a) = \mathrm{Min}(\frac{1}{2} - a + a^2,\ 2a^2) \tag{3.70}$$

For $a \leq \frac{1}{2}$, the HT rule is degenerate, so the ET rule is superior. Except for the value of $a = \frac{1}{2}$, for which the signal energies are equal (making the two rules equivalent), the ET rule is everywhere superior to the HT rule.

Although there appears to be no way in general to specify what quantities are the best to estimate, those parameters with rank equal to that of the "signal" set are superior to others considered here. Thus, the conditional means of the "signal" seem to provide appropriate parameters of full rank (dimension) in all cases.

5. Summary

The purpose of this chapter is to compare the performance of the hypothesis tester's (HT) and estimation theorist's (ET) rules. In summary, the results of the comparison are listed and briefly discussed. Results originally obtained by Gardner [3,7] are marked with a (G).

1) (G) Neither rule is always equivalent or superior to the other. The main function of this chapter, then, is to establish some conditions under which one is equivalent or superior to the other.
2) (G) The rules are equivalent for binary hypotheses. This result is one of the most important due to its generality, although its implications are limited since any two binary tests using the same statistics are already very similar.
3) (G) The rules are equivalent when the parameters being used by the ET rule are mutually orthogonal and isonormal (OI).

This result is useful for applications in which the parameters being estimated are to be chosen. However, for most ordinary estimation problems, the parameters present in the signals, such as phases, amplitudes, or frequencies, are not vector quantities, and as such, are not orthogonal. Aside from these pragmatics, the result is useful as a benchmark in partitioning the class of detection problems with respect to the classes of parameters, into recognizable pieces so the general question of superiority can at least be partially answered.

4) The rules are equivalent for regular simplex (RS) parameters. An RS set is just a translation of an OI parameter set in the next higher dimension, so this extension of result 3 is of lesser significance.

5) (G) The estimates of the random indicator vector, which is OI, are just the posterior probability estimates, so the HT and ET are trivially equivalent for this parameter.

6) Conditions for the equivalence of linearly constrained ET and HT rules are derived from which certain observations can be made:

   a) for equivalence of the two rules, the priors and energies must be inversely related;

   b) the rank of the parameter to be estimated must be at least as great as the rank of the signal conditional means for equivalence of the two rules. However, although only proved for equal priors, it is conjectured that there is no advantage in having a parameter rank greater than that of the conditional means. This would imply that the parameters need only lie in a space of the same dimension as that of the conditional means. For this reason, in addition to specific examples which support this, the conditional means are believed to provide an appropriate, possibly optimum for some cases, parameter to estimate. If the parameter is inappropriate, the ET rule can fail completely. An example of this is the estimate of the phase for a PSK signal. The ET rule fails for $M > 3$ whereas the HT rule performs appropriately. If the in-phase and quadrature components of the signal were estimated, i.e., the conditional means (or their representations with respect to a basis) of the observations, rather than the phase, the ET rule would then be equivalent to the HT rule.

c) The ET_{μ} rule (using the conditional means as parameters) is shown to be superior to the HT rule for a specific example (ternary, $M = 3$) with unequal priors, and in general for equal priors and unequal energies at high SNR. For equal priors and equal energies, the two rules are equivalent at high SNR. This is not too surprising since there is always a parameter that can perform as well as the HT rule (see result 5 above).

7) The HT rule can be degenerate under certain conditions which are satisfied when the conditional means of the observations

are devoid of sufficient hyperellipsoidal regularity. Examples of this occur for:

a) Amplitude-shift keying (ASK):

$$S_i(t) = A_i\, S(t) \ , \tag{3.71}$$

b) Quadrature amplitude modulation (QAM):

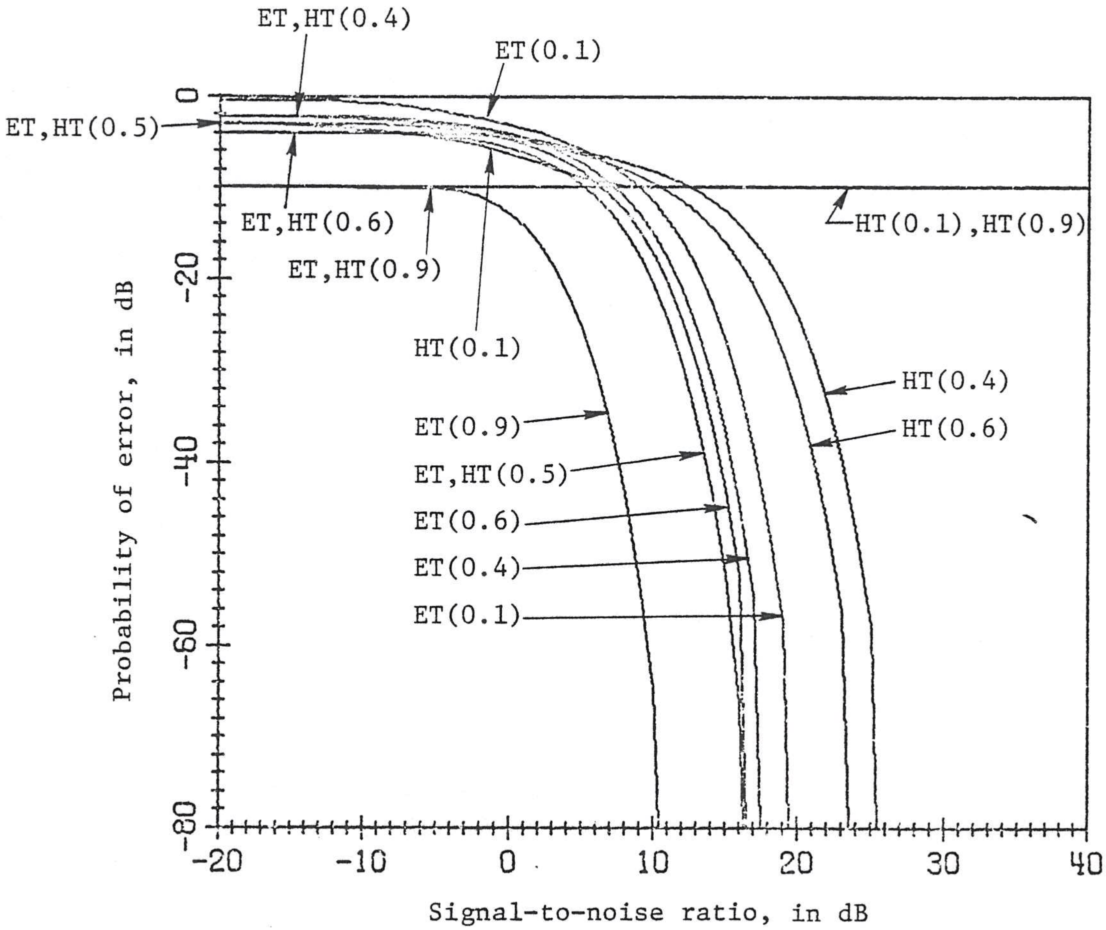
$$S_{ij}(t) = A_i S(t) \cos \omega_o t + B_j\, S(t) \sin \omega_o t \ , \tag{3.72}$$

c) Amplitude/Phase-shift keying (APK):

$$S_{ij}(t) = A_i S(t) \cos(\omega_o t + \phi_j) \quad . \tag{3.73}$$

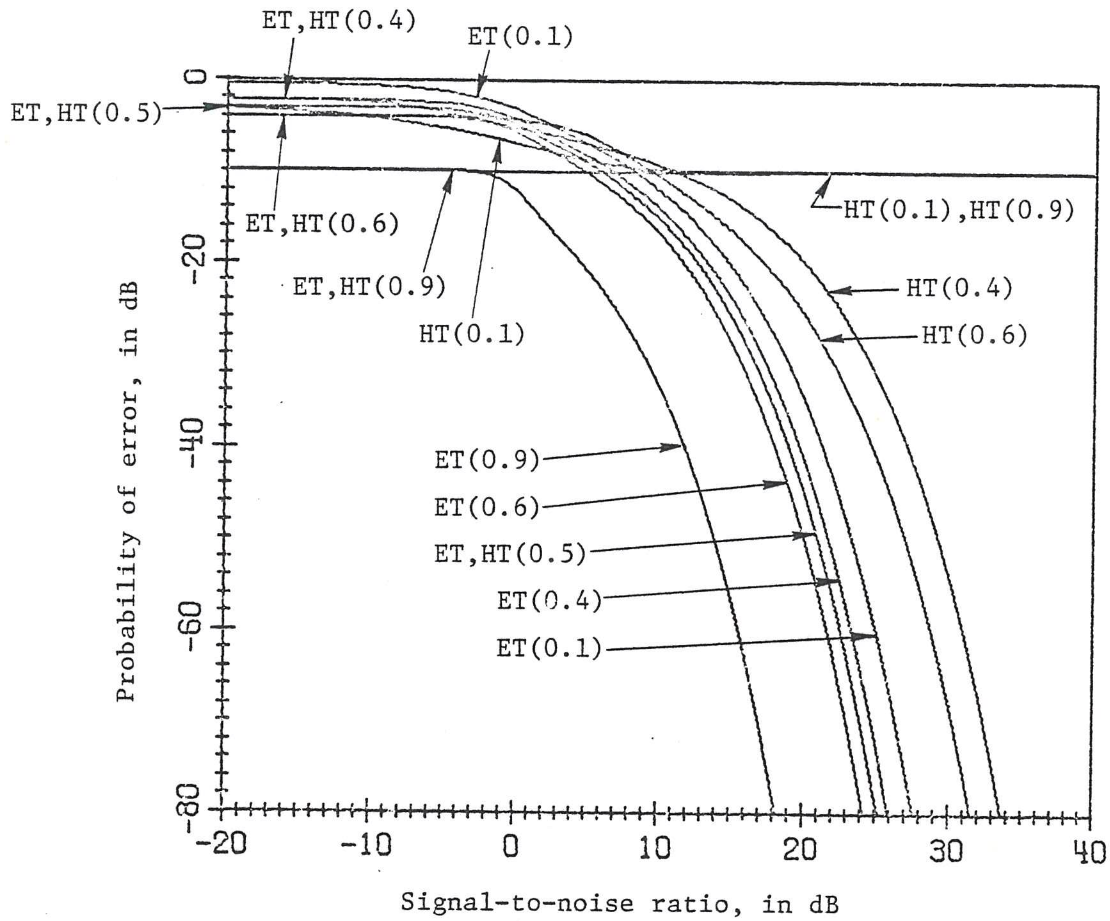
The degeneracy of linearly constrained rules can be avoided by choosing a constraint of a higher order. The minimum order constraint that has nondegenerate (HT) behavior is computed for M-ary ASK signals.

Figure 3.1. Probability of error for the hypothesis tester's ($HT(P_2)$) and estimation theorist's ($ET(P_2)$) rules for detecting a ternary random variable in (a) Gaussian (2.30), (b) Laplacian (2.12), and (c) FOB (2.31) noises.

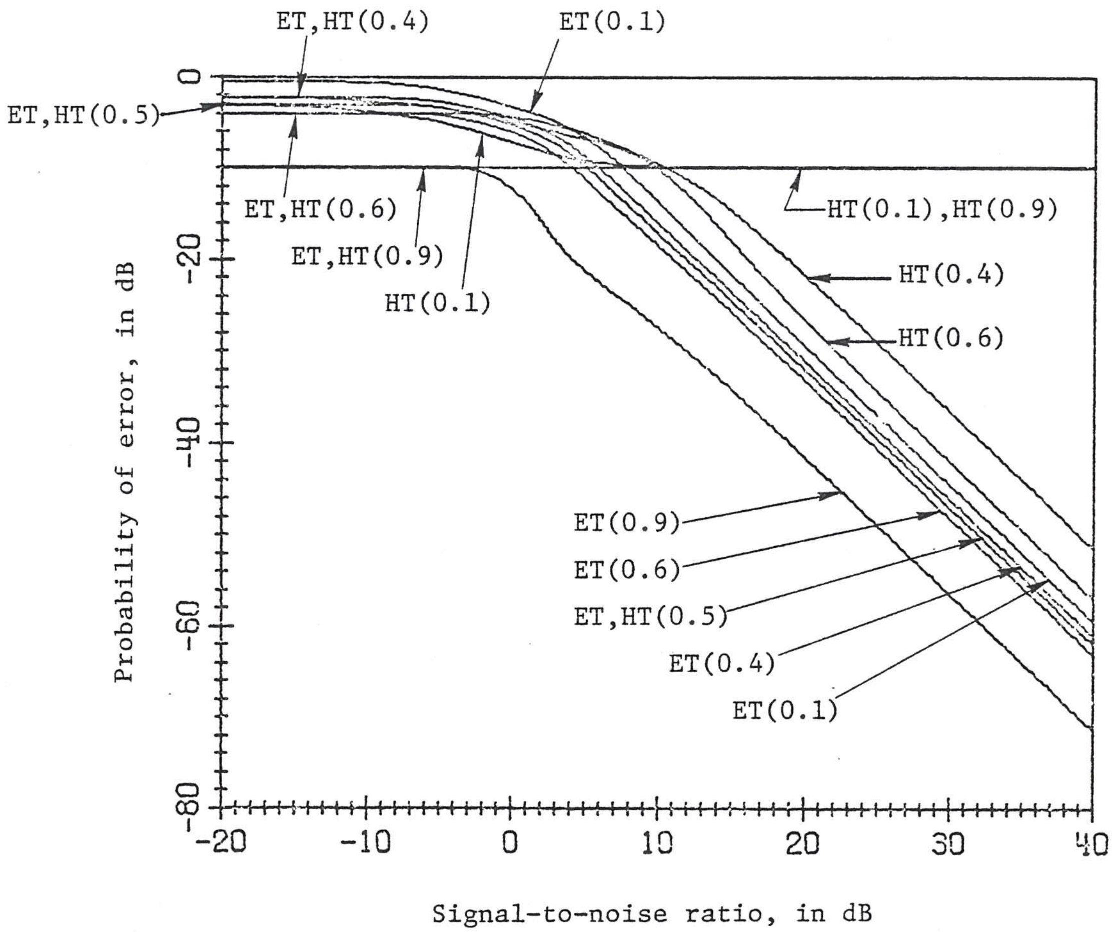


ET,HT(0.4)
ET(0.1)
ET,HT(0.5)
ET,HT(0.6)
ET,HT(0.9)
HT(0.1),HT(0.9)
HT(0.1)
HT(0.4)
HT(0.6)
ET(0.9)
ET,HT(0.5)
ET(0.6)
ET(0.4)
ET(0.1)
Probability of error, in dB
0
-20
-40
-60
-80
-20
-10
0
10
20
30
40
Signal-to-noise ratio, in dB

(a)


ET,HT(0.4)
ET(0.1)
ET,HT(0.5)
ET,HT(0.6)
ET,HT(0.9)
HT(0.1)
HT(0.1),HT(0.9)
HT(0.4)
HT(0.6)
ET(0.9)
ET(0.6)
ET,HT(0.5)
ET(0.4)
ET(0.1)
0
-20
-40
-60
-80
-20
-10
0
10
20
30
40
Probability of error, in dB
Signal-to-noise ratio, in dB

(b)


ET,HT(0.4)
ET(0.1)
ET,HT(0.5)
ET,HT(0.6)
ET,HT(0.9)
HT(0.1)
HT(0.1),HT(0.9)
HT(0.4)
HT(0.6)
ET(0.9)
ET(0.6)
ET,HT(0.5)
ET(0.4)
ET(0.1)
Probability of error, in dB
0
-20
-40
-60
-80
-20
-10
0
10
20
30
40
Signal-to-noise ratio, in dB

(c)

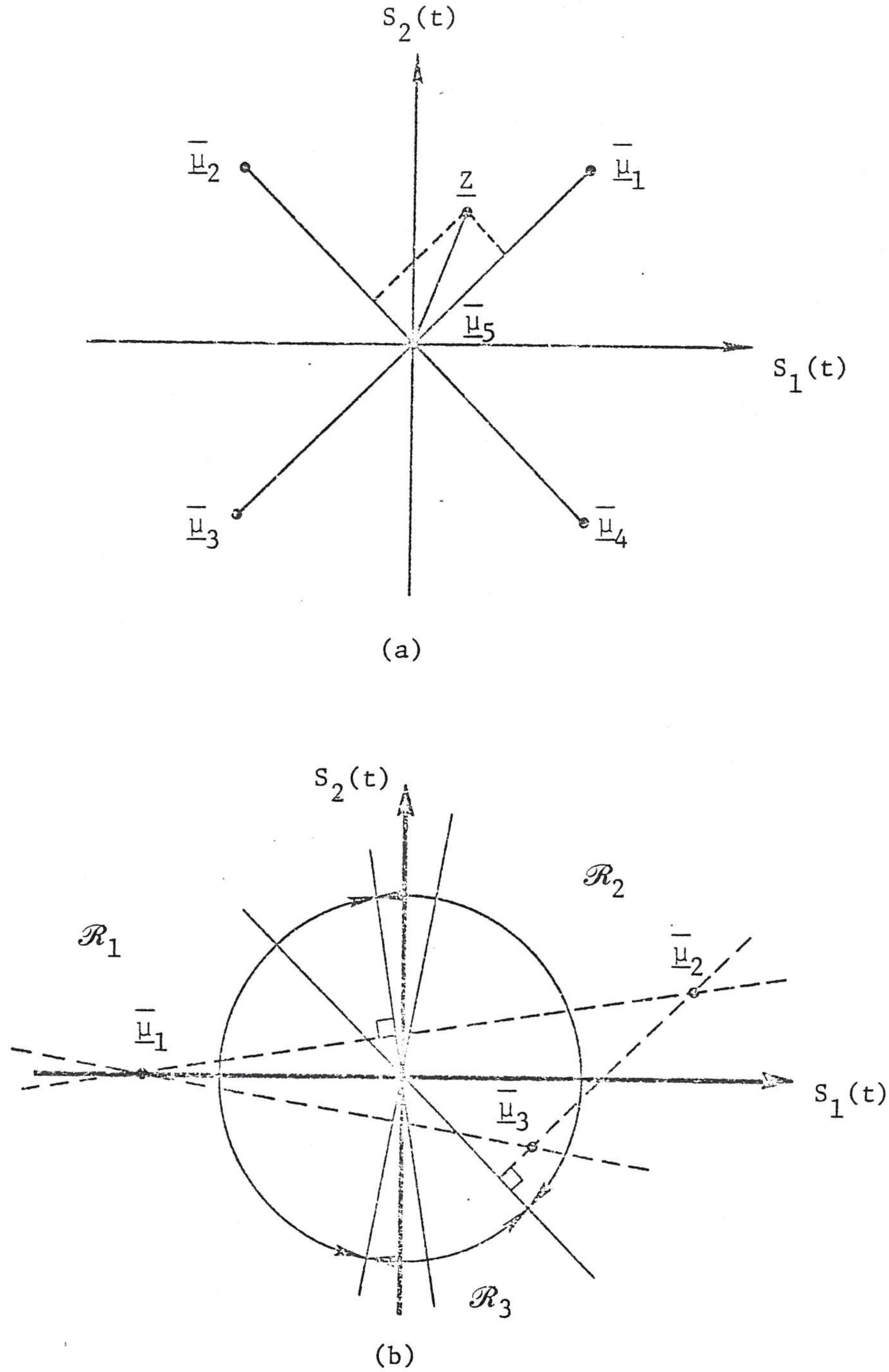


Figure 3.2. Examples of HT degeneracies
a) $\overline{\mu}_5$ cannot be detected,
b) $\overline{\mu}_3$ is not in its decision region, $\mathscr{R}_3$.

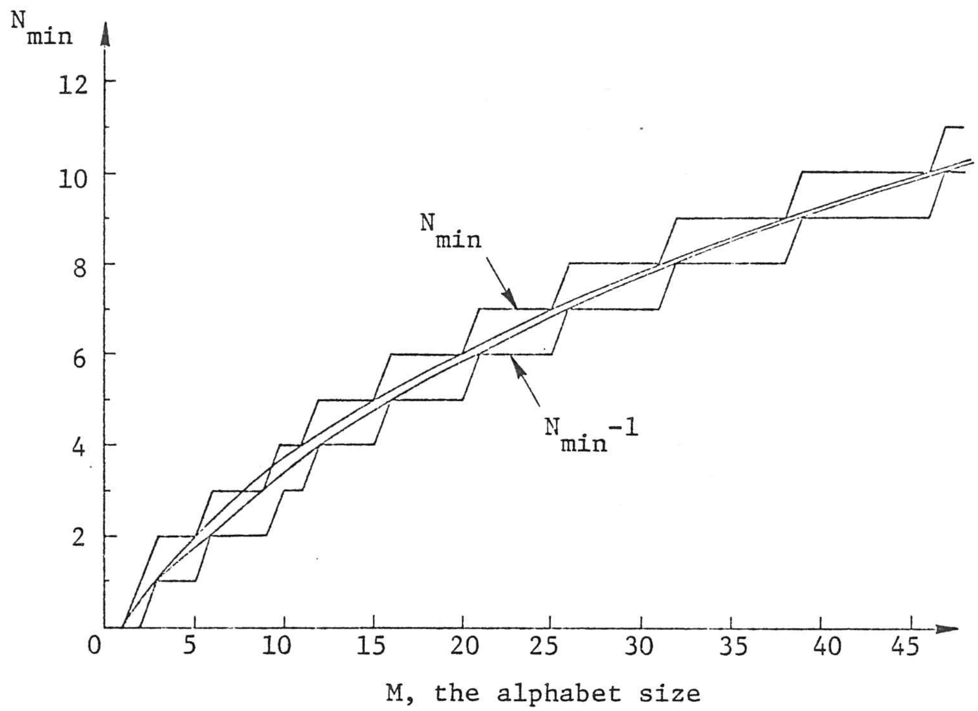


Figure 3.3. N_{min}, the minimum order of the constraint to detect M ASK signals.

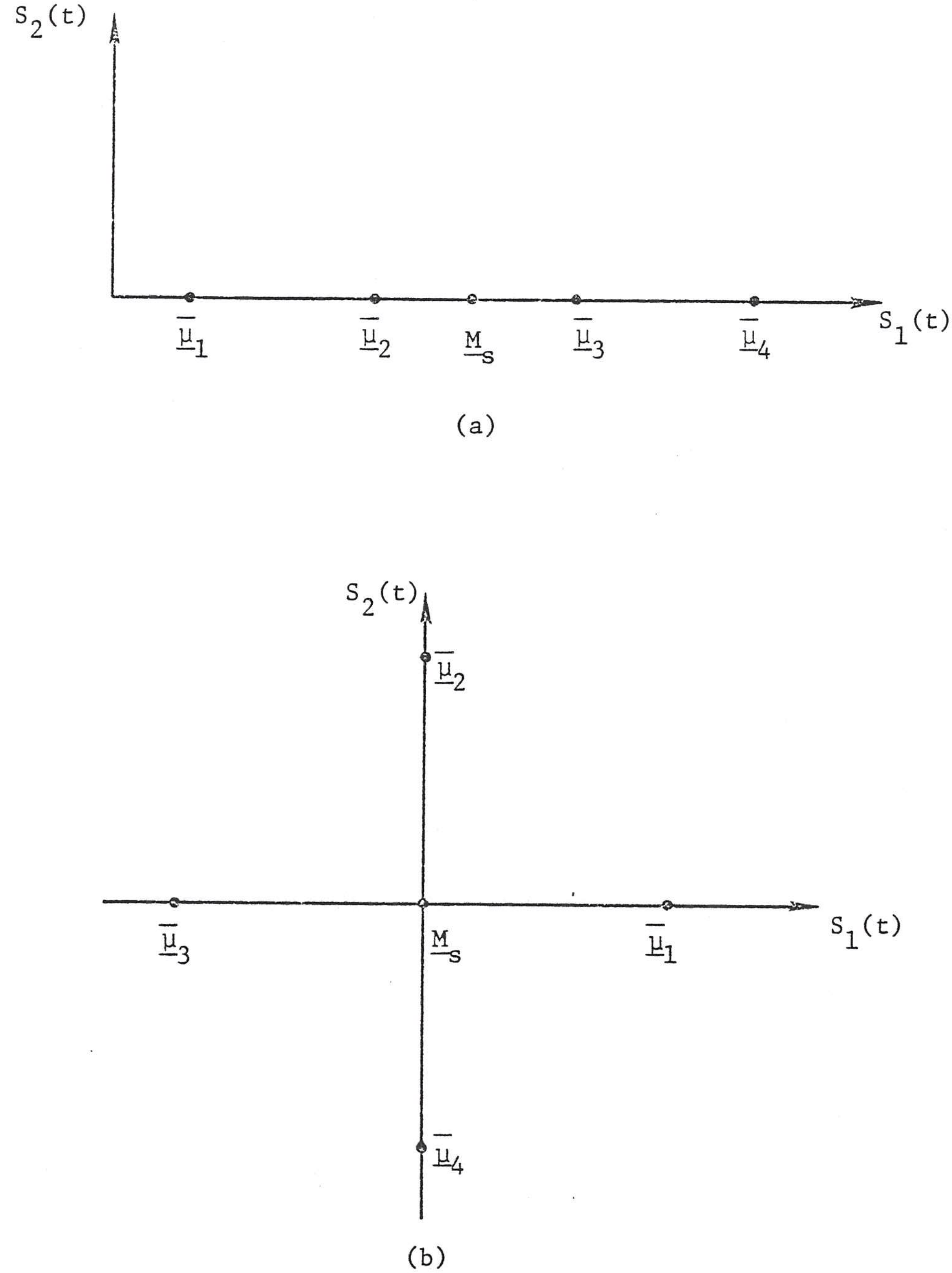


Figure 3.4. Examples of a) degenerate and b) nondegenerate signal sets.

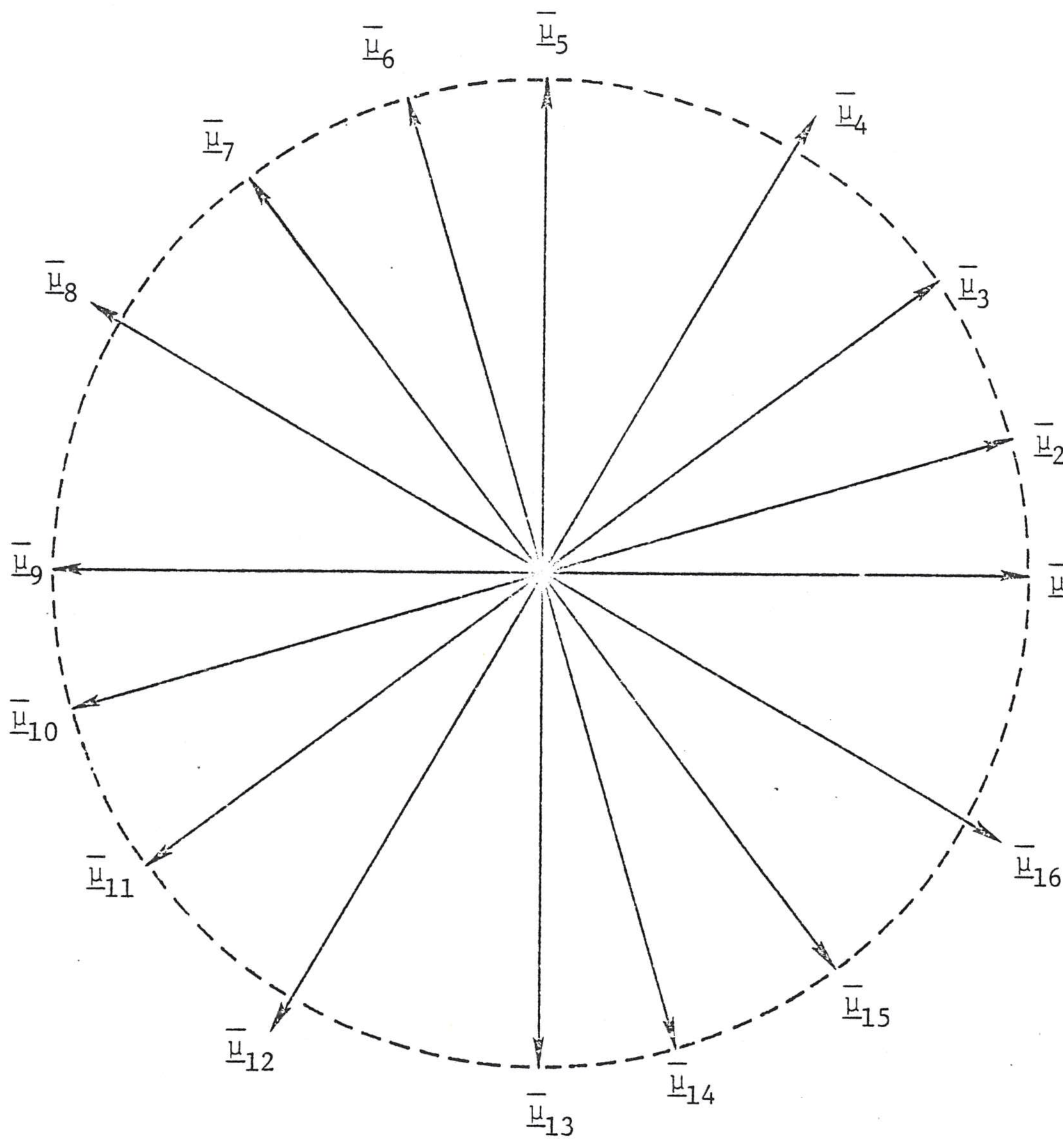


Figure 3.5. An example of a signal set with $D_{ET} = D_{HT}$, but for which the ET rule is superior. (Note: this set must be appropriately scaled to satisfy 3.60)

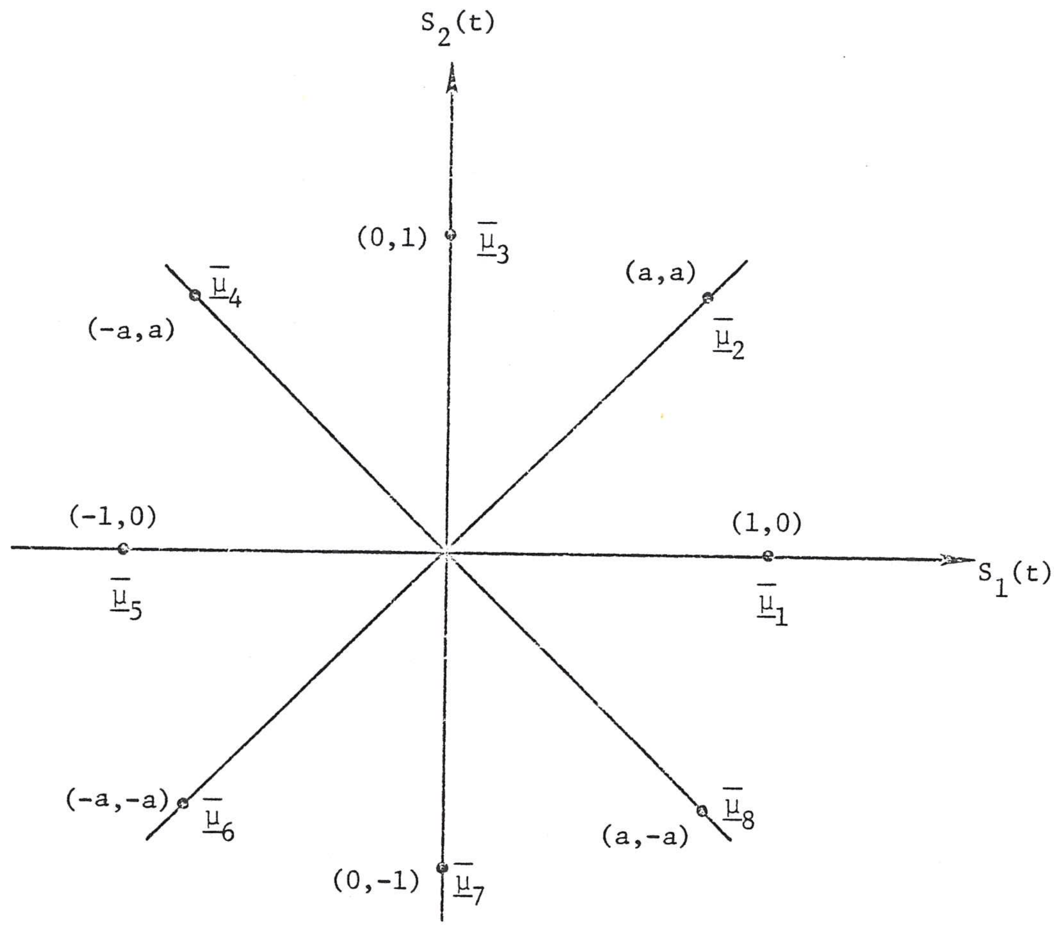


Figure 3.6. A signal set illustrating (with Figure 3.7) the use of D_{HT} and D_{ET} to demonstrate relative performance of the HT and ET rules.

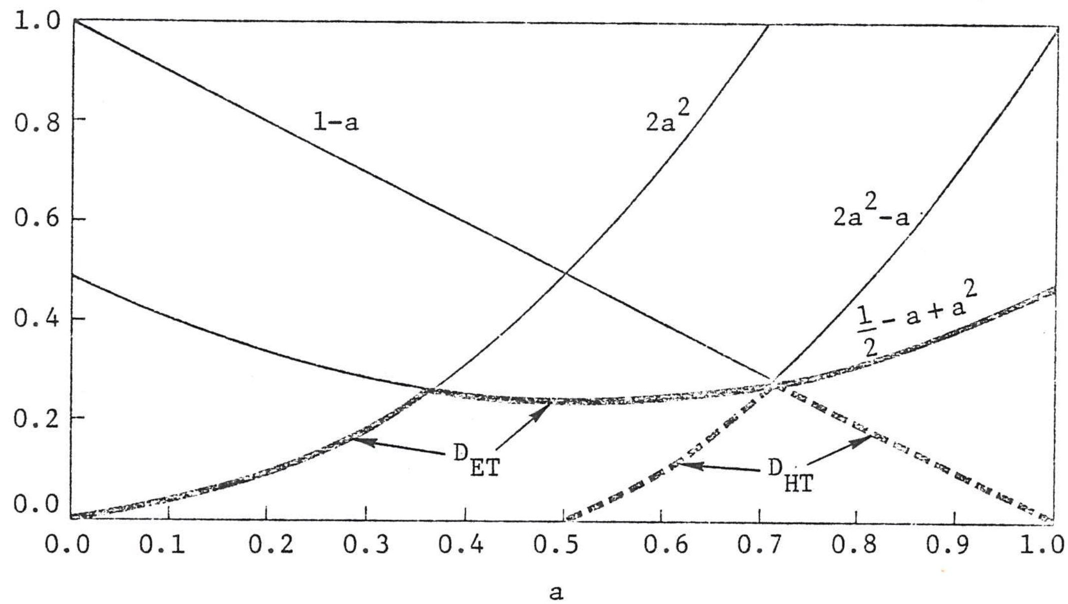


Figure 3.7. The minimum distances of the ET (3.70) and HT (3.69) rules for the signal set in Figure 3.6, as a function of the parameter a.

CHAPTER IV

LINEAR CONSTRAINTS

1. Background

The linear structure is the simplest conceivable structure and at the same time the most analytically tractable. Thus, it is natural to begin a general investigation of structural constraints with the linear structure. Following the general example of Section II.6.2, with $N=1$, the form of the linear estimate is

$$\hat{P}_{i/Y} = \phi_0^i + \int_T \phi_1^i(t)\ \overline{Y}(t)\ dt \quad , \qquad 4.1$$

where

$$\phi_0^i = P_i$$

and ϕ_1^i is the solution to the integral equation

$$P_i\ M_{\overline{y}/H_i}(t) = \int_T \phi_1^i(\tau)\ M_{\overline{y}}^{(2)}(t,\tau)\ d\tau \quad . \qquad 4.2$$

Assuming that $M_{\overline{y}}^{(2)}$ is invertible (it always will be for these applications), the inverse can be expressed as Q, where Q is the solution to

$$\int_T M_{\overline{y}}^{(2)}(t,\tau)\ Q(\tau,\sigma)\ d\tau = \delta(t,\sigma) \quad . \qquad 4.3$$

Then ϕ_1^i can be directly expressed as

$$\phi_1^i(t) = P_i \int_T Q(t,\tau)\, M_{\overline{y}/H_i}(\tau)\, d\tau \quad . \qquad 4.4$$

Thus, the estimate is (2.66)

$$\hat{P}_{i/Y} = P_i[1 + \overline{\tau}_i] \quad , \qquad 4.5$$

where

$$\overline{\tau}_i \triangleq \iint_{T\,T} Q(t,\tau)\, M_{\overline{y}/H_i}(t)\, \overline{Y}(\tau)\, dt\, d\tau \quad . \qquad 4.6$$

Note that the $\overline{\tau}_i$ and hence the conditional means of the observations, $M_{\overline{y}/H_i}$, must be distinct for the linear estimate to be useful. The block diagram of this estimate is shown in Figure 4.1.

If $\overline{y}(t)$ consists of a colored component $\overline{v}(t)$ and an independent white component $w(t)$, with power spectral density N_o, then Gardner [3,7] has shown that

$$M_{\overline{y}}^{(2)}(t,\tau) = N_o\, \delta(t - \tau) + M_{\overline{v}}^{(2)}(t,\tau) \quad , \qquad 4.7$$

and

$$Q(t,\tau) = N_o^{-1}[\delta(t - \tau) - h(t,\tau)] \quad , \qquad 4.8$$

where

$$h(t,\tau) \triangleq \int_T Q(t,\sigma)\, M_{\overline{v}}^{(2)}(\sigma,\tau)\, d\sigma \quad . \qquad 4.9$$

Furthermore,

$$\hat{\overline{V}}(t)/Y \triangleq \int_T h(t,\tau)\, \overline{Y}(\tau)\, d\tau \qquad 4.10$$

is the linear MMSE estimate of the colored component $\overline{V}$ of $\overline{Y}$, given observations $\overline{Y}$. Then

$$\overline{\tau}_i = \iint_{T\,T} N_o^{-1}[\delta(t-\tau) - h(t,\tau)]\, M_{\overline{y}/H_i}(t)\, \overline{Y}(\tau)\, dt\, d\tau$$

$$= N_o^{-1} \int_T M_{\overline{y}/H_i}(t)\, [\overline{Y}(t) - \hat{\overline{V}}(t)]\, dt \quad . \qquad 4.11$$

Thus, the estimation rule subtracts out the linear MMSE estimate of the colored component of the centered observation, and then correlates with the measure of the signal, $M_{\overline{y}/H_i}(t)$, as shown in Figure 4.2. The similarities between these receivers and those that are optimum for sure signals in additive Gaussian noise [11] are remarkable. These similarities are even stronger for the specific case of sure signals in additive noise, discussed in the following section.

2. Sure Signals in Additive Noise

Assume that the observations Y are given by

$$Y(t) = S_i(t) + N(t) \ , \quad t \in T \ , \quad \text{given } H_i \ , \qquad 4.12$$

where $N(t)$ is a sample of a zero-mean second order random process, $n(t)$, with autocovariance $M_n^{(2)}(t,\tau)$ and $\{S_i(t)\}_{i=1}^{M}$ is interpreted as the set of samples, with prior probabilities P_i, of the random signal process, $s(t)$, with autocovariance

$$M_{\overline{s}}^{(2)}(t,\tau) = \sum_{i,j=1}^{M} R_{ij}\, S_i(t)\, S_j(\tau) \quad , \qquad 4.13$$

where

$$R_{ij} \triangleq P_i \, \delta_{ij} - P_i P_j \quad . \tag{4.14}$$

Then,

$$M_{\underline{y}}^{(2)}(t,\tau) = M_{\underline{s}}^{(2)}(t,\tau) + M_n^{(2)}(t,\tau) \tag{4.15}$$

and

$$M_{\underline{y}/H_i}(t) = P_i^{-1} \sum_{j=1}^{M} R_{ij} \, S_j(t) \quad . \tag{4.16}$$

Thus, the formal solution 4.4 becomes

$$\phi_1^i(t) = \sum_{i=1}^{M} W_{ij} \, \theta_j(t) \quad , \tag{4.17}$$

where θ_i is the solution to the Fredholm equation

$$\int_T M_n^{(2)}(t,\tau) \, \theta_i(\tau) \, d\tau = S_i(t) \quad , \tag{4.18}$$

and where the matrix W is given by

$$W \triangleq R[I + VR]^{-1} \quad , \tag{4.19}$$

and where the elements of the matrix V are

$$V_{ij} \triangleq \int_T \theta_i(t) \, S_j(t) \, dt \quad , \tag{4.20}$$

and I is the identity matrix.

The linear (LC) receiver is a correlation receiver shown in Figure 4.3. The correlators are the same as those employed in the OG receiver [11]. The only differences between these receivers are that the OG receiver does not use the linear weighting network and the final biases are $\{\ell n(P_i) - \frac{1}{2} N_o^{-1} V_{ii}\}$ rather than $\{P_i\}$. Note that the receiver could be implemented without the weighting network by using the $\{\phi_1^i\}$ as correlators rather than the $\{\theta_i\}$. Moreover, if the signals are linearly dependent, then only N correlators need be used, followed by a NxM weighting network, where N is the rank of the signals (see Section III.2.3). The specific examples of amplitude-shift keying (ASK) and phase-shift keying (PSK) are discussed in Chapter III, Sections 3 and 4. In addition, the observations can be left uncentered and the final biases P_i replaced by $P_i + \int_T \phi_1^i(t)\, M_y(t)\, dt$.

Now consider the special case of white noise and mutually orthogonal signals with energies $\{E_i\}$. Then, Gardner [3,7] has shown that the solution (4.17) simplifies to

$$\phi_1^i(t) = \frac{\beta_j}{N_o} S_j(t) - \sum_{k=1}^{N} \frac{\beta_k}{\beta_0} S_k(t) \quad , \qquad 4.21$$

$$\beta_k \triangleq P_k(1 + P_k E_k/N_o)^{-1} \quad , \qquad 4.22$$

$$\beta_0 \triangleq \sum_{k=1}^{M} \beta_k \quad . \qquad 4.23$$

The similarities between the LC receiver and the OG receiver are outstanding. The LC receiver performs M-1 tests of the form

$$\sum_{i=1}^{M} L_{jk}(i)\, [\tau_i - P_i E_i] \underset{H_j}{\overset{\overline{H_k}}{\gtrless}} N_o(P_k - P_j) \quad , \qquad 4.24$$

and the OG receiver performs M-1 tests of the form

$$\sum_{i=1}^{M} G_{jk}(i)(\tau_i - E_i/2) \underset{H_j}{\overset{H_k}{\gtrless}} N_o \ln(P_k/P_j) \ , \qquad 4.25$$

where $\{\tau_j\}$ are the correlation statistics

$$\tau_j \triangleq \int_T S_j(t)\ Y(t)\ dt \ , \qquad 4.26$$

and

$$L_{jk}(i) \triangleq \beta_i \left[G_{jk}(i) - \frac{\beta_j - \beta_k}{\beta_o} \right] \ , \qquad 4.27$$

$$G_{jk}(i) = \delta_{ij} - \delta_{ik} = \begin{cases} 1, & i=j \\ -1, & i=k \\ 0, & \text{else} \end{cases} \quad (\text{note: } j \neq k) \ . \qquad 4.28$$

In fact, if the energies $\{E_i\}$ and priors $\{P_i\}$ are uniform, then the receivers are identical. For binary signal detection ($M = 2$), 4.24 and 4.25 reduce to (note that any binary signals can be made orthogonal by subtracting an appropriate constant signal):

LC rule:

$$(\tau_1 - E_1/2) - (\tau_2 - E_2/2) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{N_o}{2} [P_1^{-1} - P_2^{-1}] \triangleq \gamma_L \ , \qquad 4.29$$

OG rule:

$$(\tau_1 - E_1/2) - (\tau_2 - E_2/2) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{N_o}{2} [\ln(P_1^{-1}) - \ln(P_2^{-1})] \triangleq \gamma_G \ . \qquad 4.30$$

Thus, these two receivers differ only in the sensitivity of their thresholds, γ, to the priors. Plotted in Figure 4.4, both thresholds exhibit odd symmetry about $P_1 = P_2 = \frac{1}{2}$, and approach ∞ $(-\infty)$ as P_1 approaches 0 (1). However, the LC receiver is more sensitive to the priors since $|\gamma_L| \geq |\gamma_G|$ for all P_1 and P_2, with equality if and only if $P_1 = P_2 = \frac{1}{2}$, at which point they equal zero. Evaluation of PE (Subsection 4.1) shows that the effect of this increased sensitivity is negligible for priors within an order of magnitude of each other.

For additive noise with heavy-tailed density functions, the performance of the LC rule will be significantly inferior to the optimum rule, since these rules are highly nonlinear. A properly chosen L-constraint might perform comparably, e.g.,

$$\hat{P}_{i/Y} = \phi_0^i + \int_T \phi_1^i(t)\ G[Y(t)]\ dt \quad , \tag{4.31}$$

where G is, for example, a clipping or limiting nonlinearity. Note, however, that this estimate requires knowledge of the moments $E\{G[y(t)]G[y(\tau)]\}$ and $E\{G[y(t)]/H_i\}$. This form is investigated in Chapter VI, where the receiver structure and performance are analyzed for a number of cases.

It should be noted that considerably more general models for M-ary signal detection in additive noise also lead to explicit solutions for the LC receiver. In Reference [5], Gardner obtains the solution for repetitive detection of signals with intersymbol interference and additive noise. This problem is discussed in Section 4. In reference [4], he obtains the solution for repetitive detection for marked and filtered doubly stochastic Poisson processes in additive noise.

3. Sure Signals with Random Parameters in Additive Noise

If the model considered in 4.12 is generalized to include random "nuisance" parameters in the signals, then the hypotheses are composite:

$$H_i: \quad Y(t) = S_i(t,\theta) + N(t) \ , \tag{4.32}$$

where θ is a sample of a set of random parameters with joint probability density function $f_\theta(\cdot)$. The effect of these parameters on the linear receiver is that the signal portion, 4.13, of the kernel $M_{\overline{y}}^{(2)}$ is replaced with its average over the parameter space, and the measure, $M_{\overline{y}/H_i}(t)$, of the signal 4.16 is replaced with its average:

$$M_{\overline{s}}^{(2)}(t,\tau) = M_s^{(2)}(t,\tau) - M_s(t)\, M_s(\tau)$$

$$= \sum_{i,j=1}^{M} \left[P_i \delta_{ij}\, M_{S_i S_j}^{(2)}(t,\tau) - P_i P_j M_{S_i}(t)\, M_{S_j}(\tau) \right] , \tag{4.33}$$

$$M_{\overline{y}/H_i}(t) = M_{S_i}(t) - M_s(t)$$

$$= P_i^{-1} \sum_{j=1}^{M} R_{ij}\, M_{S_j}(t) \quad . \tag{4.34}$$

More generally, for arbitrary random signals in additive noise, the mean signal under hypothesis H_i plays the role of a sure signal and the autocovariance of the signal plays the role of the autocovariance of an additional noise--these roles being defined by the optimum receiver for sure signals in additive colored Gaussian noise.

In cases where the averaged conditional means are equal or zero, the LC estimate is just the prior probability, and therefore makes no use of the observations. An example of this is the case of bandpass signals with a uniformly distributed random phase. For this situation, a higher order (nonlinear) constraint is necessary. This is examined in further detail in the next chapter. As nontrivial examples for which the LC receiver is useful, the following sections investigate random amplitude and partially coherent phase.

3.1 Random Amplitude

Consider a sure signal with a slowly time-varying amplitude, i.e., $S_i(t,A) = A\,S_i(t)$, for all t in the observation interval T. That is, A is random, but constant over T. Then, it is easily seen that the form of the solution for ϕ_1^i is unchanged, the only difference being that V and R are replaced by

$$V' \triangleq M_A V \quad , \tag{4.35}$$

$$R' \triangleq M_A^2 R + \mathrm{Var}\,\{A\}P \quad , \tag{4.36}$$

where

$$P_{ij} \triangleq \delta_{ij} P_i \quad . \tag{4.37}$$

For amplitudes that are not slowly time-varying, the effect is a multiplicative noise, which can be treated in either one of two ways, one of which (generalizing Eqs. 36-42 in [3]) is described here, and the other in Chapter V. Given the model

$$Y(t) = A(t)\,S_i(t,\theta) + N(t) \quad , \quad \text{given } H_i \quad , \tag{4.38}$$

reexpress Y as

$$Y(t) = M_A(t)\, M_{S_i}(t) + [A(t)\, S_i(t,\theta) - M_A(t)\, M_{S_i}(t) + N(t)]$$

$$\triangleq S_i'(t) + N'(t) \quad , \qquad 4.39$$

where S_i' is the "new" deterministic signal, and N' is the "new" noise. If a, θ, and N are independent, then

$M_{n'/H_i}(t) = M^{(2)}_{n's'/H_i}(t,\tau) = 0$, and

$$M^{(2)}_{n'/H_i}(t,\tau) = M^{(2)}_n(t,\tau) + M^{(2)}_A(t,\tau)\, M^{(2)}_{S_i}(t,\tau)$$

$$- M_A(t)\, M_A(\tau)\, M_{S_i}(t)\, M_{S_i}(\tau) \quad . \qquad 4.40$$

Thus, using this approach, linear constraints for multiplicative noise or any nuisance parameters can be treated entirely using only the results obtained for additive noise. This does not carry over, however, for quadratic constraints with zero-mean signals. Suppose that $M_{S_i}(t) = 0$, as would be true for a sinusoid with a uniformly distributed random phase. Then, this approach would have the new signal as zero and the observations as all noise. For this situation, the approach described in Chapter V, Section 3 would be necessary.

3.2 Partially Coherent Phase

Suppose the signal is phase-shift keyed with a partially coherent reference phase (e.g., the output of a phase-locked loop). Then

$$S_i(t,\theta) = A \cos(\omega_o t + \phi_i + \theta) \quad , \qquad 4.41$$

where A is a constant amplitude and θ is a random phase with PDF $f_\theta(\theta)$. For simplicity, assume that the noise is white. Furthermore, assume that the phase PDF is symmetric about zero, monotonically decreasing in $|\theta|$, and zero outside $[-\pi,\pi]$. Also, define the quantities

$$C_k \triangleq \int_{-\pi}^{\pi} f_\theta(\theta) \cos(k\theta)\, d\theta \quad . \tag{4.42}$$

In order that the phase not be completely unknown, which would render the linear receiver useless, C_1 must be positive. If $C_1 = 1$, then the phase is completely known. In addition, to distinguish among the signal phases, C_1 must be approximately no smaller than $\cos(\pi/M)$ for equally spaced $\{\phi_i\}$. For the random amplitude (A) case, the form of the solution is the same as that for the known phase case. The only difference is that the correlation term (see 3.50) is weighted by the factor C_1, which is a measure of the coherence of the phase, i.e.,

$$\phi_1^i(t) = A\, C_1(N_o + A^2T/4)^{-1} \cos(\omega_o t + \phi_i) \quad . \tag{4.43}$$

Hence, the linear receiver ignores the quadrature term of the signal, $\sin(\omega_o t + \phi_i)$. For an analysis of the quadratic receiver, refer to Section V.5.

4. Signals with Intersymbol Interference in Additive Noise

As mentioned in Section 2, Gardner [5] obtains the solution for repetitive detection of sure signals with intersymbol interference (ISI) and additive noise. Typically, ISI occurs over bandlimited (dispersive) channels, which tend to "spread out" the signal pulses in time so that adjacent pulses overlap. Gardner shows that the linearly constrained* MMSE estimate of a signal parameter a with realizations $\{\alpha_i\}_{i=1}^{M}$, is given by the sum of the outputs of a bank of matched filter-tapped delay lines (MF-TDL's) shown in Figure 4.5. Note that matched filter-samplers correspond to multiplier-integrators (correlators), which are used in this work. Thus, other than the bias term, the only difference between the receivers for signals with and without ISI is the presence of the TDL's after the MF's in place of simple gains.

This section extends this receiver structure to estimate q-dimensional parameters, with specific application to passband signals. For this application, the linearly constrained structure is shown to be identical to the conventional ad hoc structure, although the results presented here are considerably more general.

Assuming the signal model 4.12, where $S(t)$ depends on the N-dimensional vector parameter, $\underline{x}$, i.e.,

$$S_i(t) = S(t,X_i) \ , \qquad i=1,\ldots,M \ , \tag{4.44}$$

the L-MMSE estimate of $\underline{x}$ can be formed using the unbiased version of Gardner's results [5] and the general formula (3.3) for L-MMSE parameter estimates in terms of posterior estimates. The resultant structure, shown in Figure 4.6, consists of M TDL's with N sets of taps for each TDL, one set for each component of $\hat{\underline{x}}$. For each value of i, the i^{th} sets of taps of the TDL's are summed to form the estimate for

*No constant (nonrandom) term is included in this estimate, which is therefore biased. Reference [4] contains information about the conversion of the biased estimate in [5] to the unbiased estimate used here.

$(\underline{x}_n)_i$, as in Figure 4.5, where n is a discrete time index for the information sequence. Now, the ET test using $\hat{\underline{x}}$ in accordance with 3.5 follows the structure in Figure 4.6 with a qxM linear transformation, a bias for each of the M final paths, and finally a maximum selector, as shown in Figure 4.7. It is easily shown that the cascade of two linear transformations T_i, with p_i feedforward and q_i feedback taps, $i=1,2$, is a linear transformation T with $p=p_1+p_2$ feedforward and $q=q_1+q_2$ feedback taps. Thus, the Mxq TDL matrix [T] shown in Figure 4.6 can be merged with the qxM transformation shown in Figure 4.7 to yield an overall simplified structure, shown in Figure 4.8. This simplified structure has M TDL's each with M sets of taps. A further simplification arises by taking advantage of the linear dependence of the signals $\{S_i(t)\}_{i=1}^{M}$. Assuming that each signal can be written as a linear combination of N waveforms, i.e.,

$$S_i(t) = \underline{\mu}_i^T \, \underline{S}(t) \quad , \qquad i=1,\ldots,M \quad , \tag{4.45}$$

then the number of MF's, and hence the number of TDL's in Figure 4.8 drops from M to N, each still with M sets of taps. Thus, the matrix size becomes NxM. Otherwise, the structure in Figure 4.8 is unchanged. This simplification is readily shown since the MF's are linear transformations on the signals. This simplification also applies to the receiver structure in Figure 4.6, where any decision scheme could follow the estimate $\hat{\underline{x}}_n$. Thus, these structures can be directly compared to the conventional structures for passband equalization. Proakis [46] describes in baseband complex-signal notation the same structure as that given in Figure 4.6 (with N MF's and a